Routing Protocol Support

This chapter contains the following sections:

- About Routing Protocol Support, on page 1
- BGP External Routed Networks with BFD Support, on page 2
- OSPF External Routed Networks, on page 38
- EIGRP External Routed Networks, on page 42

About Routing Protocol Support

Routing within the Cisco ACI fabric is implemented using BGP (with BFD support) and the OSPF or EIGRP routing protocols.

IP source routing is not supported in the ACI fabric.

About Equal-Cost Multi-Path Routing in Cisco ACI

In Cisco Application Centric Infrastructure (ACI), all next-hops connected to a border leaf switch are considered as one equal-cost multi-path (ECMP) routing path when forwarded in hardware. Cisco ACI does not redistribute ECMP paths into BGP for directly connected next-hops, but does for recursive next-hops.

In the following example, border leaf switches 1 and 2 advertise the 10.1.1.0/24 route using the next-hop propagate and redistribute attached host features:

```
10.1.1.0/24
    through 192.168.1.1 (border leaf switch 1) -> ECMP path 1
    through 192.168.1.2 (border leaf switch 1) -> ECMP path 1
    through 192.168.1.3 (border leaf switch 1) -> ECMP path 1
    through 192.168.1.4 (border leaf switch 2) -> ECMP path 2
```

- ECMP path 1: 50% (next-hops 192.168.1.1, 192.168.1.2, 192.168.1.3)
- ECMP path 2: 50% (next-hop 192.168.1.4)

Note The percentage of traffic hashing on each next-hop is an approximation. The actual percentage varies.

This route entry on a non-border leaf switch results in two ECMP paths from the non-border leaf to each border leaf switch. This can result in disproportionate load balancing to the border leaf switches if the next-hops are not evenly distributed across the border leaf switches that advertise the routes.

Beginning with Cisco ACI release 6.0(2), you can use the next-hop propagate and redistribute attached host features to avoid sub-optimal routing in the Cisco ACI fabric. When these features are enabled, packet flows from a non-border leaf switch are forwarded directly to the leaf switch connected to the next-hop address. All next-hops are now used for ECMP forwarding from the hardware. In addition, Cisco ACI now redistributes ECMP paths into BGP for both directly connected next-hops and recursive next-hops.

In the following example, leaf switches 1 and 2 advertise the 10.1.1.0/24 route with the next-hop propagate and redistribute attached host features:

```
10.1.1.0/24
    through 192.168.1.1 (border leaf switch 1) -> ECMP path 1
    through 192.168.1.2 (border leaf switch 1) -> ECMP path 2
    through 192.168.1.3 (border leaf switch 1) -> ECMP path 3
    through 192.168.1.4 (border leaf switch 2) -> ECMP path 4
```

- ECMP path 1: 25% (next-hop 192.168.1.1)
- ECMP path 2: 25% (next-hop 192.168.1.2)
- ECMP path 3: 25% (next-hop 192.168.1.3)
- ECMP path 4: 25% (next-hop 192.168.1.4)

BGP External Routed Networks with BFD Support

The following sections provide more information on BGP external routed networks with BFD support.

Guidelines for Configuring a BGP Layer 3 Outside Network Connection

When configuring a BGP external routed network, follow these guidelines:

- The BGP direct route export behavior changed after release 3.2(1), where ACI does not evaluate the originating route type (such as static, direct, and so on) when matching export route map clauses. As a result, the "match direct" deny clause that is always included in the outbound neighbor route map no longer matches direct routes, and direct routes are now advertised based on whether or not a user-defined route map clause matches.

  Therefore, the direct route must be advertised explicitly through the route map. Failure to do so will implicitly deny the direct route being advertised.

- The **AS override** option in the **BGP Controls** field in the BGP Peer Connectivity Profile for an L3Out was introduced in release 3.1(2). It allows Cisco Application Centric Infrastructure (ACI) to overwrite a remote AS in the AS_PATH with ACI BGP AS. In Cisco ACI, it is typically used when performing transit routing from an eBGP L3Out to another eBGP L3Out with the same AS number.

  However, an issue arises if you enable the **AS override** option when the eBGP neighbor has a different AS number. In this situation, strip the peer-as from the AS_PATH when reflecting it to a peer.

- The **Local-AS Number** option in the BGP Peer Connectivity Profile is supported only for eBGP peering. This enables Cisco ACI border leaf switches to appear to be a member of another AS in addition to its real AS assigned to the fabric MP-BGP Route Reflector Policy. This means that the local AS number

must be different from the real AS number of the Cisco ACI fabric. When this feature is configured, Cisco ACI border leaf switches prepend the local AS number to the AS_PATH of the incoming updates and append the same to the AS_PATH of the outgoing updates. Prepending of the local AS number to the incoming updates can be disabled by the **no-prepend** setting in the **Local-AS Number Config**. The **no-prepend + replace-as** setting can be used to prevents the local AS number from being appended to the outgoing updates in addition to not prepending the same to the incoming updates.

- A router ID for an L3Out for any routing protocols cannot be the same IP address or the same subnet as the L3Out interfaces such as routed interface, sub-interface or SVI. However, if needed, a router ID can be the same as one of the L3Out loopback IP addresses. Starting from Cisco APIC Release 6.0(4), this restriction has been removed when the "Use Router ID for Loopback Address" option is not enabled.
- If you have multiple L3Outs of the same routing protocol on the same leaf switch in the same VRF instance, the router ID for those must be the same. If you need a loopback with the same IP address as the router ID, you can configure the loopback in only one of those L3Outs.
- There are two ways to define the BGP peer for an L3Out:
  - Through the BGP peer connectivity profile (**bgpPeerP**) at the logical node profile level (**l3extLNodeP**), which associates the BGP peer to the loopback IP address. When the BGP peer is configured at this level, a loopback address is expected for BGP connectivity, so a fault is raised if the loopback address configuration is missing.
  - Through the BGP peer connectivity profile (**bgpPeerP**) at the logical interface profile level (**l3extRsPathL3OutAtt**), which associates the BGP peer to the respective interface or sub-interface.
- It is recommended to use BGP default timers and leverage bidirectional forwarding detection (BFD) to get sub-second failure detection. Aggressive timers can cause BGP sessions to flap unexpectedly during CPU intense operations.
- You must configure an IPv6 address to enable peering over loopback using IPv6.
- Tenant networking protocol policies for BGP `l3extOut` connections can be configured with a maximum prefix limit that enables monitoring and restricting the number of route prefixes received from a peer. After the maximum prefix limit is exceeded, a log entry can be recorded, further prefixes can be rejected, the connection can be restarted if the count drops below the threshold in a fixed interval, or the connection is shut down. You can use only one option at a time. The default setting is a limit of 20,000 prefixes, after which new prefixes are rejected. When the reject option is deployed, BGP accepts one more prefix beyond the configured limit and the Cisco Application Policy Infrastructure Controller (APIC) raises a fault.

Note Cisco ACI does not support IP fragmentation. Therefore, when you configure Layer 3 Outside (L3Out) connections to external routers, or Multi-Pod connections through an Inter-Pod Network (IPN), it is recommended that the interface MTU is set appropriately on both ends of a link. On some platforms, such as Cisco ACI, Cisco NX-OS, and Cisco IOS, the configurable MTU value does not take into account the Ethernet headers (matching IP MTU, and excluding the 14-18 Ethernet header size), while other platforms, such as IOS-XR, include the Ethernet header in the configured MTU value. A configured value of 9000 results in a max IP packet size of 9000 bytes in Cisco ACI, Cisco NX-OS, and Cisco IOS, but results in a max IP packet size of 8986 bytes for an IOS-XR untagged interface.

For the appropriate MTU values for each platform, see the relevant configuration guides.

We highly recommend that you test the MTU using CLI-based commands. For example, on the Cisco NX-OS CLI, use a command such as `ping 1.1.1.1 df-bit packet-size 9000 source-interface ethernet 1/1`.

- When you stretch the L3Out across the parent leaf and the tier 2 leaf, ECMP is not supported

BGP Connection Types and Loopback Guidelines

The ACI supports the following BGP connection types and summarizes the loopback guidelines for them:

BGP Connection Type	Loopback required	Loopback same as Router ID	Static/OSPF route required
iBGP direct	No	Not applicable	No
iBGP loopback peering	Yes, a separate loopback per L3Out	No, if multiple Layer 3 out are on the same node	Yes
eBGP direct	No	Not applicable	No
eBGP loopback peering (multi-hop)	Yes, a separate loopback per L3Out	No, if multiple Layer 3 out are on the same node	Yes

BGP Protocol Peering to External BGP Speakers

ACI supports peering between the border leaves and the external BGP speakers using iBGP and eBGP. ACI supports the following connections for BGP peering:

- iBGP peering over OSPF
- eBGP peering over OSPF
- iBGP peering over direct connection

- eBGP peering over direct connection
- iBGP peering over static route

Note When OSPF is used with BGP peering, OSPF is only used to learn and advertise the routes to the BGP peering addresses. All route control applied to the Layer 3 Outside Network (EPG) are applied at the BGP protocol level.

ACI supports a number of features for iBGP and eBGP connectivity to external peers. The BGP features are configured on the **BGP Peer Connectivity Profile**.

The BGP peer connectivity profile features are described in the following table.

Note ACI supports the following BGP features. NX-OS BGP features not listed below are not currently supported in ACI.

Table 1: BGP Peer Connectivity Profile Features

BGP Features	Feature Description	NX-OS Equivalent Commands
Allow Self-AS	Works with Allowed AS Number Count setting.	**allowas-in**
Disable peer AS check	Disable checking of the peer AS number when advertising.	**disable-peer-as-check**
Next-hop self	Always set the next hop attribute to the local peering address.	**next-hop-self**
Send community	Send the community attribute to the neighbor.	**send-community**
Send community extended	Send the extended community attribute to the neighbor.	**send-community extended**
Password	The BGP MD5 authentication.	**password**
Allowed AS Number Count	Works with Allow Self-AS feature.	**allowas-in**
Disable connected check	Disable connected check for the directly connected EBGP neighbors (allowing EBGP neighbor peering from the loopbacks).	
TTL	Set the TTL value for EBGP multihop connections. It is only valid for EBGP.	**ebgp-multihop <TTL>**

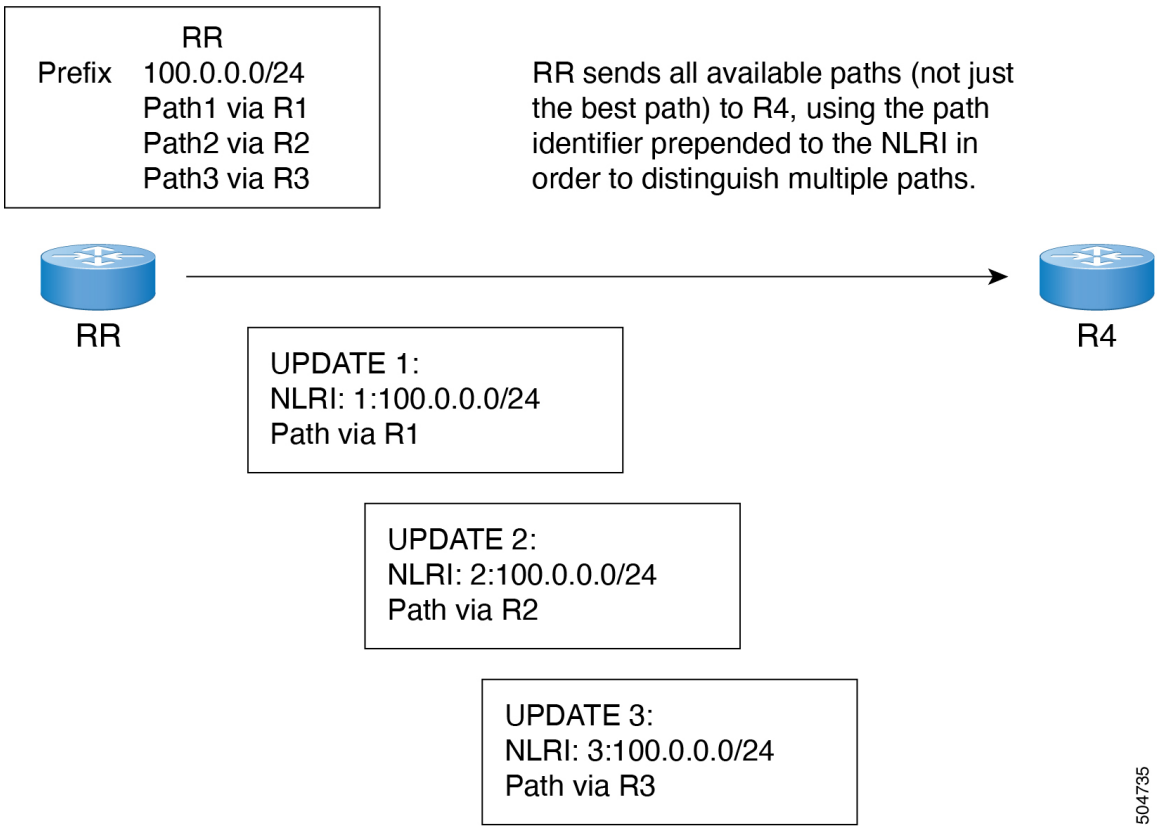
BGP Features	Feature Description	NX-OS Equivalent Commands
Autonomous System Number	Remote Autonomous System number of the peer.	**neighbor <x.x.x.x> remote-as**
Local Autonomous System Number Configuration	Options when using the Local AS feature. (No Prepend+replace-AS+dual-AS etc).	
Local Autonomous System Number	The local AS feature used to advertise a different AS number than the AS assigned to the fabric MP-BGP Route Reflector Profile. It is only supported for the EBGP neighbors and the local AS number must be different than the route reflector policy AS.	**local-as xxx <no-prepend> <replace-as> <dual-as>**
Site of Origin	The site-of-origin (SoO) is a BGP extended community attribute that is used to uniquely identify the site from which a route is learned in order to prevent routing loops.	**soo <value>**

BGP Additional-Paths

Beginning with the Cisco Application Policy Infrastructure Controller (APIC) 6.0(2) release, BGP supports the additional-paths feature, which allows the BGP speaker to propagate and receive multiple paths for the same prefix without the new paths replacing any previous paths. This feature allows BGP speaker peers to negotiate whether the peers support advertising and receiving multiple paths per prefix and advertising such paths. A special 4-byte path ID is added to the network layer reachability information (NLRI) to differentiate multiple paths for the same prefix sent across a peer session.

