Simplify Outside Network Connections Using Floating L3Outs

First Published: 2023-02-07

Last Modified: 2023-07-17

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

© 2023 Cisco Systems, Inc. All rights reserved.

CONTENTS

CHAPTER 1 **New and Changed Information** 1

New and Changed Information 1

CHAPTER 2 **About Cisco Floating L3Outs** 3

About Cisco Floating L3Outs 3

Configuring L3Outs for Virtual Environments 3

Configuring L3Outs for Physical Domains 4

Scenarios That Benefit from Floating L3Out 4

Floating L3Out Topology and Terminology 5

CHAPTER 3 **Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out** 11

Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out 11

Support for Multi-Protocol Recursive Route Resolutions 15

Alternative Deployment Model With Dedicated Devices For External Routes Advertisement 18

CHAPTER 4 **Configuring Floating L3Outs** 23

Configuring Floating L3Outs 23

Workflow for Configuring Floating L3Out 24

CHAPTER 5 **Configuring Domains Using the GUI** 25

Configuring Domains Using the GUI 25

Create a Layer 3 Domain Using the GUI 27

Configuring a VLAN Range for an Existing L3Out Domain VLAN Pool Using the GUI 28

Creating a Physical Domain Using the GUI 28

Configuring a VLAN Range For an Existing Physical Domain VLAN Pool Using the GUI 29

Create a VMM Domain for VMware VDS Using the GUI 29

Configuring a VLAN Range for an Existing VMM Domain VLAN Pool Using the GUI 30

CHAPTER 6 **Creating a Floating L3Out 33**

Creating a Floating L3Out 33

Configuration Steps With VMM Domain 35

Configuring a Secondary IP 36

Configuration Example: Different L3Outs for OSPF and BGP Using GUI 36

CHAPTER 7 **Configuring the Avoidance of Suboptimal Traffic From an ACI Internal Endpoint to a Floating L3Out Using the Cisco APIC GUI 39**

Workflow for Configuring Floating L3Out with Avoidance of Suboptimal Traffic 39

Configuring Next Hop Propagation on L3Out 42

Configuring Direct-Attached Host Route Advertising on L3Out 44

Increasing the Local Maximum ECMP Paths in a BGP Address Family Context Policy 45

Configuring Next Hop Propagation and Multipath on an L3Out 46

Verifying the L3Out Configuration 48

Verify the Floating L3Out Port Group in the VMware vCenter 48

Verify Floating L3Out on the Leaf Nodes Using the Cisco APIC GUI 48

CHAPTER 8 **Support for Multiple Encapsulation for L3Outs With SVI 51**

Support for Multiple Encapsulation for L3Outs with SVI 51

Single Floating SVI With Different VLAN Encapsulations on Non-Anchor Nodes 53

Guidelines and Limitations of Multiple Encapsulation for L3Outs with SVI 57

Configuring Multiple Encapsulation for L3Outs With SVI and Floating SVIs Using the GUI 58

Creating a Floating SVI 60

Procedure for Multiple Encapsulations With Regular SVIs 61

Procedure for Multiple Encapsulations With Both Floating and Regular SVIs 62

Configuring Multiple Encapsulation for L3Outs With SVI Using the CLI 63

Configuring Multiple Encapsulation for L3Outs With SVI Using the REST API 64

CHAPTER 9 **Floating L3Out Consideration and Restrictions 67**

Summary of Floating L3Out Deployment Considerations and Restrictions 67

APPENDIX A **Topology Examples with Avoidance of Suboptimal Traffic and ECMP 73**

Topology Examples with Avoidance of Suboptimal Traffic and ECMP 73

APPENDIX B **Configuring Floating L3Outs Using the CLI 81**

Creating a VLAN Pool for Floating L3Out Using the CLI 81

Configuring a VMM Domain Profile for VMware VDS Using the CLI 82

Configuring a Floating L3Out Using the CLI 82

Configuring a Secondary IP Using the CLI 83

Configuring the Avoidance of Suboptimal Traffic From an ACI Internal EP to a Floating L3Out Using the CLI 84

Configuring Maximum Number of Paths for Redistribution of Routes in Fabric Using the CLI 85

Configuring Multiple Next-Hops Using the CLI 85

Verifying Floating L3Out Using the CLI 85

APPENDIX C **Configuring Floating L3Out Using the REST API 87**

Configuring a VLAN Pool for Floating L3Out Using the REST API 87

Configuring a VMM Domain Profile for VMware VDS Using the REST API 88

Configuring a Layer 3 Domain Using the REST API 88

Configuring a Floating L3Out Using the REST API 89

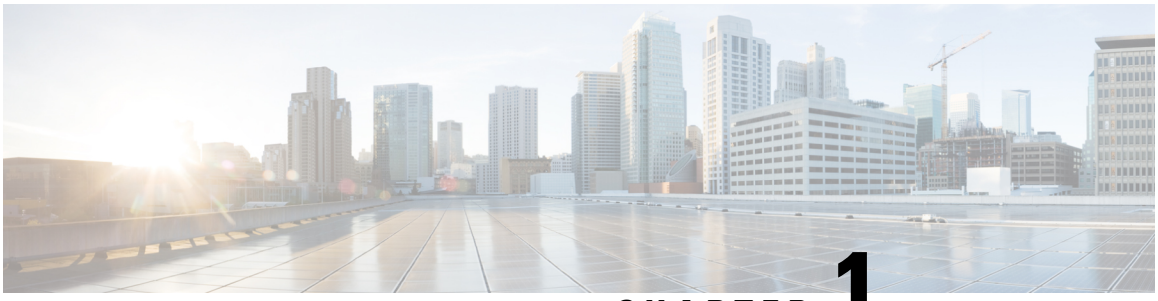
Configuring a Secondary IP Using the REST API 89

Configuring Same VLAN Encapsulation for IPv4 and IPv6 Using the REST API 90

Configuring the Avoidance of Suboptimal Traffic From an ACI Internal Endpoint to a Floating L3Out Using the REST API 90

Configuring Maximum Number of Paths for Redistribution of Routes in Fabric Using the REST API 91

Configuring Multiple Next-Hops Using the REST API 92



CHAPTER 1

New and Changed Information

• New and Changed Information, on page 1

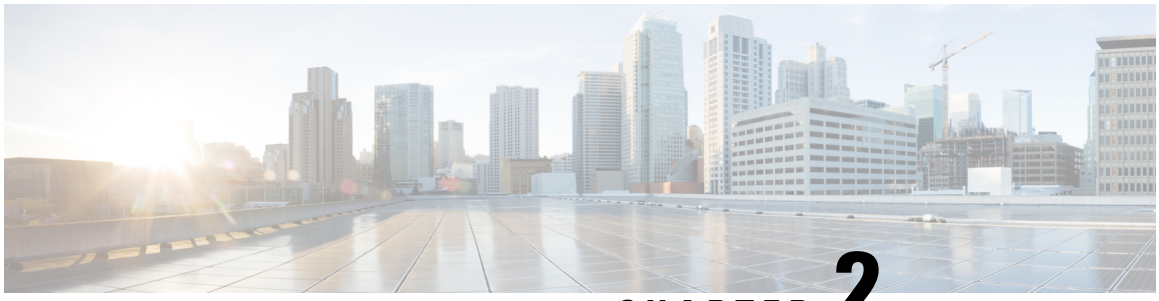
New and Changed Information

The following table provides an overview of the significant changes up to this current release. The table does not provide an exhaustive list of all changes or of the new features up to this release.

Table 1: New Features and Changed Information for Floating L3Out

Cisco APIC Release Version	Feature	Description
5.2(4)	Support for using the same encapsulation for IPv4 and IPv6 floating SVIs for the same L3Out.	The same encapsulation can be used for IPv4 and IPv6 address families with the same VMM domain. While deploying a VMM domain, if both the address families are available, both are deployed.
5.2(3)	Support for using different external VLAN encapsulations, where all of the different external encapsulation instances are treated as part of a single Layer 2 domain	Support is now available where a single L3Out SVI/Bridge domain can be deployed across different leaf switches, with each leaf switch using a different VLAN encapsulation for the same SVI/Bridge domain.
5.2(1)	Next-hop propagation supported with OSPF and static routes redistributed in BGP	For releases prior to release 5.2(1), next-hop propagation is supported with BGP only. Beginning with release 5.2(1), next-hop propagation is also supported with OSPF and static routes redistributed in BGP.
	Support for multiple next-hops to be propagated in the ACI fabric for redistributed routes in BGP	Support is available for multiple next-hops to be propagated in the ACI fabric for redistributed routes in BGP.

Cisco APIC Release Version	Feature	Description
5.0(1)	Secondary IP and floating secondary IP	You can use a secondary IP as a common IP for anchor leaf nodes. A floating secondary IP enables additional floating IP subnets on the same floating SVI.
	Physical domains	Physical domains enable you to use the floating L3Out feature with virtual routers without VMM domain integration or to use a physical router without L3Out logical interface path configurations.
	Avoiding suboptimal traffic from ACI internal endpoints to floating L3Out	Prior to Cisco ACI Release 5.0(1), even if an external router is connected under a non-anchor leaf node, traffic from an ACI internal endpoint to a floating L3Out goes to an anchor leaf node and then goes to the external router through the non-anchor leaf node, which is not an optimal traffic path. Beginning in Cisco ACI Release 5.0(1), you can avoid this suboptimal traffic path by using next-hop propagation and direct-attached host route advertising.
4.2(1)	Floating Layer 3 outside network connection (L3Out) for VMware VDS Virtual Machine Manager (VMM) domains	This new feature enables you to configure an L3Out without the need to define a logical interface. Doing so allows a virtual device that requires routing (for instance, a virtual router) to move from one server to a different one. As a result, the virtual device also moves from one leaf switch to another.



CHAPTER 2

About Cisco Floating L3Outs

- About Cisco Floating L3Outs, on page 3
- Configuring L3Outs for Virtual Environments, on page 3
- Configuring L3Outs for Physical Domains, on page 4
- Scenarios That Benefit from Floating L3Out, on page 4
- Floating L3Out Topology and Terminology, on page 5

About Cisco Floating L3Outs

Beginning with the Cisco Application Centric Infrastructure (ACI) release 4.2(1), you no longer need to specify multiple Layer 3 outside network connection (L3Out) logical interface paths to connect external network devices.

The floating L3Out feature enables you to configure an L3Out without specifying any L3Out interface on the local leaf. The feature saves you from having to configure multiple L3Out logical interfaces to maintain routing when virtual machines (performing a specific virtual network function) move from one host to another. Floating L3Out has been supported since Cisco ACI release 4.2(1) for VMM domains with VMware vSphere Distributed Switch (VDS).

Beginning with the Cisco ACI release 5.0(1), physical domains are also supported. This means that the same simplified configuration can be used for physical routers deployments as well or for virtual routers that are not part of a VMM domain.

Configuring L3Outs for Virtual Environments

When you connect to an external virtual router, you must configure the L3Out logical interface path from the border leaf switch to the uplink of the hypervisor where a virtual device resides. However, when hypervisor resources are aggregated into clusters, there is no guarantee that the virtual machine for the virtual function always runs on the same host.

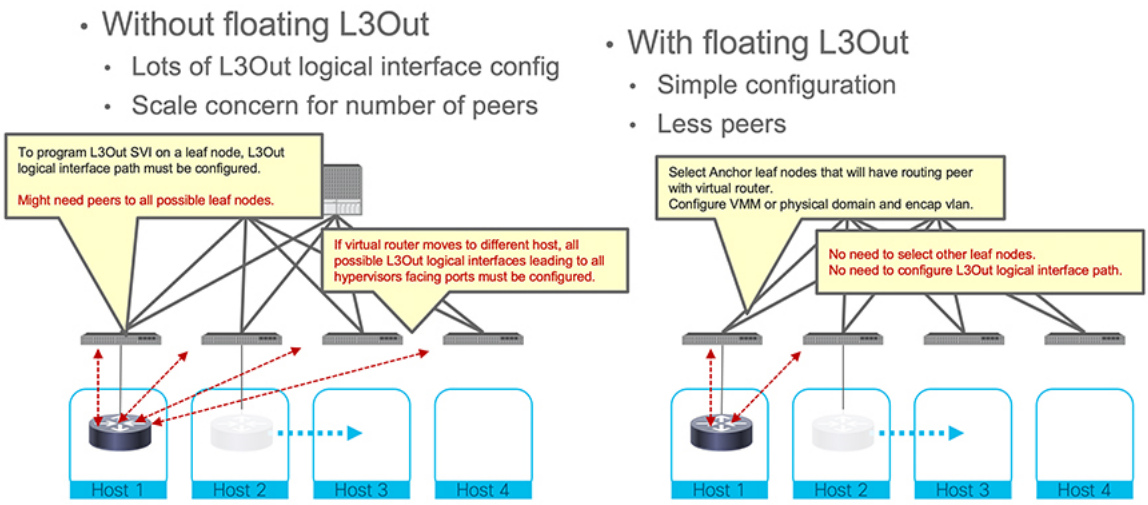
Before Cisco Application Policy Infrastructure Controller (APIC) release 4.2(1), to maintain the routing function when virtual machines moved, you had to configure all possible L3Out logical interfaces from the border leaf switches to all the hypervisors that could host the virtual machines. This extra configuration was required because L3Out switched virtual interface (SVI) and VLAN programming were not done automatically.

For example, if you had a hypervisor cluster stretched across 12 leaf switches, virtual machines could potentially move to every one of those 12 leaf switches. That meant that you had to create a policy to deploy an L3Out from every leaf node interface to every corresponding server.

However, when you configure floating L3Out, you simplify the entire process. After you configure the floating L3Out, you no longer need to configure each L3Out logical interface on all the leaf switches where the hypervisor cluster is connected.

Another benefit of floating L3Out is that only specific leaf nodes (called anchor leaf nodes in this document) will establish routing adjacencies with the external routers. This approach is beneficial for the peering scale of both Cisco ACI leaf switches and the external network devices.

Figure 1: Benefits of Floating L3Out Deployment



Configuring L3Outs for Physical Domains

Beginning with the Cisco Application Policy Infrastructure Controller (APIC) release 5.0(1), support for the floating L3Out functionality is also extended to physical domains. This enhancement enables you to use floating L3Out feature with virtual routers without VMM domain integration or to connect physical routers without L3Out logical interface path configurations.

Scenarios That Benefit from Floating L3Out

The following list provides examples of scenarios where the floating Layer 3 outside network connection (L3Out) is useful. The configuration for floating L3Out is the same for each scenario. These use cases provide additional examples to those discussed in Configuring L3Outs for Virtual Environments, on page 3.

- Physical domain: Even though a physical router does not move between different leaf nodes, using floating L3Out enables you to simplify the configuration because an L3Out logical interface configuration is not required on all the leaf switches where the external physical routers may be connected.
- Virtual firewall or virtual router that is hosted in a hypervisor cluster: Resource scheduling and allocation can be dynamically managed—for example, with the use of VMware Distributed Resource Scheduler

(DRS). The virtual machine (VM) hosting boundary therefore becomes the cluster itself, not a single host.

- Virtual firewall or virtual router with high-availability (HA): The hypervisor HA mechanism allows to restart the failed VM on any available host within the hypervisor cluster. (For example, VMware HA is a specific example of such capability. It is worth noting that this HA capability is in addition to the firewall native redundancy deployment model, such as active/active or active/standby.)
- ECMP load balancing to multiple routers: Using floating L3Out enables you to connect multiple routers to different leaf switches without requiring an L3Out logical interface configuration on each switch.
- Maintenance mode: When a hypervisor must be upgraded, VM administrators evacuate the host. That is, they perform live migration of the firewall or router VM to another host in the hypervisor cluster.
- Disaster Avoidance: In a stretched cluster, outage is expected on some nodes. The VMs, such as virtual routers, virtual firewalls or any other virtual devices, can be migrated to different hosts that are not expected to experience the outage.

Floating L3Out Topology and Terminology

This section describes an example topology for using the floating Layer 3 outside network connection (L3Out) feature. The example uses a VMM domain and a virtual port channel (vPC) deployment, but physical domains are also supported and the use of vPC is not mandatory.

- **Virtual Routers**: Virtual routers can be a router, a virtual firewall, or any other virtual device that is used as a next-hop of a static route on the Cisco Application Centric Infrastructure (ACI) fabric or establishes a routing adjacency with the ACI fabric.
- **Anchor Leaf Nodes**: In this example, there are two leaf switches acting as anchor leaf nodes (Leaf1 and Leaf2) and establishing Layer 3 adjacencies with the external routers. As of Cisco ACI release 6.0(1), the verified scalability number of anchor leaf nodes is 6 per L3Out.

  Anchor leaf nodes make use of the primary IP addresses and floating IP addresses. They can also have secondary IP and floating secondary IP addresses if needed (the purpose of all those IP addresses will be clarified later in this section of the document).
- **Non-anchor Leaf Nodes**: In this example, there are two leaf switches acting as the non-anchor leaf nodes (Leaf3 and Leaf4). The non-anchor leaf nodes do not create any adjacency with the external routers. They acts as a “pass-through” for traffic flowing between directly connected external routers and the anchor node. As of ACI release 6.0(1), the verified scalability number of non-anchor leaf nodes is 32 per L3Out.

  Non-anchor leaf nodes use a floating IP address and can have floating secondary IP addresses, if needed (those IP addresses are shared by all the non-anchor leaf nodes). If it is a VMware vDS VMM domain, the floating IP address is deployed only when the virtual router is connected to the leaf node. If it is a physical domain, the floating IP address is deployed if the leaf port uses an AEP that has an L3Out domain associated to the floating L3Out. The floating IP address is the common IP address for non-anchor leaf nodes.

Configuring an L3Out creates an L3Out bridge domain on the anchor node switches. This L3Out bridge domain is usually referred to as the “L3Out’s SVIs subnet”. In the case of a floating L3Out with VMM domain, once a virtual router moves to a host connected to a non-anchor switch, Cisco Application Policy Infrastructure Controller (APIC) deploys the L3Out bridge domain on the non-anchor leaf switch as well. It also installs the

floating IP address (and floating secondary IP addresses, when needed) on the non-anchor leaf switch. If an external EPG under the L3Out has a contract with another EPG, the routes to the EPG and policy enforcement rules for the contract are also installed on the non-anchor leaf switch. Although the location of the virtual router is changed, it still can maintain the routing adjacencies with the SVI interfaces deployed on the anchor leaf nodes because of the ACI capability of extending connectivity for the L3Out bridge domain across anchor and non-anchor leaf nodes.

Figure 2: Example of a Floating L3Out Topology (external routers are connected to a pair of anchor leaf nodes)

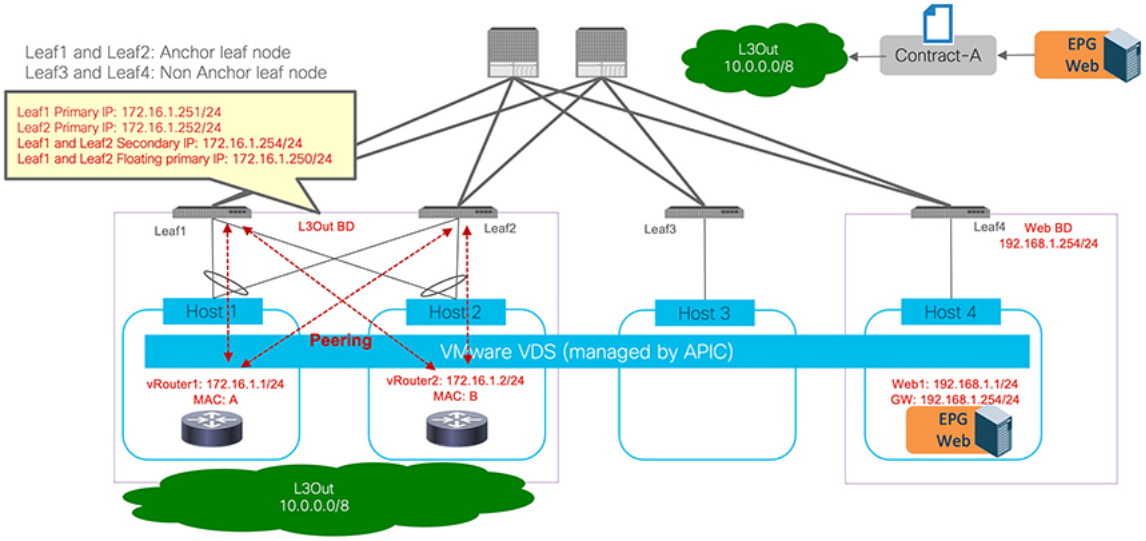
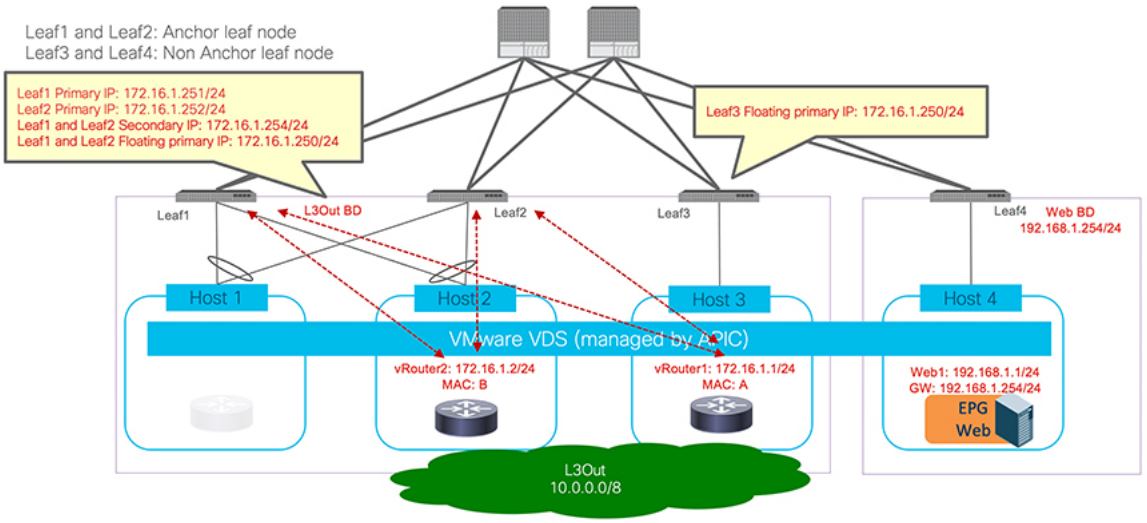


Figure 3: Example of a Floating L3Out Topology (an external router is moved to a non-anchor leaf node)



As mentioned above, anchor and non-anchor leaf nodes that are defined as part of the floating L3Out configuration make use of the following IP addresses.

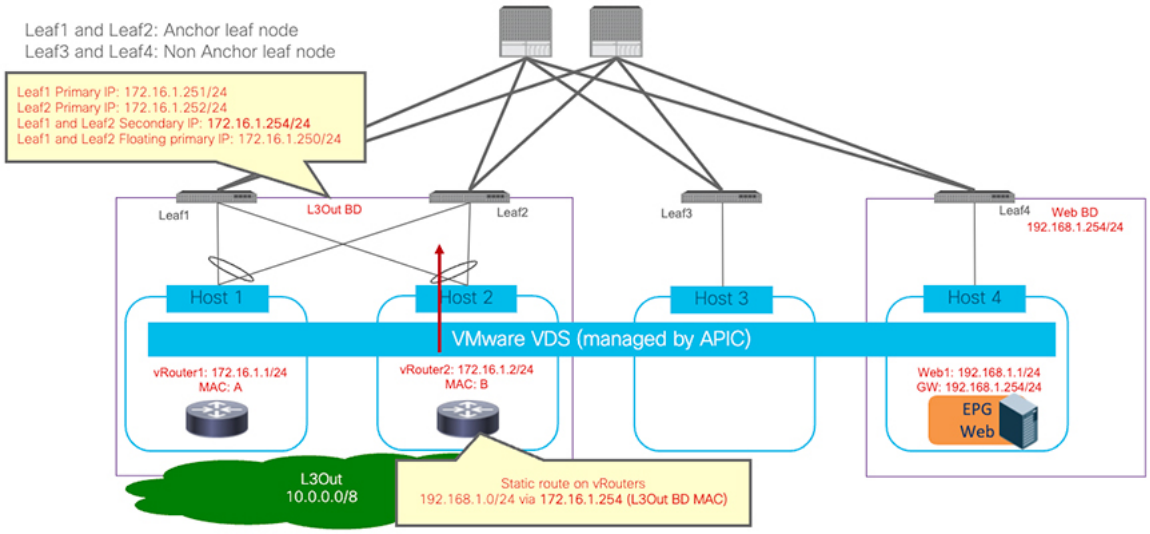
- Primary IP address: the unique IP address assigned to an SVI interface on each leaf node part of the L3Out (it is the real IP of the leaf node). In the case of floating L3Out, the provisioning of an SVI interface with a unique primary IP address is required on each anchor leaf node and it's used for establishing L3Out peering adjacencies with the external routers.

- Secondary IP address (optional): additional IP address assigned to the SVI interface for anchor leaf nodes, which can be used for the use cases below:
  - Common IP address shared by the anchor leaf nodes, acting as a virtual IP and used when external network devices are connected using static routing. The external network devices will set the next-hop gateway of the static routes to this specific IP.
  - Unique IP addresses per anchor leaf node for a secondary IP subnet provisioned in addition to the primary IP subnet.
  - Common IP address shared by the anchor leaf nodes for the secondary IP subnet (to be used for static routing, as described above for the primary IP subnet).
- Floating (primary) IP address: the floating IP is programmed on anchor and non-anchor nodes in order to program a layer-3 interface on the floating SVI. This will program the same MAC address on all anchor and non-anchor switches and allows the switch to directly forward traffic received from external routers into the fabric. It is used for ARP resolution from an anchor leaf node.
- Floating secondary IP address (optional): a common IP provisioned on anchor and non-anchor leaf nodes. Used only if multiple subnets are used in the same external bridge domain (SVI). Floating secondary IP is not supposed to be used for external communication.

The figure below illustrates an example.

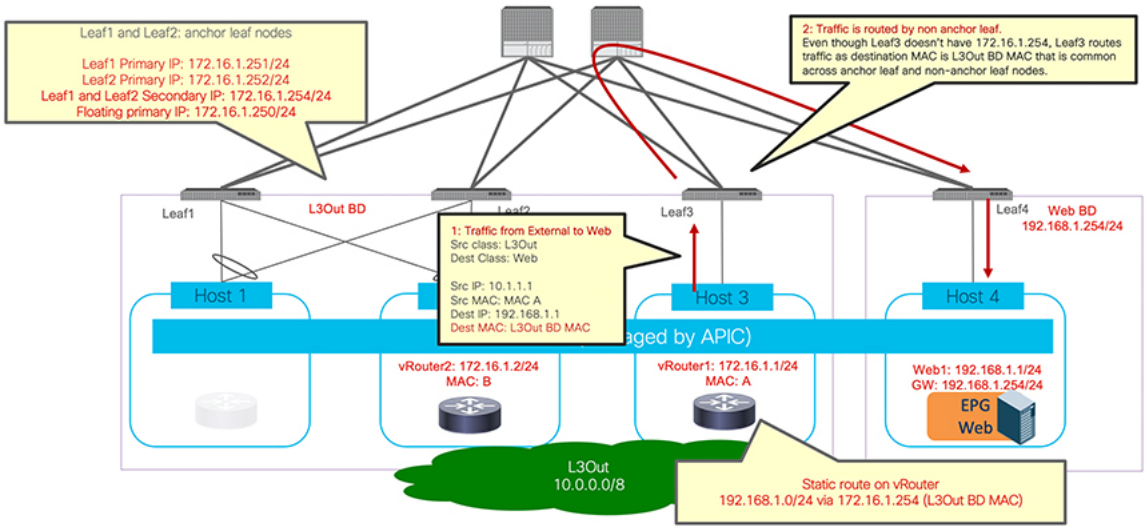
- Primary IP address: 172.16.1.251 is Leaf1 primary IP address and 172.16.1.252 is Leaf2 primary IP address (those are the anchor leaf nodes).
- Secondary IP address (optional): 172.16.1.254 is the secondary IP address of Leaf1 and Leaf2. 172.16.1.254 is used as the next-hop of the static route on the external devices to reach the IP subnet 192.168.1.0/24 deployed inside the ACI fabric.
- Floating (primary) IP address: 172.16.1.250/24 is the floating IP address that is used for ARP resolution.
- If another subnet is required using the same SVI VLAN encapsulation, additional secondary IP addresses and floating secondary IP addresses can be added under the same floating SVI. For example, 172.16.2.254 as the secondary IP address and 172.16.2.250 as the floating secondary IP address.

