Tenant Routed Multicast

This chapter contains the following sections:

Tenant Routed Multicast

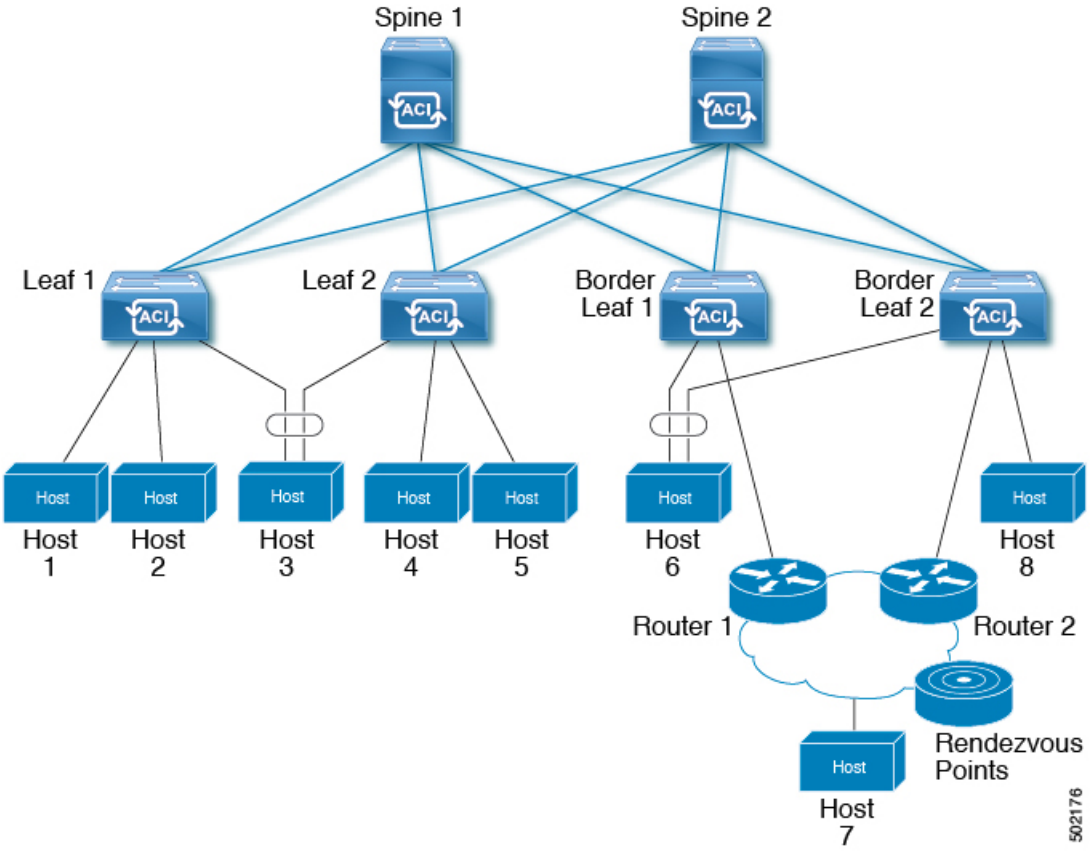
Cisco Application Centric Infrastructure (ACI) Tenant Routed Multicast (TRM) enables Layer 3 multicast routing in Cisco ACI tenant VRF instances. TRM supports multicast fowarding between senders and receivers within the same or different subnets. Multicast sources and receivers can be connected to the same or different leaf switches or external to the fabric using L3Out connections.

In the Cisco ACI fabric, most unicast and IPv4/IPv6 multicast routing operate together on the same border leaf switches, with the IPv4/IPv6 multicast protocol operating over the unicast routing protocols.

In this architecture, only the border leaf switches run the full Protocol Independent Multicast (PIM) or PIM6 protocol. Non-border leaf switches run PIM/PIM6 in a passive mode on the interfaces. They do not peer with any other PIM/PIM6 routers. The border leaf switches peer with other PIM/PIM6 routers connected to them over L3Outs and also with each other.

The following figure shows border leaf switch 1 and border leaf switch 2 connecting to router 1 and router 2 in the IPv4/IPv6 multicast cloud. Each virtual routing and forwarding (VRF) instance in the fabric that requires IPv4/IPv6 multicast routing will peer separately with external IPv4/IPv6 multicast routers.

Figure 1: Overview of Multicast Cloud



Support for Layer 3 Multicast with Remote Leaf Switches

Prior to Release 5.1(3), support has been available for Layer 3 multicast routing in single-pod, multi-pod, and multi-site topologies on local leaf switches. Beginning with Release 5.1(3), support is also available for Layer 3 multicast routing on remote leaf switches. As part of this support, the remote leaf switch can act as a border leaf switch or non-border leaf switch.