The following figure illustrates the BGP additional-paths receive capability:

Figure 1: BGP Route Advertisement with the Additional Paths Capability



The following limitations apply:

- Cisco APIC supports only the receive functionality.
- If you configure the BGP additional-paths receive feature after the session is established, the configuration takes effect on the next session flap.

Prior to the additional-paths receive feature, BGP advertised only one best path, and the BGP speaker accepted only one path for a given prefix from a given peer. If a BGP speaker received multiple paths for the same prefix within the same session, BGP used the most recent advertisement.

Configuring BGP External Routed Networks

Use the procedures in the following sections to configure BGP external routed networks.

Configuring BGP L3Out Using the GUI

Before you begin

The tenant, VRF, and bridge domain where you configure the BGP L3Out is already created, and you selected the **Configure BGP Policies** option when you were creating the VRF.

Procedure

Step 1 In the **Menu** bar, choose **Tenants** > **All Tenants**.

Step 2 In the Work pane, double-click the tenant's name.

Step 3 In the **Navigation** pane, expand ***Tenant_name*** > **Networking** > **L3Outs**.

Step 4 Right-click **L3Outs** and choose **Create L3Out**.

The **Create L3Out** wizard appears.

Step 5 Enter the necessary information in the **Identity** page of the **Create L3Out** wizard.

a) Enter the necessary information in the **Name**, **VRF** and **L3 Domain** fields.
b) In the area with the routing protocol check boxes, choose **BGP**.
c) Click **Next** to move to the **Nodes and Interfaces** window.

Step 6 Enter the necessary information in the **Nodes and Interfaces** page of the **Create L3Out** wizard.

a) In the **Layer 3** area, select **Routed**.
b) From the **Node ID** field drop-down menu, choose the node for the L3Out.

For the topology in these examples, use node `103`.

c) In the **Router ID** field, enter the router ID.
d) (Optional) You can configure another IP address for a loopback address, if necessary.

The **Loopback Address** field is automatically populated with the same entry that you provide in the **Router ID** field. This is the equivalent of the **Use Router ID for Loopback Address** option in previous builds. Enter a different IP address for a loopback address, if you don't want to use route ID for the loopback address, or leave this field empty if you do not want to use the router ID for the loopback address.

e) Enter necessary additional information in the **Nodes and Interfaces** page.

The fields shown in this page varies, depending on the options that you select in the **Layer 3** and **Layer 2** areas.

f) When you have entered the remaining additional information in the **Nodes and Interfaces** page, click **Next**.

The **Protocols** page appears.

Step 7 Enter the necessary information in the **Protocols** page of the **Create L3Out** wizard.

a) In the **BGP Loopback Policies** and **BGP Interface Policies** areas, enter the following information:

- **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.

Note
ACI does not support asdot or asdot+ format AS numbers.

b) Click **Next**.

The **External EPG** page appears.

Step 8 Enter the necessary information in the **External EPG** page of the **Create L3Out** wizard.

a) In the **Name** field, enter a name for the external network.
b) In the **Provided Contract** field, enter the name of a provided contract.
c) In the **Consumed Contract** field, enter the name of a consumed contract.
d) In the **Default EPG for all external networks** field, uncheck if you don't want to advertise all the transit routes out of this L3Out connection.

The Subnets area appears if you uncheck this box. Specify the desired subnets and controls as described in the following steps.

e) Click the **+** icon to expand **Subnet**, then perform the following actions in the **Create Subnet** dialog box.
f) In the **IP address** field, enter the IP address and network mask for the external network.

Note
Enter an IPv4 or IPv6 address depending upon what you have entered in earlier steps.

When creating the external subnet, you must configure either both the BGP loopbacks in the prefix EPG or neither of them. If you configure only one BGP loopback, then BGP neighborship is not established.

g) In the **Name** field, enter the name of the subnet.
h) In the **Scope** field, check the check boxes for **Export Route Control Subnet**, **Import Route Control Subnet**, and **Security Import Subnet**. Click **OK**.

Note
Check the **Import Route Control Subnet** check box if you wish to enforce import control with BGP.

i) Click **OK** when you have completed the necessary configurations in the **Create Subnet** window.
j) Click **Finish** to complete the necessary configurations in the **Create L3Out** wizard.

Step 9 (Optional) Navigate to the **BGP Peer Connectivity Profile** window to make additional configurations for the BGP external routed network, if necessary:

Tenants > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Logical Node Profiles** > ***log_node_prof_name*** > **BGP Peer <*address*>**

The **BGP Peer Connectivity Profile** page for this L3Out appears.

a) In the **BGP Controls** field, check the desired controls.

The peer controls specify which Border Gateway Protocol (BGP) attributes are sent to a peer. The peer control options are:

- **Allow Self AS**: Enables the autonomous number check on itself. This allows BGP peer to inject updates if the same AS number is being used.
- **AS override**: Enables the BGP AS override feature to override the default setting. The AS override function will replace the AS number from the originating router with the AS number of the sending BGP router in the AS Path of outbound routes. This feature can be enabled per feature per address family (IPv4 or IPv6).

  The **Disable Peer AS Check** check box must also be checked in order to enable the AS override feature.
- **Disable Peer AS Check**: Disables the peer autonomous number check. When the check box is checked, if the advertising router finds the AS number of the receiver in the AS path, it will not send the route to the receiver.

  The **Disable Peer AS Check** check box must be checked in order to enable the AS override feature.
- **Next-hop Self**: Sends the BGP next hop attribute to itself.

- **Send Community**: Sends the BGP community attribute to a peer.
- **Send Extended Community**: Sends the BGP extended community attribute to a peer.
- **Send Domain Path**: Sends the BGP domain path to a peer.

b) Put a check in the **Receive Additional Paths** check box to enable this eBGP L3Out peer to receive additional paths per prefix from other eBGP peers.

Without the **Receive Additional Paths** feature, eBGP allows a leaf switch to receive only one next hop from peers for a prefix.

Alternatively, you can configure all eBGP peers within the tenant's VRF instance to receive additional paths per prefix from other eBGP peers. For more information, see Configuring BGP Max Path Using the GUI, on page 13.

c) In the **Password** and **Confirm Password** field, enter the administrative password.

d) In the **Allow Self AS Number Count** field, choose the allowed number of occurrences of a local Autonomous System Number (ASN).

The range is from 1 to 10. The default is 3.

e) In the **Peer Controls** field, enter the neighbor check parameters.

The options are:

- **Bidirectional Forwarding Detection**: Enables BFD on the peer.
- **Disable Connected Check**: Disables the check for peer connection.

f) In the **Address Type Controls** field, configure the BGP IPv4/IPv6 address-family feature, if desired.

- **AF Mcast**: Check to enable the multicast address-family feature.
- **AF Ucast**: Check to enable the unicast address-family feature.

g) Note the entry in the **Routing Domain ID**, if necessary.

The value in the **Routing Domain ID** field reflects the global Domain ID Base value that was entered in the **BGP Route Reflector Policy** page. See About the BGP Domain-Path Feature for Loop Prevention for more information.

h) In the **EBGP Multihop TTL** field, 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.

i) In the **Weight for routes from this neighbor** field, choose the allowed weight for routes from the peer.

The weight assigned locally to the router is used to select the best path. The range is from 0 to 65535.

j) In the **Private AS Control** field, configure the private AS control.

These options are valid only when ACI BGP AS is a public AS number, or when the **Local-AS Number Config** with the **no-Prepend+replace-as** option is configured using a public AS number on the given BGP peer connectivity profile (the BGP neighbor configuration). The **replace-as** option is used to remove the actual local private AS from the AS_PATH because the **Private AS Control** feature does not remove its own local private AS.

The options are:

- **Remove all private AS**: In outgoing eBGP route updates to this neighbor, this option removes all private AS numbers from the AS_PATH. Use this option if you have private and public AS numbers in the eBGP route. The public AS number is retained.

If the neighbor remote AS is in the AS_PATH, this option is not applied.

To enable this option, **Remove private AS** must be enabled.

- **Remove private AS**: In outgoing eBGP route updates to this neighbor, this option removes all private AS numbers from the AS_PATH when the AS_PATH has only private AS numbers. Use this option, if you have *only* private AS numbers in the eBGP route.

  If the neighbor remote AS is in the AS_PATH, this option is not applied.

- **Replace private AS with local AS**: In outgoing eBGP route updates to this neighbor, this option replaces all private AS numbers in the AS_PATH with ACI local AS, regardless of whether a public AS or the neighbor remote AS is included in the AS_PATH.

  To enable this option, **Remove all private AS** must be enabled.

k) In the **BGP Peer Prefix Policy** field, select an existing peer prefix policy or create a new one.

The peer prefix policy defines how many prefixes can be received from a neighbor and the action to take when the number of allowed prefixes is exceeded. This feature is commonly used for external BGP peers, but can also be applied to internal BGP peers.

l) In the **Site of Origin** field, enter an extended community value to identify this peer.

The site-of-origin (SoO) extended community is a BGP extended community attribute that is used to identify routes that have originated from a site so that the readvertisement of that prefix back to the source site can be prevented. The SoO extended community uniquely identifies the site from which a router has learned a route. BGP can use the SoO value associated with a route to prevent routing loops.

Valid formats are:

- extended:as2-nn2:<2-byte number>:<2-byte number>

  For example: `extended:as2-nn2:1000:65534`

- extended:as2-nn4:<2-byte number>:<4-byte number>

  For example: `extended:as2-nn4:1000:6554387`

- extended:as4-nn2:<4-byte number>:<2-byte number>

  For example: `extended:as4-nn2:1000:65504`

- extended:ipv4-nn2:<IPv4 address>:<2-byte number>

  For example: `extended:ipv4-nn2:1.2.3.4:65515`

Note
When configuring the SoO for the User Tenant L3Outs, make sure not to configure the same SoO value as that of the global Fabric, Pod, or Multi-Site SoO configured within the ACI fabric. You can view the Fabric, Pod, and Multi-Site SoO values configured within the fabric by executing the following command on the switch:

```
show bgp process vrf overlay-1 | grep SOO
```

m) In the **Remote Autonomous System Number** field, choose a number that uniquely identifies the neighbor autonomous system.

The Autonomous System Number can be in 4-byte asplain format from 1 to 4294967295.

Note

ACI does not support asdot or asdot+ format AS numbers.

n) In the **Local-AS Number Config** field, choose the local Autonomous System Number (ASN) configuration.

Using a local AS number rather than the Global AS permits the routing devices in the associated network to appear to belong to the former AS. The configuration can be:

- **no-Prepend+replace-as+dual-as**: Does not allow prepending on local AS and is replaced with both AS numbers.

  You can prepend one or more autonomous system (AS) numbers at the beginning of an AS path. The AS numbers are added at the beginning of the path after the actual AS number from which the route originates has been added to the path. Prepending an AS path makes a shorter AS path look longer and therefore less preferable to BGP.

- **no-prepend**: Does not allow prepending on local AS.
- **no options**: Does not allow alteration of local AS.
- **no-Prepend+replace-as**: Does not allow prepending on local AS and is replaces AS number.

o) In the **Local-AS Number** field, choose the desired value.

Optionally required for the local autonomous system feature for eBGP peers. The local Autonomous System Number can be in 4-byte asplain format from 1 to 4294967295. ACI does not support asdot or asdot+ format AS numbers.

p) In the **Admin State** field, select **Disabled** or **Enabled**.

The **Admin State** field allows you to shut down the corresponding BGP neighbor. Using this feature shuts down the BGP sessions without the need to delete the BGP peer configuration.

Options are:

- **Disabled**: Disables the BGP neighbor's admin state.
- **Enabled**: Enables the BGP neighbor's admin state.

q) In the **Route Control Profile** field, configure route control policies per BGP peer.

Click + to configure the following:

- **Name**: The route control profile name.
- **Direction**: Choose one of the following options:
  - **Route Import Policy**
  - **Route Export Policy**

r) Click **Submit**.

Step 10 Navigate to **Tenants** > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** .

Step 11 Click the **Policy/Main** tab and perform the following actions:

a) (Optional) In the **Route Control Enforcement** field, check the **Import** check box.

Note
Check this check box if you wish to enforce import control with BGP.

b) Expand the **Route Control for Dampening** field, and choose the desired address family type and route dampening policy. Click **Update**.