Figure 4: Example of IP addresses: secondary IP is used as the next-hop of the static route



Because non-anchor leaf nodes instantiate the floating (primary) IP with the same MAC address as the next-hop IP (172.16.1.254 in the example above) used in the static route, even if the external router is moved under a non-anchor leaf node, traffic is directly routed by the non-anchor leaf node.

Figure 5: Example of IP addresses: secondary IP is used for the next-hop of the static route (an external router is connected to a non-anchor leaf node)



- **Traffic Flow**: Regardless of the use of dynamic peering or static routing between the external routers and anchor leaf node, before any virtual routers move, external to internal (L3Out-to-Web) traffic through an anchor node goes to the spine switch and then to the web endpoint in Host 4. Return traffic (Web-to-L3Out) goes back to the virtual router through an anchor leaf node.

  If a virtual router moves to Host 3 under non-anchor Leaf3 as illustrated in Figure 5: Example of IP addresses: secondary IP is used for the next-hop of the static route (an external router is connected to a non-anchor leaf node), on page 8, external-to-internal (L3Out-to-Web) traffic comes to the fabric through Leaf3 and then to the web endpoint in Host 4 through the spine switch.

The return traffic (Web-to-L3Out) goes back to an anchor leaf node, and then goes back to the virtual router through the non-anchor leaf node as illustrated in Figure 6: Return Traffic Flow Steered Toward the Anchor Leaf Node, on page 9 and Figure 7: Traffic Flow Bouncing between Anchor and Non-Anchor Leaf Nodes, on page 9. It's because the anchor leaf nodes learn the external route via the virtual router and redistributes the route to the other leaf nodes.

Figure 6: Return Traffic Flow Steered Toward the Anchor Leaf Node

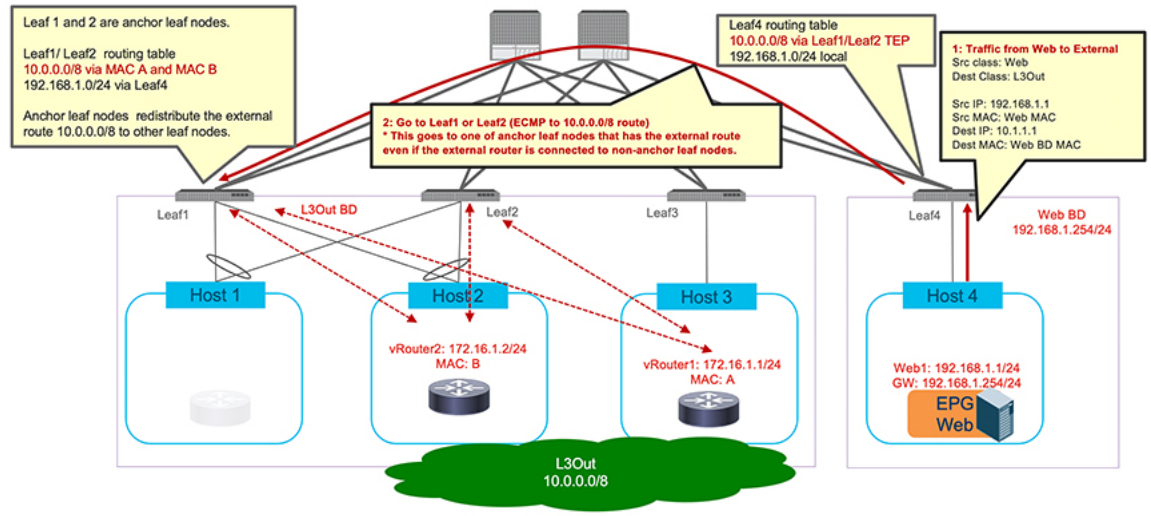
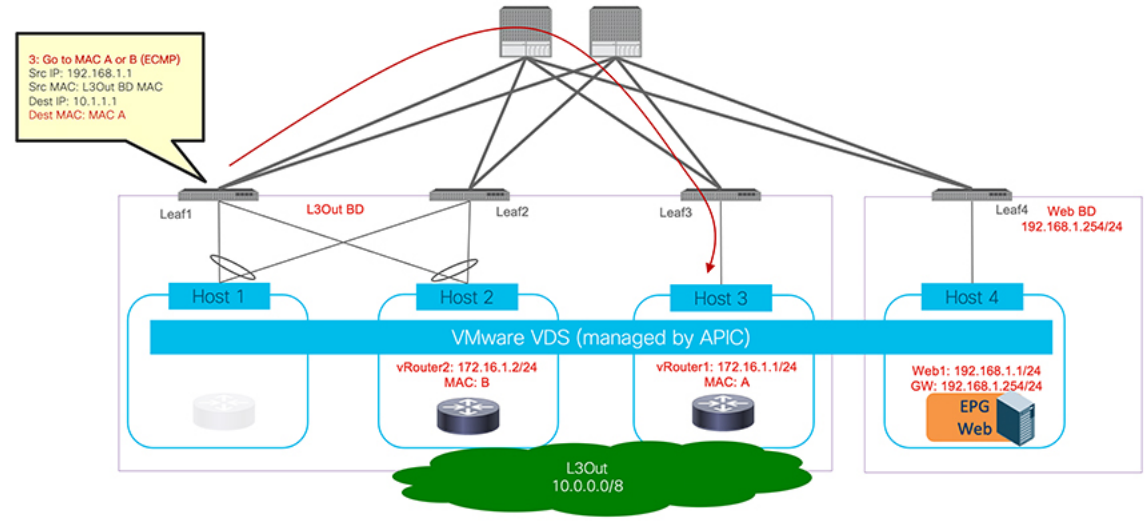
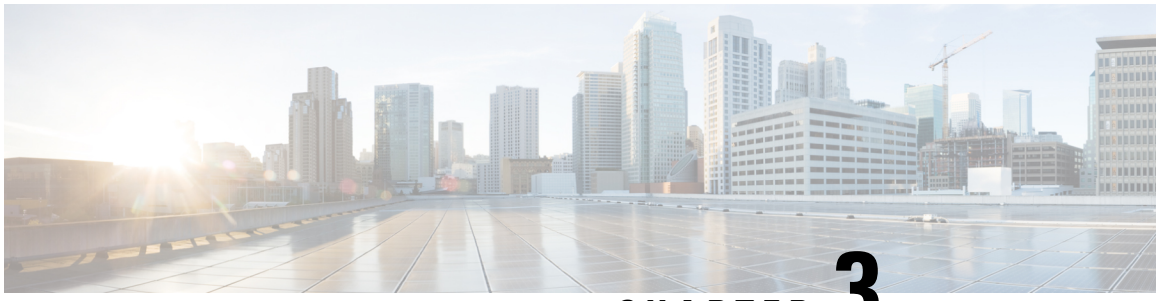


Figure 7: Traffic Flow Bouncing between Anchor and Non-Anchor Leaf Nodes



Note You can avoid this suboptimal path by using Cisco ACI Release 5.0. For more information, see Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out, on page 11.



CHAPTER 3

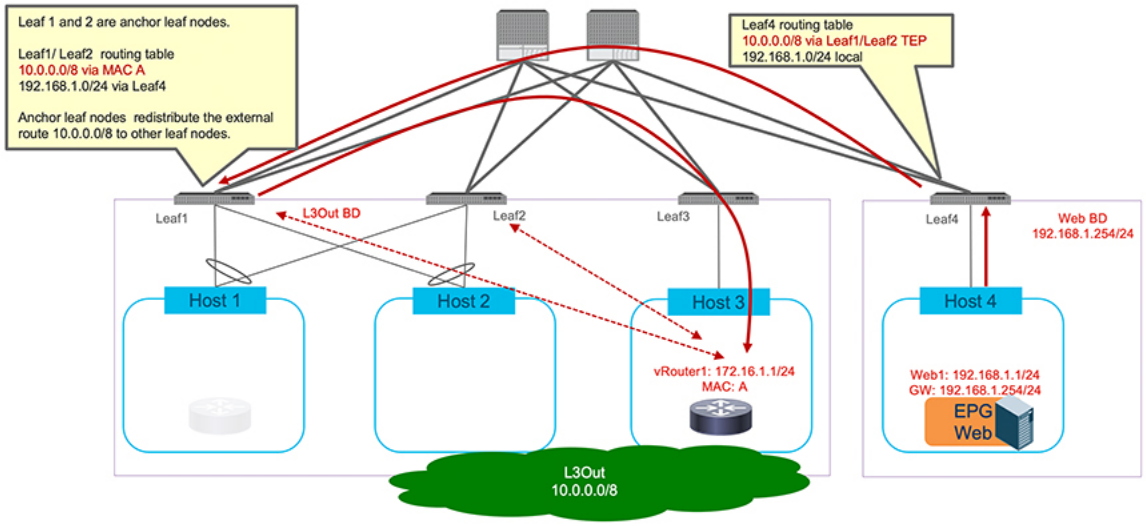
Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out

- Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out, on page 11
- Support for Multi-Protocol Recursive Route Resolutions, on page 15
- Alternative Deployment Model With Dedicated Devices For External Routes Advertisement, on page 18

Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out

Before the Cisco Application Centric Infrastructure (ACI) release 5.0(1), even if an external router is connected under a non-anchor leaf node, traffic from a Cisco ACI internal endpoint sent to the external destination goes to an anchor leaf node and then is redirected to the external router through the non-anchor leaf node, which represents a suboptimal traffic path.

Figure 8: Example of Suboptimal Traffic Flow between Internal EPG and External Network



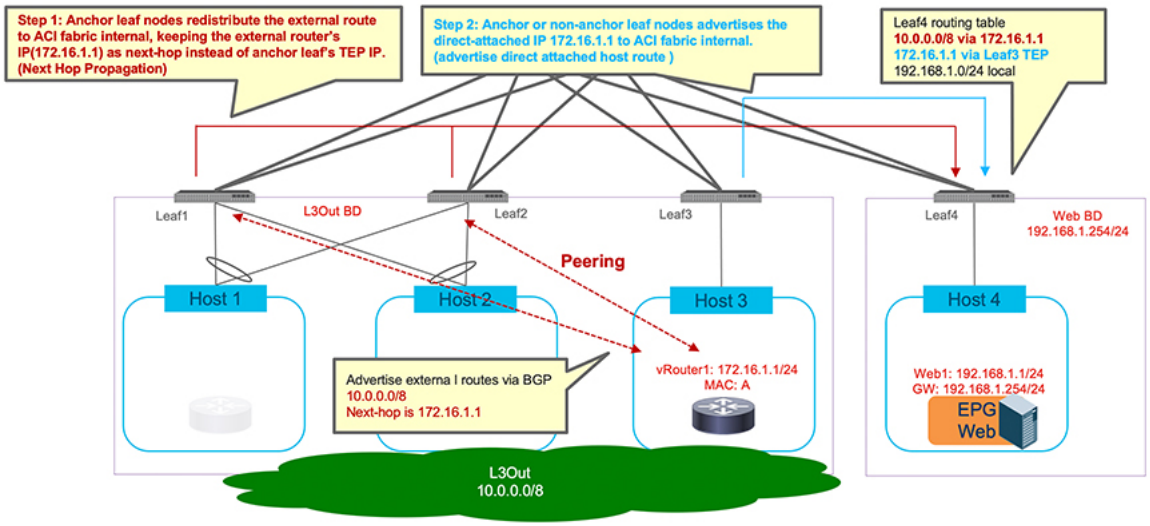
Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out

Beginning with Cisco ACI release 5.0(1), you can avoid this outbound suboptimal traffic path behavior by configuring the following two features:

- Next-hop propagation: this configuration is applied only to the anchor leaf nodes and enables them to redistribute the external prefixes inside the Cisco ACI fabric with the next-hop IP address of the external router announcing these external prefixes. That way, the compute leaf nodes (as Leaf4 in the example in Figure 9: Advertisement of External Prefixes with Next-Hop Propagation Enabled, on page 12) receive and install in their forwarding tables the external prefixes with the external router's IP address as the next-hop (10.0.0.0/8 reachable using 172.16.1.1 in the example below).
- Direct host advertisement route-control profiles: this configuration is applied on all the anchor and the non-anchor leaf nodes where the external routers are connected. It enables those leaf nodes to redistribute the directly attached host route (representing the external router's IP) inside the Cisco ACI fabric (172.16.1.1 using the Leaf3 TEP in the example below). This is critical to ensure that the compute nodes can perform a recursive lookup and send the outbound flows directly to the leaf nodes where the external routers are connected, no matter if they are anchor or non-anchor leaf nodes.

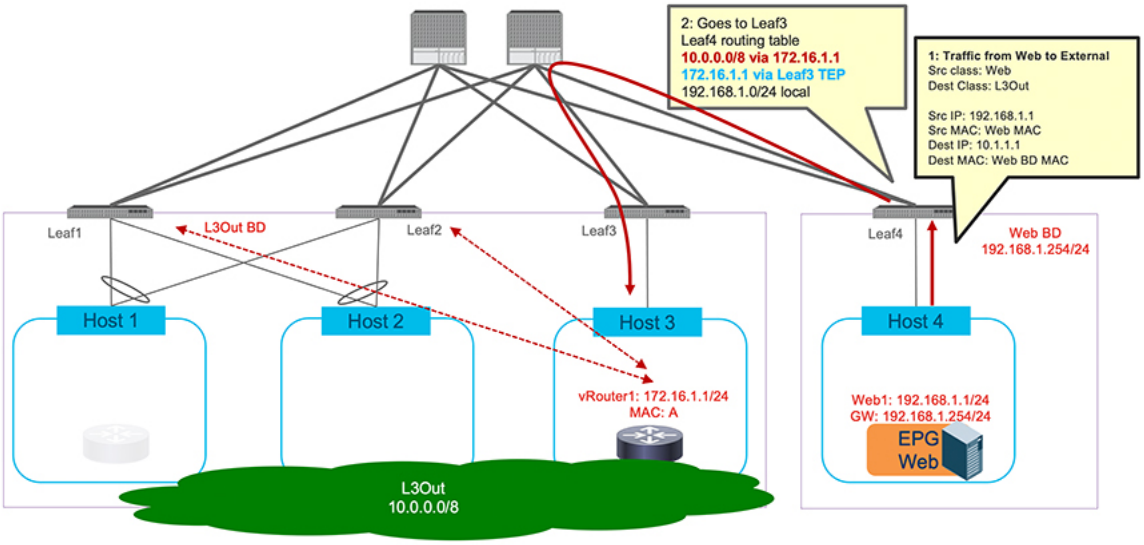
Note The functionalities listed above are supported for floating L3Outs with physical domains only, not with VMM domains.

Figure 9: Advertisement of External Prefixes with Next-Hop Propagation Enabled



Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out

Figure 10: Optimized Traffic Flow between Internal EPG and External Network



Between Cisco ACI releases 5.0(1) to 5.2(1), you can use the functionality described above to avoid this suboptimal path when using eBGP for peering between the external devices and the anchor nodes. Starting from Cisco ACI release 5.2(1), avoiding this suboptimal path is also supported when using OSPF for peering or even with static routing.

As of Cisco ACI Release 6.0(2), the outbound traffic optimization functionality described above is supported for intra-VRF traffic only. Inter-VRF traffic where the consumer and the provider are in different VRFs is not supported. This consideration is applied to both EPG to external EPG and external EPG to external EPG contracts.

Although the example above uses a single external device for peering with the anchor leaf nodes and forwarding traffic from/to the external network domain, the use of different external devices is also possible. Figure 11: ECMP for external prefixes when deploying multiple external routers with OSPF or static route, on page 14 illustrates an example using OSPF or static route, and Figure 12: Lack of ECMP for external prefixes when deploying multiple external routers with BGP, on page 14 illustrates an example using BGP. Each external device can establish routing peering with the anchor leaf nodes to propagate external prefix information into the fabric.

Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out

Figure 11: ECMP for external prefixes when deploying multiple external routers with OSPF or static route

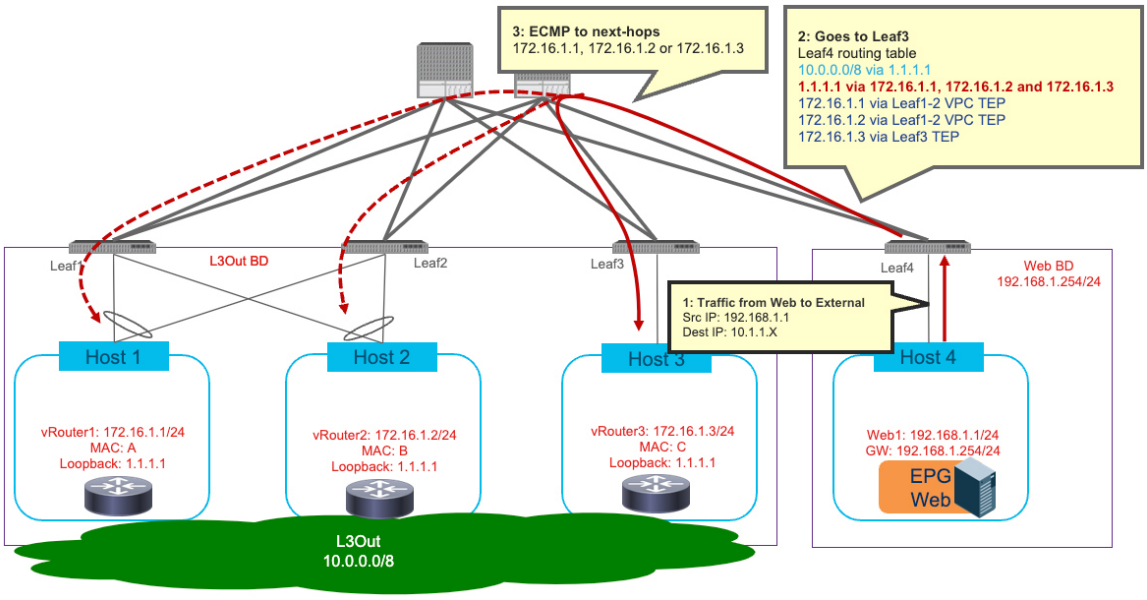
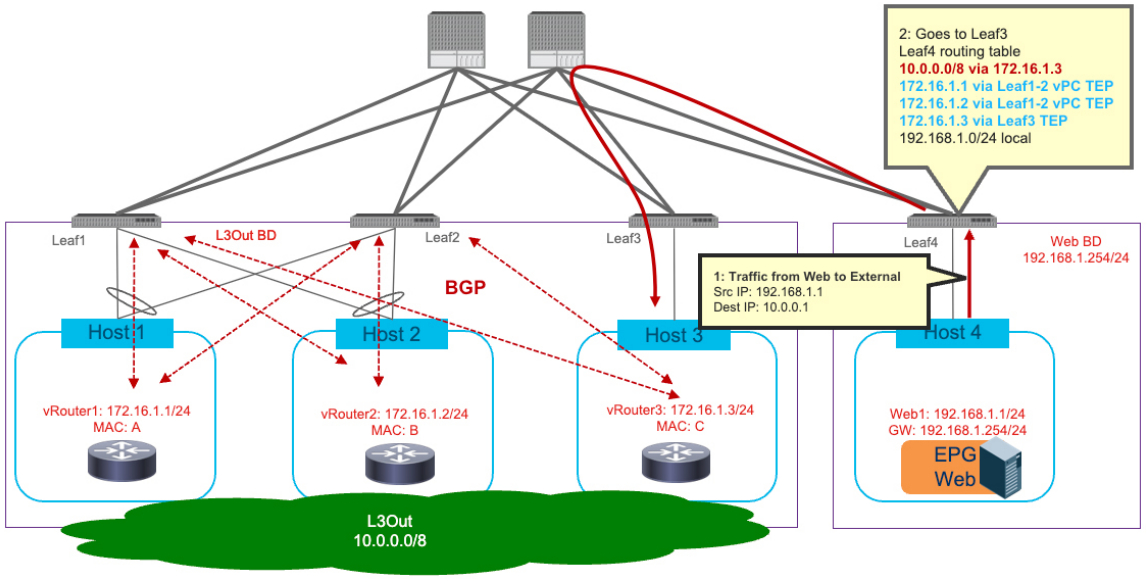


Figure 12: Lack of ECMP for external prefixes when deploying multiple external routers with BGP



When BGP is used to learn the same external prefix, even if each external router advertises the same external prefixes (10.0.0.0/8, in this example), the compute leaf nodes receiving them only install a single next-hop for each prefix. In other words, it is not possible to leverage ECMP for reaching the same external prefix. This restriction is due to the fact that, as of Cisco ACI release 6.0(1) and in the specific case of Cisco ACI floating L3Out deployments, only one IP address can be installed on the Cisco ACI leaf nodes as external next-hop to reach a prefix learned using BGP. This consideration is applicable to both IPv4 and IPv6.

Note As shown in the figure above, a single path to the external prefix is installed on the compute leaf node. This is not the case for the anchor leaf nodes, where all the paths to the external prefix received by the external routers can be successfully installed.

A possible solution to benefit from ECMP for reaching external prefixes consists of deploying the same loopback IP address on all the forwarder nodes so that the following can happen:

- All the external routers can advertise a specific external prefix to the anchor leaf nodes via BGP using the same address (the IP address of the loopback interface) as next-hop for the prefix.
- The anchor leaf nodes receives the external prefix and performs two functions:
  1. Redistribute the external prefix information into the ACI fabric, with a single next-hop represented by the common loopback IP address configured on each external router.
  2. Redistribute the common loopback IP address into the ACI fabric, with next-hops the IP addresses of the external routers part of the directly connected L3Out SVIs subnet. This information must be learned on the anchor nodes via OSPF adjacencies established with the external routers or configured via static routes.

A compute leaf node receiving the control plane information described above from the anchor nodes can then leverage the "Recursive Route Resolution" feature introduced in Cisco ACI release 5.2(1) to benefit of ECMP for outbound flows destined to the external prefix. This behavior will be discussed in greater detail in the next section: Support for Multi-Protocol Recursive Route Resolutions, on page 15.

Support for Multi-Protocol Recursive Route Resolutions

Beginning with Cisco Application Centric Infrastructure(ACI) release 5.2(1) and later, support is available for enabling recursive route resolution for external prefixes received using BGP and redistributed into the fabric. To achieve this, it is required to define a loopback IP address on the external router to be used as next-hop for the external prefixes learned on the anchor leaf nodes through BGP.

Figure 13: Advertisement of external prefixes with next-hop propagation with multipath enabled, on page 16 illustrates an example of such configuration. The external route 10.0.0.0/8 is advertised using BGP from the external router. The next hop for the external prefix is no longer the IP address of the external router that is connected to the L3Out SVIs subnet (172.16.1.1) but it is instead a loopback address defined on the external router and advertised using OSPF (or through static routing). This results in a multi-level recursion, where the BGP route next hop is resolved using the OSPF route, and the OSPF route ultimately gets resolved using the direct adjacency formed within the Cisco ACI fabric.

In such a case, you must enable an extra functionality in addition to next-hop propagation and direct host advertisement route-control profiles described in the previous section:

- Next-hop propagation with Multipath: it enables the anchor leaf nodes to redistribute the external prefix into the fabric with the next-hop IP address being the external router's loopback address learned via OSPF (or static routing), which is used for an extra recursive lookup. The compute leaf nodes (Leaf4 in the example below) receive the external router's loopback IP address with the external router's directly connected IP address as the next-hop (1.1.1.1/32 via 172.16.1.1 in the example below), which is used to reach the external route behind the external device (10.0.0.0/8 via 1.1.1.1 in the example below).

Figure 13: Advertisement of external prefixes with next-hop propagation with multipath enabled

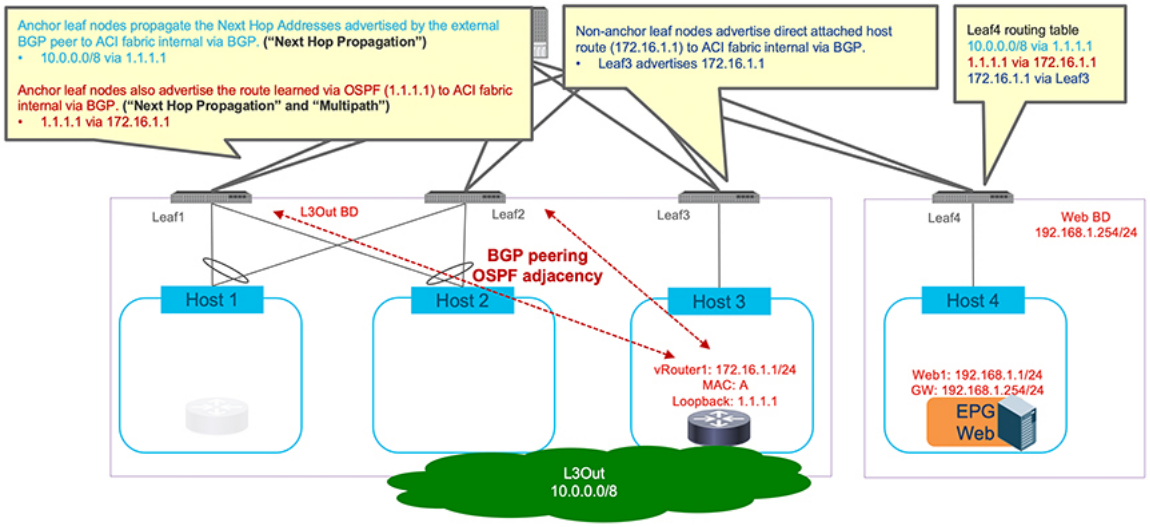
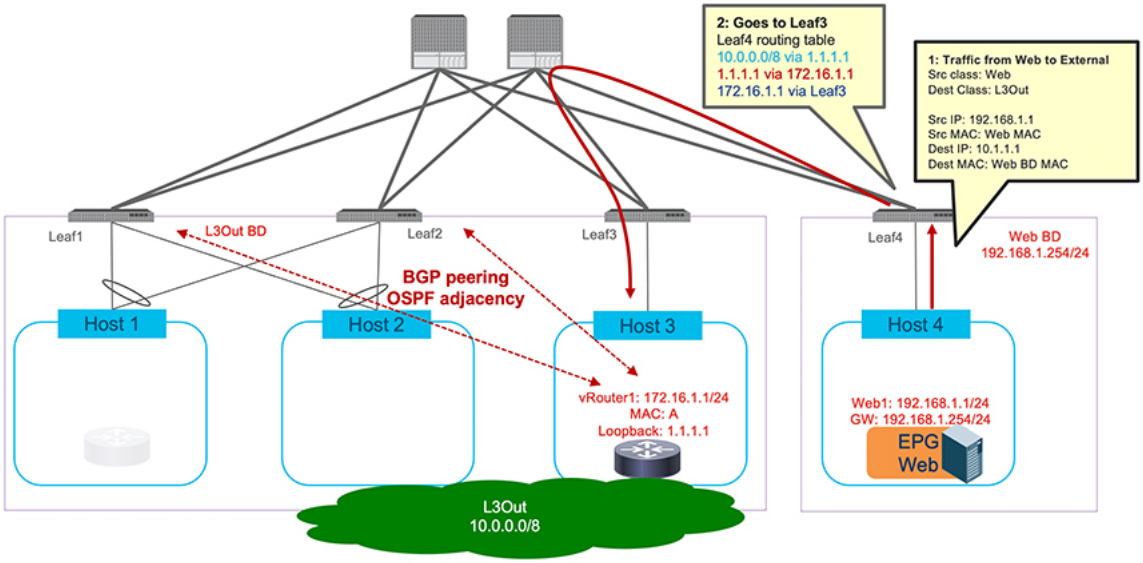


Figure 14: Optimized traffic flow between internal EPG and external network



Note Although the same external device may be used to advertise the external prefixes using BGP and the next-hop IP address using OSPF, two L3Outs are required for the design above because different routing protocols (OSPF and BGP) are used. For example, L3Out-OSPF and L3Out-BGP, which can both have the same floating SVI with the same VLAN encap by using "VRF" Encap Scope. The external EPG with the proper classification subnet (for example 10.0.0.0/8 in this specific example) needs to be configured only for one of the two L3Outs. For "VRF" Encap Scope, please see ACI Fabric L3Out Guide for detail.