There is no difference between the newly-supported remote leaf switches and the previously-supported local leaf switches with regards to the implementation of Layer 3 multicast routing through Cisco APIC or Cisco ACI Multi-Site Orchestrator. The main difference between the two is based on how traffic is forwarded:

- Layer 3 multicast between local leaf switches in a single fabric is forwarded as VXLAN multicast packets where the outer destination IP address is the VRF GIPo multicast address

- Layer 3 multicast packets sent to or sent by remote leaf switches are encapsulated as VXLAN unicast head-end replicated packets

When Layer 3 multicast routing is enabled for a VRF, the VRF GIPo multicast address is programmed on all leaf switches where the VRF is deployed. Layer 3 multicast packets will be forwarded across the pod or between pods as multicast packets and will be received by all leaf switches where the VRF is deployed. For remote leaf switches, the Layer 3 multicast packets will be forwarded using head-end replication to all remote leaf switches where the VRF is deployed. This head-end replication occurs on the pod or remote leaf where the multicast source is connected. For example, if the multicast source is connected to a local leaf switch, one of the spine switches in that pod will be selected to replicate these multicast packets to every remote leaf switch where the VRF is deployed, even if these remote leaf switches are associated with other pods. When a Layer 3 multicast source is connected to a remote leaf switch, the remote leaf switch will also use head-end replication to send a copy of the multicast packet to a spine in every pod as well as all other remote leaf switches where the VRF is deployed.

Multicast forwarding using head-end replication replicates the multicast packet as a separate unicast packet for every head-end replication tunnel. Layer 3 multicast in a remote leaf switch design should ensure that the IP network (IPN) where the remote leaf switches are connected has sufficient bandwidth to support multicast traffic requirements.

Remote leaf switches support L3Out connections with or without PIM enabled. All leaf switches in a VRF that have PIM-enabled L3Outs are eligible to send PIM joins from the fabric towards external sources and rendezvous points. When a multicast receiver connected to the fabric sends an IGMP join for a group, the fabric will select one of the PIM-enabled border leaf switches to send the join (known as the stripe winner). A remote leaf switch with a PIM-enabled L3Out can be selected as the stripe winner for a group even when the receivers for that group are connected to local leaf switches in the main pod. Due to potential sub-optimal forwarding of Layer 3 multicast traffic, deploying PIM-enabled L3Outs on remote leaf switches is not recommended.

Guidelines and Limitations

- Pod redundancy is supported for Layer 3 multicast forwarding with remote leaf switches. If all spine switches in the pod where the remote leaf switch is associated fail, the remote leaf switch can establish control plane connectivity to spine switches in another pod.
- Remote leaf switches must have connectivity to at least one spine switch in any pod. Layer 3 multicast traffic will not be forwarded if the remote leaf switches lose connectivity to all spine switches. This includes Layer 3 multicast traffic between senders and receivers on the same leaf switch.

About the Fabric Interface

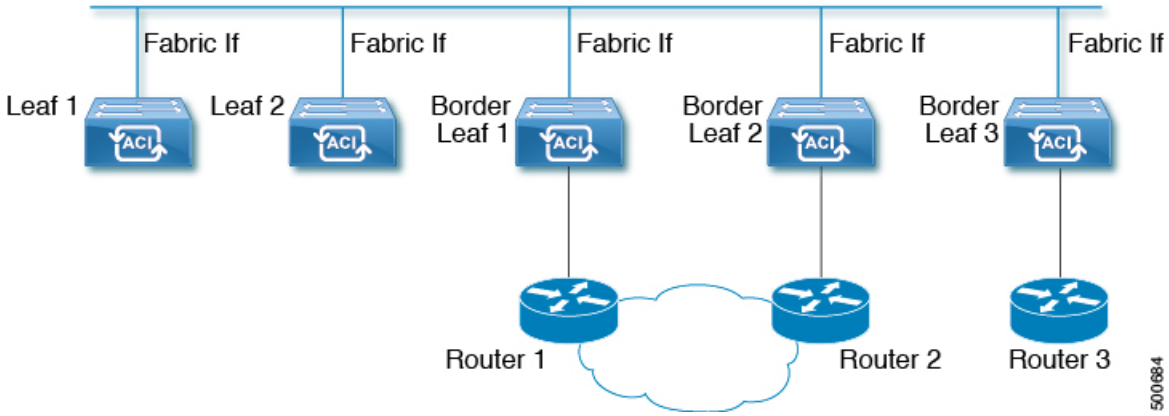
The fabric interface is a virtual interface between software modules and represents the fabric for IPv4/IP6 multicast routing. The interface takes the form of a tunnel interface with the tunnel destination being the VRF GIPo (Group IP outer address)[1]. PIM6 shares the same tunnel that PIM4 uses. For 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).

[1] 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.

Note Each multicast-enabled VRF requires one or more border leaf switches configured with a loopback interface. You must configure a unique IPv4 loopback address on all nodes in a PIM-enabled L3Out. The Router-ID loopback or another unique loopback address can be used.

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 IPv4/IP6 multicast routing.

Figure 2: Fabric for IPv4/IP6 Multicast Routing



Enabling IPv4/IPv6 Tenant Routed Multicast

The process to enable or disable IPv4 or IPv6 multicast routing in a Cisco ACI fabric occurs at three levels:

- **VRF level**: Enable multicast routing at the VRF level.
- **L3Out level**: Enable PIM/PIM6 for one or more L3Outs configured in the VRF.
- **Bridge domain level**: Enable PIM/PIM6 for one or more bridge domains where multicast routing is needed.

