Cisco ACI Support for Layer 3 IPv6 Multicast

New and Changed Information 2
Layer 3 IPv6 Multicast 2
About the Fabric Interface 3
Enabling IPv6 Multicast Routing 4
Allocating VRF GIPo 4
Multiple Border Leaf Switches as Designated Forwarder 5
PIM6 Designated Router Election 6
Non-Border Leaf Switch Behavior 6
Active Border Leaf Switch List 6
Overload Behavior On Bootup 6
First-Hop Functionality 6
The Last-Hop 7
Fast-Convergence Mode 7
About Rendezvous Points 7
Guidelines, Limitations, and Expected Behaviors for Configuring Layer 3 IPv6 Multicast 7
Configuring Layer 3 IPv6 Multicast Using the GUI 9
Configuring Layer 3 IPv6 Multicast Using REST API 10
Configuring Layer 3 IPv6 Multicast Using the NX-OS Style CLI 12
New and Changed Information


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

Table 1: New Features and Changed Behavior

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco APIC Release 4.2(1i)</td>
<td>The Layer 3 IPv6 multicast support was introduced with PIM6.</td>
</tr>
</tbody>
</table>

Layer 3 IPv6 Multicast

Note This document describes the Layer 3 IPv6 multicast and related features that are supported in Cisco ACI. For information about Layer 3 IPv4 multicast, see the most recent version of the Cisco APIC Layer 3 Networking Configuration Guide at this URL: https://www.cisco.com/c/en/us/support/cloud-systems-management/application-policy-infrastructure-controller-apic/tsd-products-support-series-home.html.

The use case for IPv6 multicast is to enable IPv6 multicast applications to send multicast over IPv6 from senders in the fabric to external receivers.

Cisco APIC release 4.2(1) adds support to connect multicast applications using IPv6 multicast. The use cases currently supported are for multicast senders located in the Cisco ACI fabric sending IPv6 multicast to external receivers.

IPv6 multicast routing is supported by enabling PIM6 on BDs connected to the IPv6 multicast servers (senders) in the fabric and on border leaves connected to external IPv6 multicast enabled networks. IPv6 multicast routing to receivers inside the fabric is not supported in this release. Border leaf switches will run full PIM6 protocols. The border leaf switches peer with other PIM6 routers connected to them over L3 Outs and also with each other. Non-border leaf switches will run PIM6 in passive mode on the PIM6 enabled BDs. Passive mode enables the BD for IPv6 multicast routing but will not form any PIM adjacencies.

The following figure shows the border leaf (BL) switches (BL1 and BL2) connecting to routers (R1 and R2) in the IPv6 multicast cloud. Each virtual routing and forwarding (VRF) in the fabric that requires IPv6 multicast routing will peer separately with external IPv6 multicast routers.
**About the Fabric Interface**

The fabric interface is a virtual interface between software modules and represents the fabric for IPv6 multicast routing. The interface takes the form of a tunnel interface with the tunnel destination being the VRF GIPo (Group IP outer address). PIM6 shares the same tunnel that PIM4 uses. For example, if a border leaf is the designated forwarder responsible for forwarding traffic for a group, then the fabric interface would be in the outgoing interface (OIF) list for the group. There is no equivalent for the interface in hardware. The operational state of the fabric interface should follow the state published by the intermediate system-to-intermediate system (IS-IS).

The user must configure a unique IPv6 loopback address on each border leaf on each VRF that enables multicast routing.

---

**Note**

The GIPo (Group IP outer address) is the destination multicast IP address used in the outer IP header of the VXLAN packet for all multi-destination packets (Broadcast, Unknown unicast, and Multicast) packets forwarded within the fabric.

Any loopback configured for unicast routing can be reused. This loopback address must be routed from the external network and will be injected into the fabric MP-BGP (Multiprotocol Border Gateway Protocol) routes for the VRF. The fabric interface source IP will be set to this loopback as the loopback interface. The following figure shows the fabric for IPv6 multicast routing.
Enabling IPv6 Multicast Routing

IPv6 multicast is enabled or disabled at three levels, the VRF, L3 Out, and the bridge domain (BD). At the top level, IPv6 multicast routing must be enabled on the VRF that has any multicast-enabled BDs. On an IPv6 multicast-enabled VRF, there can be a combination of IPv6 multicast routing-enabled BDs and BDs where IPv6 multicast routing is disabled. BD with IPv6 multicast-routing disabled will not show on VRF IPv6 multicast panel. L3 Out with IPv6 multicast routing-enabled will show up on the panel as well, but any BD that has IPv6 multicast routing-enabled will always be a part of a VRF that has IPv6 multicast routing-enabled.