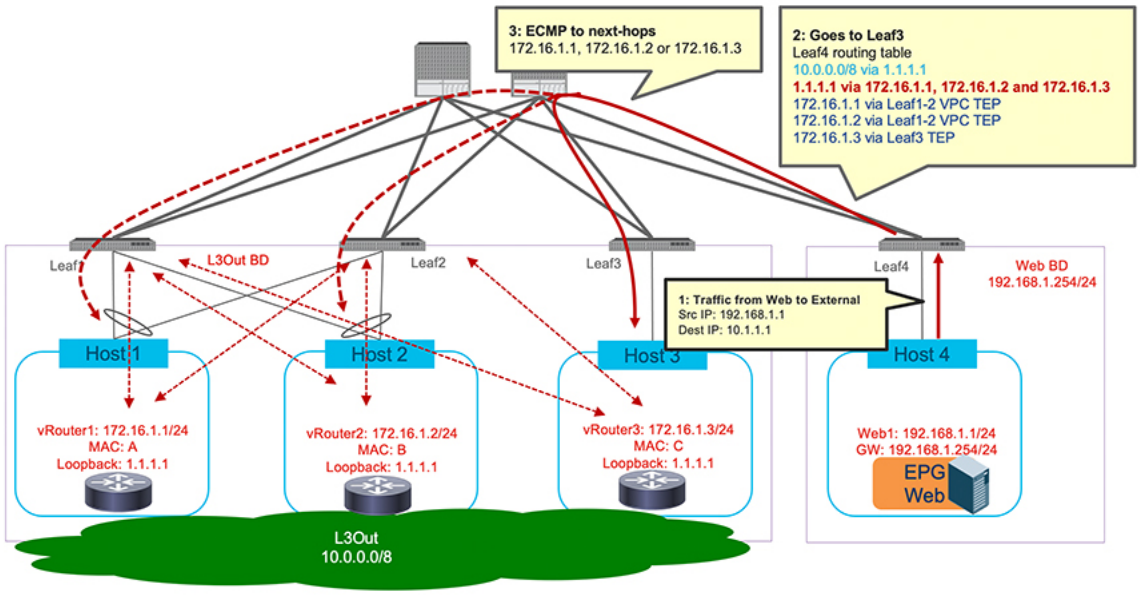
Although this example uses OSPF, static route can be used instead of OSPF. If it's a combination of BGP for external route and static route for next-hop loopback IP address, it can be configured under the same L3Out.

The use of a loopback IP address as next-hop to reach the external prefixes becomes useful if there is a possibility to add more external devices connected to the same external network, as it allows you to increase the forwarding scale leveraging the ECMP functionality to reach the same external prefix.

As previously mentioned, in the latest Cisco ACI release 6.0(1) available at the time of this writing, only one next-hop for external prefixes that are received on a floating L3Out via BGP can be installed on the Cisco ACI compute leaf nodes. If ECMP is needed to reach the external prefixes, it is possible to do the following:

- BGP sends one primary next-hop for the prefix, represented by the same loopback IP address that is configured on all the external routers. Regarding the example in Figure 15: Multiple recursive lookups on a compute leaf node to reach an external prefix, on page 17, the external prefix 10.0.0.0/8 is learned via BGP using the next-hop 1.1.1.1.
- The 1.1.1.1 next-hop for the external prefixes is preserved when the anchor leaf nodes redistribute this information in the MP-BGP VPNv4 fabric control plane.
- At the same time, the anchor leaf nodes learn the 1.1.1.1 loopback address using OSPF adjacencies established with the external routers. Each router advertises as a next-hop for such prefix the IP address of its local interface connected to the L3Out's SVI subnet (172.16.1.1, 172.16.1.2 and 172.16.1.3).
- The end result, when looking at the routing table of a generic compute leaf node (Leaf4 in the example in figure below) is the use of multiple recursive routing lookups:
  - A single loopback next-hop address (1.1.1.1) is installed to reach the external prefix (10.0.0.0/8).
  - Multiple paths (using the 172.16.1.1, 172.16.1.2 and 172.16.1.3 addresses identifying the different external routers) are available to reach the loopback next-hop address if OSPF or static route instead of BGP is used for the loopback route.
  - The VTEP addresses of different leaf nodes are used to reach the external routers' IP addresses directly connected to the L3Out's SVI subnet.

Figure 15: Multiple recursive lookups on a compute leaf node to reach an external prefix



Alternative Deployment Model With Dedicated Devices For External Routes Advertisement

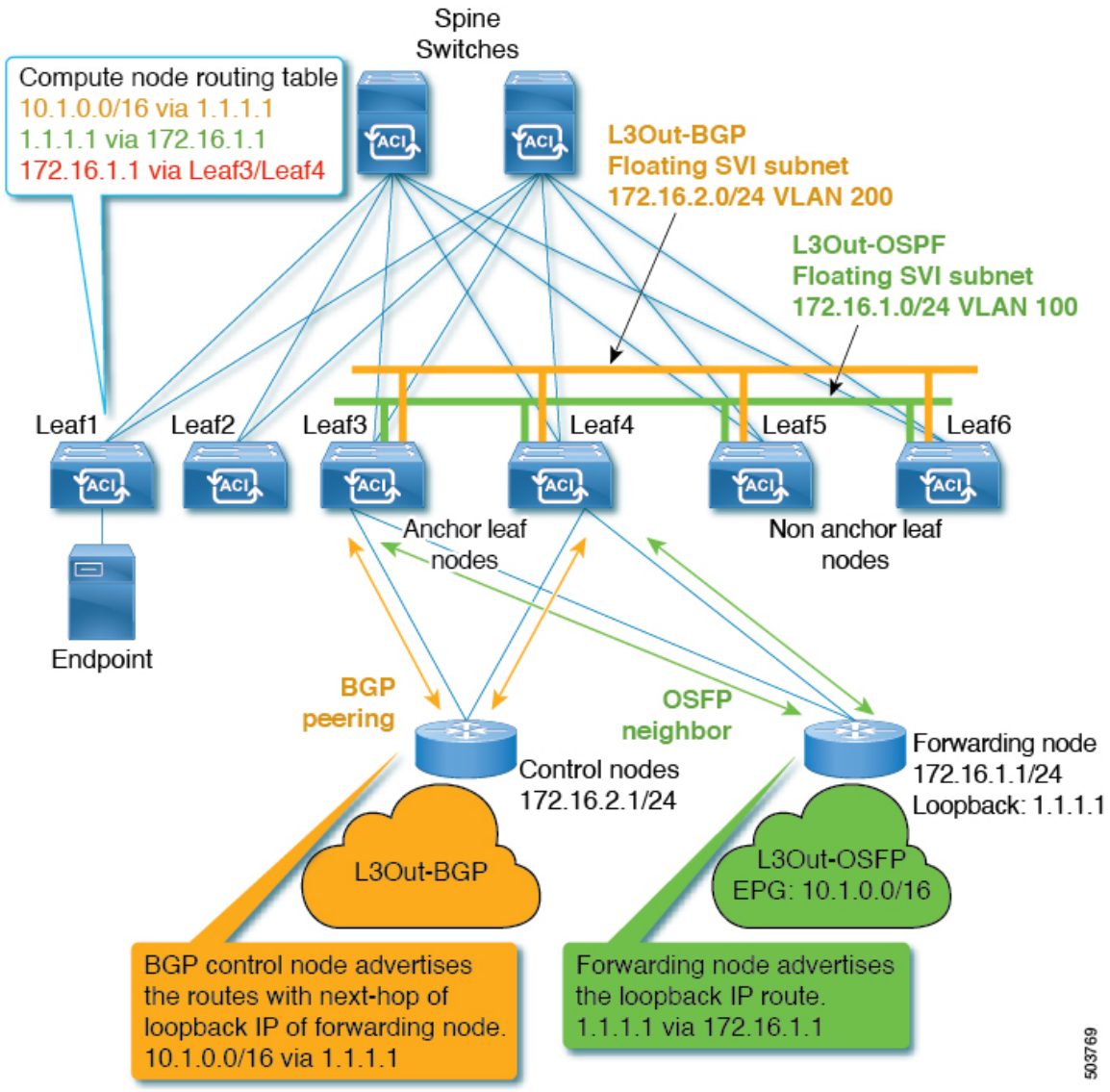
In the topology shown in previous Figure 15: Multiple recursive lookups on a compute leaf node to reach an external prefix, on page 17, the external devices are used to establish the control plane adjacencies with the anchor leaf nodes and also to forward traffic between the Cisco Application Centric Infrastructure (ACI) fabric and the external network. As an alternative deployment model, dedicated control node devices can be deployed for advertising the external prefixes to the Cisco ACI fabric (using the BGP routing protocol), while separate forwarding nodes are used for actual data-path forwarding. This design provides the possibility to increase the number of nodes for scaling up the forwarding capacity without the establishment of additional control plane peerings (BGP in this example) for the external prefixes regardless of the number of forwarding nodes that are deployed.

Figure 16: Example of separate external devices for Avoid Suboptimal Traffic with a loopback address (First Stage), on page 19 and Figure 17: Example of separate external devices for Avoid Suboptimal Traffic with a loopback address (Second Stage), on page 20 below illustrates an example with a loopback next-hop address that is advertised using OSPF. The control node is for BGP route advertisement, and the forwarding nodes are for OSPF routes advertisement and are used for actual data-path forwarding.

In Figure 16: Example of separate external devices for Avoid Suboptimal Traffic with a loopback address (First Stage), on page 19, the preliminary stage of the configuration is shown. In this case:

- Leaf3 and Leaf4 are anchor leaf nodes for both L3Out-BGP and L3Out-OSPF.
- Leaf5 and Leaf6 are non-anchor leaf nodes.
- The orange and green lines spanning all four leaf nodes (Leaf3 through Leaf6) show the L3Outs' SVI subnets reachability (i.e. they represent the external bridge domains associated to each L3Out).
- The BGP and OSPF sessions are between the external routers (control and forwarding nodes) and the anchor leaf nodes (Leaf3 and Leaf4).

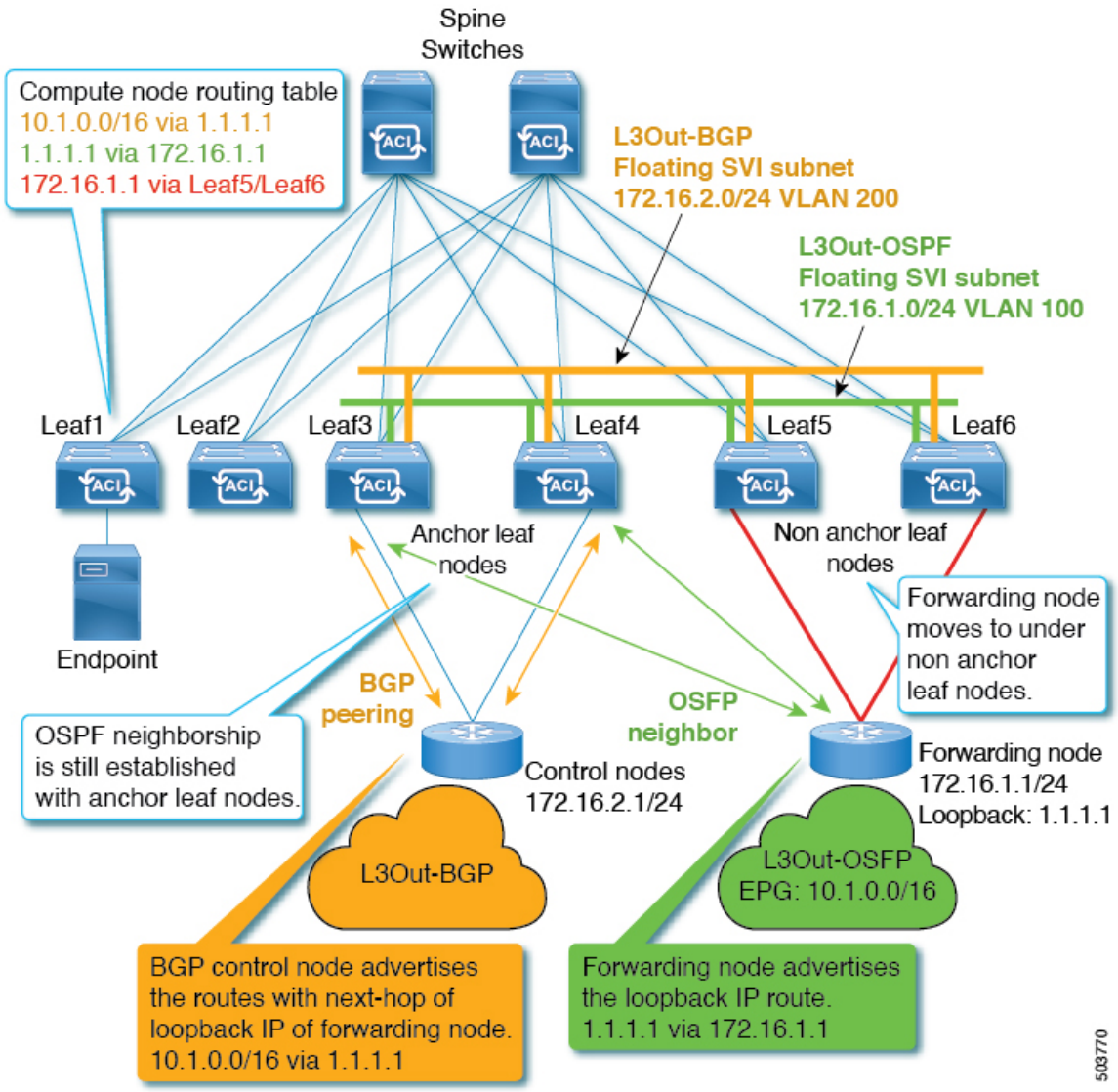
Figure 16: Example of separate external devices for Avoid Suboptimal Traffic with a loopback address (First Stage)



In the following Figure 17: Example of separate external devices for Avoid Suboptimal Traffic with a loopback address (Second Stage), on page 20, the next stage of the process is shown. At this stage of the process:

- The forwarding node has moved to the non-anchor leaf node pair (Leaf5 and Leaf6) using the floating SVI behavior.
- The BGP and OSPF protocol sessions are still established between the external routers and the anchor leaf nodes (Leaf3 and Leaf4) even after the move of the forwarding node.
- The compute leaf node is now pointing to the 172.16.1.1 address (next-hop to reach the loopback address 1.1.1.1) using the non-anchor leaf node pair (Leaf5 and Leaf6) where the forwarding node is connected.

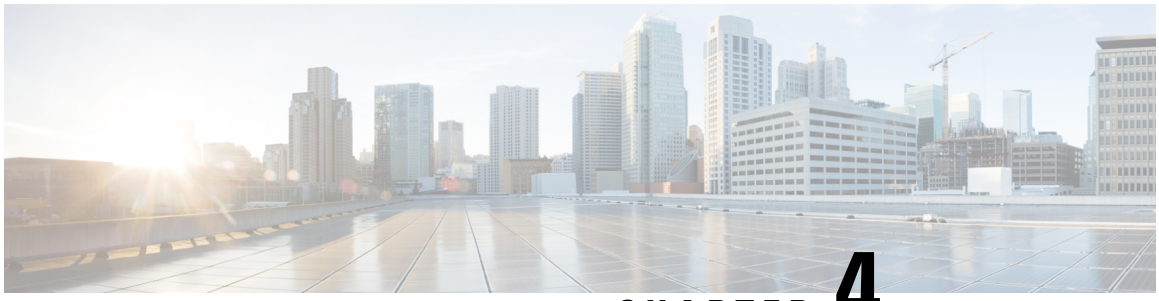
Figure 17: Example of separate external devices for Avoid Suboptimal Traffic with a loopback address (Second Stage)



This design has the following considerations:

- The forwarding node advertises the loopback IP address with the forwarding node IP (that is, the IP address of the forwarding node directly connected to the L3Out's SVI subnet) as the next-hop. Although OSPF is used in the example above, static route can be used to add the route for the loopback IP address on the anchor leaf nodes.
- The control nodes need to advertise the external prefix with the loopback IP address as the next-hop. Typically, eBGP is the option used for this.
- The control nodes can be connected to a floating L3Out or to a regular L3Out. You may use a regular L3Out instead of a floating L3Out if those control nodes are physical devices that are not moved around the fabric.

- If you want to use BGP and OSPF, you need to configure two different L3Outs (one for BGP and the other for OSPF) even if the same floating SVI subnet is used for BGP peering and OSPF neighbors. This is a general consideration for L3Outs, not specific to floating L3Outs.
- As of Cisco ACI release 6.0(1), in the case of Cisco ACI floating L3Out, only one IP address can be installed on the Cisco ACI leaf nodes as external next-hop to reach a prefix learned using BGP. This consideration is applicable to both IPv4 and IPv6. Leveraging ECMP for reaching the external prefix is possible by deploying the same loopback IP address on all the forwarder nodes. For example:
  - BGP running on the control nodes sends one primary next-hop for the external prefix to the anchor leaf nodes (for example, 10.1.0.0/16 via 1.1.1.1).
  - The anchor and non-anchor leaf nodes (for example, Leaf1, Leaf2, and Leaf3) redistribute into the fabric the direct routes that can be used to reach that loopback next-hop address (for example, 1.1.1.1 is reachable using 172.16.1.1, 172.16.1.2 and 172.16.1.3).
- When Next-hop propagation is enabled, the deployment of multiple control nodes advertising the same external prefixe with the same ECMP paths is affected by CSCwd28918. Please see the section Summary of Floating L3Out Deployment Considerations and Restrictions, on page 67 for detail.



CHAPTER 4

Configuring Floating L3Outs

- Configuring Floating L3Outs, on page 23
- Workflow for Configuring Floating L3Out, on page 24

Configuring Floating L3Outs

In this section we are going to cover the specific configuration steps required to leverage the Floating L3Out functionality. The different sections below will discuss Software and Hardware requirements, the deployment of Floating L3Outs with external devices part of a VMM or a physical domain, the specific configuration required to avoid sub-optimal traffic path toward the external destination and how to leverage multiple equal cost path for this type of communication.

Software Requirements

- Floating L3Out with VMM domain requires Cisco Application Policy Infrastructure Controller (APIC) Release 4.2(1) or later.
- Floating L3Out with physical domain requires Cisco APIC Release 5.0(1) or later.
- Avoiding a suboptimal path requires the following Cisco APIC Releases:
  - If the next-hop for the external prefix is the directly attached IP:
    - Next-hop propagation with BGP requires Cisco APIC Release 5.0(1) or later.
    - Next-hop propagation with OSPF and static route requires Cisco APIC Release 5.2(1) or later.
  - If the next-hop for the external prefixis not the directly attached IP but instead a loopback address (multi-protocol recursive route resolution), Cisco APIC Release 5.2(1) or later is required.

Hardware Requirements

Ensure that you have the correct leaf switches. Floating L3Out does not support the following top-of-rack switches:

- Cisco Nexus 9332PQ
- Cisco Nexus 9372PX
- Cisco Nexus 9372TX

- Cisco Nexus 9396PX
- Cisco Nexus 9396TX
- Cisco Nexus 93120TX
- Cisco Nexus 93128TX
- Cisco Nexus 9372PX-E
- Cisco Nexus 9372TX-E

Floating L3Out mandates the use of generation-2 leaf switches as anchor and non-anchor nodes. The generation-1 switches listed above cannot be configured as anchor or non-anchor nodes. However, you can use a generation-1 switch as a non-border leaf or a compute leaf switch where floating L3Out is not provisioned.

Workflow for Configuring Floating L3Out

This section provides a high-level description of the tasks that you must perform to configure a floating Layer 3 outside network connection (L3Out) when using a VMware vSwitch Distributed Switch (VDS) or a physical domain.

- Read and understand the guidelines and considerations.

  See the section Floating L3Out Consideration and Restrictions, on page 67.

- Fulfill all the prerequisites, which include meeting all software and hardware requirements.

  See the section Configuring Floating L3Outs, on page 23.

- Configure Domains.

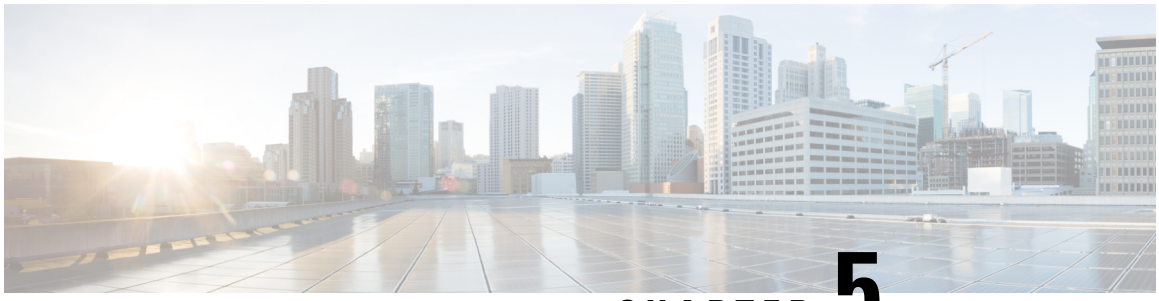
  See the section Configuring Domains Using the GUI, on page 25.

- Configure the floating L3Out.

  See the procedure Creating a Floating L3Out, on page 33.

- Verify that the L3Out configuration is correct.

  See the section Verifying the L3Out Configuration, on page 48.



CHAPTER 5

Configuring Domains Using the GUI

- Configuring Domains Using the GUI, on page 25
- Create a Layer 3 Domain Using the GUI, on page 27
- Configuring a VLAN Range for an Existing L3Out Domain VLAN Pool Using the GUI, on page 28
- Creating a Physical Domain Using the GUI, on page 28
- Configuring a VLAN Range For an Existing Physical Domain VLAN Pool Using the GUI, on page 29
- Create a VMM Domain for VMware VDS Using the GUI, on page 29
- Configuring a VLAN Range for an Existing VMM Domain VLAN Pool Using the GUI, on page 30

Configuring Domains Using the GUI

This section provides a high-level description of the tasks that you must perform to configure domains. This document doesn't cover how to configure interface access policies and AEP (Attachable Entity Profile). Please refer to other documents for this. The assumption is that you already have an AEP associated to your interface access policy for the interfaces where the external routers will be connected. The AEP needs to include the domains configured or modified in this section.

- Create a Layer 3 domain and associate a VLAN pool with a static VLAN range to it that you want to use for the floating L3Out.

  See the procedure Create a Layer 3 Domain Using the GUI, on page 27.

Note If you have already created a Layer 3 domain that you want to use, but it lacks an available static VLAN range for a floating L3Out, add the VLAN pool with the static VLAN range using the procedure Configuring a VLAN Range for an Existing L3Out Domain VLAN Pool Using the GUI, on page 28.

- Create a physical domain or a VMware VDS Virtual Machine Manager (VMM) domain and specify the VLAN pool that includes the static encap VLAN ranges that you want to use for the floating L3Out. The use of a physical domain or a VMM domain depends on the specific form factors of the external routers (deploying a mix of physical and virtual external routers is also possible).

  For creating a physical domain, see the procedure Creating a Physical Domain Using the GUI, on page 28.

  For creating a VMM domain, see the procedure Create a VMM Domain for VMware VDS Using the GUI, on page 29.

Configuring Domains Using the GUI

Note If you have already created a VMM domain or physical domain that you want to use, but it lacks an available static VLAN range for a floating L3Out, create the VLAN pool with the static VLAN range using the procedure Configuring a VLAN Range for an Existing VMM Domain VLAN Pool Using the GUI, on page 30.

Domain configuration has the following considerations:

- The VLAN pools associated to all the domains must include a static VLAN range. It is possible to define different VLAN pools for each domain, or a common VLAN pool used by all the domains. Also, the static VLAN range used for the L3Out domain must be the same as the VLAN pool of the VMM domain or physical domain. For example, both the VLAN range for the L3Out domain and the Virtual Machine Manager (VMM) or physical domain must include VLAN 200-209.
- Floating L3Out with physical domains requires Cisco Application Policy Infrastructure Controller (APIC) release 5.0(1) or later.
- If it is a physical domain, the floating IP address is deployed if the leaf port uses an AEP that, in addition to the physical domain, has an L3Out domain associated to the floating L3Out.

Figure 18: Use of an L3Out domain and a VMM domain for a floating L3Out, on page 26 below illustrates an example of an AEP that has an L3Out domain and a VMM domain for a floating L3Out.

Figure 18: Use of an L3Out domain and a VMM domain for a floating L3Out

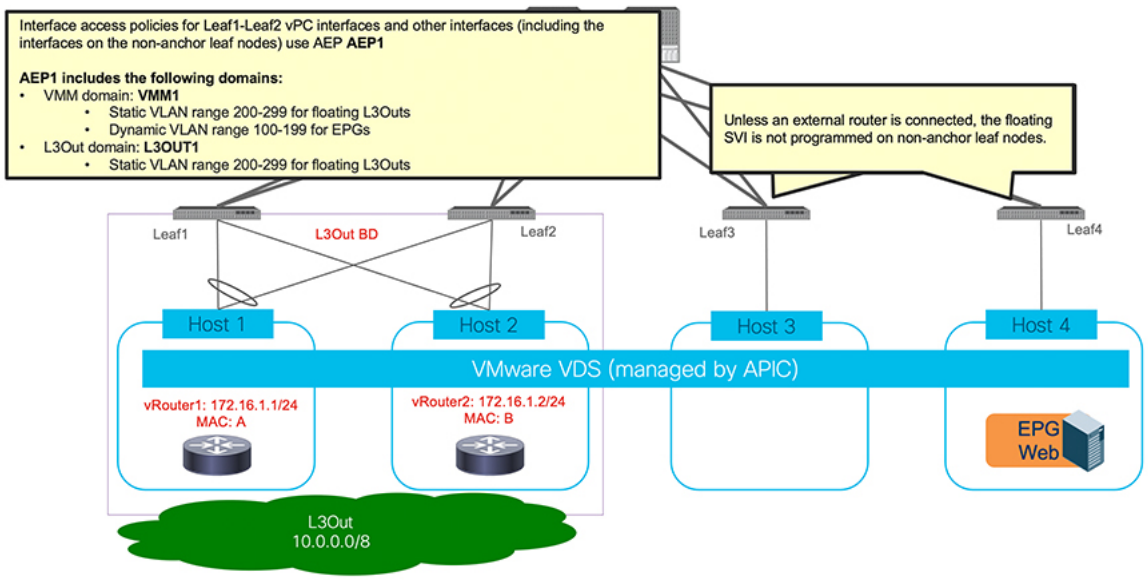
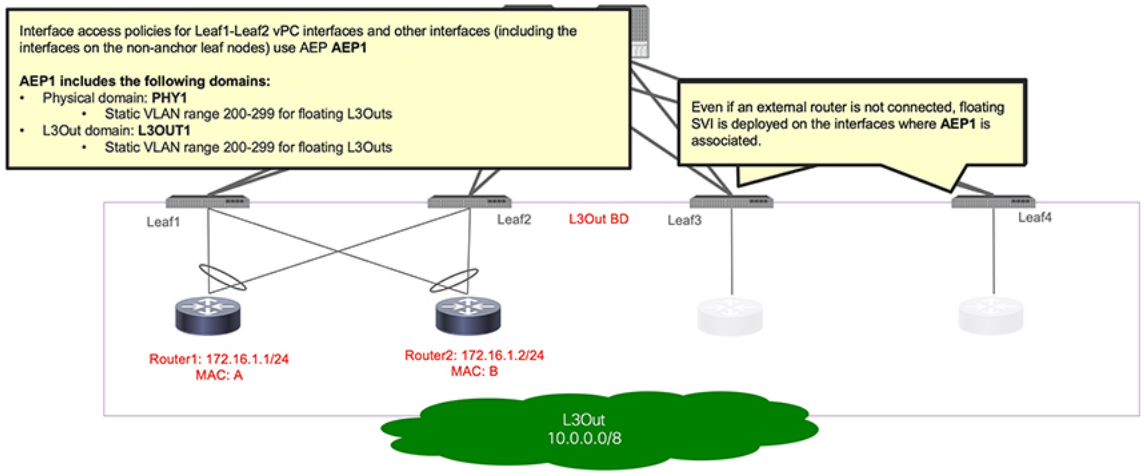


Figure 19: Use of an L3Out domain and a Physical domain for a floating L3Out, on page 27 below illustrates an example of an AEP that has an L3Out domain and a physical domain for a floating L3Out.