In this step, the policy can be created either with step 4 or there is also an option to **Create route profile** in the drop-down list where the policy name is selected.

Step 12 Navigate to **Tenants** > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** .

Step 13 Right-click **Route map for import and export route control** and choose **Create Route map for import and export route control**.

Step 14 Enter the necessary information in this window, then click + in the Context area to bring up the **Create Route Control Context** window.

a) In the **Name** field, enter a name for the route control VRF instance.
b) From the **Set Attribute** drop-down list, choose **Create Action Rule Profile**.

When creating an action rule, set the route dampening attributes as desired.

Configuring BGP Max Path

The following feature enables you to add the maximum number of paths to the route table to enable equal cost, multipath load balancing.

Configuring BGP Max Path Using the GUI

Before you begin

The appropriate tenant and the BGP external routed network are created and available.

Procedure

Step 1 In the **Menu** bar, choose **Tenants** > **All Tenants**.

Step 2 In the Work pane, double-click the tenant's name.

Step 3 In the **Navigation** pane, expand ***Tenant_name*** > **Policies** > **Protocol** > **BGP** > **BGP Address Family Context**.

Step 4 Right-click **BGP Address Family Context** and choose **Create BGP Address Family Context Policy**.

Step 5 In the **Create BGP Address Family Context Policy** dialog box, perform the following tasks:

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

a) In the **Name** field, enter a name for the policy.
b) In the **eBGP Distance** field, enter a value for the administrative distance of eBGP routes.
c) In the **iBGP Distance** field, enter a value for the administrative distance of iBGP routes.
d) In the **Local Distance** field, enter a value for the local distance.
e) In the **eBGP Max ECMP** field, enter a value for the maximum number of equal-cost paths for eBGP load sharing.
f) In the **iBGP Max ECMP** field, enter a value for the maximum number of equal-cost paths for iBGP load sharing.
g) Put a check in the **Enable Host Route Leak** checkbox if you want to enable distributing EVPN type-2 (MAC/IP) host routes to the DCIG.

h) Put a check in the **BGP Add-Path Capability: Receive** checkbox if you want all eBGP peers within the tenant's VRF instance to receive additional paths per prefix from other eBGP peers.

Without the **BGP Add-Path Capability: Receive** feature, eBGP allows a leaf switch to receive only one next hop from peers per prefix.

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

Step 6 Click **Tenants** > ***tenant_name*** > **Networking** > **VRFs** > ***vrf_name***

Step 7 Review the configuration details of the subject VRF.

Step 8 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 9 Access the BGP Address Family Context you created in the **BGP Address Family Context** drop-down list and associate it with the subject VRF.

Step 10 Click **Submit**.

Configuring AS Path Prepend

Use the procedures in the following sections to configure AS Path Prepend.

Configuring AS Path Prepend

A BGP peer can influence the best-path selection by a remote peer by increasing the length of the AS-Path attribute. AS-Path Prepend provides a mechanism that can be used to increase the length of the AS-Path attribute by prepending a specified number of AS numbers to it.

AS-Path prepending can only be applied in the outbound direction using route-maps. AS Path prepending does not work in iBGP sessions.

The AS Path Prepend feature enables modification as follows:

Prepend	Appends the specified AS number to the AS path of the route matched by the route map. **Note** • You can configure more than one AS number. • 4 byte AS numbers are supported. • You can prepend a total 32 AS numbers. You must specify the order in which the AS Number is inserted into the AS Path attribute.
Prepend-last-as	Prepends the last AS numbers to the AS path with a range between 1 and 10.

The following table describes the selection criteria for implementation of AS Path Prepend:

Prepend	1	Prepend the specified AS number.
Prepend-last-as	2	Prepend the last AS numbers to the AS path.
DEFAULT	Prepend(1)	Prepend the specified AS number.

Configuring AS Path Prepend Using the GUI

Before you begin

A configured tenant.

SUMMARY STEPS

1. Log in to the APIC GUI, and on the menu bar, click **Tenants** > ***tenant_name*** > **Policies** > **Protocol** > **Set Rules** and right click **Create Set Rules for a Route Map**.
2. In the **Create Set Rules For A Route Map** dialog box, perform the following tasks:
3. Select the criterion **Prepend AS**, then click **+** to prepend AS numbers.
4. Enter the AS number and its order and then click **Update**. Repeat by clicking **+** again if multiple AS numbers must be prepended.
5. When you have completed the prepend AS number configurations, select the criterion **Prepend Last-AS** to prepend the last AS number a specified number of times.
6. Enter **Count** (1-10).
7. Click **OK**.
8. In the **Create Set Rules For A Route Map** window, confirm the listed criteria for the set rule based on AS Path and click **Finish.**
9. On the APIC GUI menu bar, click **Tenants** > ***tenant_name*** > **Policies** > **Protocol** > **Set Rules** and right click your profile.
10. Confirm the **Set AS Path** values the bottom of the screen.

DETAILED STEPS

Procedure

Step 1 Log in to the APIC GUI, and on the menu bar, click **Tenants** > ***tenant_name*** > **Policies** > **Protocol** > **Set Rules** and right click **Create Set Rules for a Route Map**.

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

Step 2 In the **Create Set Rules For A Route Map** dialog box, perform the following tasks:

a) In the **Name** field, enter a name.
b) Check the **Set AS Path** checkbox, then click **Next**.
c) In the **AS Path** window, click + to open the **Create Set AS Path** dialog box.

Step 3 Select the criterion **Prepend AS**, then click **+** to prepend AS numbers.

Step 4 Enter the AS number and its order and then click **Update**. Repeat by clicking **+** again if multiple AS numbers must be prepended.

Step 5 When you have completed the prepend AS number configurations, select the criterion **Prepend Last-AS** to prepend the last AS number a specified number of times.

Step 6 Enter **Count** (1-10).

Step 7 Click **OK**.

Step 8 In the **Create Set Rules For A Route Map** window, confirm the listed criteria for the set rule based on AS Path and click **Finish.**

Step 9 On the APIC GUI menu bar, click **Tenants** > *tenant_name* > **Policies** > **Protocol** > **Set Rules** and right click your profile.

Step 10 Confirm the **Set AS Path** values the bottom of the screen.

BGP External Routed Networks with AS Override

Use the procedures in the following sections to configure BGB external routed networks with AS override.

About BGP Autonomous System Override

Loop prevention in BGP is done by verifying the Autonomous System number in the Autonomous System Path. If the receiving router sees its own Autonomous System number in the Autonomous System path of the received BGP packet, the packet is dropped. The receiving router assumes that the packet originated from its own Autonomous System and has reached the same place from where it originated initially. This setting is the default to prevent route loops from occurring.

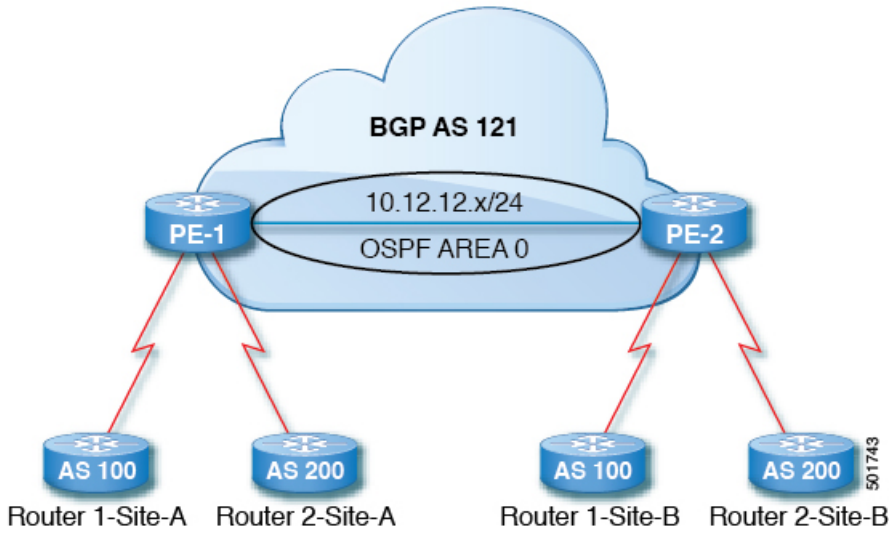
The default setting to prevent route loops from occurring could create an issue if you use the same Autonomous System number along various sites and disallow user sites with identical Autonomous System numbers to link by another Autonomous System number. In such a scenario, routing updates from one site is dropped when the other site receives them.

To prevent such a situation from occurring, beginning with the Cisco APIC Release 3.1(2m), you can now enable the BGP Autonomous System override feature to override the default setting. You must also enable the Disable Peer AS Check at the same time.

The Autonomous System override function replaces the Autonomous System number from the originating router with the Autonomous System number of the sending BGP router in the AS Path of the outbound routes. This feature can be enabled per feature per address family (IPv4 or IPv6).

The Autonomous System Override feature is supported with GOLF Layer 3 configurations and Non-GOLF Layer 3 configurations.

Figure 2: Example Topology Illustrating the Autonomous System Override Process



Router 1 and Router 2 are the two customers with multiple sites (Site-A and Site-B). Customer Router 1 operates under AS 100 and customer Router 2 operates under AS 200.

The above diagram illustrates the Autonomous System (AS) override process as follows:

1. Router 1-Site-A advertises route 10.3.3.3 with AS100.
2. Router PE-1 propagates this as an internal route to PE2 as AS100.
3. Router PE-2 prepends 10.3.3.3 with AS121 (replaces 100 in the AS path with 121), and propagates the prefix.
4. Router 2-Site-B accepts the 10.3.3.3 update.

Configuring BGP External Routed Network with Autonomous System Override Enabled Using the GUI

Before you begin

- The Tenant, VRF, Bridge Domain are created.
- The External Routed Network that is in a non-GOLF setting, a logical node profile, and the BGP peer connectivity profile are created.

Procedure

Step 1 On the menu bar, choose **Tenants** > ***Tenant_name*** > **Networking** > **L3Outs** > ***Non-GOLF Layer 3 Out_name*** > **Logical Node Profiles**.

Step 2 In the **Navigation** pane, choose the appropriate **BGP Peer Connectivity Profile**.

Step 3 In the **Work** pane, under **Properties** for the **BGP Peer Connectivity Profile**, in the **BGP Controls** field, perform the following actions:

a) Check the check box for the **AS override** field to enable the **Autonomous System override** function.

b) Check the check box for the **Disable Peer AS Check** field.

Note
You must check the check boxes for **AS override** and **Disable Peer AS Check** for the AS override feature to take effect.

c) Choose additional fields as required.

Step 4 Click **Submit**.

BGP Neighbor Shutdown and Soft Reset

Use the procedures in the following sections to configure BGB neighbor shutdown and soft reset.

About BGP Neighbor Shutdown and Soft Reset

Beginning with Release 4.2(1), support is now available for the following features:

• BGP Neighbor Shutdown, on page 18
• BGP Neighbor Soft Reset, on page 18

BGP Neighbor Shutdown

The BGP neighbor shutdown feature is similar to the neighbor shutdown command in NX-OS, which shuts down the corresponding BGP neighbor. Use this policy to disable and enable the BGP neighbor's admin state. Using this feature shuts down the BGP sessions without the need to delete the BGP peer configuration.

BGP Neighbor Soft Reset

Using the BGP route refresh capability, the BGP neighbor soft reset feature provides automatic support for a dynamic soft reset of inbound and outbound BGP routing table updates that are not dependent upon stored routing table update information. Use this policy to enable the soft dynamic inbound reset and soft outbound reset.

Configuring BGP Neighbor Shutdown Using the GUI

The following procedure describes how to use the BGP neighbor shutdown feature using the GUI.

Before you begin

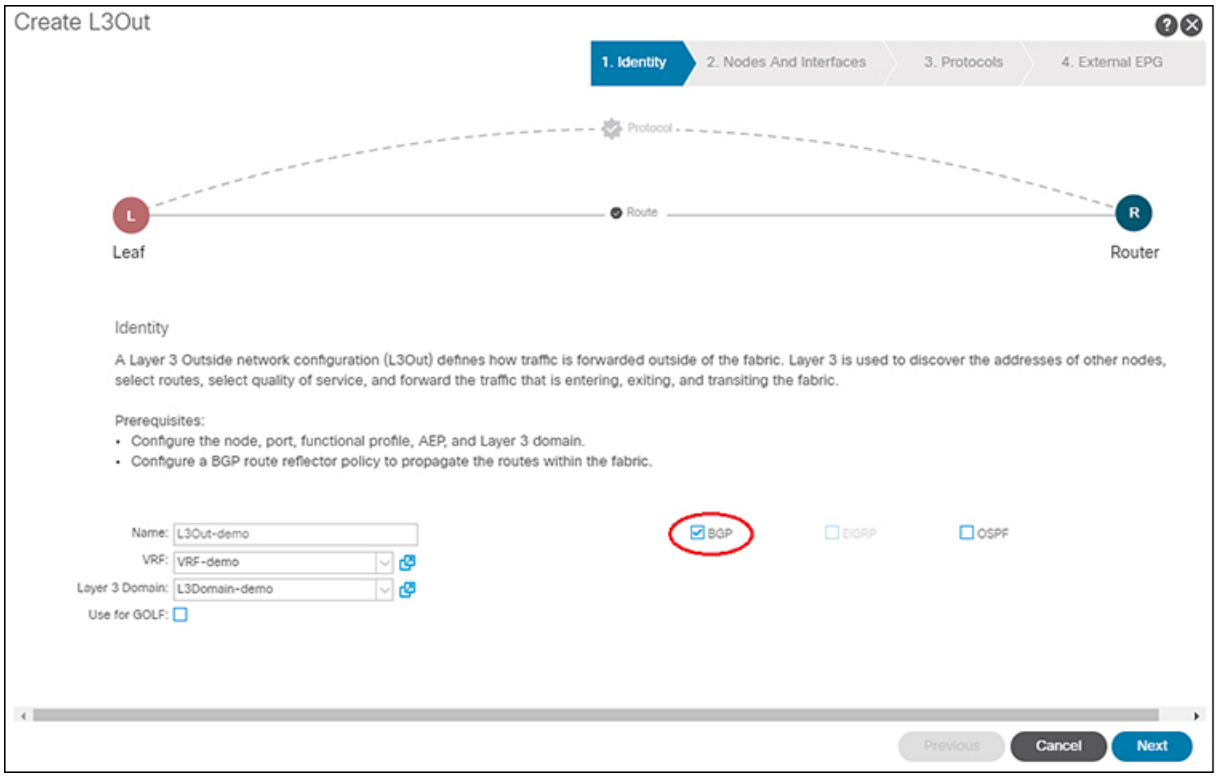
Complete the standard prerequisites before configuring an L3Out, such as:

- Configure the node, port, functional profile, AEP, and Layer 3 domain.
- Configure a BGP Route Reflector policy to propagate the routes within the fabric.

