WAN and Other External Network Forwarding

This chapter contains the following sections:

Networking Domains

A fabric administrator creates domain policies that configure ports, protocols, VLAN pools, and encapsulation. These policies can be used exclusively by a single tenant, or shared. Once a fabric administrator configures domains in the ACI fabric, tenant administrators can associate tenant endpoint groups (EPGs) to domains.

The following networking domain profiles can be configured:

- VMM domain profiles (`vmmDomP`) are required for virtual machine hypervisor integration.
- Physical domain profiles (`physDomP`) are typically used for bare metal server attachment and management access.
- Bridged outside network domain profiles (`l2extDomP`) are typically used to connect a bridged external network trunk switch to a leaf switch in the ACI fabric.
- Routed outside network domain profiles (`l3extDomP`) are used to connect a router to a leaf switch in the ACI fabric.
- Fibre Channel domain profiles (`fcDomP`) are used to connect Fibre Channel VLANs and VSANs.

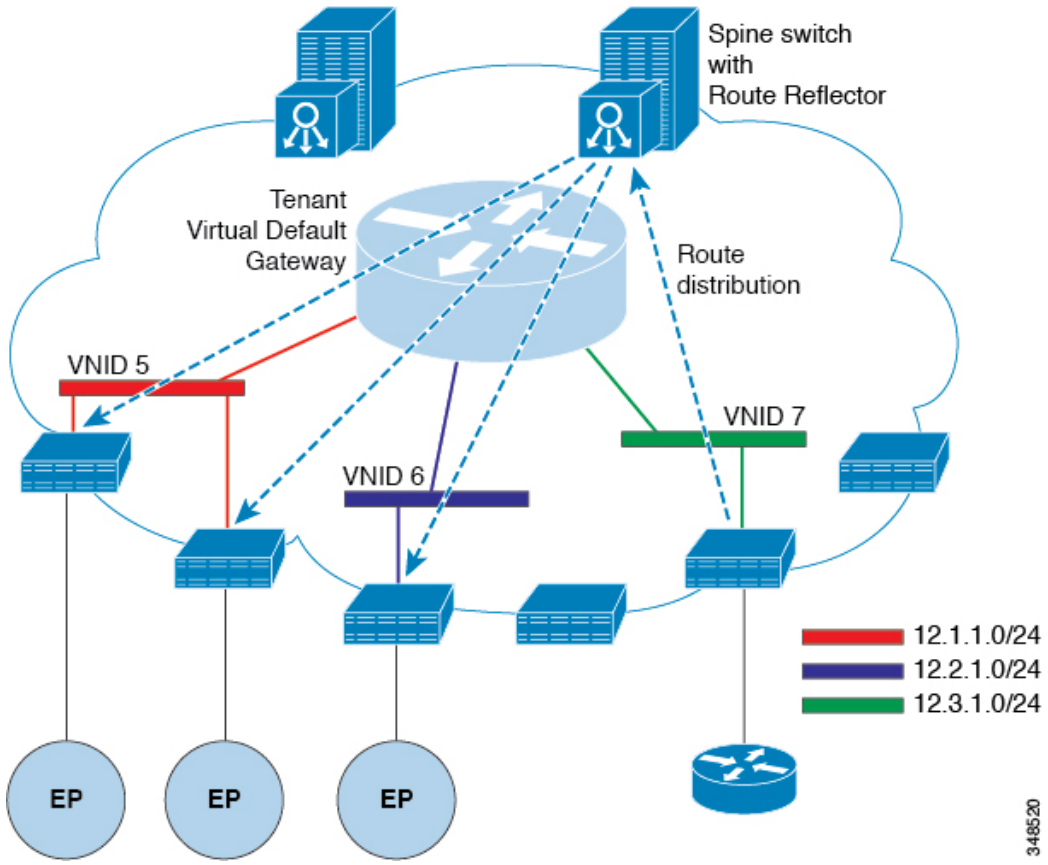
A domain is configured to be associated with a VLAN pool. EPGs are then configured to use the VLANs associated with a domain.

Note EPG port and VLAN configurations must match those specified in the domain infrastructure configuration with which the EPG associates. If not, the APIC will raise a fault. When such a fault occurs, verify that the domain infrastructure configuration matches the EPG port and VLAN configurations.