Figure 19: Use of an L3Out domain and a Physical domain for a floating L3Out



Create a Layer 3 Domain Using the GUI

Create a Layer 3 domain before you create the Layer 3 outside network connection (L3Out).

Step 1 Log in to Cisco Application Policy Infrastructure Controller (APIC).

Step 2 Go to **Fabric** > **Access Policies** > **.**

Step 3 In the **Policies** navigation pane, expand the **Physical and External Domains**, right-click the **L3 Domains** folder, and then click **Create L3 Domain**.

Step 4 In the **Create L3 Domain** dialog box, complete the following steps:

a) In the **Name** field, enter a name for the profile.

b) From the **Associated Attachable Entity Profile** drop-down list, create or choose an attachable entity profile (AEP).

If creating an attachable entity profile, enter the appropriate values in the **Create Attachable Entity Profile** dialog fields. You can click the **?** icon to view a description of each field in the online help file.

Note - For floating SVI deployment using a VMM domain, the AEP for the interfaces (anchor and non-anchor) must have both a Layer 3 domain and a VMM domain.

- For floating SVI deployment using a physical domain, the AEP for the anchor node interfaces must have both a Layer 3 domain and a physical domain.

c) From the **VLAN Pool** drop-down list, choose **Create VLAN Pool**.

d) In the **Create VLAN Pool** dialog box, in the **Name** field, enter a name for the VLAN pool.

e) In the **Allocation Mode** field, choose a mode.

f) In the **Encap Blocks** area, click the **+** (plus) icon.

g) In the **Create Ranges** dialog box, enter a range for the VLAN pool.

Note See the note about configuring the VLAN pool range at the beginning of this procedure.

h) In the **Allocation Mode** field, choose **Static Allocation**.

i) Click **OK**.

j) In the **Create VLAN Pool** dialog box, click **OK**.

Step 5 In the **Create L3 Domain** dialog box, click **Submit**.

Configuring a VLAN Range for an Existing L3Out Domain VLAN Pool Using the GUI

Use this procedure to configure the VLAN range for a floating Layer 3 outside network connection (L3Out) if you have already created the Layer 3 domain that you want to use. To use a floating L3Out, you must configure a VLAN pool for the Layer 3 domain that has the correct settings.

Before you begin

You must have created a Layer 3 domain. See the procedure

Step 1 Log in to Cisco Application Policy Infrastructure Controller (APIC).

Step 2 Go to **Fabric** > **Access Policies**.

Step 3 In the **Policies** navigation pane, expand the **Physical and External Domains** and the **L3 Domains** folder, and choose the Layer 3 domain.

Step 4 In the central **L3 Domain** work pane, from the **VLAN Pool** drop-down list, choose an existing VLAN pool.

Step 5 In the **Encap Blocks** area, click the **+** (plus) icon.

Step 6 In the **Create Ranges** dialog box, enter a range for the VLAN pool.

Step 7 In the **Allocation Mode** field, choose **Static Allocation**.

Step 8 Click **OK**.

Step 9 In the **Create VLAN Pool** dialog box, click **Submit**.

Step 10 In the central **Domain** work pane click **Submit**.

What to do next

Configuring a VLAN Range for an Existing L3Out Domain VLAN Pool Using the GUI, on page 28 or
Configuring a VLAN Range For an Existing Physical Domain VLAN Pool Using the GUI, on page 29

Creating a Physical Domain Using the GUI

This section explains how to create a physical domain using the Cisco Application Centric Infrastructure (ACI) GUI.

Step 1 From the menu bar click **Fabric** > **Access Policies**

Step 2 From the navigation bar, expand **Physical and External Domains**.

The **Physical Domains** folder appears in the navigation bar.

Step 3 From the navigation bar, right-click the **Physical Domains** folder and choose **Create Physical Domain**.

The **Create Physical Domain** dialog appears in the work pane.

Step 4 Enter the appropriate values in each field of the **Create Physical Domain** dialog.

Note Click the **?** icon to view a description of each field in the online help file.

Step 5 When finished, click **Submit**.

What to do next

Create a VMM Domain for VMware VDS Using the GUI, on page 29.

Configuring a VLAN Range For an Existing Physical Domain VLAN Pool Using the GUI

Use this procedure to configure the VLAN range for a physical domain. You must configure a VLAN pool for the domain that has the correct settings. As previously mentioned, it is also possible to reuse the VLAN pool previously created for the L3Out domain.

Step 1 Log in to Cisco Application Policy Infrastructure Controller (APIC).

Step 2 Go to **Fabric** > **Access Policies**.

Step 3 In the **Policies** navigation pane, expand the **Physical and External Domains** and the **Physical Domains** folder, and choose the domain.

Step 4 In the central **Physical Domain** work pane, from the **VLAN Pool** drop-down list, choose an existing VLAN pool.

Step 5 In the **Encap Blocks** area, click the **+** (plus) icon.

Step 6 In the **Create Ranges** dialog box, enter a range for the VLAN pool.

Step 7 In the **Allocation Mode** field, choose **Static Allocation**.

Step 8 Click **OK**.

Step 9 In the **Create VLAN Pool** dialog box, click **Submit**.

Step 10 In the central **Domain** work pane click **Submit**.

What to do next

Configuring a VLAN Range for an Existing VMM Domain VLAN Pool Using the GUI, on page 30.

Create a VMM Domain for VMware VDS Using the GUI

Step 1 Log in to Cisco Application Policy Infrastructure Controller (APIC).

Step 2 Go to **Virtual Networking** > **Inventory**.

Step 3 In the **Inventory** navigation pane, expand **VMM Domains**, right-click **VMware**, and then choose **Create vCenter Domain**.

Alternatively, in the **Inventory** navigation pane, you can choose **Quick Start** and in the central work pane choose **(VMware hypervisor) Create a vCenter Domain Profile**.

Step 4 In the **Create vCenter Domain** dialog box, complete the following steps:

a) In the **Virtual Switch Name** field, enter a name.
b) In the **Virtual Switch Area**, choose **VMware vSphere Distributed Switch**.
c) From the **Associated Attachable Entity Profile** (AEP) drop-down list, create a new AEP or choose a profile that you created earlier.

See "Create a Global Attachable Access Entity Profile" in the Cisco APIC Basic Configuration Guide for instructions.

d) From the VLAN Pool drop-down list, choose or create a VLAN pool.
e) In the **vCenter Credentials** area, click the + (plus) icon, and in the **Create vCenter credential** dialog box, do the following: Enter the VMware vCenter account profile name in the **Name** field, the VMware vCenter username in the **Username** field, enter and confirm the VMware vCenter password, and then click **OK**.
f) In the **vCenter** area, click the + (plus) icon, and in the **Create vCenter Controller** dialog box, do the following: Enter the VMWare vCenter controller name, the VMWare vCenter host name or IP address, the DVS version, data center name (which must match the data center name configured in VMware vCenter), select the credentials created in the previous step, and then click **OK**.
g) Fill out the remaining fields depending on your setup.
h) In the **Create vCenter Domain** dialog box, click **Submit**.

In the VMware work pane, you should see the newly created VMM domain, which is pushed to the VMware vCenter.

What to do next

Create a Layer 3 domain profile if you have not already done so. See the procedure Create a Layer 3 Domain Using the GUI, on page 27.

Configuring a VLAN Range for an Existing VMM Domain VLAN Pool Using the GUI

Use this procedure to configure the VLAN range for an existing floating Layer 3 outside network connection (L3Out) if you have already created the Virtual Machine Manager (VMM) domain that you want to use. To use a floating L3Out, you must configure a VLAN pool for the VMM domain that has the correct settings.

Before you begin

You must have created a VMM domain profile. See the procedure Create a VMM Domain for VMware VDS Using the GUI, on page 29.

Step 1 Log in to Cisco Application Policy Infrastructure Controller (APIC).

Step 2 Go to **Virtual Networking** > **Inventory**.

Step 3 In the **Inventory** navigation pane, expand the **VMM Domains** and **VMware** folders, and then choose the VMM domain.

Step 4 In the central **Domain** work pane, from the **VLAN Pool** drop-down list, choose an existing VLAN pool.

Step 5 In the **Encap Blocks** area, click the + (plus) icon.

Step 6 In the **Create Ranges** dialog box, enter a range for the VLAN pool.

Note See the note about configuring the VLAN pool range at the beginning of this procedure.

Step 7 In the **Allocation Mode** field, choose **Static Allocation**.

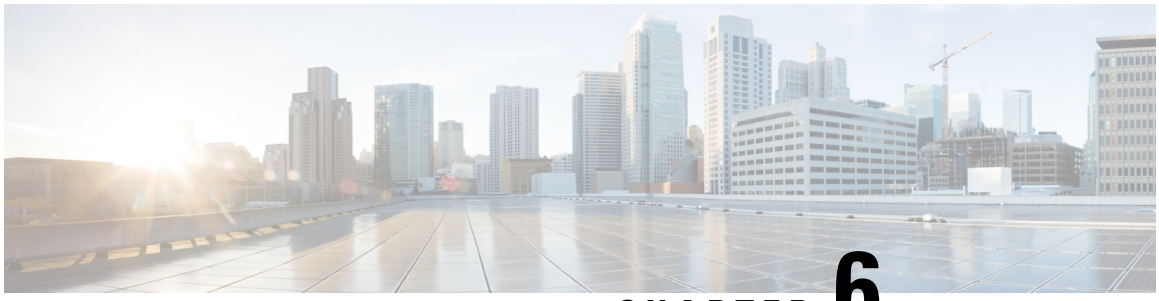
Step 8 Click **OK**.

Step 9 In the **Create VLAN Pool** dialog box, click **Submit**.

Step 10 In the central **Domain** work pane, click **Submit**.

What to do next

Create a Layer 3 domain if you have not done so already. See the section Create a Layer 3 Domain Using the GUI, on page 27.



CHAPTER 6

Creating a Floating L3Out

- Creating a Floating L3Out, on page 33
- Configuration Steps With VMM Domain, on page 35
- Configuring a Secondary IP, on page 36
- Configuration Example: Different L3Outs for OSPF and BGP Using GUI, on page 36

Creating a Floating L3Out

Follow the steps in this procedure to create a Layer 3 outside network connection (L3Out) with the floating L3Out feature. When configuring the Floating L3Out for a VMM domain, the configuration creates a port-group on the VMware VDS with the name of **<Tenant> | <L3Out name> | < VLAN Number >**. The administrator managing the virtual router would then attach the virtual router vNIC to this port-group.

Before you begin

You must have performed the following tasks before creating the L3Out.

- Configured the interface access policies, accessible attachable entity profile (AEP), and Layer 3 domain.
- Created the VMM domain or physical domain.

Configuration Steps with Physical Domain:

Step 1 Log in to the Cisco Application Policy Infrastructure Controller (APIC).

Step 2 Go to **Tenants** > ***your tenant.***

Step 3 In the tenant navigation pane, expand the **Networking** folder, right-click the **L3Outs** folder and choose **Create L3Out**.

Step 4 In the **Create L3Out** dialog box, **1. Identity** dialog box, complete the following steps.

a) In the **Name** field, enter a name for the L3Out.

b) From the **VRF** drop-down list, choose or create a virtual routing and forwarding (VRF) instance.

c) From the **L3 Domain** drop-down list, choose the Layer 3 domain that you created earlier.

d) On the right side of the dialog box, click one or more routing protocols, and accept the defaults or configure routing as appropriate for your setup.

You can choose **BGP** or **OSPF**.

e) Click **Next**.

Step 5 In the **Create L3Out** dialog box, **2. Nodes and Interfaces** dialog box, complete the following steps:

a) For the **Use Defaults** check box, leave it checked to accept the default interface and node policy names, or uncheck it to create custom names.

User can choose defaults or create custom names for the interface and node policies.

b) In the **Interface Types** field, choose **Floating SVI**.
c) From the **Domain Type** options, choose **Physical**.
d) From the **Domain** drop-down list, choose a domain.
e) In the **Floating Address** field, enter the floating IP address.

The floating IP address is the common IP address for non-anchor leaf nodes. It is used to locate the router if it is connected to any non-anchor top-of-rack switch through the data path.

f) In the **Encap** area, in the **Integer Value** field, enter the desired VLAN from the static VLAN range.
g) In the **MTU** field, enter the maximum transmission unit (MTU) of the external network.

The range is 1500 to 9216. The value inherit can be used. This will inherit the value configured for the Fabric L2 MTU policy. The default value is 9000. To inherit the value, enter "inherit" in the **MTU** field.

h) In the **Nodes** area, from the **Node ID** drop-down list, choose a node for the anchor leaf switch.
i) In the **Router ID** field, add the address for the router to be used for OSPF or BGP.
j) In the **Loopback Address** field, accept the default—which is the same as the router ID—or add a different loopback address.
k) In the **IP Address Primary** field, enter the primary IP address for the anchor leaf switch.

Note If the external router is connected behind virtual port channel (vPC) leaf anchor nodes, make sure to add a vPC peer leaf as the second anchor node.

l) (Optional) Click the + (plus sign) next to the **Loopback Address** field to add additional anchor leaf nodes.

Step 6 In the **Create L3Out** dialog box, **3. Protocols** dialog box, complete the following steps.

a) In the **Protocol Associations** area, configure items if necessary.

For example, if you chose the OSPF protocol, choose an OSPF policy:

If you chose the BGP protocol, the following configurations are available:

- **Peer Address**: Enter the peer IP address
- **EBGP Multihop TTL**: Enter the connection time to live (TTL). The range is from 1 to 255 hops; if zero, no TTL is specified. The default is 1.
- **Remote ASN**: Enter a number that uniquely identifies the neighbor autonomous system. The Autonomous System Number can be in 4-byte as plain format from 1 to 4294967295.

  Please refer Layer 3 Configuration guide for detail.

b) Click **Next**.

Step 7 In the **Create L3Out** dialog box, **4. External EPG** dialog box, complete the following steps.

a) In the **Name** field, enter a name for the external EPG.
b) From the Provided Contract drop-down list, choose or create a contract.
c) From the **Consumed Contract** drop-down list, choose or create a contract.
d) For the **Default EPG for all external networks** check box, leave it checked or uncheck it and add specific subnets.

e) Click **Finish**.

What to do next

Verify that the floating L3Out exists in Cisco APIC and that the port group exists on the VMware VDS. See the procedure Verifying the L3Out Configuration, on page 48.

Configuration Steps With VMM Domain

Step 1 Log in to the Cisco APIC.

Step 2 Go to **Tenants** > ***your tenant.***

Step 3 In the tenant navigation pane, expand the **Networking** folder, right-click the **L3Outs** folder and choose **Create L3Out or use an existing L3Out**.

Step 4 Select the folder **Logical Interface Profiles**, right click and select **Create Interface Profile**.

a) Assign it a name.
b) Select **Floating L3Out SVI** and then select +.
c) In the **Select Floating L3Out** dialog box, select the Anchor node(s) and assign the Primary IP Address. In the **IP Address Primary** field, enter the primary IP address for the anchor leaf switch.

Note: If the external router is connected behind virtual port channel (vPC) leaf anchor nodes, make sure to add a vPC peer leaf as the anchor node. In the case of the floating L3Out each vPC peer leaf must be entered individually as a separate anchor node.

d) Enter the **Path Attributes**, this is the configuration where you associate the Floating L3Out with the VMM domain.
e) Select the **Domain Type** as **Physical**.
f) Select the VMM domain.
g) Select the **Enhanced LAG Policy** for the port-group. This is the ELACP teaming configuration for the port-group on the virtualized host: if the virtualized host Vmnics are configured with ELACP and the **Path Attributes** configuration doesn't select the correct ELACP, traffic forwarding between the vRouter and the ACI leaf nodes won't work.
h) Enter the encap VLAN from the VLAN range (only a VLAN defined in the static range of the VMM domain VLAN pool is allowed).
i) In the **Floating Address** field, enter the floating IP address. The floating IP address is the common IP address for non-anchor leaf nodes. It is used to locate the router if it is connected to any non-anchor top-of-rack switch through the data path.

Step 5 In the **Create L3Out** dialog box, **4. External EPG** dialog box, complete the following steps:

a) In the **Name** field, enter a name for the external EPG.
b) From the **Provided Contract** drop-down list, choose or create a contract.
c) From the **Consumed Contract** drop-down list, choose or create a contract.
d) For the **Default EPG for all external networks** check box, leave it checked or uncheck it and add specific subnets.
e) Click **Finish**.

What to do next

Verify that the floating L3Out exists in Cisco APIC and that the port group exists on the VMware VDS. See the procedure Verifying the L3Out Configuration, on page 48.

Configuring a Secondary IP

This section explains how to create an optional secondary IP by creating a logical interface profile for floating SVI.

Step 1 From the navigation pane, go to **Tenants** > **tenant_name** > **Networking** > **L3Outs** > **L3Out_name** > **Logical Node Profiles** > **logical_node_profile_name** > **Logical Interface Profiles** > **logical_interface_profile_name**.

The logical interface profile screen appears in the work pane.

Step 2 From the work pane, click the **Floating SVI** tab.

Step 3 Double-click on an anchor leaf node.

The **Floating SVI** dialog appears.

Step 4 Locate the **IPv4 Secondary / IPv6 Additional Addresses** field and click the **+** to enable the **Address** and **IPv6 DAD** fields and enter the appropriate values.

Note
- Starting in Cisco Application Policy Infrastructure Controller (APIC) Release 5.0(1), **IPv6 DAD** and **ND RA Prefix** are disabled by default.
- Click the **?** icon to open the help file and view a description of each field.

Step 5 When finished, click **OK**.

Configuration Example: Different L3Outs for OSPF and BGP Using GUI

Use this procedure to configure OSPF and BGP for a floating Layer 3 outside network connection (L3Out) if you have already created the Layer 3 domain that you want to use. If both OSPF and BGP need to redistribute routes into the Cisco Application Centric Infrastructure (ACI) fabric, then you would have to configure OSPF and BGP under two different L3Outs, in this fashion:

Note Note that this scenario could apply to a standard L3Out as well as a floating L3Out.

Before you begin

You must have created a Layer 3 domain. See the procedure Create a Layer 3 Domain Using the GUI, on page 27.

Create the first L3Out with only the BGP protocol enabled (**l3out-bgp**), with the following settings:

Example

1. Under **l3out-bgp**, navigate to the **Select SVI** or **Select Floating SVI** page:

   Tenants > ***tenant_name*** > **Networking** > **L3Outs** > **l3out-bgp** > **Logical Node Profiles** > ***log_node_prof_name*** > **Logical Interface Profiles** > ***log_int_prof_name***, then **+** in the **SVI** or **Floating SVI** tab and configure the following:

   - In the **Encap** field, configure the VLAN settings
   - In the **Encap Scope** field, select **VRF** if the L3Out for BGP and the L3Out for OSPF use the same VLAN encapsulation ID for the floating SVI.

2. Configure the loopback under **l3out-bgp** (**Tenants** > ***tenant_name*** > **Networking** > **L3Outs** > **l3out-bgp** > **Logical Node Profiles** > ***log_node_prof_name*** > **Configured Nodes**).
3. Create a BGP peer between the **l3out-bgp** anchor leaf nodes and the external router. This example uses BGP peering between the loopback IP addresses. If loopback IP address is used for BGP peering, ensure the route to the loopback IP addresses are available on the anchor leaf nodes and external routers by using OSPF or static routes.
4. Configure the *import* route control profile at the BGP peer connectivity profile area:

   Tenants > ***tenant_name*** > **Networking** > **L3Outs** > ***BGP_L3Out*** > **Logical Node Profiles** > ***logical_node_profile_name*** > **Logical Interface Profiles** > ***logical_interface_profile_name*** > ***bgp_peer_connectivity_profile_name***

   Ensure that the external router's loopback IP addresses are not imported using this route map.
5. Configure the *export* route control profile at the BGP peer connectivity profile area:

   Tenants > ***tenant_name*** > **Networking** > **L3Outs** > ***BGP_L3Out*** > **Logical Node Profiles** > ***logical_node_profile_name*** > **Logical Interface Profiles** > ***logical_interface_profile_name*** > ***bgp_peer_connectivity_profile_name***

   This external route control profile should include all of the required routes to be exported out of the fabric, but should not export the **l3out-bgp**'s loopback IP address in this route map (either the **l3out-bgp**'s loopback IP address should not be part of the match rule, or there should be an explicit deny entry in the route map to deny the **l3out-bgp**'s loopback IP address).
6. Create the second L3Out with only the OSPF protocol enabled (**l3out-ospf**), with the following settings:
   a. Under **l3out-ospf**, navigate to the **Select SVI** or **Select Floating SVI** page:

      Tenants > ***tenant_name*** > **Networking** > **L3Outs** > **l3out-ospf** > **Logical Node Profiles** > ***log_node_prof_name*** > **Logical Interface Profiles** > ***log_int_prof_name***, then **+** in the **SVI** or **Floating SVI** tab

      and configure the following:

      - In the **Encap** field, configure the VLAN settings

- In the **Encap Scope** field, select **VRF** if the L3Out for BGP and the L3Out for OSPF use the same VLAN encapsulation ID for the floating SVI.

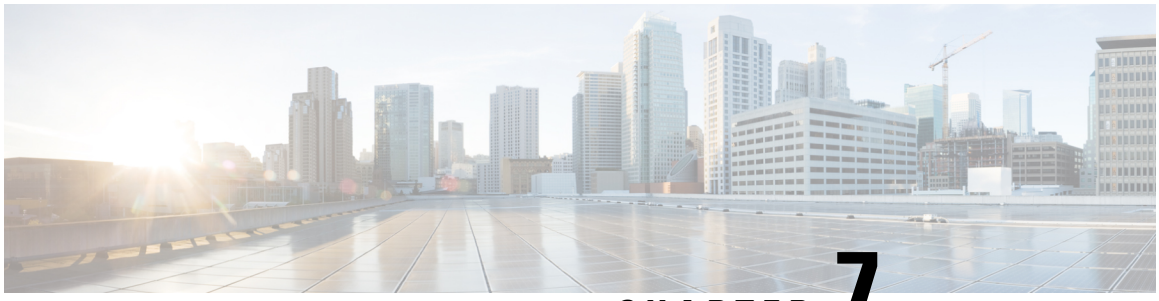
b. Create an export route map to export all of the required direct routes, including the loopback of **l3out-bgp** out of **l3out-ospf**:

Tenants > ***tenant_name*** > **Networking** > **L3Outs** > **l3out-ospf**, then right-click and choose **Create Route Map For Import and Export Route Control**, then select **default-export**

The match condition for this **default-export** should be all of the required routes and the **l3out-bgp**'s loopback IP addresses.

c. Unless the **Import Route Control Enforcement** is explicitly configured under the L3Out, the external node's loopback IP addresses are learned on the ACI fabric. If **Import Route Control Enforcement** is configured under the L3Out, the match condition for this default-import should have all the required routes and the external node's loopback IP addresses:

Tenants > ***tenant_name*** > **Networking** > **L3Outs** > **l3out-ospf** > **Route map for import and export route control** > **default-import** > **Contexts** > ***context_name*** > **Associated Matched Rules**



CHAPTER 7

Configuring the Avoidance of Suboptimal Traffic From an ACI Internal Endpoint to a Floating L3Out Using the Cisco APIC GUI

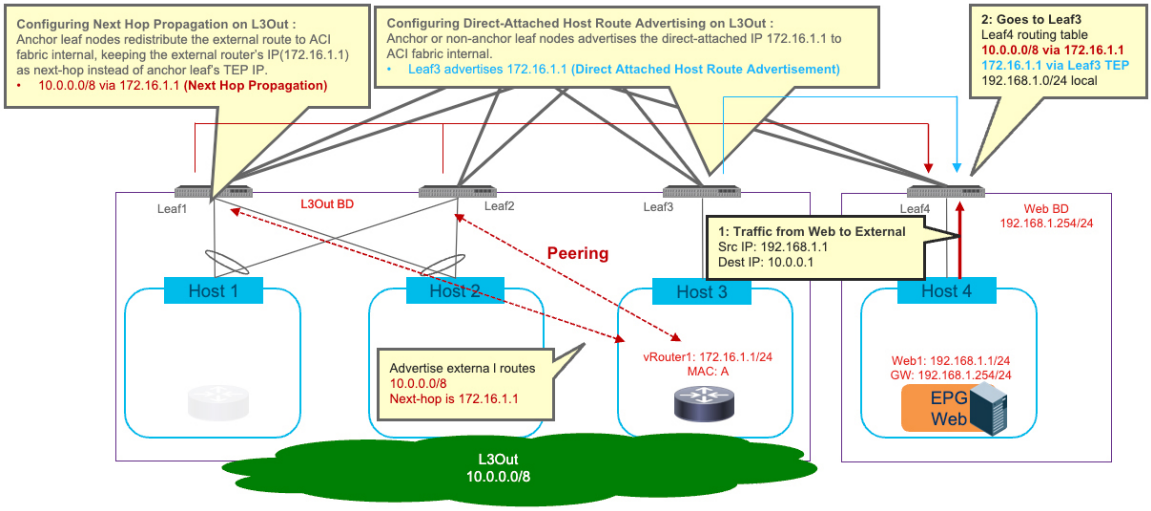
- Workflow for Configuring Floating L3Out with Avoidance of Suboptimal Traffic, on page 39
- Configuring Next Hop Propagation on L3Out, on page 42
- Configuring Direct-Attached Host Route Advertising on L3Out, on page 44
- Increasing the Local Maximum ECMP Paths in a BGP Address Family Context Policy, on page 45
- Configuring Next Hop Propagation and Multipath on an L3Out, on page 46
- Verifying the L3Out Configuration, on page 48

Workflow for Configuring Floating L3Out with Avoidance of Suboptimal Traffic

This section provides high-level configuration steps for a floating L3Out with avoidance of suboptimal traffic using topologies illustrated in Figure 20: Example of Configuration to Avoid Suboptimal Traffic Flow with directly attached next-hop, on page 40 and Figure 21: Example of Configuration to Avoid Suboptimal Traffic Flow with recursive lookup, on page 41 as examples. For other topology examples, please see Topology Examples with Avoidance of Suboptimal Traffic and ECMP, on page 73.

The figure below illustrates an example with directly attached next-hop, pointing out that even if the external prefix is learned from the external router on the anchor leaf, the outbound traffic flows originated from the internal EPG Web are sent directly to the leaf where the external router is physically connected.

Figure 20: Example of Configuration to Avoid Suboptimal Traffic Flow with directly attached next-hop



To achieve the optimal traffic forwarding behavior shown above, the following configuration steps are required:

1. Configuring Next Hop Propagation on L3Out: it enables the anchor leaf nodes to redistribute the external route with the next-hop IP address (that is the external device's IP address in this example) instead of the TEP address of the anchor leaf nodes. Therefore, the compute leaf nodes (Leaf4 in the example above) receives the external route with the external router's IP address as the next-hop (10.0.0.0/8 via 172.16.1.1 in this specific example).

   More detail is available in the section Configuring Next Hop Propagation on L3Out, on page 42.

2. Defining a match rule for a route-map to match the external prefix for which next hop propagation should be enabled (10.0.0.0/8 in the example above).

3. Defining a set rule for the same route-map to enable Next Hop Propagation for the external prefix.

4. Creating the route-map for route control referencing the match rule and the set rule to enable next hop propagation for the external prefix.