Procedure

Step 1 Create the L3Out and configure the BGP for the L3Out:

a) On the **Navigation** pane, expand **Tenant** and **Networking**.
b) Right-click **L3Outs** and choose **Create L3Out**.
c) Enter the necessary information to configure BGP for the L3Out.
You will select **BGP** in the **Identity** page in the L3Out creation wizard to configure the BGP protocol for this L3Out.



d) Continue through the remaining pages (**Nodes and Interfaces**, **Protocols**, and **External EPG**) to complete the configuration for the L3Out.

Step 2 After you have completed the L3Out configuration, configure the BGP neighbor shutdown:

a) Navigate to the BGP Peer Connectivity Profile screen:

Tenants > ***tenant*** > **Networking** > **L3Outs** > ***L3out-name*** > **Logical Node Profiles** > ***logical-node-profile-name*** > **Logical Interface Profiles** > ***logical-interface-profile-name*** > **BGP Peer Connectivity Profile** ***IP-address***

b) Scroll down to the **Admin State** field and make the appropriate selection in this field.

- **Disabled**: Disables the BGP neighbor's admin state.
- **Enabled**: Enables the BGP neighbor's admin state.

Configuring BGP Neighbor Soft Reset Using the GUI

The following procedure describes how to use the BGP neighbor soft reset feature using the GUI.

Before you begin

Complete the standard prerequisites before configuring an L3Out, such as:

- Configure the node, port, functional profile, AEP, and Layer 3 domain.
- Configure a BGP Route Reflector policy to propagate the routes within the fabric.

Procedure

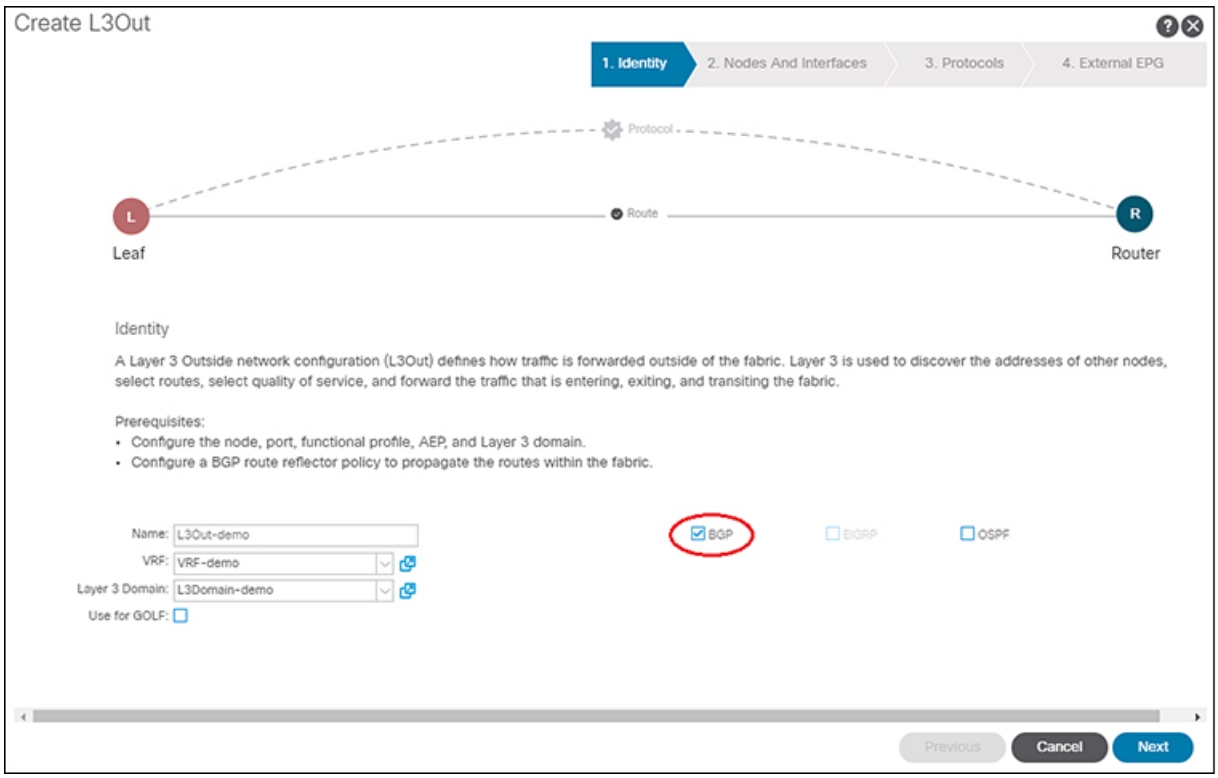
Step 1 Create the L3Out and configure the BGP for the L3Out:

a) On the **Navigation** pane, expand **Tenant** and **Networking**.

b) Right-click **L3Outs** and choose **Create L3Out**.

c) Enter the necessary information to configure BGP for the L3Out.

You will select **BGP** in the **Identity** page in the L3Out creation wizard to configure the BGP protocol for this L3Out.



d) Continue through the remaining pages (**Nodes and Interfaces**, **Protocols**, and **External EPG**) to complete the configuration for the L3Out.

Step 2 After you have completed the L3Out configuration, configure the BGP neighbor soft reset:

a) Navigate to the BGP Peer Entry screen:

Tenants > ***tenant*** > **Networking** > **L3Outs** > ***L3out-name*** > **Logical Node Profiles** > ***logical-node-profile-name*** > **Configured Nodes** > ***node*** > **BGP for VRF-*vrf-name*** > **Neighbors**

b) Right-click on the appropriate neighbor entry and select **Clear BGP Peer**.

The **Clear BGP** page appears.

c) In the **Mode** field, select **Soft**.

The **Direction** fields appear.

d) Select the appropriate value in the **Direction** field:

- **Incoming**: Enables the soft dynamic inbound reset.

- **Outgoing**: Enables the soft outbound reset.

Configuring Per VRF Per Node BGP Timer Values

Use the procedures in the following sections to configure per VRF per node BGP timer values.

Per VRF Per Node BGP Timer Values

Prior to the introduction of this feature, for a given VRF, all nodes used the same BGP timer values.

With the introduction of the per VRF per node BGP timer values feature, BGP timers can be defined and associated on a per VRF per node basis. A node can have multiple VRFs, each corresponding to a `fvCtx`. A node configuration (`l3extLNodeP`) can now contain configuration for BGP Protocol Profile (`bgpProtP`) which in turn refers to the desired BGP Context Policy (`bgpCtxPol`). This makes it possible to have a different node within the same VRF contain different BGP timer values.

For each VRF, a node has a `bgpDom` concrete MO. Its name (primary key) is the VRF, `<fvTenant>:<fvCtx>`. It contains the BGP timer values as attributes (for example, `holdIntvl, kaIntvl, maxAsLimit`).

All the steps necessary to create a valid Layer 3 Out configuration are required to successfully apply a per VRF per node BGP timer. For example, MOs such as the following are required: `fvTenant, fvCtx, l3extOut, l3extInstP, LNodeP, bgpRR`.

On a node, the BGP timer policy is chosen based on the following algorithm:

- If `bgpProtP` is specified, then use `bgpCtxPol` referred to under `bgpProtP`.
- Else, if specified, use `bgpCtxPol` referred to under corresponding `fvCtx`.
- Else, if specified, use the default policy under the tenant, for example, `uni/tn-<tenant>/bgpCtxP-default`.
- Else, use the `default` policy under tenant `common`, for example, `uni/tn-common/bgpCtxP-default`. This one is pre-programmed.

Configuring a Per VRF Per Node BGP Timer Using the Advanced GUI

When a BGP timer is configured on a specific node, then the BGP timer policy on the node is used and the BGP policy timer associated with the VRF is ignored.

Before you begin

A tenant and a VRF are already configured.

Procedure

Step 1 On the menu bar, choose **Tenant** > ***Tenant_name*** > **Policies** > **Protocol** > **BGP** > **BGP Timers**, then right click **Create BGP Timers Policy**.

Step 2 In the **Create BGP Timers Policy** dialog box, perform the following actions:

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

b) In the available fields, choose the appropriate values as desired. Click **Submit**.

A BGP timer policy is created.

Step 3 Navigate to **Tenant** > ***Tenant_name*** > **Networking** > **L3Outs**, and right-click **Create L3Out**.

The **Create L3Out** wizard appears. Create an L3Out with BGP enabled by performing the following actions.

Step 4 Enter the necessary information in the **Identity** window of the **Create L3Out** wizard.

a) In the **Name** field, enter a name for the L3Out.
b) From the **VRF** drop-down list, choose the VRF.
c) From the **L3 Domain** drop-down list, choose an external routed domain.
d) In the area with the routing protocol check boxes, check the **BGP** box.
e) Click **Next** to move to the **Nodes and Interfaces** window.
f) Continue through the remaining windows in the **Create L3Out** wizard to complete the L3Out creation process.

Step 5 After you have created the L3Out, navigate to the logical node profile in the L3Out that you just created: **Tenant** > ***Tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Logical Node Profiles** > ***LogicalNodeProfile-name*** .

Step 6 In the **Logical Node Profile** window, check next to **Create BGP Protocol Profile**.

The **Create Node Specific BGP Protocol Profile** window appears.

Step 7 In the **BGP Timers** field, from the drop-down list, choose the BGP timer policy that you want to associate with this specific node. Click **Submit**.

A specific BGP timer policy is now applied to the node.

Note
To associate an existing node profile with a BGP timer policy, right-click the node profile, and associate the timer policy.

If a timer policy is not chosen specifically in the **BGP Timers** field for the node, then the BGP timer policy that is associated with the VRF under which the node profile resides automatically gets applied to this node.

Step 8 To verify the configuration, in the **Navigation** pane, perform the following steps:

a) Expand **Tenant** > ***Tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Logical Node Profiles** > ***LogicalNodeProfile-name*** > **BGP Protocol Profile**.
b) In the **Work** pane, the BGP protocol profile that is associated with the node profile is displayed.

Troubleshooting Inconsistency and Faults

The following inconsistencies or faults could occur under certain conditions:

If different Layer 3 Outs (`l3Out`) are associated with the same VRF (`fvCtx`), and on the same node, the `bgpProtP` is associated with different policies (`bgpCtxPol`), a fault will be raised. In the case of the example below, both Layer 3 Outs (`out1` and `out2`) are associated with the same VRF (`ctx1`). Under `out1`, `node1` is associated with the BGP timer protocol `pol1` and under `out2`, `node1` is associated with a different BGP timer protocol `pol2`. This will raise a fault.

```
tn1
  ctx1
  out1
    ctx1
    node1
      protp pol1
```

```
out2
  ctx1
  node1
    protp pol2
```

If such a fault is raised, change the configuration to remove the conflict between the BGP timer policies.

BGP Graceful Restart Helper

In Border Gateway Protocol (BGP), Graceful Restart helps maintain routing stability by allowing routers to retain their forwarding state while the control plane is being restarted. Graceful Restart ensures uninterrupted packet forwarding during the control plane protocol restart. During this process, routing protocol updates are temporarily suppressed, allowing for a seamless transition without dropping existing connections. Graceful Restart is crucial for maintaining network reliability and performance during maintenance activities.

Cisco Application Centric Infrastructure (ACI) switches function as Graceful Restart (GR) Helpers rather than initiators. When a peer router undergoes operations that trigger GR, such as SSO, ISSU, or a BGP process restart, the ACI switch assists to ensure continued network operation by maintaining forwarding during the peer's BGP control plane downtime.

Graceful Restart Controls

When Graceful Restart is in progress, one router may be restarting its routing process and triggering Graceful Restart. Also, its peer may be stable but simply helping the Graceful Restart operation with the restarting router. The latter is called a Graceful Restart Receiving device or Graceful Restart Helper.

Graceful Restart Helper

The Graceful Restart helper supports nonstop forwarding during a router restart. When a neighbor router restarts, the helper allows neighboring routers to maintain forwarding of packets using stale routing information temporarily, preventing traffic disruption until the neighbor fully recovers and routing adjacencies are reestablished. Within a VRF instance, graceful restart helper capability is enabled by default.