IPv6 multicast routing is not supported on the leaf switches such as Cisco Nexus 93128TX, 9396PX, and 9396TX. All the IPv6 multicast routing and any IPv6 multicast-enabled VRF should be deployed only on the switches with -EX and -FX in their product IDs.

Note

Layer 3 Out ports and sub-interfaces are supported while external SVIs are not supported. Since external SVIs are not supported, PIM6 cannot be enabled in L3-VPC.

Allocating VRF GI Po

VRF GI Po is allocated implicitly based on configuration. There will be one GI Po for the VRF and one GI Po for every BD under that VRF. Additionally, any given GI Po might be shared between multiple BDs or multiple VRFs, but not a combination of VRFs and BDs. APIC will be required to ascertain this. In order to handle the VRF GI Po in addition to the BD GI P os already handled and build GI Po trees for them, IS-IS is modified.

Note

For the same VRF, VRF GI Po is common for both IPv4 and IPv6.

All multicast traffic for PIM6 enabled BDs will be forwarded using the VRF GI Po. This includes both Layer 2 and Layer 3 IPv6 multicast. Any broadcast or unicast flood traffic on the multicast enabled BDs will continue to use the BD GI Po. Non-IPv6 multicast enabled BDs will use the BD GI Po for all multicast, broadcast, and unicast flood traffic.
The APIC GUI will display a GIPo multicast address for all BDs and VRFs. The address displayed is always a /28 network address (the last four bits are zero). When the VXLAN packet is sent in the fabric, the destination multicast GIPo address will be an address within this /28 block and is used to select one of 16 FTAG trees. This achieves load balancing of multicast traffic across the fabric.

Table 2: GIPo Usage

<table>
<thead>
<tr>
<th>Traffic</th>
<th>Non-MC Routing-enabled BD</th>
<th>MC Routing-enabled BD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broadcast</td>
<td>BD GIPo</td>
<td>BD GIPo</td>
</tr>
<tr>
<td>Unknown Unicast Flood</td>
<td>BD GIPo</td>
<td>BD GIPo</td>
</tr>
<tr>
<td>Multicast</td>
<td>BD GIPo</td>
<td>VRF GIPo</td>
</tr>
</tbody>
</table>

**Multiple Border Leaf Switches as Designated Forwarder**

When there are multiple border leaf (BL) switches in the fabric doing IPv6 multicast routing, only one of the border leafs is selected as the designated forwarder for attracting traffic from the external IPv6 multicast network and forwarding it to the fabric. This prevents multiple copies of the traffic and it balances the load across the multiple BL switches.

This is done by striping ownership for groups across the available BL switches, as a function of the group address and the VRF virtual network ID (VNID). A BL that is responsible for a group sends PIM6 joins to the external network to attract traffic into the fabric on behalf of receivers in the fabric.

Each BL in the fabric has a view of all the other active BL switches in the fabric in that VRF. So each of the BL switches can independently stripe the groups consistently. Each BL monitors PIM6 neighbor relations on the fabric interface to derive the list of active BL switches. When a BL switch is removed or discovered, the groups are re-striped across the remaining active BL switches. The striping is similar to the method used for hashing the GIPos to external links in multi-pod deployment, so that the group-to-BL mapping is sticky and results in fewer changes on up or down.

*Figure 3: Model for Multiple Border Leafs as Designated Forwarder*
PIM6 Designated Router Election

For Layer 3 IPv6 multicast on ACI fabric, the PIM6 DR (designated router) mechanism for different interface types is as follows:

- PIM6-enabled L3 Out interfaces: Follows standard PIM6 DR mechanism in these type of interfaces.
- Fabric interface: DR election on this interface is not of much significance as the DR functionality is determined by the striping. PIM DR election continues unaltered on this interface.
- IPv6 multicast routing-enabled Pervasive BDs: The pervasive BDs in the fabric are all stubs with respect to IPv6 multicast routing. Hence, on all the leaf switches, the SVI interfaces for pervasive BDs including vPC, are considered DR on the segment.