5. Applying the route-map to the L3Out. This is done in different ways, depending on the specific peering mechanism (BGP, OSPF or static routing):

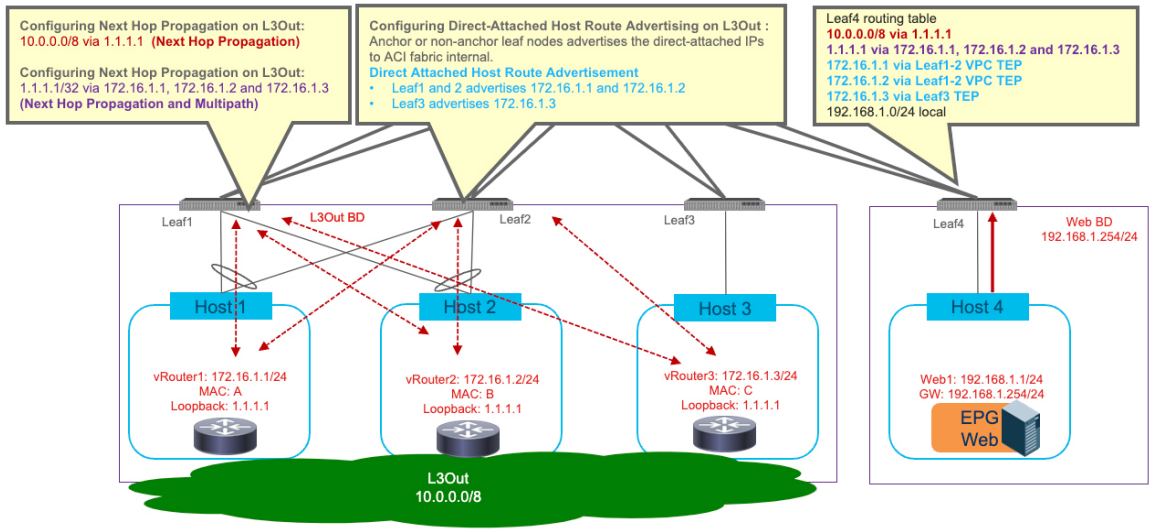
   - If the external prefix (10.0.0.0/8 in the example above) is learned via BGP: selecting the route-map at the Route Control Profile on the BGP Peer Connectivity profile of the L3Out where the external routes are exchanged via BGP.
   - If the external prefix (10.0.0.0/8 in the example above) is learned via OSPF: selecting the route-map at the Route Profile for Interleak on the L3Out where the external routes are exchanged via OSPF.
   - If the external prefix (10.0.0.0/8 in the example above) is known via static routing configuration: selecting the route-map at the Route Profile for Redistribution on the L3Out where the static route is configured.

6. Configuring Direct-Attached Host Route Advertising on L3Out: it enables the leaf nodes where the external routers are physically connected to redistribute the directly attached host routes to the Cisco Application Centric Infrastructure (ACI) fabric (172.16.1.1, 172.16.1.2, and 172.16.1.3 in the example above). See Configuring Direct-Attached Host Route Advertising on L3Out, on page 44.

- Creating a match rule for a route-map to specify the prefix of the L3Out's SVIs subnet where the external router is directly connected (172.16.1.0/24 in the example above).
- Creating the route-map for route control referencing the match rule to enable host route advertisement for the address of the router directly connected to the L3Out's SVIs subnet.
- Selecting the route-map at the Route Profile for Redistribution box on the L3Out where the external devices are connected.

If the next-hop for the external prefix is not directly attached to the L3Out's SVIs subnet (for example when using a loopback on the external router as next-hop), additional configurations are required to enable an additional recursive lookup. Figure 21: Example of Configuration to Avoid Suboptimal Traffic Flow with recursive lookup, on page 41 illustrates an example. See Increasing the Local Maximum ECMP Paths in a BGP Address Family Context Policy, on page 45.

Figure 21: Example of Configuration to Avoid Suboptimal Traffic Flow with recursive lookup



The following configuration steps are required to enable this functionality:

- Increase "Local Max ECMP" in BGP Address Family Context Policy: it enables the Cisco ACI fabric to increase Maximum Number of Paths when redistributing external routes from the anchor leaf nodes into the Fabric.
- Configure Next Hop Propagation and Multipath on L3Out:
  - Creating a match rule for a route-map to specify the prefix to enable next hop propagation and multipath (loopback address 1.1.1.1/32 in the example above).
  - Creating a set rule for a route-map to enable Next Hop Propagation and Multipath.
  - Creating the route-map for route control referencing the match rule and the set rule to enable next hop propagation and multipath for the specific loopback IP prefix.
  - If the next-hop address (1.1.1.1 in the example above) is learned via OSPF: selecting the route-map at the Route Profile for Interleak box on the L3Out where the external routes are exchanged via OSPF.

- If the next-hop address (1.1.1.1 in the example above) is learned via static route: selecting the route-map at the Route Profile for Redistribution box on the L3Out where the static route is configured.

The enablement of the functionality required to avoid sub optimal flow has the following considerations:

- Cisco ACI release 5.0(1) or later is required.
- If the next-hop for the external prefix is not directly attached to the L3Out's SVIs subnet but require recursive lookup, Cisco ACI release 5.2(1) or later is required.
- For Next Hop Propagation to work, the floating L3Out must be in a physical domain, not in a VMM domain.
- If both OSPF and BGP need to redistribute routes into the ACI fabric, then you would have to configure OSPF and BGP under two different L3Outs. For example, the OSPF L3Out is required for the fabric to learn the next-hop loopback address, whereas the BGP L3Out is used to receive the external prefixes reachable via the next-hop address.

Configuring Next Hop Propagation on L3Out

This section explains how to create match and set rules for a route map, create a route map, configure route control profiles at a BGP peer connectivity profile, and configure a route profile for redistribution. The configuration steps described below refer the network topology and example shown in Figure 20: Example of Configuration to Avoid Suboptimal Traffic Flow with directly attached next-hop, on page 40.

Before you begin

The following must be configured:

- Floating L3Out with a physical domain (For Next Hop Propagation, the floating L3Out must be in a physical domain, not in a VMM domain.)
- A BD, EPG, and a contract between the EPG and L3Out EPG

Step 1 To create a match rule for a route map:

a) From the navigation pane, go to **Tenants** > **tenant_name** > **Policies** > **Protocol**.

b) Right-click on **Match Rules** and choose **Create Match Rules for Route Map**.

The **Create Match Rule** dialog appears in the work pane.

c) Enter a name in the **Name** field

d) Locate the **Match Prefix** summary table and click the **+** to access the **IP**, **Description**, **Aggregate**, **Greater Than Mask** and **Less Than Mask** fields and enter the appropriate values.

Note
- The IP subnet should include the subnet that the external router advertises (in our example: 10.0.0.0/8).
- Click the **?** icon to open the help file and view a description of each field.

e) When finished, click **Submit**.

Step 2 To create a set rule for a route map:

a) From the navigation pane, go to **Tenants** > **tenant_name** > **Policies** > **Protocol**.
b) Right-click on **Set Rules** and choose **Create Set Rules for Route Map**.

The **Create Set Rules for Route Map** dialog appears in the work pane.

c) Enter a name in the **Name** field.
d) Click to place a check in the **Next Hop Propagation** checkbox.
e) When finished, click **Finish**.

Step 3 To create a route map for route control:

a) From the navigation pane, go to **Tenants** > **tenant_name** > **Policies** > **Protocol**.
b) Right-click on **Route Maps for Route Control** and choose **Create Route Maps for Route Control**.

The **Create Route Maps for Route Control** dialog appears in the work pane.

c) Enter a name in the **Name** field.
d) From the **Contexts** summary table in the **Create Route Maps for Route Control** dialog, click the **+** to access the **Create Route Control Context** dialog.
e) From the **Create Route Control Context** dialog, click the **Associated Match Rules +** symbol to access the **Rule Name** field and choose the **Match Rule** created in Step 1.
f) From the **Create Route Control Context** dialog, click the **Set Rule** drop-down menu and choose the **Set Rule** created in Step 2.
g) When finished, click **Submit**.

If the next-hop for the external prefix is learned via BGP, go Step 4. If the next-hop for the external prefix is learned via OSPF, go Step 5. If the next-hop for the external prefix is learned via static route, go Step 6.

Step 4 To configure a route control profile in the BGP peer connectivity profile:

Note The next hop propagation policy must be applied on the BGP peer connection policy in the L3Out for BGP.

a) From the navigation pane, go to **Tenants** > **tenant_name** > **Networking** > **L3Outs** > **L3Out_name** > **Logical Node Profiles** > **logical_node_profile_name** > **Logical Interface Profiles** > **logical_interface_profile_name** > **bgp_peer_connectivity_profile_name**

The **BGP Peer Connectivity Profile** properties appear in the work pane.

b) From the **Route Control Profile** option in the **BGP Peer Connectivity Profile** window, click the **+**.

The **Name** and **Direction** options are enabled.

c) Click the **Name** drop-down menu to specify the route map created in Step 3.
d) Click the **Direction** drop-down menu and choose **Route Import Policy**.
e) When finished, click **Update**.

Step 5 To configure a Route Profile for Interleak on the L3Out where the OSPF route is learned.

a) From the navigation pane, go to **Tenants>tenant name>Networking >L3Outs>L3Out name.**
b) Click the **Policy** tab, then click the **Main** subtab.
c) Locate the **Route Profile for Interleak** field and choose the **route-map** from step 3.
d) Click **Submit**.

Step 6 To configure a Route Profile for Redistribution on the L3Out where the static route is configured.

a) From the navigation pane, go to **Tenants> tenant name>NetworkingL3Outs>L3Out name.**
b) Click the **Policy** tab, then click the **Main** subtab.

c) From the **Route Profile for Redistribution** summary table, click the +. The **Source** and **Route Map** options are enabled.
d) Click the **Source** drop-down menu to specify static.
e) Click the **Route Map** drop-down menu to specify the route map from Step 3.
f) When finished, click **Update**.

What to do next

Configuring Direct-Attached Host Route Advertising on L3Out, on page 44

Configuring Direct-Attached Host Route Advertising on L3Out

This section explains how to create match and set rules for a route map, create a route map, configure route control profile for redistribution in order to advertise inside the fabric the specific IP addresses defined on the external routers for their connection to the SVIs L3Out network. The configuration steps described below refer to the network topology and example shown in Figure 20: Example of Configuration to Avoid Suboptimal Traffic Flow with directly attached next-hop, on page 40.

Before you begin

The following must be configured:

- Floating L3Out with a physical domain (For Direct-Attached Host route advertisement, the floating L3Out must be in a physical domain, not in a VMM domain.)
- A BD, EPG, and a contract between the EPG and L3Out EPG

Step 1 To create a match rule for a route map:

a) From the navigation pane, go to **Tenants** > **tenant_name** > **Policies** > **Protocol**.
b) Right-click on **Match Rules** and choose **Create Match Rules for Route Map**.

The **Create Match Rule** dialog appears in the work pane.

c) Enter a name in the **Name** field.
d) Locate the **Match Prefix** summary table and click the + to access the **IP**, **Description**, **Aggregate**, **Greater Than Mask** and **Less Than Mask** fields and enter the appropriate values to configure the external router IPs. In our example it is IP address 172.16.1.1.

Note Click the **?** icon to open the help file and view a description of each field.

e) When finished, click **Submit**.

Step 2 To create a route map for route control:

a) From the navigation pane, go to **Tenants** > **tenant_name** > **Policies** > **Protocol**.
b) Right-click on **Route Maps for Route Control** and choose **Create Route Maps for Route Control**.

The **Create Route Maps for Route Control** dialog appears in the work pane.

c) Enter a name in the **Name** field.

d) From the **Contexts** summary table in the **Create Route Maps for Route Control** dialog, click the **+** to access the **Create Route Control Context** dialog.
e) From the he **Create Route Control Context** dialog, click the **Associated Match Rules +** symbol to access the **Rule Name** field and choose the **Match Rule** created in Step 1.
f) When finished, click **Submit**.

Step 3 To configure a route control profile for redistribution on an L3Out:

a) From the navigation pane, go to **Tenants** > **tenant_name** > **Networking** > **L3Outs** > **l3Out_name**.

The **l3_outside_name** window appears in the work pane.

b) From the **Route Profile for Redistribution** summary table, click the **+**.

The **Source** and **Route Map** options are enabled.

c) Click the **Source** drop-down menu to specify attached-host
d) Click the **Route Map** drop-down menu to specify the route-map from Step 2.
e) When finished, click **Update**.

Increasing the Local Maximum ECMP Paths in a BGP Address Family Context Policy

This procedure configures the maximum number of paths for the redistributed routes in the Cisco Application Centric Infrastructure (ACI) fabric. Perform this procedure if the next-hop for the external prefix is not a directly attached host, but requires recursive lookup. The configuration steps described below refer to the network topology and example shown in Figure 21: Example of Configuration to Avoid Suboptimal Traffic Flow with recursive lookup, on page 41.

Step 1 Configure the maximum number of paths for the redistribution of routes in the Cisco ACI fabric. This is a necessary step before you can configure the **Multipath** field. See step 3 of Configuring Next Hop Propagation and Multipath on an L3Out, on page 46.

Step 2 Navigate to **Tenants > tenant name > Policies > Protocol> BGP > BGP Adress Family Context**.

Step 3 In the **Create BGP Address Family Context Policy** dialog box, perform the following steps:

a) In the **Name** field, enter a name for the policy.

For example, enter **`redistr-mpath`**.

b) Locate the **Local Max ECMP** field and enter the maximum number of paths (ECMP next hops) that should be selected when redistributing static or OSPF protocol routes learned on the border leaf switch into the MP-BGP fabric.

The default value for this field is 0, which indicates that the **Local Max ECMP** setting is disabled. Enter a value greater than 0 to enable this setting. The valid range is from 1 to 16.

Note Do not select the **Enable Host Route Leak** option in this scenario. This is not supported when configuring multi-protocol recursive next hop propagation.

c) Click **Submit** after you have updated your entries.

Step 4 Navigate to **Tenants** > ***tenant name*** > **Networking** > **VRFs** > ***vrf name***.

Step 5 Review the configuration details of the subject VRF.

Step 6 Locate the **BGP Context Per Address Family** field and, in the **BGP Address Family Type** area, select either **IPv4 unicast address family** or **IPv6 unicast address family**.

Step 7 Associate the BGP Address Family Context you created in the **BGP Address Family Context** drop-down list and associate it with the subject VRF.

Step 8 Click **Submit.**

Configuring Next Hop Propagation and Multipath on an L3Out

This procedure creates match and set rules for a route map, creates a route map, and configures route control profiles for the L3Out (using OSPF or static routing) where the next hop for the external prefix is learned (or reachable). This step is required if the next-hop for the external prefix is not directly attached host but requires recursive lookup. The configuration steps described below refer the network topology and example shown in Figure 21: Example of Configuration to Avoid Suboptimal Traffic Flow with recursive lookup, on page 41.

Before you begin

Configure the following things:

- Floating L3Out with a physical domain. For next hop propagation, the floating L3Out must be in a physical domain, not in a VMM domain.
- If both OSPF and BGP must redistribute routes into the Cisco Application Centric Infrastructure (ACI) fabric, then configure OSPF and BGP under two different L3Outs.
- A BD, EPG, and a contract between the EPG and the L3Out EPG for the external prefixes.

Step 1 To create a match rule for a route map:

a) From the navigation pane, go to **Tenants >*tenant_name* > Policies > Protocol**.
b) Right-click **Match Rules** and choose **Create Match Rules for Route Map**.

The **Create Match Rule** dialog appears in the work pane.

c) Enter a name in the **Name** field.
d) In the **Match Prefix** summary table and click **+** to access the **IP, Description, Aggregate, Greater Than Mask** and **Less Than Mask** fields and enter values to match the external router's loopback IP address that is utilized as next-hop to reach the external prefixes.
e) Click **Submit**.

Step 2 To create a set rule for a route map:

a) From the navigation pane, go to **Tenants >*tenant_name* > Policies > Protocol**.
b) Right-click **Set Rules** and choose **Create Set Rules for Route Map**.

The **Create Set Rules for Route Map** dialog appears in the work pane.

c) Enter a name in the **Name** field.
d) Click to place a check in the **Next Hop Propagation** and **Multipath** checkbox.

- **Next Hop Propagation**: Select this option to propagate the next hop address advertised by the external BGP peer into the fabric. If you do not enable this option, the tunnel endpoint (TEP) of the border leaf switch is used as the next hop on the other leaf switches.
- **Multipath**: Select this option to specify if multiple paths (ECMP next hops) need to be picked for redistribution for a particular route when performing a next hop unchanged redistribution. The number of paths used is based on the value that you entered in the **Local Max ECMP** field configured at "Increase Local Max ECMP in BGP Address Family Context Policy".

e) When finished, click **Finish**.

Step 3 To create a route map for route control:

a) From the navigation pane, go to **Tenants >*tenant_name* > Policies > Protocol**.
b) Right-click on **Route Maps for Route Control** and choose **Create Route Maps for Route Control**.

The **Create Route Map for Route Control** dialog appears in the work pane.

c) Enter a name in the **Name** field.
d) From the **Contexts** summary table in the **Create Route Map for Route Control** dialog, click **+** to access the **Create Route Control Context** dialog.
e) From the **Create Route Control Context** dialog, click the **Associate Map Rules+** symbol to access the **Rule Name** field and choose the **Match Rule** created in Step 1.
f) From the **Create Route Control Context** dialog, click the **Set Rule** drop-down menu and choose the **Set Rule** created in Step 2.
g) When finished, click **Submit**.

If the next hop for the external prefix is learned from OSPF, go to Step 4. If the next hop for the external prefix is learned from a static route, go to Step 5.

Step 4 To configure a route profile for interleak on the L3Out where the OSPF route is learned:

a) From the Navigation pane, choose **Tenants >*tenant_name*** > **Networking** > **L3Outs** > ***L3Out name***.
b) Click the **Policy** tab, then click the **Main** subtab.
c) Locate the **Route Profile for Interleak** field and choose the route map from Step 3.
d) Click **Submit**.

Step 5 To configure a route profile for redistribution on the L3Out where the static route is configured:

a) From the Navigation pane, choose **Tenants >*tenant_name*** > **Networking** > **L3Outs** > ***L3Out name***.
b) Click the **Policy** tab, then click the **Main** subtab.
c) From the **Route Profile for Redistribution** summary table, click the **+**.

The **Source** and **Route Map** options are enabled.

d) Click the **Source** drop-down menu to specify **static**.
e) Click the **Route Map** drop-down menu to specify the route map from Step 3.
f) When finished, click **Update**.

Verifying the L3Out Configuration

After you configure the floating Layer 3 outside network connection (L3Out), verify the creation of the port group in the VMware vCenter and the leaf node configuration on the Cisco Application Policy Infrastructure Controller (APIC).

Verify the Floating L3Out Port Group in the VMware vCenter

Verify that the port group has been generated for the Layer 3 outside connection (L3Out) in the VMware vCenter if it is a floating L3Out with a VMM domain.

Before you begin

You must have configured a floating L3Out with a VMM domain.

Step 1 Log into the VMware vCenter.

Step 2 Navigate to data center and the VMware VDS, and then expand the VMware VDS to view the port-groups.

Step 3 In the left navigation pane, find the port group that has been generated for the L3Out.

The port group has a name in the following format: Tenant_name|L3Out_name|VLAN-number.

For example, if your tenant name is Floating, your L3Out name is ExtConnect1, and the VLAN number is 205, the port group name is Floating|ExtConnect1|205.

Step 4 Under the **Summary** tab, ensure that the VLAN ID is the same as the last number of the VLAN range and that the information in the **Distributed Port Group Details** is correct.

What to do next

Verify Floating L3Out on the Leaf Nodes Using the Cisco APIC GUI, on page 48

Verify Floating L3Out on the Leaf Nodes Using the Cisco APIC GUI

Verify that the leaf nodes have the correct IP addresses.

Before you begin

You must have configured a floating L3Out.

Step 1 Log in to Cisco Application Policy Infrastructure Controller (APIC).

Step 2 Go to **Fabric** > **Inventory**.

Step 3 In the **Inventory** navigation pane, expand the ***pod***, the ***anchor_leaf_node***, the **Interfaces** folder, and the **External SVI Interfaces** folder.

Step 4 Click the interface for the floating L3Out.

The name should have the format vlan-*VLAN_ID*. For example, if the VLAN that you configured for the Layer 3 and VMM domains is 205, the interface is named `vlan-205`.

Step 5 In the **Routed Vlan Interface** central work pane, ensure that the IP address is the primary IP address and that the interface is up.

Step 6 Note the **Interface** VLAN ID, which belongs to the actual VLAN on the leaf switch.

The VLAN ID appears at the top of and in the **Properties** list of the central work pane.

Step 7 In the In the **Inventory** navigation pane, under the **Interfaces** folder, choose the **Physical Interfaces** folder.

Step 8 In the **Interfaces** central work pane, choose the Physical Interfaces tab.

Step 9 Choose your physical interfaces—for example, `eth 1/8` and `eth 1/9`—expand the **Oper Vlans** column, and ensure that the VLAN ID that you noted is in the list.

Step 10 Repeat this procedure for the non-anchor leaf nodes, choosing the non-anchor leaf node in step 3.

CHAPTER 8

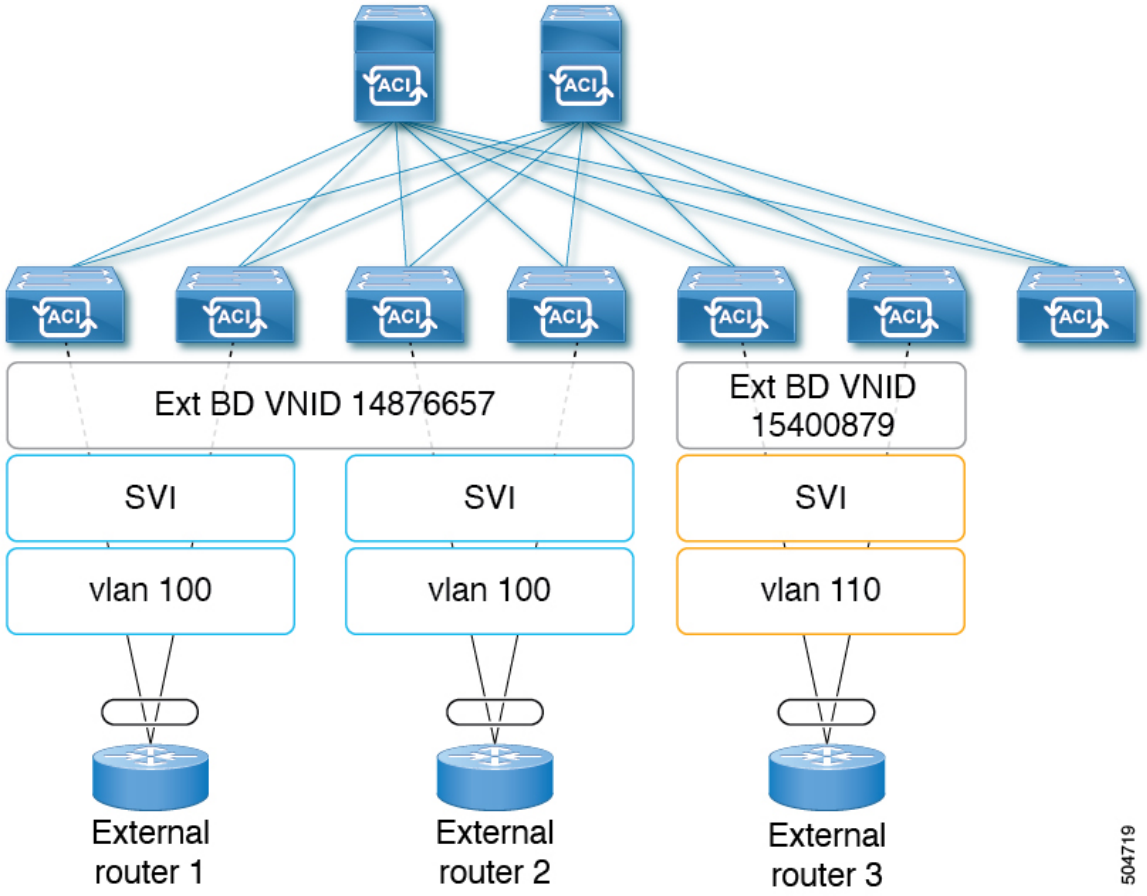
Support for Multiple Encapsulation for L3Outs With SVI

- Support for Multiple Encapsulation for L3Outs with SVI, on page 51
- Single Floating SVI With Different VLAN Encapsulations on Non-Anchor Nodes, on page 53
- Guidelines and Limitations of Multiple Encapsulation for L3Outs with SVI, on page 57
- Configuring Multiple Encapsulation for L3Outs With SVI and Floating SVIs Using the GUI, on page 58
- Configuring Multiple Encapsulation for L3Outs With SVI Using the CLI, on page 63
- Configuring Multiple Encapsulation for L3Outs With SVI Using the REST API, on page 64

Support for Multiple Encapsulation for L3Outs with SVI

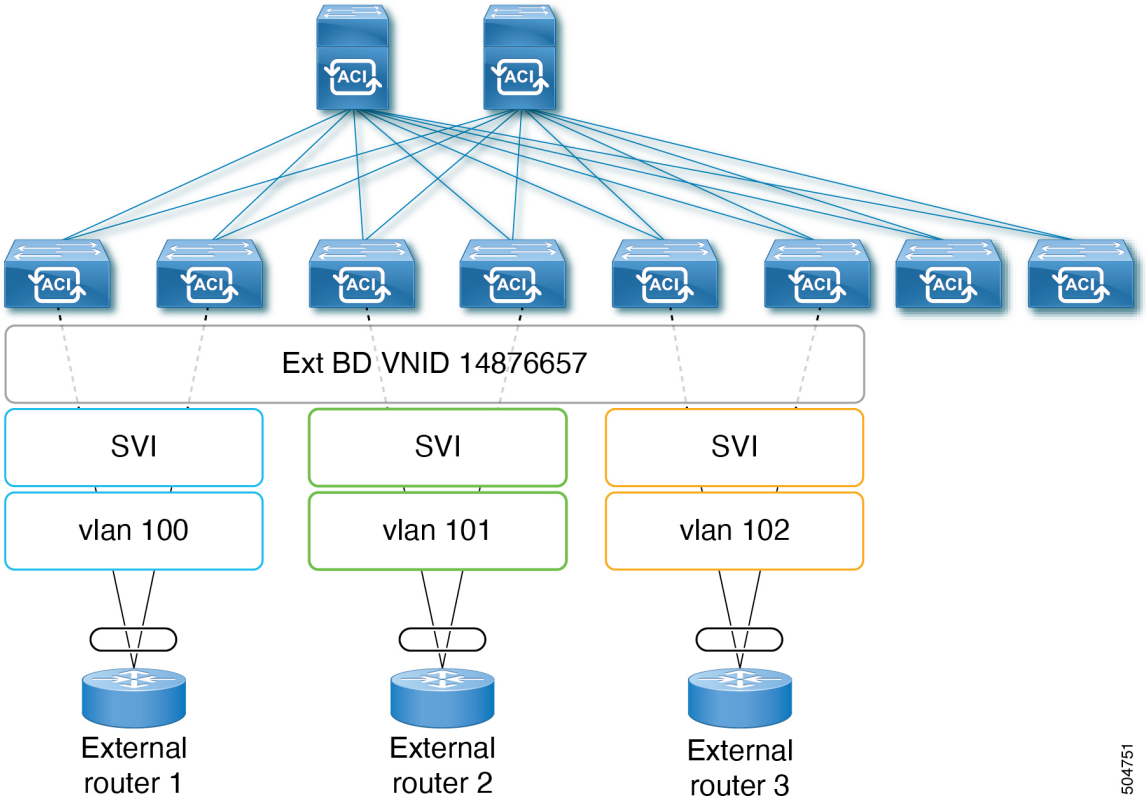
When an L3Out is configured with SVI interfaces on different leaf switches using the same encapsulation VLAN, the SVI VLAN will be mapped to the same VXLAN network identifier (VNID). This forms a single bridge domain (external bridge domain) and broadcast domain across the fabric. An SVI interface configured with a different VLAN will form a separate external bridge domain as illustrated in the diagram below. Prior to Cisco Application Centric Infrastructure (ACI) release 5.2(3) it was not possible to create a single external bridge domain with different encapsulation VLANs on different switches.

Figure 22: Separate VNID Associated to External Bridge Domains with Different Encapsulation (pre-ACI 5.2(3) Releases)



Cisco ACI Release 5.2(3) added support for configuring a single external bridge that can be configured with different encapsulation VLANs on different leaf switches. The multiple encapsulation support feature uses the floating SVI object to define the external bridge domain for floating L3Outs or an external bridge group profile for defining the external bridge domain for regular L3Outs. The use case for this feature may be where the same VLAN cannot be used on different leaf switches because it may already be in use.