Router Peering and Route Distribution

As shown in the figure below, when the routing peer model is used, the leaf switch interface is statically configured to peer with the external router's routing protocol.

Figure 1: Router Peering



The routes that are learned through peering are sent to the spine switches. The spine switches act as route reflectors and distribute the external routes to all of the leaf switches that have interfaces that belong to the same tenant. These routes are longest prefix match (LPM) summarized addresses and are placed in the leaf switch's forwarding table with the VTEP IP address of the remote leaf switch where the external router is connected. WAN routes have no forwarding proxy. If the WAN routes do not fit in the leaf switch's forwarding table, the traffic is dropped. Because the external router is not the default gateway, packets from the tenant endpoints (EPs) are sent to the default gateway in the ACI fabric.

Route Import and Export, Route Summarization, and Route Community Match

Subnet route export or import configuration options can be specified according to the scope and aggregation options described below.

For routed subnets, the following scope options are available:

- Export Route Control Subnet: Controls the export route direction.
- Import Route Control Subnet: Controls the import route direction.

Note Import route control is supported for BGP and OSPF, but not EIGRP.

- External Subnets for the External EPG (Security Import Subnet): Specifies which external subnets have contracts applied as part of a specific external L3Out EPG (`l3extInstP`). For a subnet under the `l3extInstP` to be classified as an external EPG, the scope on the subnet should be set to "import-security". Subnets of this scope determine which IP addresses are associated with the `l3extInstP`. Once this is determined, contracts determine with which other EPGs that external subnet is allowed to communicate. For example, when traffic enters the ACI switch on the Layer 3 external outside network (`L3extOut`), a lookup occurs to determine which source IP addresses are associated with the `l3extInstP`. This action is performed based on Longest Prefix Match (LPM) so that more specific subnets take precedence over more general subnets.
- Shared Route Control Subnet: In a shared service configuration, only subnets that have this property enabled will be imported into the consumer EPG Virtual Routing and Forwarding (VRF). It controls the route direction for shared services between VRFs.
- Shared Security Import Subnet: Applies shared contracts to imported subnets. The default specification is External Subnets for the external EPG.

Routed subnets can be aggregated. When aggregation is not set, the subnets are matched exactly. For example, if 11.1.0.0/16 is the subnet, then the policy will not apply to a 11.1.1.0/24 route, but it will apply only if the route is 11.1.0.0/16. However, to avoid a tedious and error prone task of defining all the subnets one by one, a set of subnets can be aggregated into one export, import or shared routes policy. At this time, only 0/0 subnets can be aggregated. When 0/0 is specified with aggregation, all the routes are imported, exported, or shared with a different VRF, based on the selection option below:

- Aggregate Export: Exports all transit routes of a VRF (0/0 subnets).
- Aggregate Import: Imports all incoming routes of given L3 peers (0/0 subnets).

Note Aggregate import route control is supported for BGP and OSPF, but not for EIGRP.

- Aggregate Shared Routes: If a route is learned in one VRF but needs to be advertised to another VRF, the routes can be shared by matching the subnet exactly, or can be shared in an aggregate way according to a subnet mask. For aggregate shared routes, multiple subnet masks can be used to determine which

specific route groups are shared between VRFs. For example, 10.1.0.0/16 and 12.1.0.0/16 can be specified to aggregate these subnets. Or, 0/0 can be used to share all subnet routes across multiple VRFs.