Non-Border Leaf Switch Behavior

On the non-border leaf switches, PIM6 runs in passive mode on the fabric interface and on the pervasive BD SVIs. PIM6 is in a new passive-probe mode where it sends only hellos. PIM6 neighbors are not expected on these pervasive BD SVIs. It is desirable to raise a fault when a PIM6 hello is heard from a router on a pervasive BD. PIM6, on the non-border leaf switches, does not send any PIM6 protocol packets except for hellos on pervasive BDs and source register packets on the fabric interface.

At the same time, PIM6 will receive and process the following PIM6 packets on the fabric interface:

PIM6 Hellos: This is used to track the active BL list on the fabric interface and on the pervasive BDs, this is used to raise faults.

Active Border Leaf Switch List

On every leaf switch, PIM6 maintains a list of active border leaf switches that is used for striping and other purposes. On the border leaf switches themselves this active border leaf list is derived from the active PIM6 neighbor relations. On non-border leaf switches, the list is generated by PIM6 using the monitored PIM6 Hello messages on the fabric interface. The source IP on the hello messages is the loopback IPv6 address assigned to each border leaf switch.

Overload Behavior On Bootup

When a border leaf switch gains connectivity to the fabric for the first time after bootup or after losing connectivity, it is not desirable to cause the border leaf switch to be part of the active border leaf switch list till the border leaf switch has had a chance to pull the COOP repo information and to bring up its southbound protocol adjacencies. This can be achieved by delaying the transmission of PIM6 hello messages for a non-configured period of time.

Note: All IPv6 multicast group membership information is stored in the COOP database on the spines. When a border leaf boots up it pulls this information from the spine switch.

First-Hop Functionality

The directly connected leaf will handle the first-hop functionality needed for PIM6 sparse mode.
The Last-Hop

The last-hop router is connected to the receiver and is responsible for doing a Shortest-Path Tree (SPT) switchover in case of PIM6 any-source multicast (ASM). The border leaf switches will handle this functionality. The non-border leaf switches do not participate in this function.

Fast-Convergence Mode

The fabric supports a configurable fast-convergence mode where every border leaf switch with external connectivity towards the root ($RP$ for $(*, G)$ and source for $(S, G)$ pulls traffic from the external network. To prevent duplicates, only one of the BL switches forwards the traffic to the fabric. The BL that forwards the traffic for the group into the fabric is called the designated forwarder (DF) for the group. The stripe winner for the group decides on the DF. If the stripe winner has reachability to the root, then the stripe winner is also the DF. If the stripe winner does not have external connectivity to the root, then that BL chooses a DF by sending a PIM6 join over the fabric interface. All non-stripe winner BL switches with external reachability to the root send out PIM6 joins to attract traffic but continue to have the fabric interface as the RPF interface for the route. This results in the traffic reaching the BL switch on the external link, but getting dropped.

The advantage of the fast-convergence mode is that when there is a stripe owner change due to a loss of a BL switch for example, the only action needed is on the new stripe winner of programming the right Reverse Path Forwarding (RPF) interface. There is no latency incurred by joining the PIM6 tree from the new stripe winner. This comes at the cost of the additional bandwidth usage on the non-stripe winners' external links.

Note

Fast-convergence mode can be disabled in deployments where the cost of additional bandwidth outweighs the convergence time saving.

About Rendezvous Points

A rendezvous point (RP) is an IPv6 address that you choose in a IPv6 multicast network domain that acts as a shared root for a IPv6 multicast shared tree. You can configure as many RPs as you like, and you can configure RPs to cover different group ranges.

Static RP—Enables you to statically configure an RP for a IPv6 multicast group range. To do so, you must configure the address of the RP on every router in the domain.

Guidelines, Limitations, and Expected Behaviors for Configuring Layer 3 IPv6 Multicast

See the following guidelines and restrictions:

• The Layer 3 IPv6 multicast feature is supported on the following leaf switches:

  • EX models:
    • N9K-93108TC-EX
    • N9K-93180LC-EX
    • N9K-93180YC-EX
• FX models:
  • N9K-93108TC-FX
  • N9K-93180YC-FX
  • N9K-C9348GC-FXP

• FX2 models:
  • N9K-C9336C-FX2
  • N9K-C9336C-FX2

• Source Specific Multicast (SSM) is supported, but RFC 3306 - Unicast-Prefix-based IPv6 Multicast Addresses specifies a fixed SSM range. Therefore, the SSM range cannot be changed in IPv6.

• For PIM6 only static RP configuration is supported. Auto-RP and BSR are not supported for PIM6.

• Enabling PIM6 and Advertise Host routes on a BD is supported.