Figure 23: Separate VNID Associated to External Bridge Domains with Different Encapsulation (pre-ACI 5.2(3) Releases)

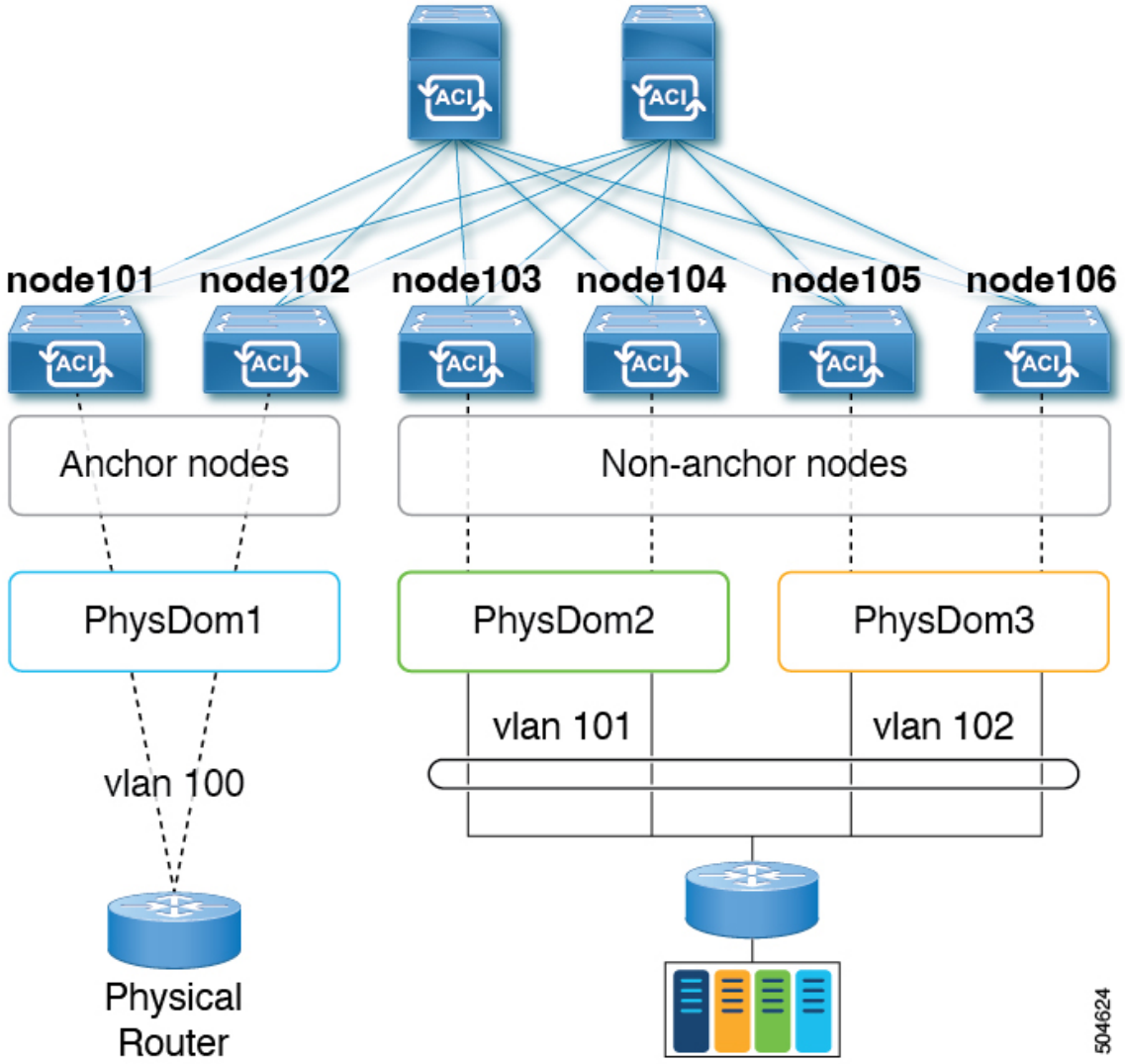


As of Cisco ACI release 6.0(1), this feature is supported for physical domain L3Outs only, not for VMM domain L3Outs.

Single Floating SVI With Different VLAN Encapsulations on Non-Anchor Nodes

The following figure shows a configuration with a single floating SVI using multiple physical domains and different VLAN encapsulations.

Figure 24: Single Floating SVI using Multiple Physical Domains and Different VLAN Encapsulations



For this use case:

- Leaf switches node101 and node102 are the anchor nodes, and leaf switches node103, node104, node105, and node106 are the non-anchor nodes for floating SVI vlif-100.
- The following leaf switches needs to use the same VLAN encapsulation because they are vPC pair:
  - node101 and node102
- The use of different physical domains is required to provision different VLAN encapsulations for the same floating SVI on different sets of leaf switches. It's because the VLANs are provisioned based on physical domains configured on path attributes for different sets of leaf switches. In this example, three physical domains are required:
  - physDom1: VLAN 100 for leaf node 101 and node 102.
  - physDom2: VLAN 101 for leaf node 103 and node 104.

- physDom3: VLAN 102 for leaf node 105 and node 106.

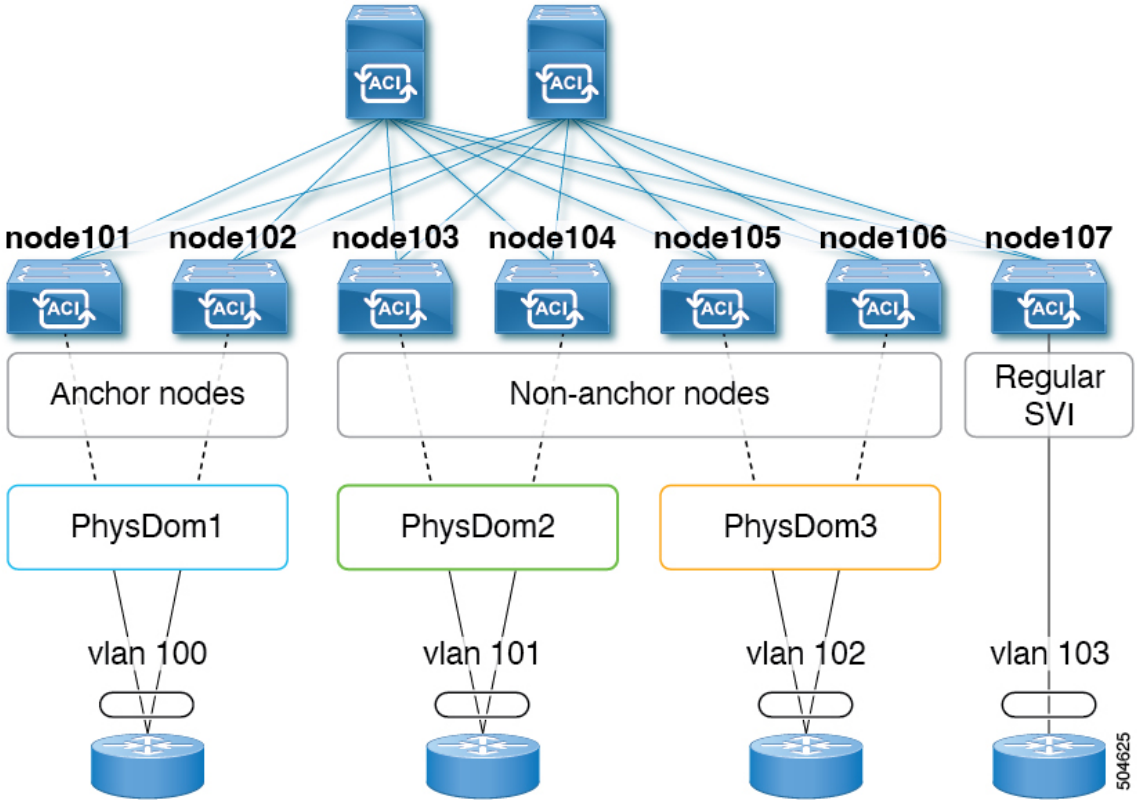
To configure the use case shown above:

1. Create a physical domain **physDom1** with an attachable entity profile (AEP) **anchor-nodes** associated with leaf switches **node101** and **node102** as anchor nodes.
2. Create additional physical domains for each of the required access encapsulation sets.
3. Create a physical domain **physDom2** with an AEP **floating-set1** associated with leaf switches **node103** and **node104**.
4. Create a physical domain **physDom3** with an AEP **floating-set2** associated with leaf switches **node105** and **node106**.
5. Create a floating SVI **vlif-100** with encapsulation **vlan100** with **node101** as the anchor node.
6. Create a floating SVI **vlif-100** with encapsulation **vlan100** with **node102** as the anchor node.
7. Add physical domain path attributes to the floating SVI:
   - Physical domain **physDom2** added with access encapsulation **vlan101**
   - Physical domain **physDom3** added with access encapsulation **vlan102**

Note Anchor leaf nodes will use the VLAN encapsulation configured for the floating SVI. Thus, anchor leaf nodes using physical domain **physDom1** vlan100 that is the same as the floating SVI vlif-100.

The following figure shows a configuration where a floating SVI and a regular SVI are grouped together with different VLAN encapsulations. For regular L3Outs, a new object called external bridge group profile needs to be configured to group SVIs under an L3Out with different VLANs together to be part of the same external bridge domain.

Figure 25: Different VLAN encapsulations with Floating and Regular SVIs



For this use case:

- Leaf switches node101 and node102 are the anchor nodes, and leaf switches node103, node104, node105, and node106 are the non-anchor nodes, for floating SVI **vlif-100**
- Leaf switch node107 has regular SVI **vlif-103**
- The following leaf switch pairs need to use the same VLAN encapsulation because they are VPC pairs:
  - node101 and node102
  - node103 and node104
  - node105 and node106
- The use of different physical domains is required to provision different VLAN encapsulations for the same floating SVI on different sets of leaf switches. It's because the VLANs are provisioned based on physical domains configured on path attribute for different sets of leaf switches. In this example, three physical domains are required:
  - physDom1: VLAN 100 for leaf node 101 and node 102.
  - physDom2: VLAN 101 for leaf node 103 and node 104.
  - physDom3: VLAN 102 for leaf node 105 and node 106.
  - For regular SVI with VLAN 103, this consideration is not applicable.

To configure the use case shown above, where you are grouping multiple SVIs into an external bridge domain:

1. Create the floating SVI **vlif-100** with encapsulation **vlan100**.
2. Configure anchor leaf switches node101 and node102 using the VLAN encapsulation **vlan100** (the same VLAN encapsulation as the **vlif-100** anchor encapsulation).
3. Configure the remaining leaf switches with a different VLAN encapsulation:
   - Configure leaf switches node103 and node104 with access encapsulation **vlan101**
   - Configure leaf switches node105 and node106 with access encapsulation **vlan102**
4. Create the regular SVI **svi-103** with encapsulation **vlan103** on leaf switch node107.
5. Group the floating SVI **vlif-100** and the regular SVI **svi-103** together to behave as part of a single external bridge domain:
   a. Create an external bridge group profile.

      The external bridge group profile is represented by the new MO *l3extBdProfile*
   b. Provide a unique name string for the external bridge group profile.
   c. Associate each of the regular and floating SVIs that need to be grouped together to the same external bridge domain.
   d. Associate the SVIs to the external bridge group profile.

      Two new MOs are available for this association: *l3extBdProfileCont*and *l3extRsBdProfile*.

Guidelines and Limitations of Multiple Encapsulation for L3Outs with SVI

- This feature is supported for physical domain L3Outs only, not for VMM domain L3Outs.
- The use case for this feature is for connectivity to external routers. Layer 2 loops are supposed to be blocked by the external device/hypervisor. Loops may occur if this feature is used with external switches that rely on spanning tree protocol to prevent loops.
- Configuring an external bridge group profile under an SVI or floating SVI causes the VLAN to be reprogrammed on the leaf switches, which causes traffic disruption.
- Create a separate physical domain and AEP for each of the different access encapsulations that you want to deploy.
- The nodes that are part of each of these AEPs should be non-overlapping. For example, AEP1 and AEP2 cannot be used on the same leaf node if VLAN 101 and VLAN102 are used for the same floating SVI.
  - AEP1 has physical domain1 with VLAN 101
  - AEP2 has physical domain1 with VLAN 102
- SVIs in the same external bridge group with different VLAN encapsulations cannot be programmed on the same leaf. For example, SVI1 and SVI2 cannot be programmed on the same leaf node.

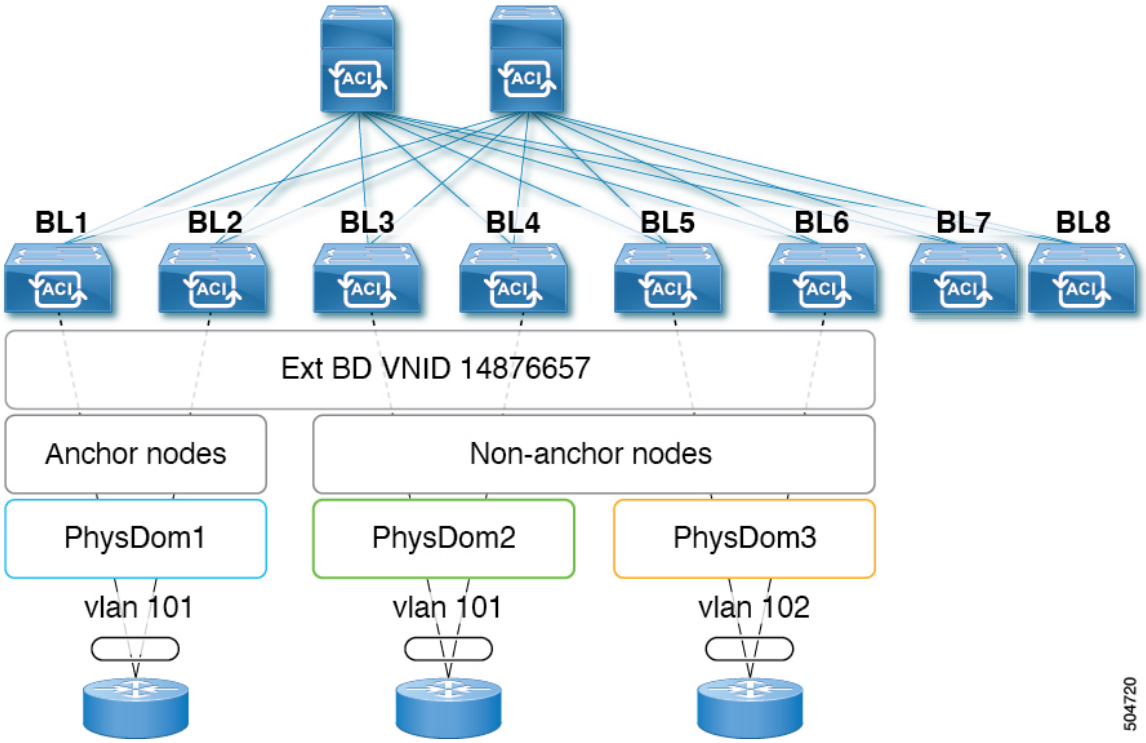
  - SVI1 on node101 with VLAN 101
  - SVI2 on node101 with VLAN 102

- The anchor nodes and the VPC pairs of these anchor nodes should be part of a single physical domain and AEP.
- Path Attribute configuration (*l3extRsDynPathAtt*) at floating SVI has the following considerations:
  - Physical domain for the VLAN used on anchor leaf nodes: Access Encap needs to be blank
  - Physical domains for other VLANs: the specific VLAN needs to be configured at Access Encap
- Do not configure the same physical domain with different VLAN encapsulation under the same floating SVI or external bridge profile group.
- The SVI Encap Scope must be set to Local. VRF Encap scope is not supported.
- If you downgrade from release Cisco Application Centric Infrastructure (ACI) 5.2(3) to a previous release where multiple VLAN encapsulation for L3Outs with SVI is not supported, the following actions will be performed on the L3Out that was configured with multiple encapsulations and/or the external bridge group profile:
  - The new allocator used for the multiple encapsulation support (*l3extBdProfileEncapAllocator*) will be deleted
  - All external bridge group profiles (new *l3extBdProfile* MOs) will be deleted
  - All new *l3extBdProfileCont* MOs will be deleted
  - All new *l3extRsBdProfile* MOs will be deleted
  - All L3Out dynamic attachments (*l3extRsDynPathAtt* MOs) with explicit encapsulation configurations will be deleted

Configuring Multiple Encapsulation for L3Outs With SVI and Floating SVIs Using the GUI

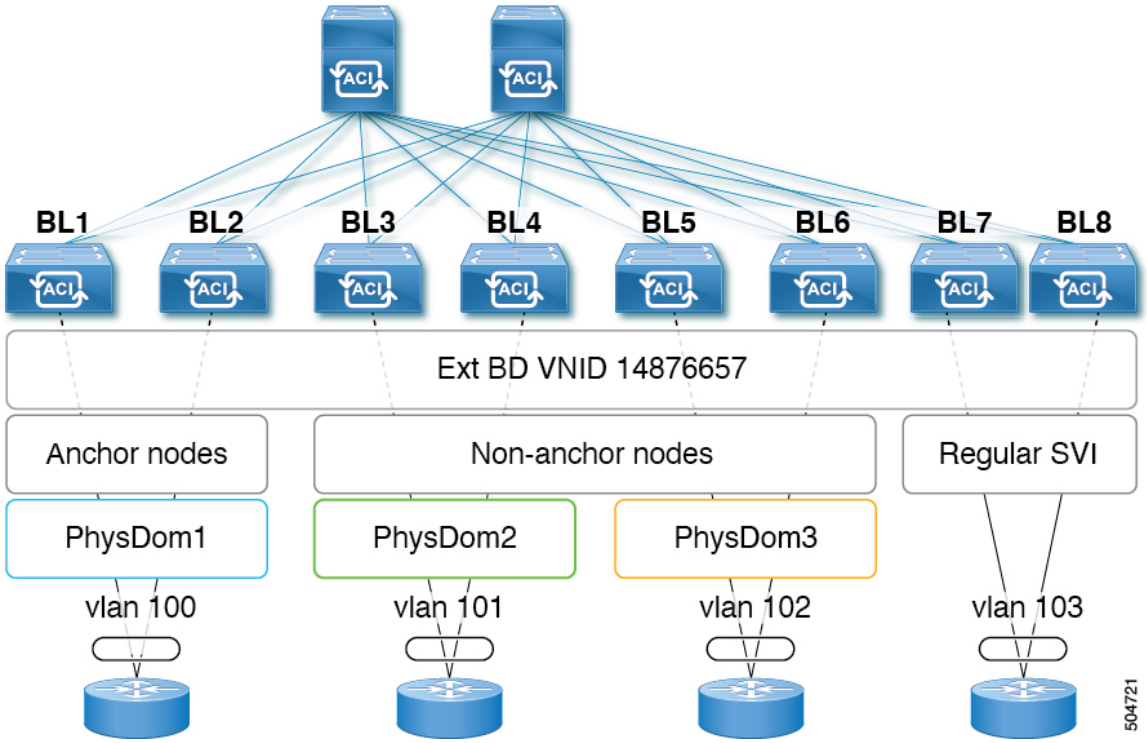
Multiple Encapsulation with SVI is supported on both regular and floating SVIs but uses different configuration. Floating SVIs support mapping the floating SVI to a physical domain using a Path Attribute settings under the floating SVI. To use multiple encapsulations with floating SVIs you would map the floating SVI to different domains each with a different VLAN encapsulation. The figure below illustrates an example used in this section.

Figure 26: Floating SVI configured with multiple encapsulations using different domains



Regular SVIs do not support automatic configuration of the SVI using a physical domain. With regular SVIs you must create a separate SVI on each leaf node or leaf node pair where the SVI needs to be deployed. Support for multiple encapsulations with regular SVIs is done by configuring multiple SVIs under the L3Out, each with different encapsulations, and then grouping all SVIs using a bridge group. This section will cover the configuration steps for both the floating and regular SVIs.

Figure 27: Regular SVI configured with multiple encapsulations using Bridge Groups



Creating a Floating SVI

Step 1 To create a floating SVI:

a) Navigate to **Tenants**> *tenant-name*> **Networking**> **L3Outs**> **L3Out name**> **Logical Node Profile**>**log-node-profile-name**>**Logical Interface Profile**>**log-int-profile-name**.

The **General** page for this logical interface profile is displayed.

b) Click on the **Floating SVI** tab.
c) Click on the + to add a floating SVI.
d) In the **Anchor Node** drop-down list, select a switch node for the anchor node.
e) In the **IPv4 Primary / IPv6 Preferred Address** field, enter an IPv4 or IPV6 address and mask.
f) In the **Encap** field, enter a VLAN number for the anchor nodes.
g) Click + in the **Path Attributes** area.

The **Create Floating Path Attributes** window appears

h) In the **Domain** drop-down list, select the physical domain associated to the anchor node and other nodes where the floating SVI will be deployed with the same VLAN as the anchor node.
i) In the **Floating Primary IPv4 / IPv6 Address** field, enter an IP address and mask for the floating nodes.
j) Leave the **Access Encap VLAN** field blank.
k) Click **OK**.
l) Click **Submit**.

Step 2 Add additional path attributes to floating SVI for nodes that will use a different VLAN encapsulation.

a) Navigate to **Tenants** > ***tenant-name*** > **Networking** > **L3Outs** > ***L3Out-name*** > **Logical Node Profile** > ***log-node-profile-name*** > **Logical Interface Profile** > ***log-int-profile-name***.

The **General** page for this logical interface profile is displayed.

b) Click on the **Floating SVI** tab.
A page showing the already-configured floating switch virtual interfaces is displayed.

c) Double-click on the floating switch virtual interface where you want to specify the separate encapsulation.

d) Click **+** in the **Path Attributes** area.

The **Create Floating Path Attributes** window appears.

e) In the **Domain** drop-down list, select the physical domain associated to nodes that will use different encapsulation.

f) In the **Floating Primary IPv4/IPv6 Address** field, enter an IP address and mask for the floating nodes.

g) In the **Access Encap** field, enter the VLAN id that will be used by the switches associated with this domain.

h) Click **OK**.

i) Click **Submit**.

Procedure for Multiple Encapsulations With Regular SVIs

Step 1 To create an external bridge group that will be used for SVI grouping:

a) Navigate to **Tenants> tenant-name > Policies > Protocol > External Bridge Group Profiles**.

b) Right click on **External Bridge Group Profiles** and choose **Create External Bridge Group Profile**.

c) Enter a name for the external bridge group profile, then click **Submit**.

The page showing the already-configured external bridge group profiles is updated with the new external bridge group profile.

Step 2 To associate a regular SVI with the external bridge group profile:

a) Navigate to **Tenants>tenant-name>Networking>L3Outs>L3Out-name> Logical Node Profile>log-node-profile-name>Logical Interface Profile>log-int-profile-name**.

The **General** page for this logical interface profile is displayed.

b) Click on the **SVI** tab.

A page showing the already-configured floating switch virtual interfaces is displayed.

c) Double-click on the switch virtual interface that you want to associate with the external bridge group profile.

General information for this switch virtual interface is displayed.

d) In the **External Bridge Group Profile** field, select the external bridge group profile that you want to associate with this switch virtual interface.

e) Click **Submit**.

Step 3 To associate another regular SVI under the same L3Out using a different encapsulation to the same bridge:

a) Navigate to **Tenants>tenant-name>Networking>L3Outs>L3Out-name> Logical Node Profile>log-node-profile-name>Logical Interface Profile>log-int-profile-name**.

The **General** page for this logical interface profile is displayed.

b) Click on the **SVI** tab.

A page showing the already-configured switch virtual interfaces is displayed.

c) Double-click on the switch virtual interface that you want to associate with the external bridge group profile.

General information for this switch virtual interface is displayed.

d) In the **External Bridge Group Profile** field, select the external bridge group profile that you want to associate with this switch virtual interface.

e) Click **Submit**.

Procedure for Multiple Encapsulations With Both Floating and Regular SVIs

Step 1 To create an external bridge group profile that will be used for SVI grouping:

a) Navigate to **Tenants>tenant-name>Policies>Protocol> External Bridge Group Profiles**.

A page showing the already-configured external bridge group profiles is displayed.

b) Right-click on **External Bridge Group** and choose **Create External Bridge Group Profile**.

The **Create External Bridge Group Profile** page is displayed.

c) Enter a name for the external bridge group profile, then click **Submit**.

The page showing the already-configured external bridge group profiles is updated with the new external bridge group profile.

Step 2 To associate a floating SVI with the external bridge group profile:

a) Navigate to **Tenants>tenant-name>Networking>L3Outs>L3Out-name>Logical Node Profile>log-node-profile-name>Logical Interface Profile>log-int-profile-name**.

The **General** page for this logical interface profile is displayed.

b) Click on the **Floating SVI** tab.

A page showing the already-configured switch virtual interfaces is displayed.

c) Double-click on the switch virtual interface that you want to associate with the external bridge group profile.

General information for this switch virtual interface is displayed.

d) In the **External Bridge Group Profile** field, select the external bridge group profile that you want to associate with this switch virtual interface.

e) Click **Submit**.

Step 3 To associate a regular SVI under with the external bridge group profile:

a) Navigate to **Tenants>tenant-name>Networking>L3Outs>L3Out-name>Logical Node Profile>log-node-profile-name>Logical Interface Profile>log-int-profile-name**.

The **General** page for this logical interface profile is displayed.

b) Click on the **SVI** tab.

A page showing the already-configured switch virtual interfaces is displayed.

c) Double-click on the switch virtual interface that you want to associate with the external bridge group profile.

General information for this switch virtual interface is displayed.

d) In the **External Bridge Group Profile** field, select the external bridge group profile that you want to associate with this switch virtual interface.

e) Click **Submit**.

Step 4 To specify the separate encapsulation for non-anchor (floating) nodes:

a) Navigate to **Tenants>tenant-name>Networking>L3Outs>L3Out-name>Logical Node Profile>log-node-profile-name>Logical Interface Profile>log-int-profile-name**.

The **General** page for this logical interface profile is displayed.

b) Click on the **Floating SVI** tab.

A page showing the already-configured floating switch virtual interfaces is displayed.

c) Double-click on the floating switch virtual interface where you want to specify the separate encapsulation.

General information for this floating switch virtual interface is displayed.

d) Click + in the **Path Attributes** area.

The **Create Floating Path Attributes** window appears.

e) In the **Access Encap** field, enter the access encapsulation for the non-anchor (floating) nodes.

f) Click **Submit**.

You are returned to the **Floating SVI** page.

g) Click **Submit**.