At the top level, IPv4/IPv6 multicast routing must be enabled on the VRF that has any multicast routing-enabled bridge domains. On an IPv4/IPv6 multicast routing-enabled VRF, there can be a combination of IPv4/IPv6 multicast routing-enabled bridge domains and bridge domains where IPv4/IPv6 multicast routing is disabled. A bridge domain with IPv4/IPv6 multicast routing disabled will not show on the VRF IPv4/IPv6 multicast panel. An L3Out with IPv4/IPv6 multicast routing-enabled will show up on the panel, but any bridge domain that has IPv4/IPv6 multicast routing enabled will always be a part of a VRF that has IPv4/IPv6 multicast routing enabled.

IPv4/IPv6 multicast routing is not supported on the leaf switches such as Cisco Nexus 93128TX, 9396PX, and 9396TX. All the IPv4/IPv6 multicast routing and any IPv4/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, PIM/PIM6 cannot be enabled in L3-VPC.

Allocating VRF GIPo

VRF GIPo is allocated implicitly based on configuration. There will be one GIPo for the VRF and one GIPo for every BD under that VRF. Additionally, any given GIPo 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 GIPo in addition to the BD GIPos already handled and build GIPo trees for them, IS-IS is modified.

Note For the same VRF, VRF GIPo is common for both IPv4 and IPv6.

All multicast traffic for PIM/PIM6 enabled BDs will be forwarded using the VRF GIPo. This includes both Layer 2 and Layer 3 IPv4/IPv6 multicast. Any broadcast or unicast flood traffic on the multicast enabled BDs will continue to use the BD GIPo. Non-IPv4/IPv6 multicast enabled BDs will use the BD GIPo 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 1: GIPo Usage

Traffic	Non-MC Routing-enabled BD	MC Routing-enabled BD
Broadcast	BD GIPo	BD GIPo
Unknown Unicast Flood	BD GIPo	BD GIPo
Multicast	BD GIPo	VRF GIPo

Multiple Border Leaf Switches as Designated Forwarder

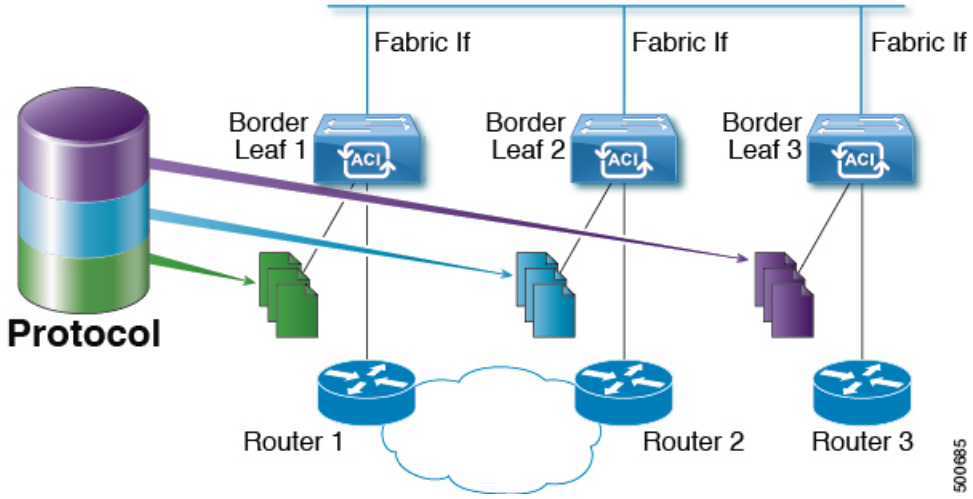
When there are multiple border leaf (BL) switches in the fabric doing IPv4/IPv6 multicast routing, only one of the border leafs is selected as the designated forwarder for attracting traffic from the external IPv4/IPv6 multicast network and forwarding it to the fabric. This prevents multiple copies of 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 PIM/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 PIM/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



PIM/PIM6 Designated Router Election

For Layer 3 IPv4/IPv6 multicast on ACI fabric, the PIM/PIM6 DR (designated router) mechanism for different interface types is as follows:

- PIM/PIM6-enabled L3 Out interfaces: Follows standard PIM/PIM6 DR mechanism in these types of interfaces.
- Fabric interface: DR election on this interface is not of much significance as the DR functionality is determined by the striping. PIM/PIM6 DR election continues unaltered on this interface.
- IPv4/IPv6 multicast routing-enabled pervasive BDs: The pervasive BDs in the fabric are all stubs with respect to IPv4/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, PIM/PIM6 runs in passive mode on the fabric interface and on the pervasive BD SVIs. PIM/PIM6 is in a new passive-probe mode where it sends only *hellos*. PIM/PIM6 neighbors are not expected on these pervasive BD SVIs. It is desirable to raise a fault when a PIM/PIM6 *hello* is heard from a router on a pervasive BD. PIM/PIM6, on the non-border leaf switches, does not send any PIM/PIM6 protocol packets except for *hellos* on pervasive BDs and source register packets on the fabric interface.

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

- **PIM/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.

- **PIM BSR, Auto-RP advertisements**: Supported only for PIM, not supported for PIM6. This is received on the fabric interface and is processed to glean the RP to group-range mapping.