In Graceful Restart, there are two major timers. One is called restart timer, which is configured and advertised by a restarting router to inform its peer of the maximum time it will take for the restarting router to finish restarting its routing protocol. A Graceful Restart Helper device will delete all stale routes after this timer is expired by assuming the restarting device failed to restart its routing protocol. Hence, this timer is not configured on ACI (Graceful Restart Helper). Another timer, called a stale timer, is configured and used by the Graceful Restart Helper device. See *Stale Interval*.

Stale Interval (sec)

When a Graceful Restart is in progress, the routes previously received from the peer are still used for forwarding but marked as stale. Once the session between two routers is re-established and route information is synced again, all the stale routes are deleted and the latest routes from the latest exchange are used. The **Stale Interval** is a timer to delete those stale routes in case the session is not re-established within this interval. This interval is applied locally. The default value is 300 seconds.

BGP Graceful Restart Helper supported scenarios

In Cisco ACI, BGP Graceful Restart Helper mode supports when a peer device undergoes these scenarios:

- Stateful switchover (SSO)
- In-Service Software Upgrade (ISSU)

- BGP process restart

The peer router can act as a Graceful Restart Helper, maintaining forwarding while the restarting device recovers its BGP state.

BGP Graceful Restart Helper guidelines

BGP Graceful Restart capability is exchanged in an open message during the initial BGP session establishment between two routers. Both the restarter, the router performing the restart, and the its helper, the ACI leaf, must exchange this capability to enable graceful restart for that specific session.

Modifying BGP Graceful Restart Helper

By default, BGP GR Helper mode is enabled on Cisco ACI leaf, but you can modify the stale interval values with this procedure.

Follow these steps to modify BGP Graceful Restart Helper:

Before you begin

Advertise the **BGP Graceful Restart** capability and **Restart Timer** with the peer and receive the capability from the neighbor.

Procedure

Step 1 On the menu bar, choose **Tenants** > **All Tenants** , then in the work pane, double click the tenant's name.

Step 2 Follow one of these sets of substeps to enable a BGP timers policy.

Follow these substeps to enable a BGP timers policy on all BGP peers in a particular VRF instance:

a) In the navigation pane, choose ***tenant_name*** > **Networking** > **VRFs** > ***vrf_name***.
b) In the **Policy** tab, choose **Create BGP Timers Policy** in the **BGP Timers** dropdown menu.

Follow these substeps to enable a BGP timers policy per node per VRF instance:

a) ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Logical Node Profiles** > ***logical_node_profile_name***.
b) In the **Policy** tab, place a checkmark in **Create BGP Protocol Profile** box.
c) Choose **Create BGP Timers Policy** in the **BGP Timers** dropdown menu.

Step 3 In the **Create BGP Timers Policy** dialog box, perform the following actions:

a) In the **Name** field, enter the BGP Timers policy name.
b) In the **Graceful Restart Controls** field, place a checkmark in the **Graceful Restart Helper** box.
c) (Optional) Enter the desired value for the **Stale Interval (sec)** field.

The default value is 300 seconds.

d) Click **Submit**.

A BGP timer policy is created.

Step 4 If you created the BGP timers policy per node per VRF instance, click **Submit** in the **Create Node Specific BGP Protocol Profile** dialog box.

Step 5 Click **Submit**.

Configuring BFD Support

Use the procedures in the following sections to configure BFD support.

Bidirectional Forwarding Detection

Use Bidirectional Forwarding Detection (BFD) to provide sub-second failure detection times in the forwarding path between Cisco Application Centric Infrastructure (ACI) fabric border leaf switches configured to support peering router connections.

BFD is particularly useful in the following scenarios:

- When the peering routers are connected through a Layer 2 device or a Layer 2 cloud where the routers are not directly connected to each other. Failures in the forwarding path may not be visible to the peer routers. The only mechanism available to control protocols is the hello timeout, which can take tens of seconds or even minutes to time out. BFD provides sub-second failure detection times.
- When the peering routers are connected through a physical media that does not support reliable failure detection, such as shared Ethernet. In this case too, routing protocols have only their large hello timers to fall back on.
- When many protocols are running between a pair of routers, each protocol has its own hello mechanism for detecting link failures, with its own timeouts. BFD provides a uniform timeout for all the protocols, which makes convergence time consistent and predictable.

Observe the following BFD guidelines and limitations:

- Beginning with Cisco APIC release 3.1(1), BFD between leaf and spine switches is supported on fabric-interfaces for IS-IS. In addition, BFD feature on spine switch is supported for OSPF and static routes.
- Beginning with Cisco APIC release 5.2(4), the BFD feature is supported for static routes that are reachable using secondary IPv4/IPv6 subnets. Static BFD session cannot be sourced from a secondary subnet of L3Out interface if there are more than one addresses configured in the subnet. Shared subnet address (used for vPC scenario) and floating IP address used for floating L3Out are allowed as additional addresses in the subnet and are automatically skipped, and are not used to source static BFD session.

Note Modifying the secondary address that is being used for sourcing the session is allowed by adding a new address in the same subnet and later removing the previous one.

- BFD is supported on modular spine switches that have -EX and -FX line cards (or newer versions), and BFD is also supported on the Nexus 9364C non-modular spine switch (or newer versions).
- BFD between vPC peers is not supported.
- Beginning with Cisco APIC release 5.0(1), BFD multihop is supported on leaf switches. The maximum number of BFD sessions is unchanged, as BFD multihop sessions are now included in the total.
- Beginning with Cisco APIC release 5.0(1), Cisco ACI supports C-bit-aware BFD. The C-bit on incoming BFD packets determines whether BFD is dependent or independent of the control plane.
- BFD over iBGP is not supported for loopback address peers.

- BFD sub interface optimization can be enabled in an interface policy. One sub-interface having this flag will enable optimization for all the sub-interfaces on that physical interface.
- BFD for BGP prefix peer not supported.

Note Cisco ACI does not support IP fragmentation. Therefore, when you configure Layer 3 Outside (L3Out) connections to external routers, or Multi-Pod connections through an Inter-Pod Network (IPN), it is recommended that the interface MTU is set appropriately on both ends of a link. On some platforms, such as Cisco ACI, Cisco NX-OS, and Cisco IOS, the configurable MTU value does not take into account the Ethernet headers (matching IP MTU, and excluding the 14-18 Ethernet header size), while other platforms, such as IOS-XR, include the Ethernet header in the configured MTU value. A configured value of 9000 results in a max IP packet size of 9000 bytes in Cisco ACI, Cisco NX-OS, and Cisco IOS, but results in a max IP packet size of 8986 bytes for an IOS-XR untagged interface.

For the appropriate MTU values for each platform, see the relevant configuration guides.

We highly recommend that you test the MTU using CLI-based commands. For example, on the Cisco NX-OS CLI, use a command such as `ping 1.1.1.1 df-bit packet-size 9000 source-interface ethernet 1/1`.

Optimizing BFD on Subinterfaces

You can optimize BFD on subinterfaces. BFD creates sessions for all configured subinterfaces. BFD sets the subinterface with the lowest configured VLAN ID as the master subinterface and that subinterface uses the BFD session parameters of the parent interface. The remaining subinterfaces use the slow timer.

If the optimized subinterface session detects an error, BFD marks all subinterfaces on that physical interface as down.

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

Note If one of the subinterfaces flap, subinterfaces on that physical interface are impacted and will go down for a second.

Configuring Bidirectional Forwarding Detection on a Secondary IP Address Using the GUI

This procedure configures bidirectional forwarding detection (BFD) on a secondary IP address using the GUI.

Procedure

Step 1 On the menu bar, choose **Tenants** > **All Tenants**.

Step 2 In the Work pane, double-click the tenant's name.

Step 3 In the Navigation pane, choose ***tenant_name*** > **Networking** > **L3Outs** > ***l3out_name*** > **Logical Node Profiles** > ***node_profile_name*** > **Logical Interface Profiles** > ***interface_profile_name***.

Step 4 In the Work pane, choose **Policy** > **Routed Sub-interfaces**, **Policy** > **Routed Interfaces**, or **Policy** > **SVI**, as appropriate.

Step 5 Double-click an interface to edit its properties.

Step 6 Perform one of the following substeps depending on the type of interface:

a) If the interface is a routed sub-interface or routed interface, or a Switch Virtual Interface (SVI) with the **Path Type** set to **Port** or **Direct Port Channel**, in the **IPv4 secondary/IPv6 Additional Addresses** table, click the **+**, enter the IP address and subnet, and click **Submit**.

b) If the interface is a Switch Virtual Interface (SVI) with the **Path Type** set to **Virtual Port Channel**, in the **Side B IPv4 Secondary/IPv6 Additional Addresses** table, click the **+**, enter the IP address and subnet, and click **OK**.

Step 7 In the Navigation pane, choose ***tenant_name*** > **Networking** > **L3Outs** > ***l3out_name*** > **Logical Node Profiles** > ***node_profile_name*** > **Configured Nodes** > ***node_name***.

Step 8 In the **Static Routes** table, click the **+** and perform the following substeps:

a) In the **Prefix** field, enter the static route IP address and mask that is assigned to the outside network.

b) Put a check in the **BFD** check box.

c) In the **Next Hop Addresses** table, click the **+** and in the **Next Hop Address** field, enter an IP address that is reachable by the secondary IP address that you specified for the interface.

d) Fill out the remaining fields as necessary.

e) Click **OK**.

Step 9 Fill out the remaining fields as necessary.

Step 10 Click **Submit**.

Configuring BFD Globally on Leaf Switch Using the GUI

Procedure

Step 1 On the menu bar, choose **Fabric** > **Access Policies**.

Step 2 In the **Navigation** pane, expand **Policies** > **Switch** > **BFD**.
There are two types of bidirectional forwarding detection (BFD) configurations available:

- BFD IPV4
- BFD IPV6

For each of these BFD configurations, you can choose to use the default policy or create a new one for a specific switch (or set of switches).

Note
By default, the APIC controller creates default policies when the system comes up. These default policies are global, bi-directional forwarding detection (BFD) configuration polices. You can set attributes within that default global policy in the **Work** pane, or you can modify these default policy values. However, once you modify a default global policy, note that your changes affect the entire system (all switches). If you want to use a specific configuration for a particular switch (or set of switches) that is not the default, create a switch profile as described in the next step.

Step 3 To create a switch profile for a specific global BFD policy (which is not the default), in the **Navigation** pane, expand **Switches** > **Leaf Switches** > **Profiles**.
The **Leaf Switches - Profiles** screen appears in the **Work** pane.

Step 4 On the right side of the **Work** pane, under the acttions icon, select **Create Leaf Profile**.

The **Create Leaf Profile** dialog box appears.

Step 5 In the **Create Leaf Profile** dialog box, perform the following actions:

a) In the **Name** field, enter a name for the leaf switch profile.
b) (Optional) In the **Description** field, enter a description of the profile.
c) (Optional) In the **Leaf Selectors** toolbar, click **+**
d) Enter the appropriate values for **Name** (name the switch), **Blocks** (select the switch), and **Policy Group** (select **Create Access Switch Policy Group**).
The **Create Access Switch Policy Group** dialog box appears where you can specify the Policy Group identity properties.

Step 6 (If configuring a leaf selector) In the **Create Access Switch Policy Group** dialog box, perform the following actions:

a) In the **Name** field, enter a name for the policy group.
b) (Optional) In the **Description** field, enter a description of the policy group.
c) Choose a BFD policy type (**BFD IPV4 Policy** or **BFD IPV6 Policy**), then select a value (**default** or **Create BFD Global Ipv4 Policy** for a specific switch or set of switches).
d) Click **Update**.

Step 7 Click **Next** to advance to **Associations**.

(Optional) In the **Associations** menu, you can associate the leaf profile with leaf interface profiles and access module profiles.

Step 8 Click **Finish**.

Another way to create a BFD global policy is to right-click on either **BFD IPV4** or **BFD IPV6** in the **Navigation** pane.

Step 9 To view the BFD global configuration you created, in the **Navigation** pane, expand **Policies** > **Switch** > **BFD**.

Configuring BFD Globally on Spine Switch Using the GUI

Procedure

Step 1 On the menu bar, choose **Fabric** > **Access Policies**.

Step 2 In the **Navigation** pane, expand **Policies** > **Switch** > **BFD**.
There are two types of bidirectional forwarding detection (BFD) configurations available:

- BFD IPV4
- BFD IPV6

For each of these BFD configurations, you can choose to use the default policy or create a new one for a specific switch (or set of switches).

Note
By default, the APIC controller creates default policies when the system comes up. These default policies are global, bi-directional forwarding detection (BFD) configuration polices. You can set attributes within that default global policy in the **Work** pane, or you can modify these default policy values. However, once you modify a default global policy, note that your changes affect the entire system (all switches). If you want to use a specific configuration for a particular switch (or set of switches) that is not the default, create a switch profile as described in the next step.

Step 3 To create a spine switch profile for a specific global BFD policy (which is not the default), in the **Navigation** pane, expand **Switches** > **Spine Switches** > **Profiles**.
The **Spine Switches - Profiles** screen appears in the **Work** pane.

Step 4 On the right side of the **Work** pane, under the acttions icon, select **Create Spine Profile**.
The **Create Spine Profile** dialog box appears.