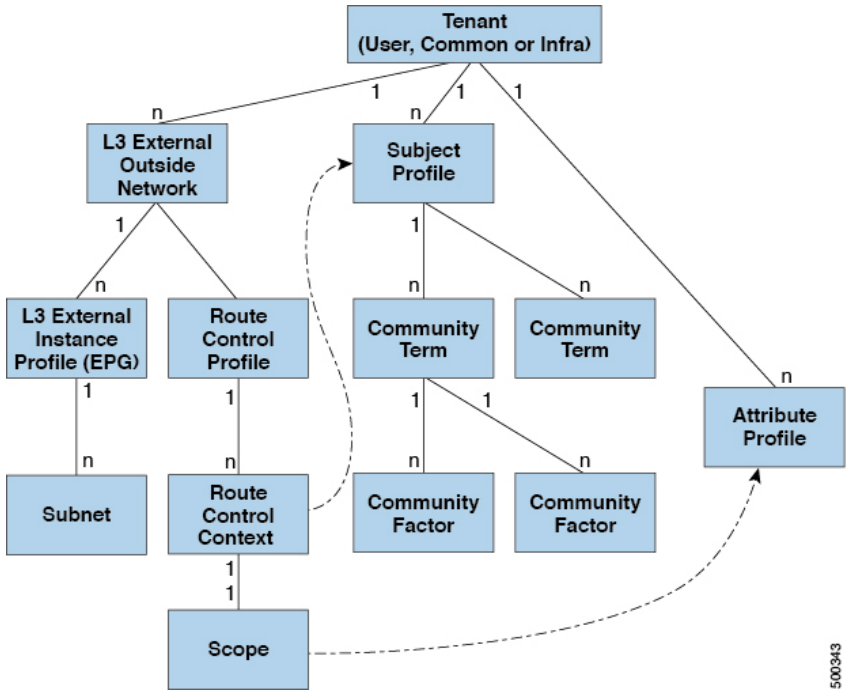
Note Routes shared between VRFs function correctly on Generation 2 switches (Cisco Nexus N9K switches with "EX" or "FX" on the end of the switch model name, or later; for example, N9K-93108TC-EX). On Generation 1 switches, however, there may be dropped packets with this configuration, because the physical ternary content-addressable memory (TCAM) tables that store routes do not have enough capacity to fully support route parsing.

Route summarization simplifies route tables by replacing many specific addresses with an single address. For example, 10.1.1.0/24, 10.1.2.0/24, and 10.1.3.0/24 are replaced with 10.1.0.0/16. Route summarization policies enable routes to be shared efficiently among border leaf switches and their neighbor leaf switches. BGP, OSPF, or EIGRP route summarization policies are applied to a bridge domain or transit subnet. For OSPF, inter-area and external route summarization are supported. Summary routes are exported; they are not advertised within the fabric. In the example above, when a route summarization policy is applied, and an EPG uses the 10.1.0.0/16 subnet, the entire range of 10.1.0.0/16 is shared with all the neighboring leaf switches.

Note When two `L3extOut` policies are configured with OSPF on the same leaf switch, one regular and another for the backbone, a route summarization policy configured on one `L3extOut` is applied to both L3extOut policies because summarization applies to all areas in the VRF.

As illustrated in the figure below, route control profiles derive route maps according to prefix-based and community-based matching.

Figure 2: Route Community Matching



The route control profile (`rtctrtlProfile`) specifies what is allowed. The Route Control Context specifies what to match, and the scope specifies what to set. The subject profile contains the community match specifications, which can be used by multiple `l3extOut` instances. The subject profile (`SubjP`) can contain multiple community terms each of which contains one or more community factors (communities). This arrangement enables specifying the following boolean operations:

- Logical `or` among multiple community terms
- Logical `and` among multiple community factors

For example, a community term called northeast could have multiple communities that each include many routes. Another community term called southeast could also include many different routes. The administrator could choose to match one, or the other, or both. A community factor type can be regular or extended. Care should be taken when using extended type community factors, to ensure there are no overlaps among the specifications.

The scope portion of the route control profile references the attribute profile (`rtctrlAttrP`) to specify what set-action to apply, such as preference, next hop, community, and so forth. When routes are learned from an `l3extOut`, route attributes can be modified.

The figure above illustrates the case where an `l3extOut` contains a `rtctrtlProfile`. A `rtctrtlProfile` can also exist under the tenant. In this case, the `l3extOut` has an interleak relation policy (`L3extRsInterleakPol`) that associates it with the `rtctrtlProfile` under the tenant. This configuration enables reusing the `rtctrtlProfile` for multiple `l3extOut` connections. It also enables keeping track of the routes the fabric learns from OSPF to which it gives BGP attributes (BGP is used within the fabric). A `rtctrtlProfile` defined under an L3extOut has a higher priority than one defined under the tenant.