• Any Source Multicast (ASM) is supported.

• OIF and VRF scale numbers are the same as they are for IPv4.

• The supported route scale is 2,000. The multicast scale number is a combined scale that includes both IPv4 and IPv6. The total route limit is defined as route counts. Each IPv4 route is counted as 1, and each IPv6 route is counted as 4. Even with node profiles that support more multicast scales, the IPv6 route scale will remain at 2,000.

• PIM6 is supported on Layer 3 Out routed interfaces and routed subinterfaces including Layer 3 port-channel interfaces. PIM6 is not supported on Layer 3 Out SVI interfaces.

• Receivers inside the fabric are not supported. MLD Snoop Policy must be disabled when enabling IPv6 multicast. MLD snooping and PIM6 cannot be enabled in the same VRF.

• Currently, Layer 3 Multicast Listener Discovery (MLD) is not supported with Cisco ACI.

• Cisco Multi-Site Orchestrator support is not available.

• Custom QoS policy is not supported for Layer 3 IPv6 multicast traffic sourced from outside the Cisco ACI fabric (received from L3Out).

• Layer 3 IPv6 multicast is not currently supported on VRFs that are configured with a shared L3 Out.

• Bidirectional PIM6 is currently not supported.

• Fabric Rendezvous Point (RP) is not supported for IPv6 multicast.

• IPv6 Multicast routers are not supported in pervasive bridge domains.

• Layer 3 IPv6 multicast is enabled at the VRF level and the IPv6 multicast protocols will function within the VRF. Each VRF can have IPv6 multicast enabled or disabled independently.

• Once a VRF is enabled for IPv6 multicast, the individual bridge domains (BDs) and L3 Outs under the enabled VRF can be enabled for IPv6 multicast configuration. By default, IPv6 multicast is disabled in all BDs and Layer 3 Outs.

• Enabling PIM6 on an L3Out causes an implicit external network to be configured. This action results in the L3Out being deployed and protocols potentially coming up even if you have not defined an external network.
• For Layer 3 IPv6 multicast support, when the ingress leaf switch receives a packet from a source that is attached on a bridge
domain, and the bridge domain is enabled for IPv6 multicast routing, the ingress leaf switch sends only a routed VRF copy to
the fabric (routed implies that the TTL is decremented by 1, and the source-mac is rewritten with a pervasive subnet MAC). The
egress leaf switch also routes the packet into receivers. The egress leaf also decrements the TTL in the packet by 1. This results
in TTL being decremented two times. Also, for ASM the multicast group must have a valid RP configured.

Cisco ACI does not support IP fragmentation. Therefore, when you configure Layer 3 Outside (L3Out) connections to external
routers, or multipod connections through an Inter-Pod Network (IPN), it is critical 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.

Configuring Layer 3 IPv6 Multicast Using the GUI

Before you begin
• The desired VRF, bridge domains, Layer 3 Out interfaces with IPv6 addresses must be configured to enable PIM6. For Layer
  3 Out, for IPv6 multicast to work, an IPv6 loopback address is configured for the node in the logical node profile.
  • Basic unicast network must be configured.

Procedure

Step 1  On the menu bar, navigate to Tenants > Tenant_name > Networking > VRFs > VRF_name > Multicast IPv6.
In the Work pane, a message is displayed as follows: PIM6 is not enabled on this VRF. Would you like to enable
PIM6?.

Step 2  Click YES, ENABLE MULTICAST IPv6.

Step 3  Configure interfaces:
  a) From the Work pane, click the Interfaces tab.
  b) Expand the Bridge Domains table to display the Create Bridge Domain dialog, and choose the appropriate BD
     from drop-down list.
  c) Click Select.
  d) Expand the Interfaces table to display the Select an L3 Out dialog box.
  e) Click the L3 Out drop-down arrow to choose an L3 Out.
  f) Click Select.

Step 4  Configure a rendezvous point (RP).
  a) In the Work pane, click the Rendezvous Points tab, choose Static RP.
Configuring Layer 3 IPv6 Multicast Using REST API

Before you begin

- The desired VRF, bridge domains, Layer 3 Out interfaces with IPv6 addresses must be configured to enable PIM6. For Layer 3 Out, for IPv6 multicast to work, an IPv6 loopback address is configured for the node in the logical node profile.

- Basic unicast network must be configured.

Procedure

Step 1  Enable PIM6 on the VRF.