Active Border Leaf Switch List

On every leaf switch, PIM/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 PIM/PIM6 neighbor relations. On non-border leaf switches, the list is generated by PIM/PIM6 using the monitored PIM/PIM6 *Hello* messages on the fabric interface. The source IP on the *hello* messages is the loopback IPv4/IPv6 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[2] information and to bring up its southbound protocol adjacencies. This can be achieved by delaying the transmission of PIM/PIM6 *hello* messages for a non-configured period of time.

First-Hop Functionality

The directly connected leaf switch will handle the first-hop functionality needed for PIM/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 PIM/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 PIM/PIM6 join over the fabric interface. All non-stripe winner BL switches with external reachability to the root send out PIM/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.

[2] All IPv4/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

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 PIM/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 IP address that you choose in a IPv4/IPv6 multicast network domain that acts as a shared root for a IPv4/IPv6 multicast shared tree. You can configure as many RPs as you like, and you can configure RPs to cover different group ranges. When multiple RPs are configured, each RP must be configured for a unique group range.

PIM-enabled border leaf switches are required for VRFs where multicast routing is enabled. PIM is enabled for a border leaf switch by enabling PIM at the L3Out level. When PIM is enabled for an L3Out, this will enable PIM for all nodes and interfaces configured under that L3Out.

ACI supports the following RP configurations:

- **External RP-** The RP is external to the ACI fabric.
  - **Static RP**—Enables you to statically configure an RP for a IPv4/IPv6 multicast group range. To do so, you must configure the address of the RP on every router in the domain.
  - **Auto-RP**—Enables the ACI border leaf to act as an Auto-RP forwarder, Auto-RP listener, and apply Auto-RP mapping agent route-maps.
  - **BSR**—Enables the ACI border leaf to act as a BSR forwarder, BSR listener, and apply route-map to filter BSR messages.
- **Fabric RP**—Applies only for IPv4 multicast; fabric RP is not supported for IPv6 multicast. Enables a PIM anycast RP loopback interface on all PIM-enabled border leaf switches in the VRF. A PIM-enabled L3Out (with loopback interfaces) is required for fabric RP configuration. When configured, external routers can use the fabric RP using static RP configuration. Auto-RP and BSR are not supported with Fabric RP. Fabric RP peering with an external anycast RP member is not supported.

Note Fabric RP has the following restrictions:

- Fabric RP does not support fast-convergence mode.
- The fabric IP:
  - Must be unique across all the static RP entries within the static RP and fabric RP.
  - Cannot be one of the Layer 3 out router IDs

For information about configuring an RP, see the following sections:

- Configuring Layer 3 IPv4 Multicast Using the GUI, on page 18
- Configuring Layer 3 Multicast Using the NX-OS Style CLI
- Configuring Layer 3 Multicast Using REST API

About Inter-VRF Multicast

Note Inter-VRF multicast is not supported for IPv6 multicast.

In typical data center with multicast networks, the multicast sources and receivers are in the same VRF, and all multicast traffic is forwarded within that VRF. There are use cases where the multicast sources and receivers may be located in different VRFs:

- Surveillance cameras are in one VRF while the people viewing the camera feeds are on computers in a different VRF.
- A multicast content provider is in one VRF while different departments of an organization are receiving the multicast content in different VRFs.

ACI release 4.0 adds support for inter-VRF multicast, which enables sources and receivers to be in different VRFs. This allows the receiver VRF to perform the reverse path forwarding (RPF) lookup for the multicast route in the source VRF. When a valid RPF interface is formed in the source VRF, this enables an outgoing interface (OIF) in the receiver VRF. All inter-VRF multicast traffic will be forwarded within the fabric in the source VRF. The inter-VRF forwarding and translation is performed on the leaf switch where the receivers are connected.

Note

- For any-source multicast, the RP used must be in the same VRF as the source.
- Inter-VRF multicast supports both shared services and share L3Out configurations. Sources and receivers can be connected to EPGs or L3Outs in different VRFs.

For ACI, inter-VRF multicast is configured per receiver VRF. Every NBL/BL that has the receiver VRF will get the same inter-VRF configuration. Each NBL that may have directly connected receivers, and BLs that may have external receivers, need to have the source VRF deployed. Control plane signaling and data plane forwarding will do the necessary translation and forwarding between the VRFs inside the NBL/BL that has receivers. Any packets forwarded in the fabric will be in the source VRF.

Inter-VRF Multicast Requirements

This section explains the inter-vrf multicast requirements.

- All sources for a particular group must be in the same VRF (the source VRF).
- Source VRF and source EPGs need to be present on all leafs where there are receiver VRFs.

- For ASM:
  - The RP must be in the same VRF as the sources (the source VRF).
  - For releases prior to 4.2(4), the source VRF must be using fabric RP. This restriction does not apply for Release 4.2(4) and later.
  - The same RP address configuration must be applied under the source and all receiver VRFs for the given group-range.

ACI Multicast Feature List

The following sections provide a list of ACI multicast features with comparisons to similar NX-OS features.