The `rtctrtlProfile` has two modes: combinable, and global. The default combinable mode combines pervasive subnets (`fvSubnet`) and external subnets (`l3extSubnet`) with the match/set mechanism to render the route map. The global mode applies to all subnets within the tenant, and overrides other policy attribute settings. A global `rtctrtlProfile` provides permit-all behavior without defining explicit (0/0) subnets. A global `rtctrtlProfile` is used with non-prefix based match rules where matching is done using different subnet attributes such as community, next hop, and so on. Multiple `rtctrtlProfile` policies can be configured under a tenant.

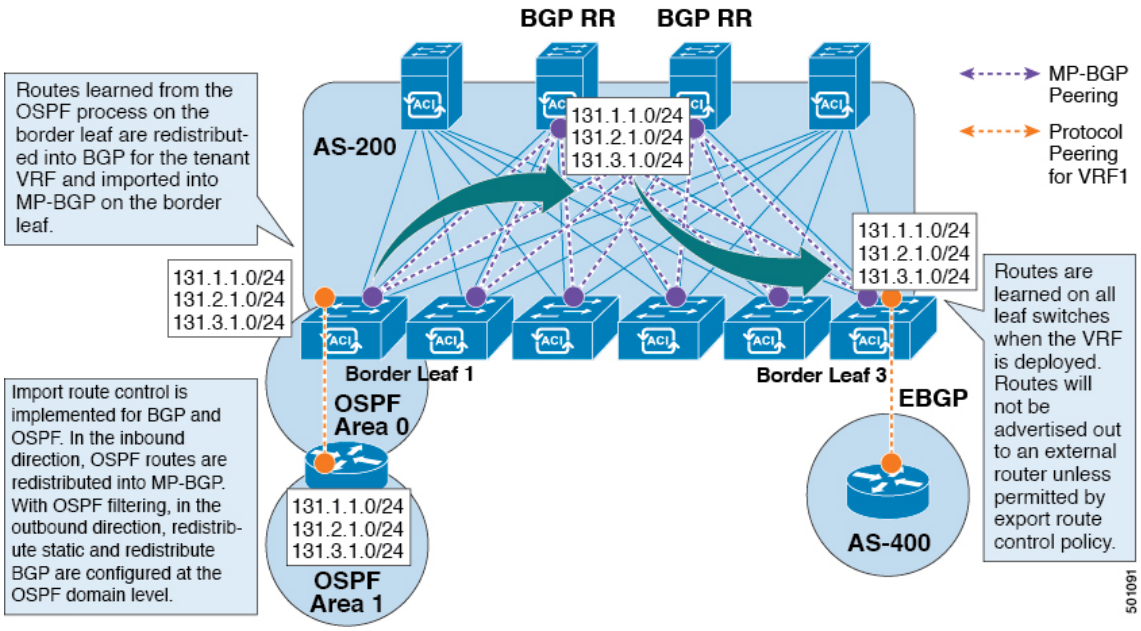
`rtctrtlProfile` policies enable enhanced default import and default export route control. Layer 3 Outside networks with aggregated import or export routes can have import/export policies that specify supported default-export and default–import, and supported 0/0 aggregation policies. To apply a `rtctrtlProfile` policy on all routes (inbound or outbound), define a global default `rtctrtlProfile` that has no match rules.

Note While multiple `l3extOut` connections can be configured on one switch, all Layer 3 outside networks configured on a switch must use the same `rtctrtlProfile` because a switch can have only one route map.

The protocol interleak and redistribute policy controls externally learned route sharing with ACI fabric BGP routes. Set attributes are supported. Such policies are supported per `L3extOut`, per node, or per VRF. An interleak policy applies to routes learned by the routing protocol in the `L3extOut`. Currently, interleak and redistribute policies are supported for OSPF v2 and v3. A route control policy `rtctrtlProfile` has to be defined as `global` when it is consumed by an interleak policy.

ACI Route Redistribution

Figure 3: ACI Route Redistribution



- The routes that are learned from the OSPF process on the border leaf are redistributed into BGP for the tenant VRF and they are imported into MP-BGP on the border leaf.
- Import route control is supported for BGP and OSPF, but not for EIGRP.
- Export route control is supported for OSPF, BGP, and EIGRP.
- The routes are learned on the border leaf where the VRF is deployed. The routes are not advertised to the External Layer 3 Outside connection unless it is permitted by the export route control.