Configuring Multiple Encapsulation for L3Outs With SVI Using the CLI

Step 1 Log into your Cisco Application Policy Infrastructure Controller (APIC) through the CLI, then go into configuration mode and tenant configuration mode.

```
apic1#
apic1# configuration
apic1(config)# tenant <tenant-name>
apic1(config-tenant)#
```

Step 2 Enter the following commands to create an external bridge profile that will be used for SVI grouping.

```
apic1(config-tenant)# external-bridge-profile <bridge-profile-name>
apic1(config-tenant-external-bridge-profile)# ?
```

Step 3 Enter the following commands to associate a floating SVI with the external bridge group profile.

```
apic1(config)# leaf <leaf-ID>
apic1(config-leaf)# virtual-interface-profile <ipv4/ipv6> vlan <vlan-num> tenant <tenant-name> vrf
<VRF-name> l3out <L3Out-name>
apic1(virtual-interface-profile)# ip address <IP-address>
apic1(virtual-interface-profile)#  physical-domain <phy-dom-name> floating-addr <IP-address>
apic1(physical-domain)# vlan <vlan-num>
apic1(physical-domain)# exit
apic1(config-tenant)# external-bridge-profile <bridge-profile-name>
apic1(config-tenant-external-bridge-profile)#
```

Step 4 Enter the following commands to associate a regular SVI with the external bridge group profile.

```
apic1(config)# leaf <leaf-ID>
apic1(config-leaf)# interface vlan <vlan-num>
apic1(config-leaf-if)# vrf member tenant <tenant-name> vrf <VRF-name>
apic1(config-leaf-if)# ip address <IP-address>
apic1(config-leaf-if)# external-bridge-profile <bridge-profile-name>
```

Configuring Multiple Encapsulation for L3Outs With SVI Using the REST API

Step 1 Enter a post such as the following example to create an external bridge profile that will be used for SVI grouping.

```
<fvTenant name="t1" dn="uni/tn-t1" >
    <l3extBdProfile name="bd100" status=""/>
</fvTenant>
```

Step 2 Enter a post such as the following example to associate a floating SVI with the external bridge group profile.

```
<fvTenant name="t1">
    <l3extOut name="l1">
        <l3extLNodeP name="n1">
            <l3extLIfP name="i1">
                <l3extVirtualLIfP addr="10.1.0.1/24"
                    encap="vlan-100"
                    nodeDn="topology/pod-1/node-101"
                    ifInstT="ext-svi">
                    <l3extBdProfileCont>
                        <l3extRsBdProfile tDn="uni/tn-t1/bdprofile-bd100"/>
                    </l3extBdProfileCont>
                </l3extVirtualLIfP>
            </l3extLIfP>
    </l3extOut>
</fvTenant>
```

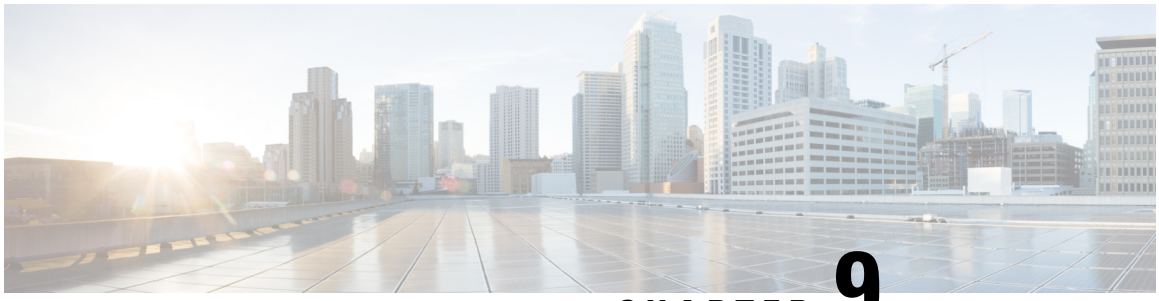
Step 3 Enter a post such as the following example to associate a regular SVI with the external bridge group profile.

```
<fvTenant name="t1">
    <l3extOut name="l1">
        <l3extLNodeP name="n1">
```

```
            <l3extLIfP name="i1">
                <l3extRsPathL3OutAtt encap="vlan-108"
                    tDn="topology/pod-1/paths-108/pathep-[eth1/10]"
                    ifInstT="ext-svi">
                    <l3extBdProfileCont>
                        <l3extRsBdProfile tDn="uni/tn-t1/bdprofile-bd100" status=""/
                    </l3extBdProfileCont>
                </l3extRsPathL3OutAtt>
            </l3extLIfP>
        </l3extLNodeP>
    </l3extOut>
</fvTenant>
```

Step 4 Enter a post such as the following example to specify the separate encapsulation for floating nodes.

```
<fvTenant name="t1">
    <l3extOut name="l1">
        <l3extLNodeP name="n1">
            <l3extLIfP name="i1">
                <l3extVirtualLIfP addr="10.1.0.1/24"
                    encap="vlan-100"
                    nodeDn="topology/pod-1/node-101"
                    ifInstT="ext-svi">
                    <l3extRsDynPathAtt floatingAddr="10.1.0.100/24"
                        encap="vlan-104"
                        tDn="uni/phys-phyDom"/>
                </l3extVirtualLIfP>
            </l3extLIfP>
    </l3extOut>
</fvTenant>
```



CHAPTER 9

Floating L3Out Consideration and Restrictions

• Summary of Floating L3Out Deployment Considerations and Restrictions, on page 67

Summary of Floating L3Out Deployment Considerations and Restrictions

The following list summarizes some of the requirements and limitations of the floating Layer 3 outside network connection (L3Out) feature.

- Floating L3Out with VMware vDS VMM domain considerations include the following:
  - When floating L3Out is deployed, the virtual router must be connected to the port-group created by the VMM domain.
  - A floating L3Out Switch Virtual Interface (SVI) is not programmed on a leaf node if at least a virtual router connected to the created port-group does not exist on one of the locally connected hypervisors.
  - Once a virtual router moves and becomes attached to the port group, the leaf node has SVI programmed. This is the same as a regular endpoint group (EPG) in a Virtual Machine Manager (VMM) domain with on-demand resolution immediacy, and deployment immediacy is immediate.
  - Secondary floating IP with a VMM domain is supported only with Cisco Application Policy Infrastructure Controller (APIC) Release 5.0(1). Secondary address configurations must be removed before downgrading to an earlier release.
- Floating L3Out with physical domains requires Cisco APIC release 5.0(1) or later.
- The functionality to avoid sub-optimal path requiring Next Hop propagation is supported for intra-VRF traffic only. Inter-VRF traffic where the consumer and the provider are in different VRFs is not supported. This consideration is applied to both EPG to external EPG and external EPG to external EPG contracts.

Note This functionality is only supported for physical domains.

- Cisco APIC supports the deployment of floating L3Out across pods that are part of a Cisco Application Centric Infrastructure (ACI) Multi-Pod fabric, but traffic from ACI internal endpoints to the floating L3Out has the following considerations:

- Avoiding sub-optimal flow (next-hop propagation and multi-path) is not supported with Floating L3Out across pods that are part of an ACI Multi-Pod fabric. This consideration is not applicable if the floating L3Out is not deployed across pods.
  - For example, this is fine because the anchor leaf node and the non-anchor leaf node are in the same pod:
    - Anchor leaf node in pod1.
    - Non-anchor leaf node in pod1.
    - Internal endpoint in any pod.
  - For example, this is NOT supported because the anchor leaf node and the non-anchor leaf node are in the different pods:
    - Anchor leaf node in pod1
    - Non-anchor leaf node in pod2.
    - Internal endpoint in any pod.

- Cisco APIC does not support floating L3Out with remote leaf switches even if remote leaf direct is enabled. A remote leaf switch can't be deployed as an anchor leaf or a non-anchor leaf for floating L3Out.
- A non-anchor node should not be configured under a logical node profile.
- Per port VLAN feature is not supported on ports that are part of floating L3Outs.
- If a virtual port channel (vPC) interface is used for external router connectivity and one leaf switch in the vPC pair is an anchor leaf node, the other leaf switch in the same vPC pair must also be an anchor leaf node. A mixture of anchor and non-anchor leaf nodes in the same vPC pair for the same VLAN encapsulation is not supported.
- For a static route, the external router must use the primary or secondary IP address of the anchor leaf switch as the next hop.
- Make sure that the primary IP address and floating primary IP address are part of the same subnet.
- In some situations, you might not be able to achieve a desired result using a single L3Out and you will have to configure two different L3Outs to achieve that result instead (for example, if you want to have both BGP and OSPF to redistribute the routes into the fabric).

  For example, consider a situation where you want to have the following configuration:
  - An eBGP session that will be established between a border leaf switch's loopback IP address and an external router's loopback IP address
  - OSPF configured to exchange the route for the loopback IP address for the border leaf switch and the external router
  - OSPF also configured to redistribute additional routes learned from the external nodes into the Cisco ACI fabric

  You will not be able to have that configuration using a single L3Out because, in order to get that configuration, you will have to use both OSPF and BGP to redistribute the routes into the fabric, which cannot be done on a single L3Out due to an existing L3Out limitation.

- Be sure to choose the recommended Bidirectional Forwarding Detection (BFD) timers while on the floating L3Out. You can expect traffic loss of 1 to 2 seconds when the virtual router moves from one host to another host within the cluster. We recommend BFD TX/RX intervals of 700 milliseconds or higher.
- A floating L3Out SVI and a non-floating L3Out SVI can exist on the same leaf switch with the same VLAN encapsulation as long as they use the same primary IP address.
- Floating L3Out requires generation-2 leaf switches. Generation-1 switches cannot be configured as an anchor or non-anchor switch. However, you can use a generation-1 switch as a non-border leaf or a compute leaf switch.
- If ARP entry for an external device's IP is manually cleared on the non-anchor leaf node, internal to external traffic path will be sub-optimal even if direct host advertisement is enabled to avoid sub-optimal path. It's because the leaf node stops advertising the external device's IP, whereas the anchor leaf nodes still have the ARP entry which still have the traffic forwarded. To avoid sub-optimal flow, it's recommended to clear the ARP entry on all anchor and non-anchor leaf nodes which will recreate the ARP entry on the leaf nodes where the external IP is connected. So that internal to external traffic path will be back to optimal forwarding. This consideration is applicable to IPv6 adjacency too.
- For Next Hop Propagation, the floating L3Out must be in a physical domain, not in a VMM domain.
- Prior to Cisco Nexus Dashboard Orchestrator (NDO) release 4.1(1), as with other logical interfaces under L3Out, the floating SVI configuration is at the Cisco APIC level and not at the NDO level. However, Cisco NDO can refer an L3Out that contains floating SVI.

  Starting from NDO release 4.1(1), the full L3Out configuration, including floating L3Out, is available on NDO. However, the floating L3Out functionalities (such as next-hop propagation, etc.) is available within each fabric.
- 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.
- The following rules apply for a floating L3Out with IPv4 and IPv6 address family:
  - For IPv4 and IPv6 address family for the same L3Out on the same leaf nodes, you need different L3Out logical interface profiles.
  - Physical domain: for the same VLAN encapsulation, you must use the same anchor leaf nodes for IPv4 and IPv6 address family. A different set of leaf nodes cannot be the anchor for an IPv4 and IPv6 address family.
  - VMM domain: Prior to Cisco APIC Release 5.2(4), you must use different VLAN encapsulations for IPv4 and IPv6 address families. Different port groups are created for IPv4 and IPv6 floating SVI interfaces.

    Beginning with Cisco APIC release 5.2(4), same encapsulation can be used. One port group is created for IPv4 and IPv6 address families for the same L3Out with a VMM domain. While deploying a floating SVI, if both address families are configured under the L3Out, both floating SVIs for IPv4 and IPv6 are deployed on the leaf nodes.
    - If you downgrade from Cisco APIC release 5.2(4) to a previous release, VMM domain dynamic attachments under a floating L3Out with the same SVI encapsulation but different address-families are deleted.

- If you upgrade from pre-5.2(4) to 5.2(4), and want to migrate from different encapsulation for IPv4 or IPv6 to same encapsulation for IPv4 or IPv6, you need to delete the existing logical interface profile, and add it using the same VLAN encapsulation for both IPv4 or IPv6 address families.

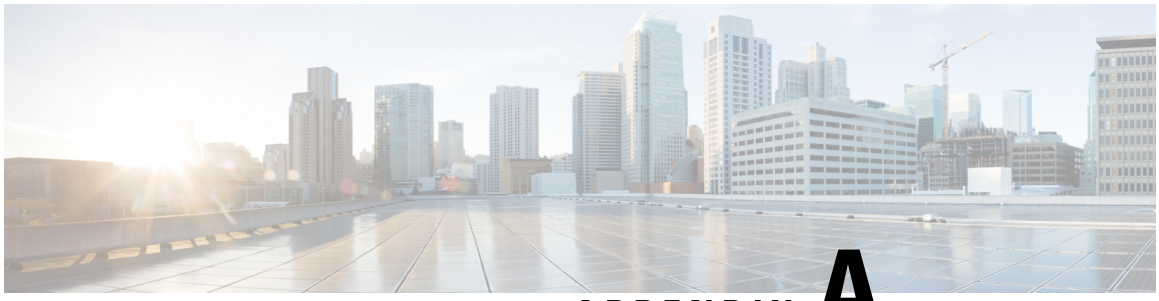
- As part of the support for avoiding suboptimal traffic from a Cisco ACI internal endpoint to a floating L3Out that was introduced in release 5.0(1), support was also available for learning the directly attached host routes (external router's IPs in the L3Out SVI subnet) on the anchor and non-anchor leaf switches and redistributing them into the ACI fabric, as described in Avoiding Suboptimal Traffic From a Cisco ACI Internal Endpoint to a Floating L3Out, on page 11 and Configuring Direct-Attached Host Route Advertising on L3Out, on page 44.

  Prior to release 5.2(1), these attached host routes could be advertised out of the ACI fabric if you configured the export rules to explicitly advertise the attached host routes out of the ACI fabric.

  Beginning with release 5.2(1), this behavior has changed and attached host routes are implicitly denied from going out of the ACI fabric.

- The following considerations and restrictions apply to the multi-protocol recursive next hop propagation feature, introduced in release 5.2(1):
  - As of Cisco ACI release 6.0(1), in the case of ACI floating L3Out with next-hop propagation enabled, only one next-hop for a prefix learned via BGP is used for forwarding. If ECMP is needed for the prefix, it's achievable by using ECMP to the next-hop below. This consideration is applicable to both IPv4 and IPv6.
    - BGP sends one primary next-hop for the prefix. (For example, 10.1.0.0/16 via 1.1.1.1).
    - Anchor leaf node redistributes multi-path for the next-hop (For example, 1.1.1.1 via 172.16.1.1 and 172.16.1.2).
  - When next-hop propagation is enabled, it’s not recommended to have multiple external routers for control nodes redundancy advertising the same external prefix with the same ECMP paths because of CSCwd28918.
    - For example, this is fine because it uses different next-hop IPs:
      - external router1, that advertises 10.1.0.0/16 via 172.16.1.1 via BGP.
      - external router2, that advertises 10.1.0.0/16 via 172.16.1.2 via BGP.
    - For example, this is NOT recommended because it has only one next-hop IP, not ECMP:
      - external router1, that advertises 10.1.0.0/16 via 1.1.1.1 via BGP.
      - external router2, that advertises 10.1.0.0/16 via 1.1.1.1 via BGP.
  - IPv6 link local addresses are not supported when configuring next hop propagation for recursive route. You must use global addresses for IPv6 in this situation.
  - Redistribution into a Not-So-Stubby Area (NSSA) creates a special type of link-state advertisement (LSA) known as type 7, which can only exist in an NSSA area. Routes should be learned as this type 7 LSA to have next hops as the global address, which means that the L3Out for the forwarding node, with only the OSPF protocol enabled (**l3out-ospf**), must use the **NSSA area** option in the **OSPF Area Type** field.

- If multiple next hops are used for a static route, the route might not be optimal if one of the next hops is down. See CSCvy10946 for more details.
- The following limitations exist when configuring the multi-protocol recursive next hop propagation feature through the CLI:
  - The next-hop unchanged route-map is not supported for BGP peers.
  - The route-profile template doesn't support match rules.



APPENDIX A

Topology Examples with Avoidance of Suboptimal Traffic and ECMP

- Topology Examples with Avoidance of Suboptimal Traffic and ECMP, on page 73

Topology Examples with Avoidance of Suboptimal Traffic and ECMP

This section provides high-level configuration requirements for a floating L3Out with avoidance of suboptimal traffic and ECMP. For configuration steps and each configuration option, please see Configuring the Avoidance of Suboptimal Traffic From an ACI Internal Endpoint to a Floating L3Out Using the Cisco APIC GUI, on page 39.

The following examples are covered in this section:

- External prefix is learned via BGP. Recursive lookup with loopback is required because we can have only one next-hop for the route learned via BGP. For example:
  - 10.0.0.0/8 is the external prefix learned via BGP with next-hop that is the external router's loopback IP address 1.1.1.1. This loopback address is learned via OSPF or configured via static routing.
  - 1.1.1.1/32 via 172.16.1.1, 172.16.1.2 and 172.16.1.3, that are learned via OSPF or static route.
- External prefix is learned via OSPF or static route. For example:
  - 10.0.0.0/8 is the external prefix learned via OSPF (or configured via static routing) with next-hop that is the external router's connected IP addresses 172.16.1.1. 172.16.1.2 and 172.16.1.3.
- External prefix is learned via BGP with multiple next-hops. This option requires Cisco Application Centric Infrastructure (ACI) Release 6.0(2) or later.
  - 10.0.0.0/8 via 1.1.1.1, 1.1.1.2 and 1.1.1.3, that are learned via BGP
  - 1.1.1.1 via 172.16.1.1, 1.1.1.2 via 172.16.1.2 and 1.1.1.3 via 172.16.1.3 that are learned via OSPF or static route.

Example 1: External prefix is learned via BGP

Topology Examples with Avoidance of Suboptimal Traffic and ECMP

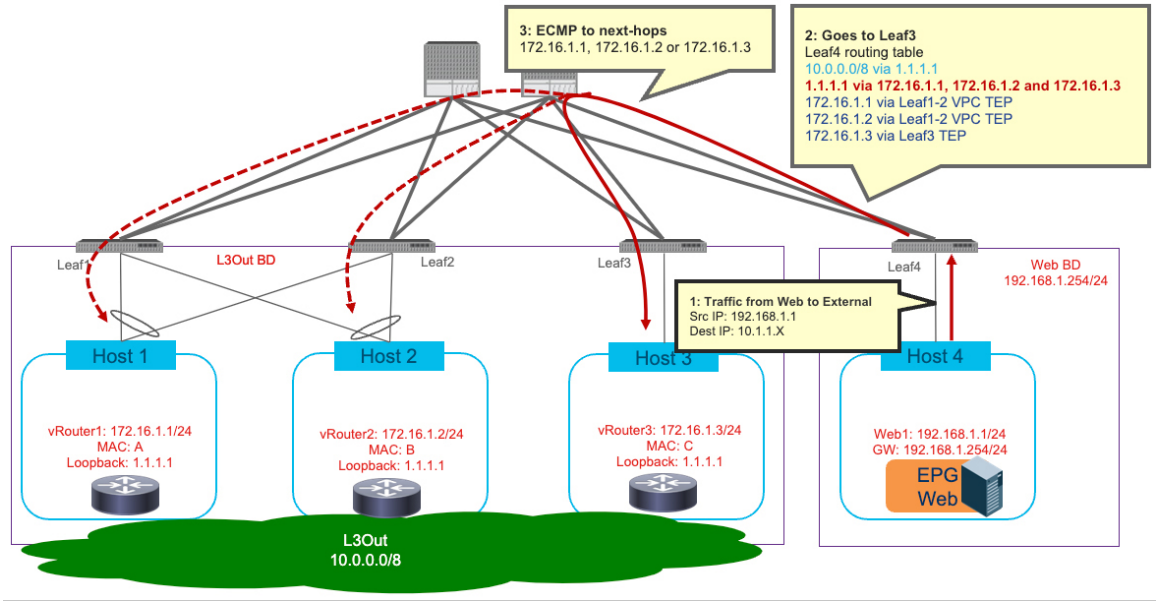
This option requires Cisco ACI Release 5.2 or later.The figures below illustrate an example. The external route 10.0.0.0/8 is advertised using BGP from the external routers. The next hop for the external prefix is not the IP address of the external routers that are connected to the L3Out SVIs subnet (172.16.1.1, 172.16.1.2 and 172.16.1.3) but it is instead a loopback address defined on the external routers and advertised using OSPF (or through static routing). This results in a multi-level recursion, where the BGP route next hop is resolved using the OSPF route, and the OSPF route ultimately gets resolved using the direct adjacency formed within the Cisco ACI fabric.

The required configurations are the following ones:

- Increase “Local Max ECMP” in BGP Address Family Context Policy.
- Configure a route-map with the following option on the L3Out for BGP:
  - Next Hop Propagation for the external prefix (10.0.0.0/8).
- Configure route-maps with the following options on the L3Out for OSPF:
  - Next Hop Propagation and Multi-path for the loopback IP (1.1.1.1).
  - Advertise Direct Attached host for the directly attached next-hop IPs (172.16.1.1, 172.16.1.2 and 172.16.1.3)

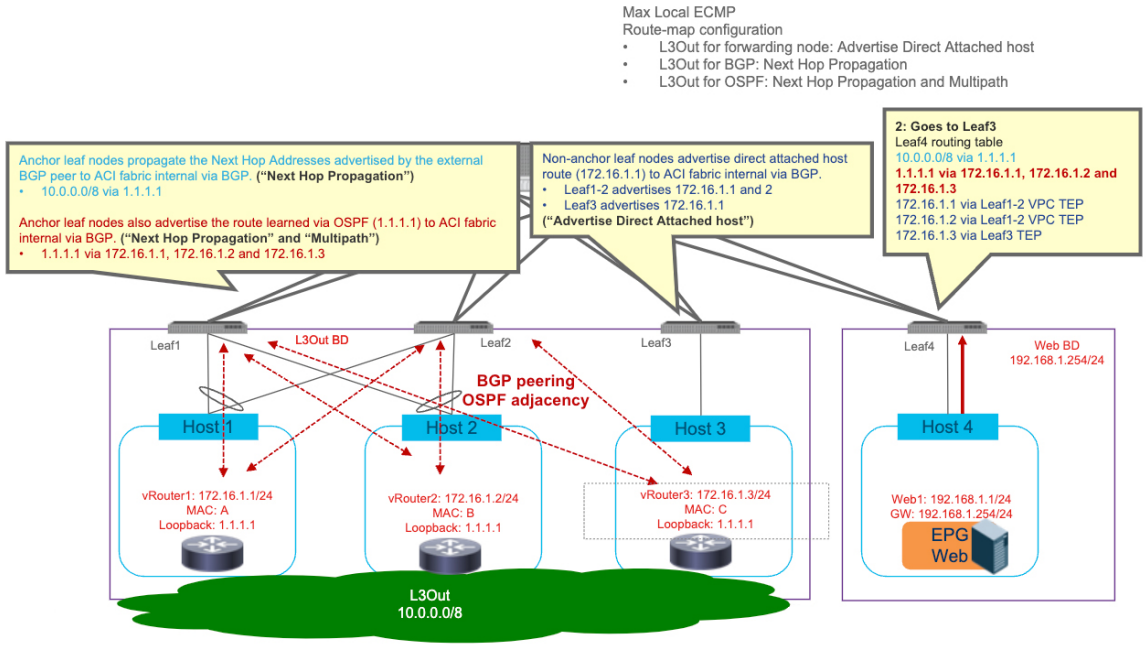
If static route is used instead of OSPF, the L3Out for BGP and the L3Out for OSPF can be combined to one L3Out.

Figure 28: External prefix is learned via BGP



Topology Examples with Avoidance of Suboptimal Traffic and ECMP

Figure 29: External prefix is learned via BGP (route-map configuration)



Example 2: External prefix is learned via OSPF or static route

This option requires Cisco ACI Release 5.2 or later. The figures below illustrate an example.The external route 10.0.0.0/8 is advertised using OSPF from the external routers (or through static routing). The next hops for the external prefix are the IP address of the external routers that are connected to the L3Out SVIs subnet (172.16.1.1, 172.16.1.2 and 172.16.1.3).

The required configurations are the following ones:

- Increase "Local Max ECMP" in BGP Address Family Context Policy.
- Configure route-maps with the following options on the L3Out for OSPF:
  - Next Hop Propagation and Multi-path for the external prefix (10.0.0.0/8).
  - Advertise Direct Attached host for the directly attached next-hop IPs (172.16.1.1, 172.16.1.2 and 172.16.1.3).

Topology Examples with Avoidance of Suboptimal Traffic and ECMP

Figure 30: External prefix is learned via OSPF or static route

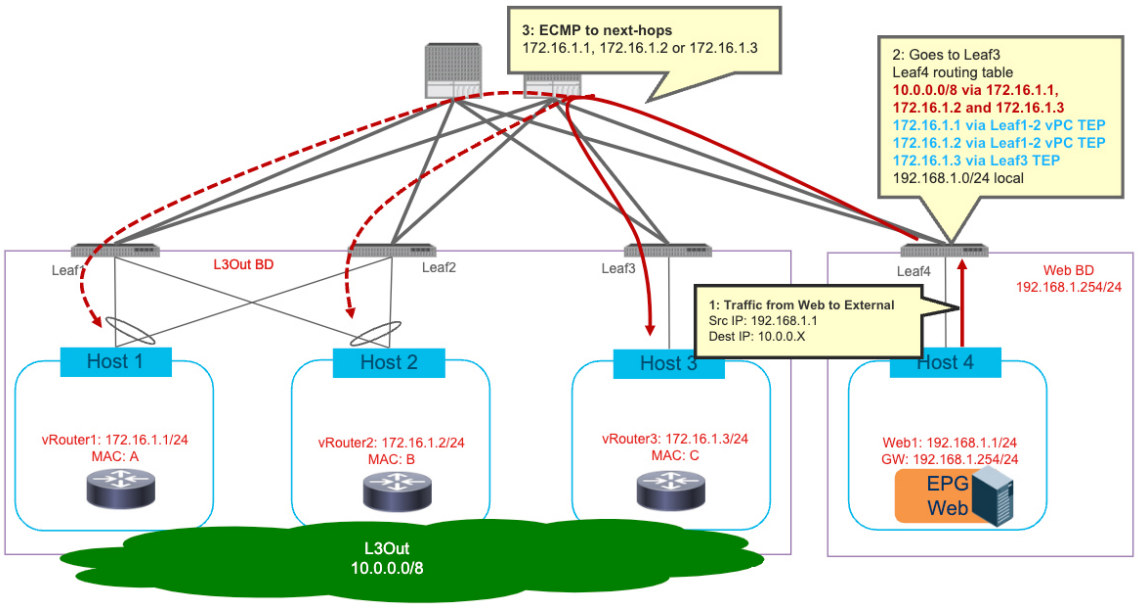
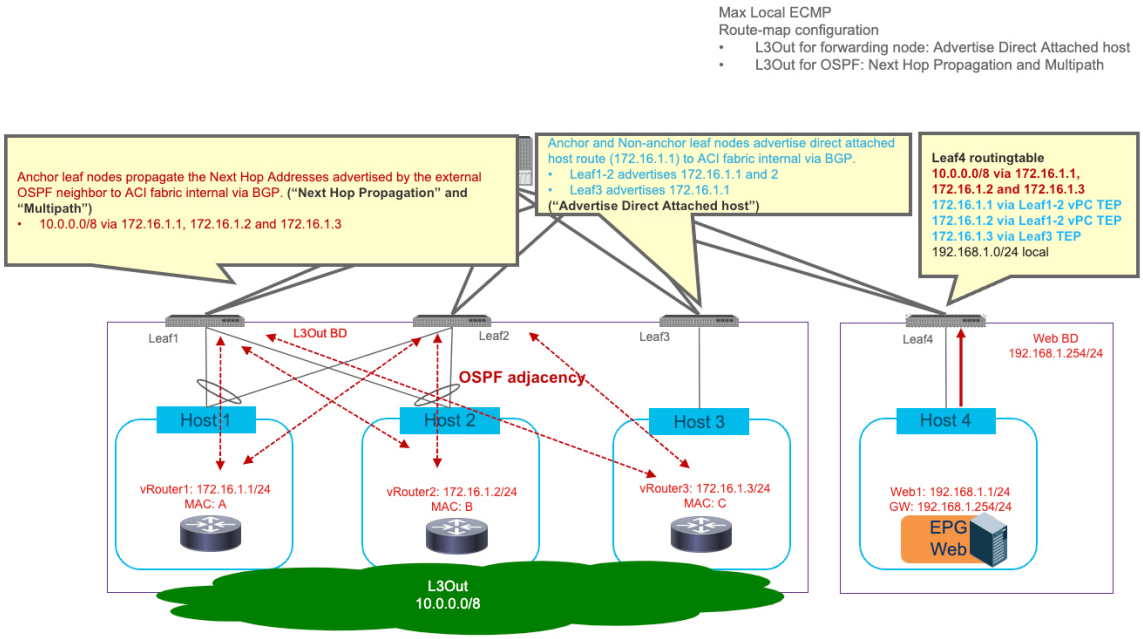


Figure 31: External prefix is learned via OSPF or static route (route-map configuration)



Example 3: External prefix learned via BGP with BGP additional paths capability and next-hops directly connected

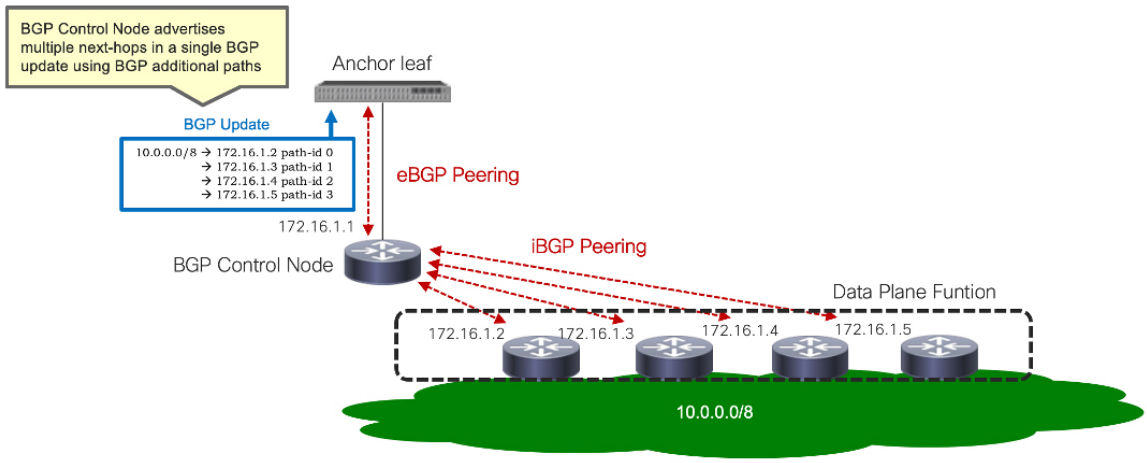
This option requires Cisco ACI Release 6.0(2) and the use of BGP additional paths **receive** capability configuration.