IGMP Features

ACI Feature Name	NX-OS Feature	Description
Allow V3 ASM	ip igmp allow-v3-asm	Allow accepting IGMP version 3 source-specific reports for multicast groups outside of the SSM range. When this feature is enabled, the switch will create an (S,G) mroute entry if it receives an IGMP version 3 report that includes both the group and source even if the group is outside of the configured SSM range. This feature is not required if hosts send (*,G) reports outside of the SSM range, or send (S,G) reports for the SSM range.
Fast Leave	ip igmp immediate-leave	Option that minimizes the leave latency of IGMPv2 group memberships on a given IGMP interface because the device does not send group-specific queries. When immediate leave is enabled, the device removes the group entry from the multicast routing table immediately upon receiving a leave message for the group. The default is disabled. Note: Use this command only when there is one receiver behind the BD/interface for a given group
Report Link Local Groups	ip igmp report-link-local-groups	Enables sending reports for groups in 224.0.0.0/24. Reports are always sent for nonlink local groups. By default, reports are not sent for link local groups.
Group Timeout (sec)	ip igmp group-timeout	Sets the group membership timeout for IGMPv2. Values can range from 3 to 65,535 seconds. The default is 260 seconds.
Query Interval (sec)	ip igmp query-interval	Sets the frequency at which the software sends IGMP host query messages. Values can range from 1 to 18,000 seconds. The default is 125 seconds.

ACI Feature Name	NX-OS Feature	Description
Query Response Interval (sec)	ip igmp query-max-response-time	Sets the response time advertised in IGMP queries. Values can range from 1 to 25 seconds. The default is 10 seconds.
Last Member Count	ip igmp last-member-query-count	Sets the number of times that the software sends an IGMP query in response to a host leave message. Values can range from 1 to 5. The default is 2.
Last Member Response Time (sec)	ip igmp last-member-query-response-time	Sets the query interval waited after sending membership reports before the software deletes the group state. Values can range from 1 to 25 seconds. The default is 1 second.
Startup Query Count	ip igmp startup-query-count	Sets the query count used when the software starts up. Values can range from 1 to 10. The default is 2.
Querier Timeout	ip igmp querier-timeout	Sets the query timeout that the software uses when deciding to take over as the querier. Values can range from 1 to 65,535 seconds. The default is 255 seconds.
Robustness Variable	ip igmp robustness-variable	Sets the robustness variable. You can use a larger value for a lossy network. Values can range from 1 to 7. The default is 2.
Version	ip igmp version <2-3>	IGMP version that is enabled on the bridge domain or interface. The IGMP version can be 2 or 3. The default is 2.
Report Policy Route Map*	ip igmp report-policy <route-map>	Access policy for IGMP reports that is based on a route-map policy. IGMP group reports will only be selected for groups allowed by the route-map
Static Report Route Map*	ip igmp static-oif	Statically binds a multicast group to the outgoing interface, which is handled by the switch hardware. If you specify only the group address, the (*, G) state is created. If you specify the source address, the (S, G) state is created. You can specify a route-map policy name that lists the group prefixes, group ranges, and source prefixes. Note A source tree is built for the (S, G) state only if you enable IGMPv3.
Maximum Multicast Entries	ip igmp state-limit	Limit the mroute states for the BD or interface that are created by IGMP reports. Default is disabled, no limit enforced. Valid range is 1-4294967295.
Reserved Multicast Entries	ip igmp state-limit <limit> reserved <route-map>	Specifies to use the route-map policy name for the reserve policy and set the maximum number of (*, G) and (S, G) entries allowed on the interface.
State Limit Route Map*	ip igmp state-limit <limit> reserved <route-map>	Used with Reserved Multicast Entries feature

IGMP Snooping Features

ACI Feature Name	NX-OS Feature	Description
IGMP snooping admin state	[no] ipigmp snooping	Enables/disables the IGMP snooping feature. Cannot be disabled for PIM enabled bridge domains

ACI Feature Name	NX-OS Feature	Description
Fast Leave	ip igmp snooping fast-leave	Option that minimizes the leave latency of IGMPv2 group memberships on a given IGMP interface because the device does not send group-specific queries. When immediate leave is enabled, the device removes the group entry from the multicast routing table immediately upon receiving a leave message for the group. The default is disabled. Note: Use this command only when there is one receiver behind the BD/interface for a given group
Enable Querier	ip igmp snooping querier <ip address>	Enables the IP IGMP snooping querier feature on the Bridge Domain. Used along with the BD subnet Querier IP setting to configure an IGMP snooping querier for bridge domains. Note: Should not be used with PIM enabled bridge domains. The IGMP querier function is automatically enabled for when PIM is enabled on the bridge domain.
Query Interval	ip igmp snooping query-interval	Sets the frequency at which the software sends IGMP host query messages. Values can range from 1 to 18,000 seconds. The default is 125 seconds.
Query Response Interval	ip igmp snooping query-max-response-time	Sets the response time advertised in IGMP queries. Values can range from 1 to 25 seconds. The default is 10 seconds.
Last Member Query Interval	ip igmp snooping last-member-query-interval	Sets the query interval waited after sending membership reports before the software deletes the group state. Values can range from 1 to 25 seconds. The default is 1 second.
Start Query Count	ip igmp snooping startup-query-count	Configures snooping for a number of queries sent at startup when you do not enable PIM because multicast traffic does not need to be routed. Values can range from 1 to 10. The default is 2.
Start Query Interval (sec)	ip igmp snooping startup-query-interval	Configures a snooping query interval at startup when you do not enable PIM because multicast traffic does not need to be routed. Values can range from 1 to 18,000 seconds. The default is 31 seconds