Note When a subnet for a bridge domain/EPG is set to Advertise Externally, the subnet is programmed as a static route on a border leaf. When the static route is advertised, it is redistributed into the EPG's Layer 3 outside network routing protocol as an external network, not injected directly into the routing protocol.

Route Distribution Within the ACI Fabric

ACI supports the following routing mechanisms:

- Static Routes
- OSPFv2 (IPv4)
- OSPFv3 (IPv6)

- iBGP
- eBGP (IPv4 and IPv6)
- EIGRP (IPv4 and IPv6) protocols

ACI supports the VRF-lite implementation when connecting to the external routers. Using sub-interfaces, the border leaf can provide Layer 3 outside connections for the multiple tenants with one physical interface. The VRF-lite implementation requires one protocol session per tenant.

Within the ACI fabric, Multiprotocol BGP (MP-BGP) is implemented between the leaf and the spine switches to propagate the external routes within the ACI fabric. The BGP route reflector technology is deployed in order to support a large number of leaf switches within a single fabric. All of the leaf and spine switches are in one single BGP Autonomous System (AS). Once the border leaf learns the external routes, it can then redistribute the external routes of a given VRF to an MP-BGP address family VPN version 4 or VPN version 6. With address family VPN version 4, MP-BGP maintains a separate BGP routing table for each VRF. Within MP-BGP, the border leaf advertises routes to a spine switch, that is a BGP route reflector. The routes are then propagated to all the leaves where the VRFs (or private network in the APIC GUI's terminology) are instantiated.

External Layer 3 Outside Connection Types

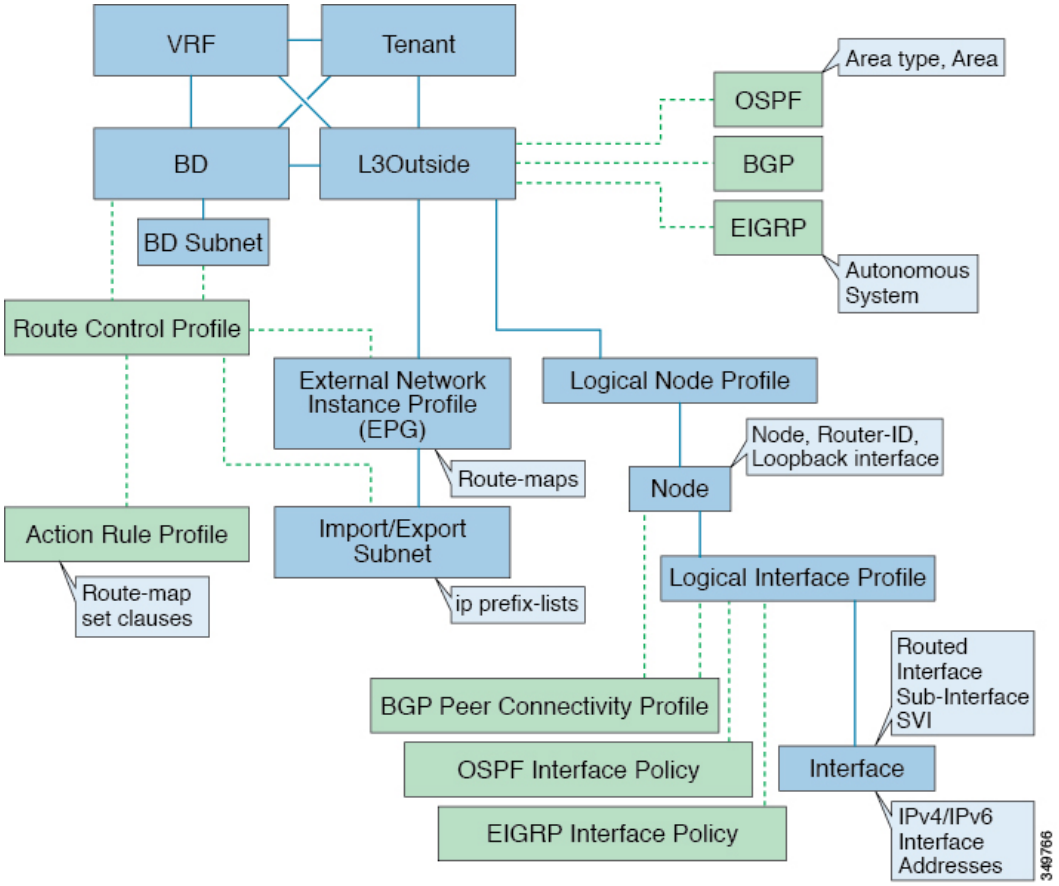
ACI supports the following external Layer 3 Outside connection options:

- Static Routing (supported for IPv4 and IPv6)
- OSPFv2 for normal and NSSA areas (IPv4)
- OSPFv3 for normal and NSSA areas (IPv6)
- iBGP (IPv4 and IPv6)
- eBGP (IPv4 and IPv6)
- EIGRP (IPv4 and IPv6)

The external Layer 3 Outside connections are supported on the following interfaces:

- Layer 3 Routed Interface
- Subinterface with 802.1Q tagging - With subinterface, you can use the same physical interface to provide a Layer 2 outside connection for multiple private networks.
- Switched Virtual Interface (SVI) - With an SVI interface, the same physical interface that supports Layer 2 and Layer 3 and the same physical interface can be used for a Layer 2 outside connection and a Layer 3 outside connection.

Figure 4: ACI Layer 3 Managed Objects



The managed objects that are used for the L3Outside connections are:

- External Layer 3 Outside (L3ext): Routing protocol options (OSPF area type, area, EIGRP autonomous system, BGP), private network, External Physical domain.
- Logical Node Profile: Profile where one or more nodes are defined for the external Layer 3 Outside connections. The configurations of the router-IDs and the loopback interface are defined in the profile.

Note Use the same router-ID for the same node across multiple external Layer 3 Outside connections.

Note Within a single L3Out, a node can only be part of one Logical Node Profile. Configuring the node to be a part of multiple Logical Node Profiles in a single L3Out might result in unpredictable behavior, such as having a loopback address pushed from one Logical Node Profile but not from the other. Use more path bindings under the existing Logical Interface Profiles or create a new Logical Interface Profile under the existing Logical Node Profile instead.

- Logical Interface Profile: IP interface configuration for IPv4 and IPv6 interfaces. It is supported on the Route Interfaces, Routed subinterfaces, and SVIs. The SVIs can be configured on physical ports, port-channels, or vPCs.
- OSPF Interface Policy: Includes details such as OSPF Network Type and priority.
- EIGRP Interface Policy: Includes details such as Timers and split horizon.
- BGP Peer Connectivity Profile: The profile where most BGP peer settings, remote-as, local-as, and BGP peer connection options are configured. You can associate the BGP peer connectivity profile with the logical interface profile or the loopback interface under the node profile. This determines the update-source configuration for the BGP peering session.
- External Layer 3 Outside EPG (l3extInstP): The external EPG is also referred to as the prefix-based EPG or InstP. The import and export route control policies, security import policies, and contract associations are defined in this profile. You can configure multiple external EPGs under a single L3Out. You may use multiple external EPGs when a different route or a security policy is defined on a single external Layer 3 Outside connections. An external EPG or multiple external EPGs combine into a route-map. The import/export subnets defined under the external EPG associate to the IP prefix-list match clauses in the route-map. The external EPG is also where the import security subnets and contracts are associated. This is used to permit or drop traffic for this L3out.
- Action Rules Profile: The action rules profile is used to define the route-map set clauses for the L3Out. The supported set clauses are the BGP communities (standard and extended), Tags, Preference, Metric, and Metric type.
- Route Control Profile: The route-control profile is used to reference the action rules profiles. This can be an ordered list of action rules profiles. The Route Control Profile can be referenced by a tenant BD, BD subnet, external EPG, or external EPG subnet.

There are more protocol settings for BGP, OSPF, and EIGRP L3Outs. These settings are configured per tenant in the ACI Protocol Policies section in the GUI.

Note When configuring policy enforcement between external EPGs (transit routing case), you must configure the second external EPG (InstP) with the default prefix 0/0 for export route control, aggregate export, and external security. In addition, you must exclude the preferred group, and you must use an any contract (or desired contract) between the transit InstPs.