Step 5 In the **Create Spine Profile** dialog box, perform the following actions:

a) In the **Name** field, enter a name for the switch profile.
b) In the **Description** field, enter a description of the profile. (This step is optional.)
c) (Optional) In the **Spine Selectors** toolbar, click **+**
d) Enter the appropriate values for **Name** (name the switch), **Blocks** (select the switch), and **Policy Group** (select **Create Spine Switch Policy Group**).
The **Create Spine Switch Policy Group** dialog box appears where you can specify the Policy Group identity properties.

Step 6 (If configuring a spine selector) In the **Create Spine Switch Policy Group** dialog box, perform the following actions:

a) In the **Name** field, enter a name for the policy group.
b) (Optional) In the **Description** field, enter a description of the policy group.
c) Choose a BFD policy type (**BFD IPV4 Policy** or **BFD IPV6 Policy**), then select a value (**default** or **Create BFD Global Ipv4 Policy** for a specific switch or set of switches).
d) Click **Update**.

Step 7 Click **Next** to advance to **Associations**.

(Optional) In the **Associations** menu, you can associate the spine profile with spine interface profiles.

Step 8 Click **Finish**.

Another way to create a BFD global policy is to right-click on either **BFD IPV4** or **BFD IPV6** in the **Navigation** pane.

Step 9 To view the BFD global configuration you created, in the **Navigation** pane, expand the **Policies** > **Switch** > **BFD**.

Configuring BFD Interface Override Using the GUI

There are three supported interfaces (routed Layer 3 interfaces, the external SVI interface, and the routed sub-interfaces) on which you can configure an explicit bi-directional forwarding detection (BFD) configuration. If you don't want to use the global configuration, yet you want to have an explicit configuration on a given interface, you can create your own global configuration, which gets applied to all the interfaces on a specific switch or set of switches. This interface override configuration should be used if you want more granularity on a specific switch on a specific interface.

Note When a BFD interface policy is configured over a parent routed interface, by default all of its routed sub-interfaces with the same address family as that of the parent interface will inherit this policy. If any of the inherited configuration needs to be overridden, configure an explicit BFD interface policy on the sub-interfaces. However, if **Admin State** or **Echo Admin State** is disabled on the parent interface, the property cannot be overridden on the sub-interfaces.

Before you begin

A tenant has already been created.

Procedure

Step 1 On the menu bar, choose **Tenant**.

Step 2 In the **Navigation** pane (under Quick Start), expand the Tenant you created ***Tenant_name*** > **Networking** > **L3Outs**.

Step 3 Right-click on **L3Outs** and select **Create L3Out**.
The **Create L3Out** wizard appears.

Step 4 Enter the necessary information in the **Identity** window of the **Create L3Out** wizard.

a) Enter the necessary information in the **Name**, **VRF** and **L3 Domain** fields.
b) In the area with the routing protocol check boxes, choose **BGP**.
c) Click **Next** to move to the **Nodes and Interfaces** window.

Step 5 Enter the necessary information in the **Nodes and Interfaces** window of the **Create L3Out** wizard.

a) In the **Layer 3** area, select **Routed**.
b) From the **Node ID** field drop-down menu, choose the node for the L3Out.

For the topology in these examples, use node `103`.

c) In the **Router ID** field, enter the router ID.
d) (Optional) You can configure another IP address for a loopback address, if necessary.

The **Loopback Address** field is automatically populated with the same entry that you provide in the **Router ID** field. This is the equivalent of the **Use Router ID for Loopback Address** option in previous builds. Enter a different IP address for a loopback address, if you don't want to use route ID for the loopback address, or leave this field empty if you do not want to use the router ID for the loopback address.

e) Enter necessary additional information in the **Nodes and Interfaces** window.

The fields shown in this window varies, depending on the options that you select in the **Layer 3** and **Layer 2** areas.

f) When you have entered the remaining additional information in the **Nodes and Interfaces** window, click **Next**.

The **Protocols** window appears.

Step 6 Enter the necessary information in the **Protocols** window of the **Create L3Out** wizard.

a) In the **BGP Loopback Policies** and **BGP Interface Policies** areas, enter the following information:

- **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 zero.
- **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.

Note
ACI does not support asdot or asdot+ format AS numbers.

b) In the **OSPF** area, choose the default OSPF policy, a previously created OSPF policy, or **Create OSPF Interface Policy**.

c) Click **Next**.

The **External EPG** window appears.

Step 7 Enter the necessary information in the **External EPG** window of the **Create L3Out** wizard.

a) In the **Name** field, enter a name for the external network.
b) In the **Provided Contract** field, enter the name of a provided contract.
c) In the **Consumed Contract** field, enter the name of a consumed contract.
d) In the **Default EPG for all external networks** field, uncheck if you don't want to advertise all the transit routes out of this L3Out connection.

The Subnets area appears if you uncheck this box. Specify the desired subnets and controls as described in the following steps.

e) Click **Finish** to complete the necessary configurations in the **Create L3Out** wizard.

Step 8 Navigate to **Tenants** > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Logical Node Profiles** > ***logical_node_profile_name*** > **Logical Interface Profiles** > ***logical_interface_profile_name***

Step 9 In the **Logical Interface Profile** window, scroll down to the **Create BFD Interface Profile** field, then check the box next to this field.

Step 10 In the **Create BFD Interface Profile** window, enter BFD details.

- In the **Authentication Type** field, choose **No authentication** or **Keyed SHA1**.

  If you choose to authenticate (by selecting Keyed SHA1), enter the **Authentication Key ID**, enter the **Authentication Key** (password), then confirm the password by re-entering it next to **Confirm Key**.

- In the **BFD Interface Policy** field, select either the **common/default** configuration (the default BFD policy), or create your own BFD policy by selecting **Create BFD Interface Policy**.

  If you select **Create BFD Interface Policy**, the **Create BFD Interface Policy** dialog box appears where you can define the BFD interface policy values.

Step 11 Click **SUBMIT**.

Step 12 To see the configured interface level BFD policy, navigate to **Policies** > **Protocol** > **BFD**.

Configuring BFD Consumer Protocols Using the GUI

This procedure provides the steps to enable bi-directional forwarding detection (BFD) in the consumer protocols (OSPF, BGP, EIGRP, Static Routes, and IS-IS), which are consumers of the BFD feature. To consume the BFD on these protocols, you must enable a flag in them.

Note These four consumer protocols are located in the left navigation pane under **Tenant** > **Policies** > **Protocol**.

Before you begin

A tenant has already been created.

Procedure

Step 1 Create an L3Out using the **Create L3Out** wizard.

Step 2 On the menu bar, choose **Tenant**.

Step 3 To configure BFD in the BGP protocol, in the **Navigation** pane (under Quick Start), expand the Tenant you created ***Tenant_name*** > **Policies** > **Protocol** > **BGP** > **BGP Peer Prefix**.

Step 4 On the right side of the **Work** pane, under **ACTIONS**, select **Create BGP Peer Prefix Policy**.
The **Create BGP Peer Prefix Policy** dialog box appears.

Note
You can also right-click on **BGP Peer Prefix** from the left navigation pane to select **Create BGP Peer Prefix** to create the policy.

Step 5 Enter a name in the **Name** field and provide values in the remaining fields to define the BGP peer prefix policy.

Step 6 Click **Submit**.
The BGP peer prefix policy you created now appears under **BGP Peer Prefix** in the left navigation pane.

Step 7 Navigate 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**.

Step 8 In the **BGP Peer Connectivity Profile** window, scroll down to the BGP Peer Prefix Policy field and select the BGP peer prefix policy that you just created.

Step 9 In the **Peer Controls** field, select **Bidirectional Forwarding Detection** to enable BFD on the BGP consumer protocol (or uncheck the box to disable BFD).

Step 10 To configure BFD in the OSPF protocol, in the **Navigation** pane, go to **Policies** > **Protocol** > **OSPF** > **OSPF Interface**.

Step 11 On the right side of the **Work** pane, under **ACTIONS**, select **Create OSPF Interface Policy**.
The **Create OSPF Interface Policy** dialog box appears.

Note
You can also right-click on **OSPF Interface** from the left navigation pane and select **Create OSPF Interface Policy** to create the policy.

Step 12 Enter a name in the **Name** field and provide values in the remaining fields to define the OSPF interface policy.

Step 13 In the **Interface Controls** section of this dialog box, you can enable or disable BFD. To enable it, check the box next to **BFD**, which adds a flag to the OSPF consumer protocol, shown as follows (or uncheck the box to disable BFD).

Step 14 Click **Submit**.

Step 15 To configure BFD in the EIGRP protocol, in the **Navigation** pane, go back to ***tenant_name*** > **Policies** > **Protocol** > **EIGRP** > **EIGRP Interface**.

Step 16 On the right side of the **Work** pane, under **ACTIONS**, select **Create EIGRP Interface Policy**.
The **Create EIGRP Interface Policy** dialog box appears.

Note
You can also right-click on **EIRGP Interface** from the left navigation pane and select **Create EIGRP Interface Policy** to create the policy.

Step 17 Enter a name in the **Name** field and provide values in the remaining fields to define the OSPF interface policy.

Step 18 In the **Control State** section of this dialog box, you can enable or disable BFD. To enable it, check the box next to **BFD**, which adds a flag to the EIGRP consumer protocol (or uncheck the box to disable BFD).

Step 19 Click **Submit**.

Step 20 To configure BFD in the Static Routes protocol, in the **Navigation** pane, go back to **Networking** > **L3Outs** > ***L3Out_name*** > **Configured Nodes**, then click on the configured node to bring up the **Node Association** window.

Step 21 In the **Static Routes** section, click the "+" (expand) button.
The **Create Static Route** dialog box appears. Enter values for the required fields in this section.

Step 22 Next to **Route Control**, check the box next to **BFD** to enable (or uncheck the box to disable) BFD on the specified Static Route.

Step 23 Click **Submit**.

Step 24 To configure BFD in the IS-IS protocol, in the **Navigation** pane go to **Fabric** > **Fabric Policies** > **Policies** > **Interface** > **L3 Interface**.

Note
Fabric BFD (BFD on ISIS) is not recommended.The key reasons are:

- In a leaf–spine fabric, ISIS peers are directly connected sub-interfaces. If a peer device (leaf or spine) fails, then the physical link also goes down and triggers Layer-1 convergence. BFD does not provide any additional convergence benefit in most failure cases.
- BFD is susceptible to false flaps even when there is no actual network path issue. This occurs when BFD packets do not receive sufficient CPU cycles, especially during periods of high CPU utilization by other processes (for example, during tech-support collections). False BFD flaps directly impact ISIS adjacencies. Since ISIS is a critical control-plane protocol for the entire fabric, leading to widespread and severe instability at this level can have a widespread and severe impact across the data-center network.

Step 25 On the right side of the **Work** pane, under **ACTIONS**, select **Create L3 Interface Policy**.
The **Create L3 Interface Policy** dialog box appears.

Note
You can also right-click on **L3 Interface** from the left navigation pane and select **Create L3 Interface Policy** to create the policy.

Step 26 Enter a name in the **Name** field and provide values in the remaining fields to define the L3 interface policy.

Step 27 To enable BFD ISIS Policy, in the BFD ISIS Policy Configuration field, click **enabled**.

Step 28 Click **Submit**.

BFD Multihop

BFD multihop provides subsecond forwarding failure detection for a destination with more than one hop and up to 255 hops. Beginning with Release 5.0(1), APIC supports BFD multihop for IPv4 and BFD multihop for IPv6 in compliance with RFC5883. BFD multihop sessions are set up between a unique source and destination address pair. A BFD multihop session is created between a source and destination rather than with an interface, as with single-hop BFD sessions.

BFD multihop sets the TTL field to the maximum limit supported by BGP, and does not check the value on reception. The ACI leaf has no impact on the number of hops a BFD multihop packet can traverse, but the number of hops is limited to 255.

Guidelines and Limitations for BFD Multihop

- The default and minimum transmit and receive interval timers for BFD multihop are 250 ms.
- The default and minimum detection multiplier is 3.
- Echo mode is not supported for BFD multihop.

Configuring a BFD Multihop Policy

You can configure a BFD multihop policy in several locations in the GUI, depending on the purpose of the policy.

- **Global Policies**: By default, the APIC controller creates default policies when the system comes up. These default policies are global BFD multihop configuration policies. You can set attributes within the default global policy in the Work pane, or you can modify these default policy values. However, once you modify a default global policy, your changes affect the entire system (all switches). If you want to use a specific configuration for a particular switch or set of switches that isn't the default, create a switch profile and modify the BFD multihop values in the switch profile.

  You can create or modify global BFD multihop configuration policies for IPv4 or IPv6 in these GUI locations:

  - **Fabric > Access Policies > Policies > Switch > BFD Multihop > BFD Multihop IPv4**: right-click and select **Create BFD Global IPv4 MH Policy**.
  - **Fabric > Access Policies > Policies > Switch > BFD Multihop > BFD Multihop IPv6**: right-click and select **Create BFD Global IPv6 MH Policy**.

- **Node Policies**: A BFD Multihop node policy applies to interfaces under a node profile.

  You can create or modify BFD multihop node policies in this GUI location:

  - **Tenants > tenant > Policies > Protocol > BFD Multihop > Node Policies**: right-click and select **Create BFD Multihop Node Policy**.

- **Inteface Policies**: A BFD Multihop interface policy applies to interfaces under an interface profile.

  You can create or modify BFD multihop interface policies in this GUI location:

  - **Tenants > tenant > Policies > Protocol > BFD Multihop > Interface Policies**: right-click and select **Create BFD Multihop Interface Policy**.