MLD Snooping Features

ACI Feature Name	NX-OS Feature	Description
MLD snooping admin state	ipv6 mld snooping	IPv6 MLD snooping feature. Default is disabled
Fast Leave	ipv6 mld snooping fast-leave	Allows you to turn on or off the fast-leave feature on a per bridge domain basis. This applies to MLDv2 hosts and is used on ports that are known to have only one host doing MLD behind that port. This command is disabled by default.
Enable Querier	ipv6 mld snooping querier	Enables or disables IPv6 MLD snooping querier processing. MLD snooping querier supports the MLD snooping in a bridge domain where PIM and MLD are not configured because the multicast traffic does not need to be routed.
Query Interval	ipv6 mld snooping query-interval	Sets the frequency at which the software sends MLD host query messages. Values can range from 1 to 18,000 seconds. The default is 125 seconds.

ACI Feature Name	NX-OS Feature	Description
Query Response Interval	ipv6 mld snooping query-interval	Sets the response time advertised in MLD queries. Values can range from 1 to 25 seconds. The default is 10 seconds.
Last Member Query Interval	ipv6 mld snooping last-member-query-interval	Sets the query response time after sending membership reports before the software deletes the group state. Values can range from 1 to 25 seconds. The default is 1 second.

PIM Features (Interface Level)

ACI Feature Name	NX-OS Feature	Description
Authentication	ip pim hello-authentication ah-md5	Enables MD5 hash authentication for PIM IPv4 neighbors
Multicast Domain Boundary	ip pim border	Enables the interface to be on the border of a PIM domain so that no bootstrap, candidate-RP, or Auto-RP messages are sent or received on the interface. The default is disabled.
Passive	ip pim passive	If the passive setting is configured on an interface, it will enable the interface for IP multicast. PIM will operate on the interface in passive mode, which means that the leaf will not send PIM messages on the interface, nor will it accept PIM messages from other devices across this interface. The leaf will instead consider that it is the only PIM device on the network and thus act as the DR. IGMP operations are unaffected by this command.
Strict RFC Compliant	ip pim strict-rfc-compliant	When configured, the switch will not process joins from unknown neighbors and will not send PIM joins to unknown neighbors
Designated Router Delay (sec)	ip pimdr-delay	Delays participation in the designated router (DR) election by setting the DR priority that is advertised in PIM hello messages to 0 for a specified period. During this delay, no DR changes occur, and the current switch is given time to learn all of the multicast states on that interface. After the delay period expires, the correct DR priority is sent in the hello packets, which retriggers the DR election. Values are from 1 to 65,535. The default value is 3. Note: This command delays participation in the DR election only upon bootup or following an IP address or interface state change. It is intended for use with multicast-access non-vPC Layer 3 interfaces only.
Designated Router Priority	ip pim dr-priority	Sets the designated router (DR) priority that is advertised in PIM hello messages. Values range from 1 to 4294967295. The default is 1.
Hello Interval (milliseconds)	ip pim hello-interval	Configures the interval at which hello messages are sent in milliseconds. The range is from 1000 to 18724286. The default is 30000.
Join-Prune Interval Policy (seconds)	ip pim jp-interval	Interval for sending PIM join and prune messages in seconds. Valid range is from 60 to 65520. Value must be divisible by 60. The default value is 60.

ACI Feature Name	NX-OS Feature	Description
Interface-level Inbound Join-Prune Filter Policy*	ip pimjp-policy	Enables inbound join-prune messages to be filtered based on a route-map policy where you can specify group, group and source, or group and RP addresses. The default is no filtering of join-prune messages.
Interface-level Outbound Join-Prune Filter Policy*	ip pim jp-policy	Enables outbound join-prune messages to be filtered based on a route-map policy where you can specify group, group and source, or group and RP addresses. The default is no filtering of join-prune messages.
Interface-level Neighbor Filter Policy*	ip pim neighbor-policy	Controls which PIM neighbors to become adjacent to based on route-map policy where you specify the source address/address range of the permitted PIM neighbors

PIM Features (VRF Level)