About the Modes of Configuring Layer 3 External Connectivity

Because APIC supports multiple user interfaces (UIs) for configuration, the potential exists for unintended interactions when you create a configuration with one UI and later modify the configuration with another UI. This section describes considerations for configuring Layer 3 external connectivity with the APIC NX-OS style CLI, when you may also be using other APIC user interfaces.

When you configure Layer 3 external connectivity with the APIC NX-OS style CLI, you have the choice of two modes:

- Implicit mode, a simpler mode, is not compatible with the APIC GUI or the REST API.
- Named (or Explicit) mode is compatible with the APIC GUI and the REST API.

In either case, the configuration should be considered read-only in the incompatible UI.

How the Modes Differ

In both modes, the configuration settings are defined within an internal container object, the "L3 Outside" (or "L3Out"), which is an instance of the **l3extOut** class in the API. The main difference between the two modes is in the naming of this container object instance:

- Implicit mode—the naming of the container is implicit and does not appear in the CLI commands. The CLI creates and maintains these objects internally.
- Named mode—the naming is provided by the user. CLI commands in the Named Mode have an additional **l3Out** field. To configure the named L3Out correctly and avoid faults, the user is expected to understand the API object model for external Layer 3 configuration.

Note Except for the procedures in the *Configuring Layer 3 External Connectivity Using the Named Mode* section, this guide describes Implicit mode procedures.

Guidelines and Restrictions

- In the same APIC instance, both modes can be used together for configuring Layer 3 external connectivity with the following restriction: The Layer 3 external connectivity configuration for a given combination of tenant, VRF, and leaf can be done only through one mode.
- For a given tenant VRF, the policy domain where the External-l3 EPG can be placed can be in either the Named mode or in the Implicit mode. The recommended configuration method is to use only one mode for a given tenant VRF combination across all the nodes where the given tenant VRF is deployed for Layer 3 external connectivity. The modes can be different across different tenants or different VRFs and no restrictions apply.
- In some cases, an incoming configuration to a Cisco APIC cluster will be validated against inconsistencies, where the validations involve externally-visible configurations (northbound traffic through the L3Outs). An Invalid Configuration error message will appear for those situations where the configuration is invalid.
- The external Layer 3 features are supported in both configuration modes, with the following exception:
  - Route-peering and Route Health Injection (RHI) with a L4-L7 Service Appliance is supported only in the Named mode. The Named mode should be used across all border leaf switches for the tenant VRF where route-peering is involved.
- Layer 3 external network objects (l3extOut) created using the Implicit mode CLI procedures are identified by names starting with "__ui_" and are marked as read-only in the GUI. The CLI partitions these external-l3 networks by function, such as interfaces, protocols, route-map, and EPG. Configuration modifications performed through the REST API can break this structure, preventing further modification through the CLI.

For the steps to remove such objects, see *Troubleshooting Unwanted _ui_ Objects* in the *APIC Troubleshooting Guide*.

Controls Enabled for Subnets Configured under the L3Out Network Instance Profile

The following controls can be enabled for the subnets that are configured under the L3Out Network Instance Profile.

Table 1: Route Control Options

Route control Setting	Use	Options
Export Route Control	Controls which external networks are advertised out of the fabric using route-maps and IP prefix lists. An IP prefix list is created on the BL switch for each subnet that is defined. The export control policy is enabled by default and is supported for BGP, EIGRP, and OSPF.	Specific match (prefix and prefix length).
Import Route Control	Controls the subnets that are allowed into the fabric. Can include set and match rules to filter routes. Supported for BGP and OSPF, but not for EIGRP. If you enable the import control policy for an unsupported protocol, it is automatically ignored. The import control policy is not enabled by default, but you can enable it on the **Create L3Out** panel. On the **Identity** tab, enable **Route Control Enforcement: Import**.	Specific match (prefix and prefix length) .
Security Import Subnet	Used to permit the packets to flow between two prefix-based EPGs. Implemented with ACLs.	Uses the ACL match prefix or wildcard match rules.
Aggregate Export	Used to allow all prefixes to be advertised to the external peers. Implemented with the 0.0.0.0/ le 32 IP prefix-list.	Only supported for 0.0.0.0/0 subnet (all prefixes).
Aggregate Import	Used to allow all prefixes that are inbound from an external BGP peer. Implemented with the 0.0.0.0/0 le 32 IP prefix-list.	Only supported for the 0.0.0.0/0 subnet (all prefixes).