- **Overriding Global Policies**: If you don't want to use the default global configuration, but you want to have an explicit configuration on a given interface, you can create your own global configuration. This configuration is then applied to all the interfaces on a specific switch or set of switches. You can use this interface override configuration when you want more granularity on a specific switch on a specific interface.

  You can create or modify BFD multihop override policies for a node profile or interface profile in these GUI locations:

  - **Tenants > tenant > Networking > L3Outs > l3out > Logical Node Profiles > logical_node_profile**: right-click, select **Create BFD Interface Protocol Profile**, specify BFD Multihop node policy.

- **Tenants > tenant > Networking > L3Outs > l3out > Logical Node Profiles > logical_node_profile > Logical Interface Profiles > logical_interface_profile**: right-click, select **Create MH-BFD Interface Protocol Profile**, specify BFD Multihop interface policy.
- **Tenants > infra > Networking > SR-MPLS Infra L3Outs > l3out > Logical Node Profiles > logical_node_profile > Logical Interface Profiles > logical_interface_profile**: right-click, select **Create MH-BFD Interface Profile**, specify BFD Multihop interface policy.

Procedure

Step 1 Navigate to the GUI location where you will create or configure the BFD multihop policy.

Step 2 Edit an existing profile or policy or launch the dialog box to create a new profile.

Step 3 In the profile, choose an **Authentication Type** for BFD multihop sessions.

You can choose to require no authentication or SHA-1 authentication.

Step 4 If you are creating a new policy, configure the settings in the dialog box:

a) Enter a **Name** for the policy.

b) Set the **Admin State** to **Enabled**.

c) Set the **Detection Multiplier** value.

Specifies the minimum number of consecutive packets that can be missed before BFD declares a session to be down. The range is from 1 to 50 packets. The default is 3.

d) Set the **Minimum Transmit Interval** value.

The minimum interval time for packets being transmitted. The range is from 250 to 999 milliseconds. The default is 250.

e) Set the **Maximum Receive Interval** value.

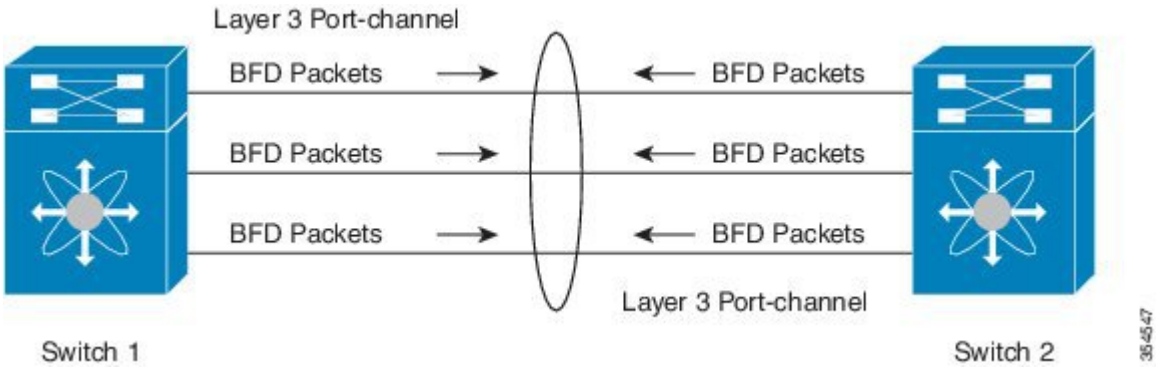
The maximum interval time for packets being received. The range is from 250 to 999 milliseconds. The default is 250.

f) Click **Submit**.

Micro BFD

Beginning with Cisco APIC Release 5.2(3), APIC supports Micro BFD, as defined in IETF RFC 7130. When Bidirectional Forwarding Detection (BFD) is configured on a port channel, keep-alive packets are sent on any available member link. The failure of a single member link might not be detected, because the keep-alive packets can merely traverse a remaining link. Micro BFD is an enhancement to BFD that establishes individual BFD sessions on each member link of a port channel, as shown in the following figure.

Figure 3: Micro BFD Sessions on a Port Channel



When a per-link BFD session senses a failure on its member link, the failed link is removed from the forwarding table. This mechanism delivers faster failure detection and assists in identifying which link has failed on the port channel.

Guidelines and Limitations for Micro BFD

- Micro BFD is supported on both LACP and non-LACP port channels.
- Micro BFD can run simultaneously with multi-hop BFD on the same port channel, but not with single-hop BFD.
- Micro BFD is a single-hop BFD implementation. It does not function if a Layer 2 switch exists between the main port channel on a switch and the switch's peer.
- Micro BFD can run simultaneously on a parent port-channel with single-hop BFD running on the (same) port-channel sub-interfaces.
- Micro BFD is not supported on first-generation leaf switches. First-generation switches are those that do not contain a suffix, such as -EX or -FX, in the PID (product identifier).
- Micro BFD is supported only on the routed interface over a port channel.
- Client protocols can run on a sub-interface on the same port-channel where Micro BFD is enabled.
- Micro BFD is not supported on FEX ports or fabric ports.
- BFD Echo is not supported on Micro BFD sessions.
- On a dual IP stack port channel (IPv4 and IPv6) with Micro BFD enabled, you must configure Micro BFD with either an IPv4 address or an IPv6 address but not both. You cannot have both IPv4 and IPv6 Micro BFD sessions.
- Beginning with Cisco APIC Release 5.2(3), Cisco APIC allows using L3 port channel main interface along with sub-interfaces on the same L3 port channel. But creating or deleting L3 port channel main interface will flap port channel physical member ports. This will cause traffic loss if the port channel sub-interfaces are already active.

Configuring Micro BFD on a Port Channel

This procedure modifies an L3Out port channel interface to enable Micro BFD. Micro BFD establishes individual BFD sessions on each member link of a port channel.

Before you begin

- A direct port channel is configured for an L3Out interface.

Procedure

Step 1 Navigate to **Tenants** > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **Logical Node Profiles** > ***logical_node_profile_name*** > **Logical Interface Profiles**

Step 2 Select the **Logical Interface Profile** that you want to modify.

Step 3 Select the **Routed Interfaces** tab.

Micro BFD is supported only on the routed interface over a port channel.

Step 4 In the **Routed Interfaces** section, double click the existing interface to modify it, or click the **+** icon to add a new interface to the Logical Interface Profile.

The remaining steps of this procedure describe only the enabling of Micro BFD on an existing logical interface. If you are adding a new interface to the Logical Interface Profile, refer to Modifying Interfaces for L3Out Using the GUI.

Step 5 In the configured properties of the selected interface, verify that the selected **Path Type** is **Direct Port Channel**.

Micro BFD is applicable only on a port channel.

Step 6 Check the checkbox for **Enable Micro BFD**.

Step 7 Type the destination IP address of the port channel in **Micro BFD Destination Address**.

Step 8 Enter a value between 60 to 3600 seconds in **Micro BFD Start Timer (sec)**.

The start timer delays activation of BFD monitoring on member links in order to allow BFD sessions to establish. The timer is optional. If the timer is not configured, activation is not delayed.

Step 9 Click **Submit**.

What to do next

You can verify the Micro BFD sessions using the CLI, as shown in the following example:

```
leaf4# show port-channel database interface port-channel 3
port-channel3
Last membership update is successful
4 ports in total, 4 ports up
First operational port is Ethernet1/44
Age of the port-channel is 0d:22h:46m:03s
Time since last bundle is 0d:22h:42m:43s
Last bundled member is Ethernet1/44
Ports: Ethernet1/41 [on] [up]
Ethernet1/42 [on] [up]
Ethernet1/43 [on] [up]
Ethernet1/44 [on] [up] *

leaf4# show bfd neighbors vrf tenant1:vrf1

OurAddr NeighAddr
LD/RD RH/RS Holdown(mult) State Int Vrf Type
```

```
2003:190:190:1::1 2003:190:190:1::2
1090519041/0 Up 6000(3) Up Po3 tenant1:vrf1 singlehop

2003:190:190:1::1 2003:190:190:1::2
1090519042/2148074790 Up 180(3) Up Eth1/44 tenant1:vrf1 singlehop

2003:190:190:1::1 2003:190:190:1::2
1090519043/2148074787 Up 180(3) Up Eth1/41 tenant1:vrf1 singlehop

2003:190:190:1::1 2003:190:190:1::2
1090519044/2148074789 Up 180(3) Up Eth1/43 tenant1:vrf1 singlehop

2003:190:190:1::1 2003:190:190:1::2
1090519045/2148074788 Up 180(3) Up Eth1/42 tenant1:vrf1 singlehop
```

OSPF External Routed Networks

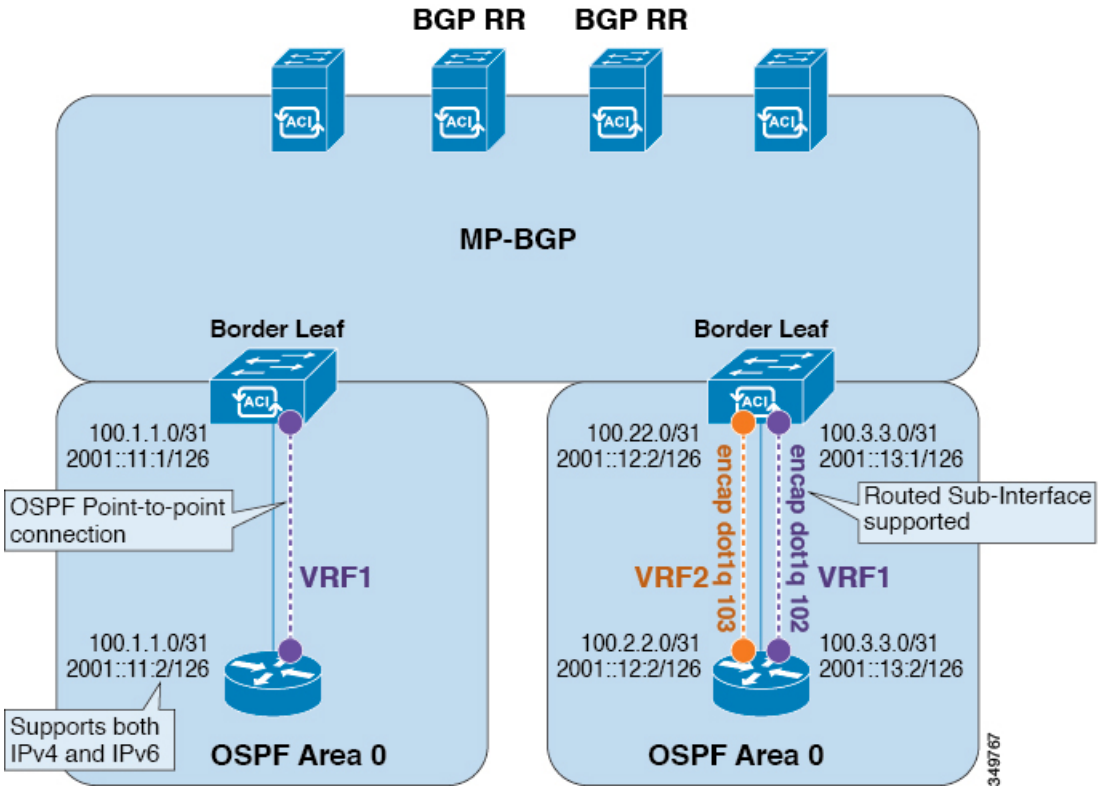
Use the procedures in the following sections to configure OSPF external routed networks.

OSPF Layer 3 Outside Connections

OSPF Layer 3 Outside connections can be normal or NSSA areas. The backbone (area 0) area is also supported as an OSPF Layer 3 Outside connection area. Cisco Application Centric Infrastructure (ACI) supports both OSPFv2 for IPv4 and OSPFv3 for IPv6. When creating an OSPF Layer 3 Outside, it is not necessary to configure the OSPF version. The correct OSPF process is created automatically based on the interface profile configuration (IPv4 or IPv6 addressing). Both IPv4 and IPv6 protocols are supported on the same interface (dual stack) but it is necessary to create two separate interface profiles.

Layer 3 Outside connections are supported for the routed interfaces, routed sub-interfaces, and SVIs. The SVIs are used when there is a need to share the physical connect for both Layer 2 and Layer 3 traffic. The SVIs are supported on ports, port channels, and virtual port channels.

Figure 4: OSPF Layer3 Out Connections



When an SVI is used for a Layer 3 Outside connection, an external bridge domain is created on the border leaf switches. The external bridge domain allows connectivity between the two vPC switches across the Cisco ACI fabric. This allows both the vPC switches to establish the OSPF adjacencies with each other and the external OSPF device.

When running OSPF over a broadcast network, the time to detect a failed neighbor is the dead time interval (default 40 seconds). Reestablishing the neighbor adjacencies after a failure may also take longer due to designated router (DR) election.

Note

- A link or port channel failure to one vPC node does not cause OSPF adjacency to go down. OSPF adjacency can stay up using the external bridge domain that is accessible through the other vPC node.
- When an OSPF time policy or a OSPF or EIGRP address family policy is applied to an L3Out, you can observe the following behaviors:
  - If the L3Out and the policy are defined in the same tenant, then there is no change in behavior.
  - If the L3Out is configured in a user tenant other than the common tenant, the L3Out VRF instance is resolved to the common tenant, and the policy is defined in the common tenant, then only the default values are applied. Any change in the policy will not take effect.
- If a border leaf switch forms OSPF adjacency with two external switches and one of the two switches experiences a route loss while the adjacent switches does not, the Cisco ACI border leaf switch reconverges the route for both neighbors.
- OSPF supports aggressive timers. However, these timers quickly bring down the adjancency and cause CPU churn. Therefore, we recommend that you use the default timers and use bidirectional forwarding detection (BFD) to get sub-second failure detection.