Example:

```xml
<fvTenant name="t0">
  <fvCtx name="ctx1" pcEnfPref="unenforced">
    <pimIPV6CtxP ctrl="" mtu="1500" />  
  </fvCtx>
</fvTenant>
```
Step 2  Enable PIM6 on the Layer 3 Out.

Example:

```
<fvTenant dn="uni/tn-t0" name="t0">
  <l3extOut enforceRtctrl="出口" name="bl_l3out_1">
    <pimExtP enabledAf="ipv6-mcast" name="pim"/>
  </l3extOut>
</fvTenant>
```

Step 3  Enable PIM6 on the BD.

Example:

```
<fvTenant name="t0">
  <fvBD name="BD_VPC5" ipv6McastAllow="yes">
    <fvRsCtx tnFvCtxName="ctx1"/>
    <fvSubnet ip="124:1:ffff:ffff:ffff:0/64" scope="public"/>
  </fvBD>
</fvTenant>
```

Step 4  Configure Static Rendezvous Point.

Example:

```
<fvTenant name="t0">
  <pimRouteMapPol dn="uni/tn-t0/rtmap-static_101_ipv6" name="static_101_ipv6">
    <pimRouteMapEntry action="permit" grp="ff00::/8" order="1" rp="2001:0:2001:1:1:1:1/128" src="::"/>
  </pimRouteMapPol>
  <fvCtx name="ctx1" pcEnfPref="unenforced">
    <pimIPV6CtxP ctrl="" mtu="1500">
      <pimStaticRPPol>
          <pimRPGrpRangePol>
            <rtmcRsFilterToRtMapPol tDn="uni/tn-t0/rtmap-static_101_ipv6"/>
          </pimRPGrpRangePol>
        </pimStaticRPEntryPol>
      </pimStaticRPPol>
    </pimIPV6CtxP>
  </fvCtx>
</fvTenant>
```

Step 5  Configure a PIM6 interface policy and apply it on the Layer 3 Out.

Example:

```
<fvTenant dn="uni/tn-t0" name="t0">
  <l3extOut enforceRtctrl="出口" name="bl_l3out_1">
    <l3extLNodeP annotation="" configIssues="" descr="" name="common_npl" nameAlias="" ownerKey="" ownerTag="" tag="yellow-green" targetDscp="unspecified">
      <l3extLIfP annotation="" descr="" name="common_intp1_v6" nameAlias="" ownerKey="" ownerTag="" prio="unspecified" tag="yellow-green">
        <pimIPv6IfP annotation="" descr="" name="" nameAlias=""/>
        <pimRsIPv6IfP annotation="" tDn="uni/tn-common/pimIPv6pol-pimv6_policy"/>
      </l3extLIfP>
    </l3extLNodeP>
  </l3extOut>
</fvTenant>
```
Configuring Layer 3 IPv6 Multicast Using the NX-OS Style CLI

Before you begin

- The desired VRF, bridge domains, Layer 3 Out interfaces with IPv6 addresses must be configured to enable PIM6. For Layer 3 Out, for IPv6 multicast to work, an IPv6 loopback address is configured for the node in the logical node profile.
- Basic unicast network must be configured.

Procedure

**Step 1** Enable PIM6 on the VRF and configure the Rendezvous Point (RP).

Example:

```bash
apic1(config)# tenant tenant1
apic1(config-tenant)# vrf context tenant1_vrf
apic1(config-tenant-vrf)# ipv6 pim
apic1(config-tenant-vrf)# ipv6 rp-address 2018::100:100:100:100 route-map ipv6_pim_routemap
```

**Step 2** Configure a PIM6 interface policy and apply it on the Layer 3 Out.

Example:

```bash
apic1(config-tenant)# l3out tenant1_l3out
apic1(config-tenant-l3out)# ipv6 pim
apic1(config-tenant-l3out)# exit
apic1(config-tenant)# exit
apic1(config)#

apic1(config)# leaf 101
apic1(config)# interface ethernet 1/125
apic1(config-if) ipv6 pim inherit interface-policy pim6_intpol1
```

**Step 3** Enable PIM6 on the BD.

Example:

```bash
apic1(config-tenant)# interface bridge-domain tenant1_bd
apic1(config-tenant-interface)# ipv6 multicast
apic1(config-tenant)# exit
apic1(config)#
```

Layer 3 IPv6 multicast with PIM6 is enabled.
Americas Headquarters
Cisco Systems, Inc.
San Jose, CA 95134-1706
USA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.