ACI Feature Name	NX-OS Feature	Description
Static RP	ippimrp-address	Configures a PIM static RP address for a multicast group range. You can specify an optional route-map policy that lists multicast group ranges for the static RP. If no route-map is configured, the static RP will apply to all multicast group ranges excluding any configured SSM group ranges. The mode is ASM.
Fabric RP	n/a	Configures an anycast RP on all multicast enabled border leaf switches in the fabric. Anycast RP is implemented using PIM anycast RP. You can specify an optional route-map policy that lists multicast group ranges for the static RP.
Auto-RP Forward Auto-RP Updates	ip pim auto-rp forward	Enables the forwarding of Auto-RP messages. The default is disabled.
Auto-RP Listen to Auto-RP Updates	ip pim auto-rp listen	Enables the listening for Auto-RP messages. The default is disabled.
Auto-RP MA Filter *	ip pim auto-rp mapping-agent-policy	Enables Auto-RP discover messages to be filtered by the border leaf based on a route-map policy where you can specify mapping agent source addresses. This feature is used when the border leaf is configured to listen for Auto-RP messages. The default is no filtering of Auto-RP messages.
BSR Forward BSR Updates	ippimbsr forward	Enables forwarding of BSR messages. The default is disabled, which means that the leaf does not forward BSR messages.
BSR Listen to BRS Updates	ip pim bsr listen	Enables listening for BSR messages. The default is disabled, which means that the leaf does not listen for BSR messages.
BSR Filter	ip pim bsr bsr-policy	Enables BSR messages to be filtered by the border leaf based on a route-map policy where you can specify BSR source. This command can be used when the border leaf is configured to listen to BSR messages. The default is no filtering of BSR messages.
ASM Source, Group Expiry Timer Policy *	ip pim sg-expiry-timer <timer> sg-list	Applies a route map to the ASM Source, Group Expiry Timer to specify a group/range of groups for the adjusted expiry timer.

ACI Feature Name	NX-OS Feature	Description
ASM Source, Group Expiry Timer Expiry (sec)	ip pim sg-expiry-timer	To adjust the (S,G) expiry timer interval for Protocol Independent Multicast sparse mode (PIM-SM) (S,G) multicast routes. This command creates persistency of the SPT (source based tree) over the default 180 seconds for intermittent sources. Range is from 180 to 604801 seconds.
Register Traffic Policy: Max Rate	ip pim register-rate-limit	Configures the rate limit in packets per second. The range is from 1 to 65,535. The default is no limit.
Register Traffic Policy: Source IP	ip pim register-source	Used to configure a source IP address of register messages. This feature can be used when the source address of register messages is routed in the network where the RP can send messages. This may happen if the bridge domain where the source is connected is not configured to advertise its subnet outside of the fabric.
SSM Group Range Policy*	ippimssm route-map	Can be used to specify different SSM group ranges other than the default range 232.0.0.0/8. This command is not required if you want to only use the default group range. You can configure a maximum of four ranges for SSM multicast including the default range.
Fast Convergence	n/a	When fast convergence mode is enabled, every border leaf in the fabric will send PIM joins towards the root (RP for (*,G) and source (S,G)) in the external network. This allows all PIM enabled BLs in the fabric to receive the multicast traffic from external sources but only one BL will forward traffic onto the fabric. The BL that forwards the multicast traffic onto the fabric is the designated forwarder. The stripe winner BL decides on the DF. The advantage of the fast-convergence mode is that when there is a changed of the stripe winner due to a BL failure there is no latency incurred in the external network by having the new BL send joins to create multicast state. Note: Fast convergence mode can be disabled in deployments where the cost of additional bandwidth outweighs the convergence time saving.
Strict RFC Compliant	ip pim strict-rfc-compliant	When configured, the switch will not process joins from unknown neighbors and will not send PIM joins to unknown neighbors
MTU Port	ippimmtu	Enables bigger frame sizes for the PIM control plane traffic and improves the convergence. Range is from 1500 to 9216 bytes
Resource Policy Maximum Limit	ip pim state-limit	Sets the maximum (*,G)/(S,G) entries allowed per VRF. Range is from 1 to 4294967295
Resource Policy Reserved Route Map*	ip pim state-limit <limit> reserved <route-map>	Configures a route-map policy matching multicast groups or groups and sources to be applied to the Resource Policy Maximum Limit reserved entries.
Resource Policy Reserved Multicast Entries	ip pim state-limit <limit> reserved <route-map> <limit>	Maximum reserved (*, G) and (S, G) entries allowed in this VRF. Must be less than or equal to the maximum states allowed. Used with the Resource Policy Reserved Route Map policy

Guidelines, Limitations, and Expected Behaviors for Configuring Layer 3 IPv4/IPv6 Multicast

See the following guidelines and restrictions:

- Guidelines and Limitations for IPv4 and IPv6 Multicast, on page 16
- Guidelines and Limitations for IPv4 Multicast, on page 17
- Guidelines and Limitations for IPv6 Multicast, on page 18

Guidelines and Limitations for IPv4 and IPv6 Multicast

The following restrictions apply for both IPv4 and IPv6 multicast:

- The Layer 3 IPv4/IPv6 multicast feature is supported on second generation leaf switches. A second generation switch is one with -EX, -FX, -FX2, -FX3, -GX, or any later suffix in the product ID.
- Custom QoS policy is not supported for Layer 3 multicast traffic sourced from outside the Cisco Application Centric Infrastructure (ACI) fabric (received from L3Out).
- Enabling PIMv4/PIM6 and Advertise Host routes on a bridge domain is supported.
- Layer 3 multicast is enabled at the VRF level and the multicast protocols will function within the VRF instance. Each VRF instance can have multicast enabled or disabled independently.
- After a VRF instance is enabled for multicast, the individual bridge domains and L3Outs under the enabled VRF instance can be enabled for multicast configuration. By default, multicast is disabled in all bridge domains and L3Outs.
- Bidirectional PIMv4/PIM6 is currently not supported.
- Multicast routers are not supported in pervasive bridge domains.
- 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.
- PIMv4/PIM6 is supported on Layer 3 Out routed interfaces, routed subinterfaces including Layer 3 port-channel interfaces, and SVI interfaces.
- Enabling PIMv4/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.
- If the multicast source is connected to Leaf-A as an orphan port and you have an L3Out on Leaf-B, and Leaf-A and Leaf-B are in a vPC pair, the EPG encapsulation VLAN tied to the multicast source will need to be deployed on Leaf-B.
- The behavior of an ingress leaf switch receiving a packet from a source that is attached to a bridge domain differs for Layer 3 IPv4 or IPv6 multicast support:
    - For Layer 3 IPv4 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 IPv4 multicast routing,