BGP additional paths capability is a BGP extension that allows BGP speakers to send multiple paths (next hops) in a single BGP update. Without this capability, the advertisement of a prefix with a different path from

the same peer will replace the previous path for that prefix. A BGP speaker can be configured to send, receive, or both send and receive additional paths. Cisco ACI 6.0(2) only supports the receiving capability.

The BGP additional paths feature is useful in a topology where control and data plane functions are performed on different nodes or virtual routers. The control node is a BGP speaker that will advertise prefixes where the next hop address is a different virtual router performing the data plane function. BGP additional paths allows the BGP control node to advertise multiple next hops for a single prefix in a single BGP update. This requires BGP additional paths **send** capability on the control node. The use of BGP additional paths allows for load balancing of data plane traffic over multiple ECMP paths without requiring separate BGP sessions for each path. The figure below illustrates an example topology where the BGP additional paths feature may be used. The BGP control node is configured to advertise routes learned from internal BGP peers in a single BGP update using the additional paths feature.

Figure 32: BGP control node advertising multiple paths using BGP additional paths feature.



The figure below illustrates an example where this feature is used with floating L3Out and propagate next hop. The external route 10.0.0.0/8 is advertised by the control node with three path-ids each with a different next hop address. The next hop addresses for the external prefix are the IP addresses of the external routers that are connected to the L3Out SVIs subnet (172.16.1.2, 172.16.1.3, and 172.16.1.4). An ECMP path for the prefix 10.0.0.0/8 will be installed in the routing table on the anchor nodes (Leaf1 and Leaf2) with all three next hops. This ECMP path will also be propagated via BGP to the non-border leaf switches where an ECMP path will be installed with all three propagated next hops.

Note The propagation of ECMP paths from the anchor leaf to non-border leaf switches does not require the BGP additional paths feature. Prior to Cisco ACI release 6.0.2, the anchor leaf would only advertise a single path to the non-border leaf switches if the anchor leaf had an ECMP path learned via BGP. This behavior was changed in Cisco ACI release 6.0(2) to take full advantage of the newly introduced additional paths feature. The anchor leaf will now advertise an ECMP path for BGP learned routes. The ECMP path on the anchor leaf can be learned using the BGP add path feature or from individual updates from directly connected BGP peers.

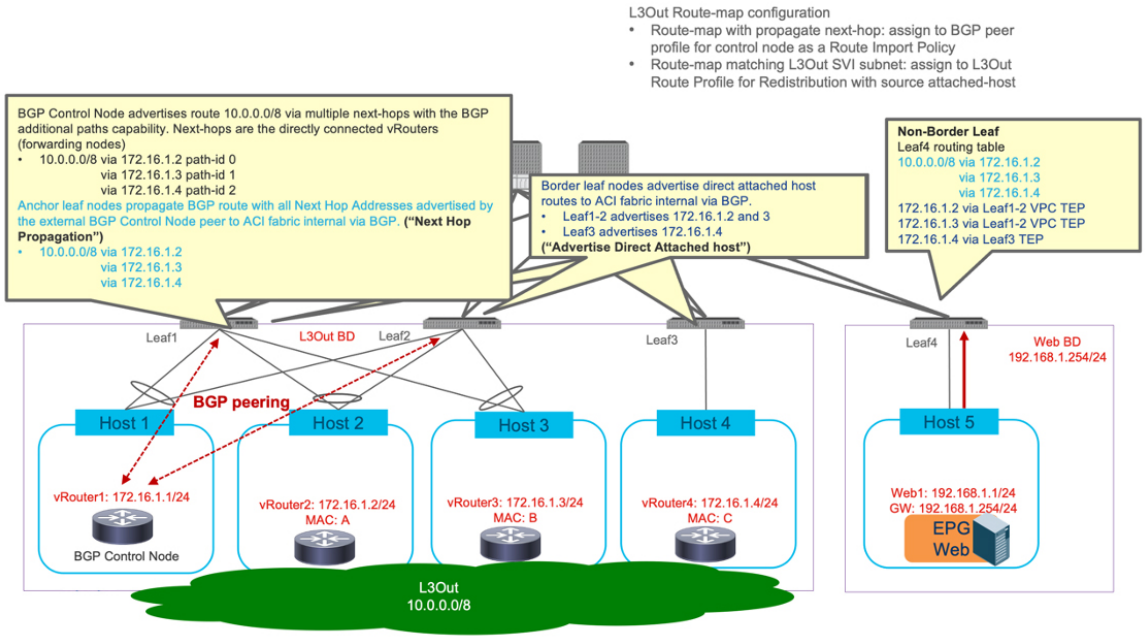
The required configurations are as follows (this configuration example only requires one L3Out):

- The External BGP speaker (control node) is configured with BGP additional paths send capability (this is not configured on the Cisco ACI fabric) and it is creating BGP adjacencies with the anchor nodes (Leaf1 and Leaf2).

Topology Examples with Avoidance of Suboptimal Traffic and ECMP

- Enable BGP additional paths receive capability in Cisco ACI. This can be enabled at the BGP peer profile level or at the VRF level.
- Configure a route-map with the following options and apply it to the BGP peers for the control node as a Route Import Policy.
  - Next Hop Propagation for the external prefix (10.0.0.0/8).
- Configure a route-map with the following options for the directly attached next-hops.
  - Advertise direct-attached host for the directly attached next-hop IPs (172.16.1.2, 172.16.1.3, and 172.16.1.4)

Figure 33: External prefix ECMP path learned via BGP control node with additional paths feature

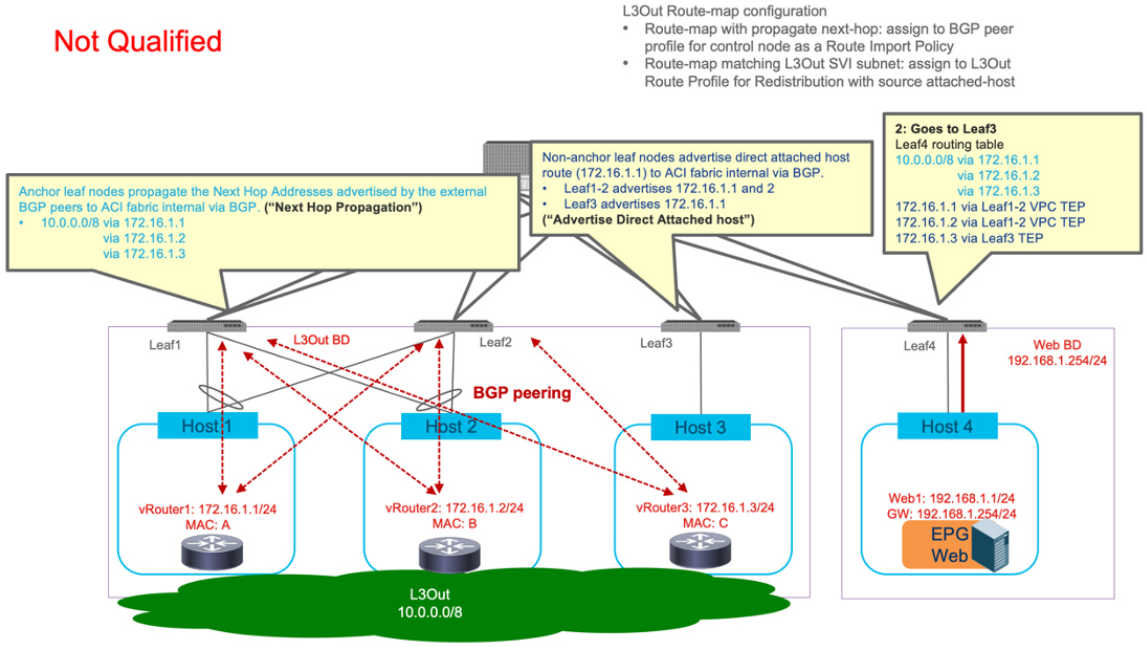


The figure below shows an example where the ECMP paths are learned via directly connected BGP peers without the use of a control node. The BGP additional paths feature is not required in this topology.

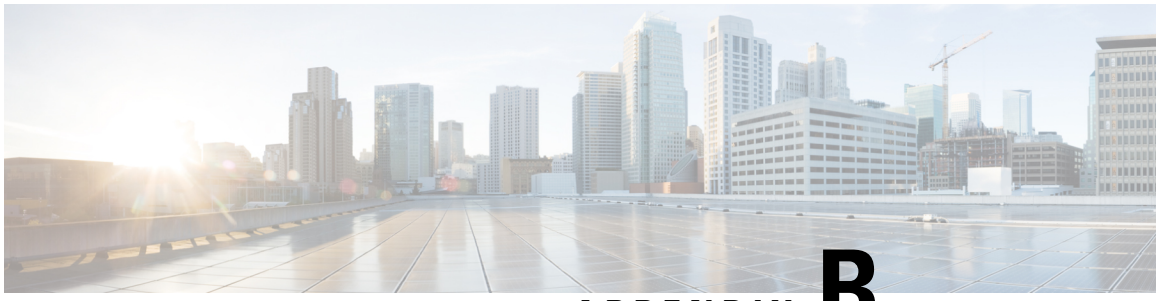
Note This configuration has not been qualified at this time.

Topology Examples with Avoidance of Suboptimal Traffic and ECMP

Figure 34: External prefix ECMP path learned via directly connected BGP speakers



Topology Examples with Avoidance of Suboptimal Traffic and ECMP



APPENDIX B

Configuring Floating L3Outs Using the CLI

- Creating a VLAN Pool for Floating L3Out Using the CLI, on page 81
- Configuring a VMM Domain Profile for VMware VDS Using the CLI, on page 82
- Configuring a Floating L3Out Using the CLI, on page 82
- Configuring a Secondary IP Using the CLI, on page 83
- Configuring the Avoidance of Suboptimal Traffic From an ACI Internal EP to a Floating L3Out Using the CLI, on page 84
- Configuring Maximum Number of Paths for Redistribution of Routes in Fabric Using the CLI, on page 85
- Configuring Multiple Next-Hops Using the CLI, on page 85
- Verifying Floating L3Out Using the CLI, on page 85

Creating a VLAN Pool for Floating L3Out Using the CLI

This section demonstrates how to configure a VLAN pool specifically to use with the floating Layer 3 outside network connection (L3Out).

Note The VLAN pool for the L3Out must have a static VLAN range. It must also be the same for the VMware vSphere Distributed Switch (VDS) Virtual Machine Manager (VMM) domain and the Layer 3 domain. After you configure the VLAN pool, you configure the VMM and Layer 3 domains, adding the same VLAN pool to each domain.

To configure a VLAN pool for floating L3Out:

Example:

```
vlan-domain dom1
    vlan 300-400
    exit
```

What to do next

Create a VMM Domain Profile for VMware VDS. See the procedure Configuring a VMM Domain Profile for VMware VDS Using the CLI, on page 82.

Configuring a VMM Domain Profile for VMware VDS Using the CLI

Use this procedure to create a Virtual Machine Manager (VMM) profile for the VMware vSphere Distributed Switch (VDS) if you have not already done so and want to use floating Layer 3 Out network communication (L3Out).

Note To use a floating Layer 3 outside network connection (L3Out), you must configure a VLAN pool that has a static VLAN range for the VMM domain. Also, the VLAN pool must be the same as the VLAN pool of the L3Out domain. For example, both the range for the L3Out domain and the Virtual Machine Manager (VMM) domain must be 200-209.

To configure a VMM domain profile for VMware VDS:

Example:

```
vmware-domain vmmdom1
    vlan-domain member dom1
    vcenter 192.168.66.2 datacenter prodDC
      username administrator password *****
    configure-dvs
      exit
    exit
```

What to do next

Configure a floating L3Out. See procedure Configuring a Floating L3Out Using the CLI, on page 82.

Configuring a Floating L3Out Using the CLI

This section demonstrates how to create a floating L3Out.

Before you begin

You must have created the following:

- A VLAN pool for floating L3Out
- A VMM domain profile for VMware VDS

To configure a floating L3Out:

Example:

```
tenant t1
    vrf context vrf1
      exit
    l3out l3out
      vrf member vrf1
      exit
    external-l3 epg instp l3out l3out
      vrf member vrf1
      exit
      exit
leaf 101
  vrf context tenant t1 vrf vrf1 l3out l3out
  exit
leaf 101
     virtual-interface-profile ipv4 vlan 680 tenant Floating vrf Floating l3out CLI
        ip address 1.68.0.3/16
            physical-domain Floating-CP-L3out floating-addr 1.68.0.9/16
            exit
         vlan-domain member CP-L3
         exit
      virtual-interface-profile ipv6 vlan 680 tenant Floating vrf Floating l3out CLI
         ipv6 address 2000:68::2/64 preferred
            physical-domain Floating-CP-L3out floating-addr 2000:68::9/16
            vlan-domain member CP-L3
         exit
```

What to do next

- (Optional) Configuring a Secondary IP Using the CLI, on page 83.
- Verifying Floating L3Out Using the CLI, on page 85and that the port group exists on the VMware VDS.

Configuring a Secondary IP Using the CLI

This section demonstrates how to configure a secondary and floating secondary IP using the CLI.

To configure a secondary and floating secondary IP:

```
leaf 101
    virtual-interface-profile vlan 100 tenant t1 vrf v1
      ip address 10.1.1.1/24
      ip address 10.1.1.3/24 secondary
      ip address 11.1.1.1/24 secondary
      ip address 11.1.1.3/24 secondary
      vmm-domain mininet floating-addr 10.1.1.100/24
        ip address 11.1.1.100/24 secondary
        exit
```

```
    exit
  exit
```

Configuring the Avoidance of Suboptimal Traffic From an ACI Internal EP to a Floating L3Out Using the CLI

This section demonstrates how to configure next hop propagation and direct-attached host route advertising using the CLI

Before you begin

The following must be configured:

- For Next Hop Propagation, the floating L3Out must be in a physical domain, not in a VMM domain.
- A BD, EPG, and a contract between the EPG and L3Out EPG

Step 1 To configure next hop propagation:

Example:

```
tenant t1 vrf v1 route-map sap match
 prefix-list p1
  leaf 101
    vrf context tenant t1 vrf v1
      route-map sap
        match prefix-list p1
          set next-hop-unchanged
          exit
        exit
      exit
    exit
```

Step 2 Configuring direct-attached host route advertising:

Example:

```
leaf 101
    router bgp 100
      vrf member tenant t1 vrf v1
        redistribute static route-map r2
        redistribute attached-host route-map r1
        exit
      exit
    exit
```

Configuring Maximum Number of Paths for Redistribution of Routes in Fabric Using the CLI

The following example provides information on how to configure the BGP Max Path feature using the CLI.

Before you begin

Refer to the *Verified Scalability Guide for Cisco APIC* on the Cisco APIC documentation page for the acceptable values for the following fields.

Enter the following commands, where the `maximum-paths local` command is used specifically to configure the maximum number of paths for redistribution of routes in the fabric using the CLI:

```
apic1(config)# leaf 101
apic1(config-leaf)# template bgp address-family newAf tenant t1
apic1(config-bgp-af)# maximum-paths local 12
apic1(config-bgp-af)# exit
apic1(config-leaf)# exit
apic1#
```

Configuring Multiple Next-Hops Using the CLI

The following example provides information on how to configure multiple next-hops using the CLI.

Enter the following commands, where the `set next-hop-unchanged` and `set redist-multipath` commands are used specifically to configure the multiple next-hops using the CLI:

```
apic1(config)# leaf 101
apic1(config-leaf)# template route-profile test_rp tenant t1
apic1(config-leaf-template-route-profile)# set next-hop-unchanged
apic1(config-leaf-template-route-profile)# set redist-multipath
apic1(config-leaf-template-route-profile)# exit
apic1(config-leaf)# exit
apic1#
```

Verifying Floating L3Out Using the CLI

This section demonstrates how to verify a floating L3Out configuration.

Step 1 To verify floating L3Out on leaf nodes (anchor leaf):

In this example, the anchor leaf has the primary IP, secondary IP, and floating primary IP.

Example:

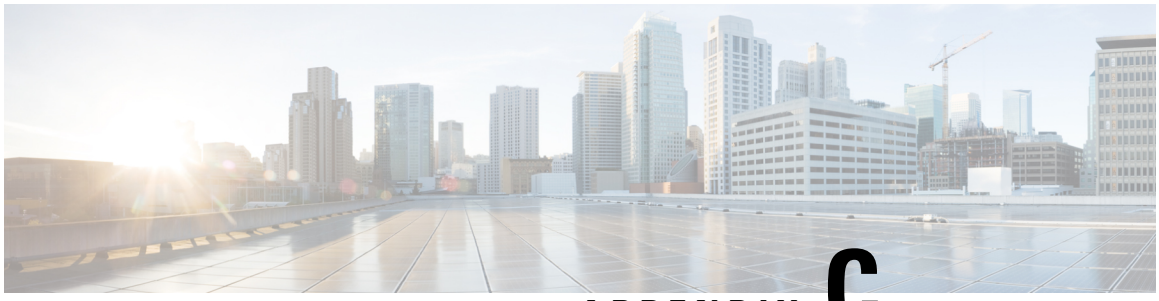
```
Switch# show ip interface brief vrf floating:vrf1
IP Interface Status for VRF "floating:vrf1"(9)
Interface Address   Interface Status
vlan14    192.168.1.254/24     protocol-up/link-up/admin-up
vlan17    192.168.2.254/24     protocol-up/link-up/admin-up
vlan49    172.16.1.251/24      protocol-up/link-up/admin-up
lo2             11.11.11.11/32       protocol-up/link-up/admin-up

Switch# show ip interface vlan49
IP Interface Status for VRF "floating:vrf1"
vlan49, Interface status: protocol-up/link-up/admin-up, iod: 110, mode: external
 IP address: 172.16.1.251, IP subnet: 172.16.1.0/24
 IP address: 172.16.1.250, IP subnet: 172.16.1.0/24 secondary anchor-floating-ip
 IP address: 172.16.1.254, IP subnet: 172.16.1.0/24 secondary
 IP broadcast address: 255.255.255.255
 IP primary address route-preference: 0, tag:

Switch# # show vlan id 49 extended

VLAN Name         Encap      Ports
--------------------------------------------------------------------------------
49 floating:vrf1:l3out- vxlan-14876650, Eth1/5, Eth1/6, Po1, Po2
        L3Out:vlan-208       vlan-208
```

Step 2 To verify floating L3Out on non-anchor leaf nodes:

When using a VMM domain, if there is no external VM connected, the non-anchor leaf does not have floating IP. When using a physical domain, the floating IP and VLAN are provisioned based on AEP. If the leaf has an AEP that contains the L3Out domain for floating L3Out, the floating IP is provisioned.

Example:

```
Switch# show ip interface brief vrf floating:vrf1
IP Interface Status for VRF "floating:vrf1"(6)
Interface Address  Interface Status
```



APPENDIX C

Configuring Floating L3Out Using the REST API

- Configuring a VLAN Pool for Floating L3Out Using the REST API, on page 87
- Configuring a VMM Domain Profile for VMware VDS Using the REST API, on page 88
- Configuring a Layer 3 Domain Using the REST API, on page 88
- Configuring a Floating L3Out Using the REST API, on page 89
- Configuring a Secondary IP Using the REST API, on page 89
- Configuring Same VLAN Encapsulation for IPv4 and IPv6 Using the REST API, on page 90
- Configuring the Avoidance of Suboptimal Traffic From an ACI Internal Endpoint to a Floating L3Out Using the REST API, on page 90
- Configuring Maximum Number of Paths for Redistribution of Routes in Fabric Using the REST API, on page 91
- Configuring Multiple Next-Hops Using the REST API, on page 92

Configuring a VLAN Pool for Floating L3Out Using the REST API

This section demonstrates how to configure a VLAN pool specifically to use with the floating Layer 3 outside network connection (L3Out).

To configure a VLAN pool for floating L3Out

Example:

```
<fvnsVlanInstP name="vlanPool1" allocMode="dynamic">
```

What to do next

Configure a VMM Domain Profile for VMware VDS. See procedure Configuring a VMM Domain Profile for VMware VDS Using the REST API, on page 88.

Configuring a VMM Domain Profile for VMware VDS Using the REST API

Use this procedure to create a Virtual Machine Manager (VMM) profile for the VMware vSphere Distributed Switch (VDS) if you have not already done so and want to use floating Layer 3 Out network communication (L3Out).

Note To use a floating Layer 3 outside network connection (L3Out), you must configure a VLAN pool that has a static VLAN range for the VMM domain. Also, the VLAN pool must be the same as the VLAN pool of the L3Out domain. For example, both the range for the L3Out domain and the Virtual Machine Manager (VMM) domain must be 200-209.

To configure a VMM domain profile for VMware VDS:

Example:

```
<polUni>
 <vmmProvP vendor="VMware">
  <vmmDomP name="FTD">
   <infraRsVlanNs tDn="uni/infra/vlanns-[vlanPool1]-dynamic" />
   <vmmUsrAccP name="creds" usr="administrator@vsphere.local" pwd="N1k@12345" />
   <vmmCtrlrP name="vcenter" rootContName="Datacenter" hostOrIp="10.197.145.212"
status="created,modified">
    <vmmRsAcc tDn="uni/vmmp-VMware/dom-FTD/usracc-creds" />
   </vmmCtrlrP>
  </vmmDomP>
 </vmmProvP>
</polUni>
```

What to do next

Configure a Layer 3 Domain. See procedure Configuring a Layer 3 Domain Using the REST API, on page 88.

Configuring a Layer 3 Domain Using the REST API

Create a Layer 3 domain before you create the Layer 3 outside network connection (L3Out).

To configure a Layer 3 domain:

Example:

```
<l3extDomP name="L3Dom">
```

```
 <infraRsVlanNs tDn="uni/infra/vlanns-[vlanPool1]-dynamic"  status=""/>
</l3extDomP>
```

What to do next

Configure a floating L3Out. See the procedure Configuring a Floating L3Out Using the REST API, on page 89.

Configuring a Floating L3Out Using the REST API

This section demonstrates how to configure a floating L3Out using the REST API.

To configure a floating L3Out:

Example:

```
<fvTenant name="t1" status="">
<fvCtx name="inb"/>
<l3extOut name="l3out" status="">
<l3extRsL3DomAtt tDn="uni/l3dom-L3Dom"/>
<l3extInstP name="instPP">
<fvRsCons tnVzBrCPName="inb-mgmt-allow-all-contract"/>
</l3extInstP>
<l3extLNodeP name="borderLeaf">
<l3extRsNodeL3OutAtt tDn="topology/pod-1/node-101" rtrId="10.10.10.11" status=""/>
<l3extRsNodeL3OutAtt tDn="topology/pod-1/node-102" rtrId="10.10.10.12" status=""/>
<l3extRsNodeL3OutAtt tDn="topology/pod-1/node-103" rtrId="10.10.10.13" status=""/>
<l3extLIfP name="phyDom">
<l3extVirtualLIfP descr="" encap="vlan-638" nodeDn="topology/pod-1/node-101" mode="regular"
addr="11.11.11.11/24" ifInstT='ext-svi' status="">
<l3extRsDynPathAtt tDn="uni/phys-Floating-L3out" floatingAddr="11.11.11.12/24" status="">
<l3extIp addr="12.12.12.100/24" status=""/>
 </l3extRsDynPathAtt>
<l3extIp addr="12.12.12.14/24" status=""/>
</l3extVirtualLIfP>
</l3extLIfP>
</l3extLNodeP>
<l3extRsEctx tnFvCtxName="inb"/>
</l3extOut>
</fvTenant>
```

What to do next

Verify the floating L3Out configuration.

Configuring a Secondary IP Using the REST API

This section demonstrates how to configure a secondary and floating secondary IP using the REST API.

To configure a secondary and floating secondary IP:

```
<l3extVirtualLIfP mtu="9000" addr="20.20.20.21/24" encap="vlan-1030" nodeDn="topology/pod-1/node-101"
 ifInstT="ext-svi" status="">
                      <l3extIp addr=”11.11.11.1/24”/>
                  <l3extRsDynPathAtt tDn="uni/phys-physDom1" floatingAddr ="20.20.20.1/24">
                    <l3extIp addr=”11.11.11.2/24”/>
                  </l3extRsDynPathAtt>
                </l3extVirtualLIfP>.
```

Configuring Same VLAN Encapsulation for IPv4 and IPv6 Using the REST API

This section demonstrates how to configure the same VLAN encap for both IPv4 and IPv6 address families using the same VMM domain.

To configure the same VLAN encap for IPv4 and IPv6:

```
<l3extOut enforceRtctrl="export" mplsEnabled="no" name="l3out1">
    <l3extLNodeP name="l3out1_nodeProfile">
    <l3extLIfP name="l3out1_interfaceProfile">
        <l3extVirtualLIfP addr="60.60.60.1/24" encap="vlan-100" nodeDn="topology/pod-1/node-101">
            <l3extRsDynPathAtt floatingAddr="60.60.60.100/24" tDn="uni/vmmp-VMware/dom-vmmDom1"/>
        </l3extVirtualLIfP>
    </l3extLIfP>

    <l3extLIfP name="l3out1_interfaceProfile2">
        <l3extVirtualLIfP addr="2021::1/64" encap="vlan-100" nodeDn="topology/pod-1/node-101">
            <l3extRsDynPathAtt floatingAddr="2021::100/64" tDn="uni/vmmp-VMware/dom-vmmDom1"/>
        </l3extVirtualLIfP>
    </l3extLIfP>
    </l3extLNodeP>
</l3extOut>
```

In the above configuration, the example addresses used are: 60.60.60.1 for IPv4 and 2021::1 for IPv6. VLAN encap is 100 for both.

Configuring the Avoidance of Suboptimal Traffic From an ACI Internal Endpoint to a Floating L3Out Using the REST API

This section demonstrate how to configure next hop propagation and direct-attached host route advertising using the REST API.

Before you begin

The following must be configured:

- For Next Hop Propagation, the floating L3Out must be in a physical domain, not in a VMM domain.
- A BD, EPG, and a contract between the EPG and L3Out EPG

Step 1 To configure next hop propagation:

Example:

```
<rtctrlSetNhUnchanged annotation="" childAction="" descr="" extMngdBy="" lcOwn="local"modTs="never"
 name="" nameAlias="" rn="nhunchanged" status="created" type="nh-unchanged" uid="0"/>
```

Step 2 Configuring direct-attached host route advertising:

Example:

```
<l3extRsRedistributePol annotation="" childAction=""
dn="uni/tn-neo/out-neoL3Out/rsredistributePol-[sap-rtmap]-am" extMngdBy="" forceResolve="yes"
lcOwn="local" modTs="never"monPolDn="" rType="mo" rn="" src="am" state="unformed" stateQual="none"
status="created"tCl="rtctrlProfile" tContextDn="" tDn="" tRn="" tType="name"
tnRtctrlProfileName="sap-rtmap" uid="0"/>
```

Configuring Maximum Number of Paths for Redistribution of Routes in Fabric Using the REST API

The two properties that enable you to configure more paths are `maxEcmp` and `maxEcmpIbgp` in the `bgpCtxAfPol` object. After you configure these two properties, they are propagated to the rest of your implementation. The ECMP policy is applied at the VRF level.

The following example provides information on how to configure the BGP Max Path feature using the REST API:

Before you begin

Refer to the *Verified Scalability Guide for Cisco APIC* on the Cisco APIC documentation page for the acceptable values for the following fields.

Post the following to configure the maximum number of paths for redistribution of routes in the fabric:

```
<fvTenant descr="" dn="uni/tn-t1" name="t1">
    <fvCtx name="v1">
        <fvRsCtxToBgpCtxAfPol af="ipv4-ucast" tnBgpCtxAfPolName="bgpCtxPol1"/>
    </fvCtx>
    <bgpCtxAfPol name="bgpCtxPol1" maxLocalEcmp="16"/>
</fvTenant>
```

Configuring Multiple Next-Hops Using the REST API

To configure multiple next-hops, post the following:

```
<fvTenant dn="uni/tn-t1">
    <rtctrlAttrP name="s1">
        <rtctrlSetRedistMultipath/>
    </rtctrlAttrP>
</fvTenant>
```