Creating an OSPF L3Out for Management Tenant Using the GUI

- You must verify that the router ID and the logical interface profile IP address are different and do not overlap.
- The following steps are for creating an OSPF L3Out for a management tenant. To create an OSPF L3Out for a tenant, you must choose a tenant and create a VRF for the tenant.
- For more details, see *Cisco APIC and Transit Routing*.

Procedure

Step 1 On the menu bar, choose **Tenants** > **mgmt**.

Step 2 In the **Navigation** pane, expand **Networking** > **L3Outs**.

Step 3 Right-click **L3Outs**, and click **Create L3Out**.

The **Create L3Out** wizard appears.

Step 4 In the **Identity** window in the **Create L3Out** wizard, perform the following actions:

a) In the **Name** field, enter a name (RtdOut).

b) In the **VRF** field, from the drop-down list, choose the VRF (inb).

Note
This step associates the routed outside with the in-band VRF.

c) From the **L3 Domain** drop-down list, choose the appropriate domain.

d) Check the **OSPF** check box.

e) In the **OSPF Area ID** field, enter an area ID.
f) In the **OSPF Area Control** field, check the appropriate check box.
g) In the **OSPF Area Type** field, choose the appropriate area type.
h) In the **OSPF Area Cost** field, choose the appropriate value.
i) Click **Next**.

The **Nodes and Interfaces** window appears.

Step 5 In the **Nodes and Interfaces** window, perform the following actions:

a) Uncheck the **Use Defaults** box.

This allows you to edit the **Node Profile Name** field.

b) In the **Node Profile Name** field, enter a name for the node profile. (borderLeaf).
c) In the **Node ID** field, from the drop-down list, choose the first node. (leaf1).
d) In the **Router ID** field, enter a unique router ID.
e) In the **Loopback Address** field, use a different IP address or leave this field empty if you do not want to use the router ID for the loopback address.

Note
The **Loopback Address** field is automatically populated with the same entry that you provide in the **Router ID** field. This is the equivalent of the **Use Router ID for Loopback Address** option in previous builds. Use a different IP address or leave this field empty if you do not want to use the router ID for the loopback address.

f) Enter the appropriate information in the **Interface**, **IP Address**, **Interface Profile Name** and **MTU** fields for this node, if necessary.
g) In the **Nodes** field, click + icon to add a second set of fields for another node.

Note
You are adding a second node ID.

h) In the **Node ID** field, from the drop-down list, choose the first node. (leaf1).
i) In the **Router ID** field, enter a unique router ID.
j) In the **Loopback Address** field, use a different IP address or leave this field empty if you do not want to use the router ID for the loopback address.

Note
The **Loopback Address** field is automatically populated with the same entry that you provide in the **Router ID** field. This is the equivalent of the **Use Router ID for Loopback Address** option in previous builds. Use a different IP address or leave this field empty if you do not want to use the router ID for the loopback address.

k) Enter the appropriate information in the **Interface**, **IP Address**, **Interface Profile Name** and **MTU** fields for this node, if necessary.
l) Click **Next**.

The **Protocols** window appears.

Step 6 In the **Protocols** window, in the **Policy** area, click **default**, then click **Next**.

The **External EPG** window appears.

Step 7 In the **External EPG** window, perform the following actions:

a) In the **Name** field, enter a name for the external network (extMgmt).

b) Uncheck the **Default EPG for all external networks** field.

The **Subnets** area appears.

c) Click **+** to access the **Create Subnet** dialog box.
d) In the **Create Subnet** dialog box, in the **IP address** field, enter an IP address and mask for the subnet.
e) In the **Scope** field, check the desired check boxes. Click **OK**.
f) In the **External EPG** dialog box, click **Finish**.

Note
In the **Work** pane, in the **L3Outs** area, the L3Out icon (RtdOut) is now displayed.

EIGRP External Routed Networks

Use the procedures in the following sections to configure EIGRP external routed networks.

About EIGRP Layer 3 Outside Connections

This example shows how to configure Enhanced Interior Gateway Routing Protocol (EIGRP) when using the Cisco APIC. The following information applies when configuring EIGRP:

- The tenant, VRF, and bridge domain must already be created.
- The Layer 3 outside tenant network must already be configured.
- The route control profile under routed outside must already be configured.
- The EIGRP VRF policy is the same as the EIGRP family context policy.
- EIGRP supports only export route control profile. The configuration related to route controls is common across all the protocols.

You can configure EIGRP to perform automatic summarization of subnet routes (route summarization) into network-level routes. For example, you can configure subnet 131.108.1.0 to be advertised as 131.108.0.0 over interfaces that have subnets of 192.31.7.0 configured. Automatic summarization is performed when there are two or more network router configuration commands configured for the EIGRP process. By default, this feature is enabled. For more information, see *Route Summarization*.

EIGRP Protocol Support

EIGRP protocol is modeled similar to other routing protocols in the Cisco Application Centric Infrastructure (ACI) fabric.

Supported Features

The following features are supported:

- IPv4 and IPv6 routing
- Virtual routing and forwarding (VRF) and interface controls for each address family

- Redistribution with OSPF across nodes
- Default route leak policy per VRF
- Passive interface and split horizon support
- Route map control for setting tag for exported routes
- Bandwidth and delay configuration options in an EIGRP interface policy
- Authentication support

Unsupported Features

The following features are not supported:

- Stub routing
- EIGRP used for BGP connectivity
- Multiple EIGRP `L3extOut`s on the same node
- Per-interface summarization (an EIGRP summary policy will apply to all interfaces configured under an L3Out)
- Per interface distribute lists for import and export

Categories of EIGRP Functions

EIGRP functions can be broadly categorized as follows:

- Protocol policies
- `L3extOut` configurations
- Interface configurations
- Route map support
- Default route support
- Transit support

Primary Managed Objects That Support EIGRP

The following primary managed objects provide EIGRP support:

- **EIGRP Address Family Context Policy** `eigrpCtxAfPol`: Address Family Context policy configured under `fvTenant` (Tenant/Protocols).
- `fvRsCtxToEigrpCtxAfPol`: Relation from a VRF to a `eigrpCtxAfPol` for a given address family (IPv4 or IPv6). There can be only one relation for each address family.
- `eigrpIfPol`: EIGRP Interface policy configured in `fvTenant`.
- `eigrpExtP`: Enable flag for EIGRP in an `L3extOut`.
- `eigrpIfP`: EIGRP interface profile attached to an `l3extLIfP`.

- `eigrpRsIfPol`: Relation from EIGRP interface profile to an `eigrpIfPol`.
- `Defrtleak`: Default route leak policy under an `l3extOut`.

EIGRP Protocol Policies Supported Under a Tenant

The following EIGRP protocol policies are supported under a tenant:

- **EIGRP Interface policy** (`eigrpIfPol`)—contains the configuration that is applied for a given address family on an interface. The following configurations are allowed in the interface policy:
  - *Hello interval* in seconds
  - *Hold interval* in seconds
  - One or more of the following interface control flags:
    - *split horizon*
    - *passive*
    - *next hop self*

- **EIGRP Address Family Context Policy (`eigrpCtxAfPol`)**—contains the configuration for a given address family in a given VRF. An `eigrpCtxAfPol` is configured under tenant protocol policies and can be applied to one or more VRFs under the tenant. An `eigrpCtxAfPol` can be enabled on a VRF through a relation in the VRF-per-address family. If there is no relation to a given address family, or the specified `eigrpCtxAfPol` in the relation does not exist, then the default VRF policy created under the `common` tenant is used for that address family.

  The following configurations are allowed in the `eigrpCtxAfPol`:
  - Administrative distance for internal route
  - Administrative distance for external route
  - Maximum ECMP paths allowed
  - Active timer interval
  - Metric version (32-bit / 64-bit metrics)

Guidelines and Limitations When Configuring EIGRP

When configuring EIGRP, follow these guidelines:

- Configuring EIGRP and BGP for the same Layer 3 outside is not supported.
- Configuring EIGRP and OSPF for the same Layer 3 outside is not supported.
- There can be one EIGRP Layer 3 Out per node per VRF. If multiple VRFs are deployed on a node, each VRF can have its own Layer 3 Out.
- Multiple EIGRP peers from a single Layer 3 Out is supported. This enables you to connect to multiple EIGRP devices from the same node with a single Layer 3 Out.

- Change the VRF transit tag value to a non-default value within the range of 1-255 when enabling EIGRP on an L3Out. If you require a larger value for the transit VRF tag, you can switch to the wide metric style from the EIGRP family context. EIGRP under the default narrow metric style only supports internal tag values within the range of 1-255.
- The Route Policy Manager (RPM) does not support IPv6 redistribution for filtering BGP route updates based on the BGP community attribute using community lists in a route map. This functionality is only available for IPv4 redistribution.

The following configurations will cause the EIGRP neighbors to flap:

- Changing administrative distances or metric style (wide/narrow) through an EIGRP address family context in the VRF
- Setting configurations of multiple EIGRP summary routes to external EPG all at once. In contrast, configuring only a single EIGRP summary route will not cause the EIGRP neighbors to flap.
- Setting the following configurations that will cause a table-map used internally to be updated:
  - Changing the route tag for the VRF
  - Setting configurations of import direction route control for an OSPF L3Out in the same VRF on the same border leaf switch as an EIGRP L3Out (for example, enabling or disabling the Route Control Enforcement “Import” option or changing routes that are allowed or denied for the import direction). Note that such configurations are not allowed in an EIGRP L3Out itself as the feature is not supported for EIGRP. However, the configurations in an OSPF L3Out still impacts EIGRP L3Outs in the same VRF and leaf switch. This is because the import route control for OSPF utilizes a table-map that is shared, for other purposes, with EIGRP in the same VRF on the same border leaf switch.

Configuring EIGRP Using the GUI

Procedure

Step 1 On the menu bar, choose **Tenants** > **All Tenants**.

Step 2 In the **Work** pane, double click a tenant.

Step 3 In the **Navigation** pane, expand the ***Tenant_name*** > **Policies** > **Protocol** > **EIGRP**.

Step 4 Right-click **EIGRP Address Family Context** and choose **Create EIGRP Address Family Context Policy**.

Step 5 In the **Create EIGRP Address Family Context Policy** dialog box, perform the following actions:

a) In the **Name** field, enter a name for the context policy.
b) In the **Active Interval (min)** field, choose an interval timer.
c) In the **External Distance** and the **Internal Distance** fields, choose the appropriate values.
d) In the **Maximum Path Limit** field, choose the appropriate load balancing value between interfaces (per node/per leaf switch).
e) In the **Metric Style** field, choose the appropriate metric style. Click **Submit**.

In the **Work** pane, the context policy details are displayed.

Step 6 To apply the context policy on a VRF, in the **Navigation** pane, expand **Networking** > **VRFs**.

Step 7 Choose the appropriate VRF, and in the **Work** pane under the **Policy** tab, expand **EIGRP Context Per Address Family**.

Step 8 In the **EIGRP Address Family Type** drop-down list, choose an IP version.

Step 9 In the **EIGRP Address Family Context** drop-down list, choose the context policy. Click **Update**, and Click **Submit**.

Step 10 To enable EIGRP within the Layer 3 Out, in the **Navigation** pane, click **Networking** > **L3Outs**, and click the desired Layer 3 outside network.

Step 11 In the **Work** pane under the **Policy** tab, check the checkbox for **EIGRP**, and enter the EIGRP Autonomous System number. Click **Submit**.

Step 12 To create an EIGRP interface policy, in the **Navigation** pane, click ***Tenant_name*** > **Policies** > **Protocol** > **EIGRP** and perform the following actions:

a) Right-click **EIGRP Interface**, and click **Create EIGRP Interface Policy**.
b) In the **Create EIGRP Interface Policy** dialog box, in the **Name** field, enter a name for the policy.
c) In the **Control State** field, check the desired checkboxes to enable one or multiple controls.
d) In the **Hello Interval (sec)** field, choose the desired interval.
e) In the **Hold Interval (sec)** field, choose the desired interval. Click **Submit**.
f) In the **Bandwidth** field, choose the desired bandwidth.
g) In the **Delay** field, choose the desired delay in tens of microseconds or pico seconds.

In the **Work** pane, the details for the EIGRP interface policy are displayed.

Step 13 In the **Navigation** pane, click the appropriate external routed network where EIGRP was enabled, expand **Logical Node Profiles** and perform the following actions:

a) Expand an appropriate node and an interface under that node.
b) Right-click the interface and click **Create EIGRP Interface Profile**.
c) In the **Create EIGRP Interface Profile** dialog box, in the **EIGRP Policy** field, choose the desired EIGRP interface policy. Click **Submit**.

Note
The EIGRP VRF policy and EIGRP interface policies define the properties that are used when EIGRP is enabled. EIGRP VRF policy and EIGRP interface policies are also available as default policies if you do not want to create new policies. So, if you do not explicitly choose either one of the policies, the default policy is automatically utilized when EIGRP is enabled.

This completes the EIGRP configuration.