You may prefer to advertise all the transit routes out of an L3Out connection. In this case, use the aggregate export option with the prefix 0.0.0.0/0. Using this aggregate export option creates an IP prefix-list entry (permit

0.0.0.0/0 le 32) that the APIC system uses as a match clause in the export route-map. Use the **show route-map <outbound route-map>** and **show ip prefix-list <match-clause>** commands to view the output.

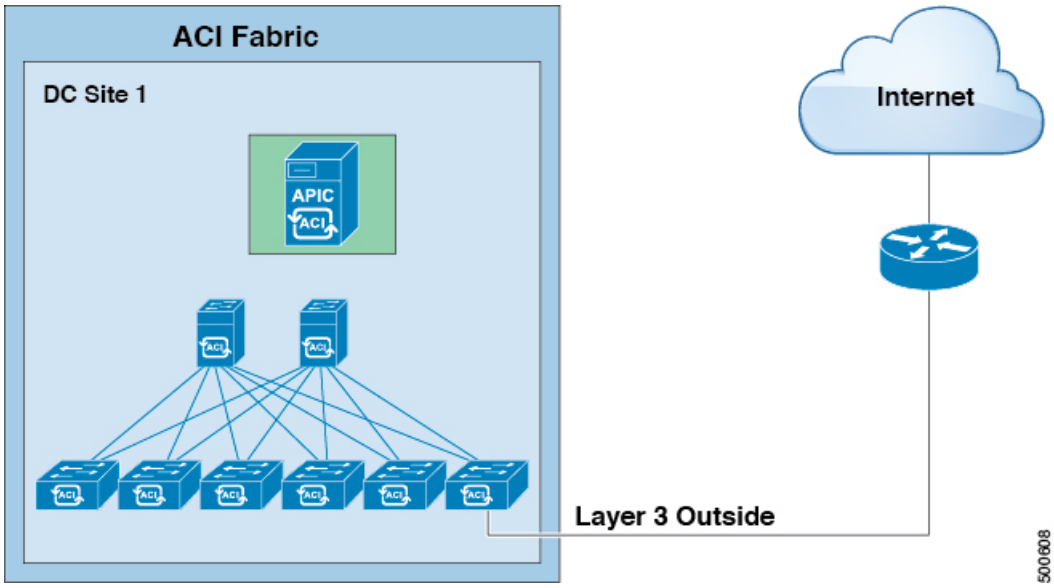
If you enable aggregate shared routes, if a route learned in one VRF must be advertised to another VRF, the routes can be shared by matching the subnet exactly, or they can be shared by using an aggregate subnet mask. Multiple subnet masks can be used to determine which specific route groups are shared between VRFs. For example, 10.1.0.0/16 and 12.1.0.0/16 can be specified to aggregate these subnets. Or, 0/0 can be used to share all subnet routes across multiple VRFs.

Note Routes shared between VRFs function correctly on Generation 2 switches (Cisco Nexus N9K switches with "EX" or "FX" on the end of the switch model name, or later; for example, N9K-93108TC-EX). On Generation 1 switches, however, there may be dropped packets with this configuration, because the physical ternary content-addressable memory (TCAM) tables that store routes do not have enough capacity to fully support route parsing.

ACI Layer 3 Outside Network Workflows

This workflow provides an overview of the steps required to configure a Layer 3 Outside (L3Out) network connection.

Figure 5: Layer 3 outside network connection



1. Prerequisites

- Ensure that you have read/write access privileges to the infra security domain.
- Ensure that the target leaf switches with the necessary interfaces are available.

Configure a Layer 3 Outside Network

Choose which of these L3Out scenarios you will use:

- For an L3Out that will be consumed within a single tenant, follow the instructions for configuring BGP or OSPF.
- For an L3Out that will be consumed (shared) among multiple tenants, follow the "Shared Layer 3 Out" guidelines.
- For an L3Out transit routing use case, follow ACI transit routing instructions.

  Note: This feature requires APIC release 1.2(1x) or later.