the ingress leaf switch sends only a routed VRF instance 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 in all the relevant bridge domains. Therefore, if a receiver is on the same bridge domain as the source, but on a different leaf switch than the source, that receiver continues to get a routed copy, although it is in the same bridge domain. This also applies if the source and receiver are on the same bridge domain and on the same leaf switch, if PIM is enabled on this bridge domain.

For more information, see details about Layer 3 multicast support for multipod that leverages existing Layer 2 design, at the following link Adding Pods.

- 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 instance 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.

- You cannot use a filter with inter-VRF multicast communication.
- Do not use the “clear ip mroute” command. This command is used for internal debugging and is not supported in a production network.

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.`

Guidelines and Limitations for IPv4 Multicast

The following restrictions apply specifically for IPv4 multicast:

- If the border leaf switches in your Cisco ACI fabric are running multicast and you disable multicast on the L3Out while you still have unicast reachability, you will experience traffic loss if the external peer is a Cisco Nexus 9000 switch. This impacts cases where traffic is destined towards the fabric (where the sources are outside the fabric but the receivers are inside the fabric) or transiting through the fabric (where the source and receivers are outside the fabric, but the fabric is transit).
- Any Source Multicast (ASM) and Source-Specific Multicast (SSM) are supported for IPv4.
- You can configure a maximum of four ranges for SSM multicast in the route map per VRF instance.

- IGMP snooping cannot be disabled on pervasive bridge domains with multicast routing enabled.
- Layer 3 multicast is supported with FEX. Multicast sources or receivers that are connected to FEX ports are supported. For further details about how to add FEX in your testbed, see Configure a Fabric Extender with Application Centric Infrastructure at this URL: https://www.cisco.com/c/en/us/support/docs/cloud-systems-management/application-policy-infrastructure-controller-apic/200529-Configure-a-Fabric-Extender-with-Applica.html. Multicast sources or receivers that are connected to FEX ports are not supported.

Guidelines and Limitations for IPv6 Multicast

The following restrictions apply specifically for IPv6 multicast:

- 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.
- You can configure a maximum of four ranges for SSM multicast in the route map per VRF instance.
- Any Source Multicast (ASM) is supported for IPv6.
- OIF and VRF scale numbers for IPv6 are the same as they are for IPv4.
- For PIM6 only static RP configuration is supported. Auto-RP and BSR are not supported for PIM6.
- 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 instance.
- Currently, Layer 3 Multicast Listener Discovery (MLD) is not supported with Cisco ACI.
- Fabric Rendezvous Point (RP) is not supported for IPv6 multicast.
- Cisco Multi-Site Orchestrator support is not available.

Configuring Layer 3 IPv4 Multicast Using the GUI

This section explains how to configure Layer 3 multicast using the Cisco APIC GUI.

Note Click the help icon (**?**) located in the top-right corner of the **Work** pane and of each dialog box for information about a visible tab or a field.

Before you begin

- The desired VRF, bridge domains, Layer 3 Out interfaces with IP addresses must be configured to enable PIM and IGMP.
- Basic unicast network must be configured.

Procedure

Step 1 Navigate to **Tenants** > *Tenant_name* > **Networking** > **VRFs** > *VRF_name* > **Multicast**.
In the **Work** pane, a message is displayed as follows: **PIM is not enabled on this VRF. Would you like to enable PIM?**.

Step 2 Click **YES, ENABLE MULTICAST**.

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 enter the appropriate value in each field.
c) Click **Select**.
d) Expand the **Interfaces** table to display the **Select an L3 Out** dialog.
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 and choose from the following rendezvous point (RP) options:

- **Static RP**
  a. Expand the **Static RP** table.
  b. Enter the appropriate value in each field.
  c. Click **Update**.
- **Fabric RP**
  a. Expand the **Fabric RP** table.
  b. Enter the appropriate value in each field.
  c. Click **Update**.
- **Auto-RP**
  a. Enter the appropriate value in each field.
- **Bootstrap Router (BSR)**
  a. Enter the appropriate value in each field.

Step 5 Configure the pattern policy:

a) From the **Work** pane, click the **Pattern Policy** tab and choose the **Any Source Multicast (ASM)** or **Source Specific Multicast (SSM)** option.
b) Enter the appropriate value in each field.

Step 6 Configure the PIM settings:

a) Click the **PIM Setting** tab.
b) Enter the appropriate value in each field.

Step 7 Configure the IGMP settings:

a) Click the **IGMP Setting** tab.
b) Expand the **IGMP Context SSM Translate Policy** table.
c) Enter appropriate value in each field.
d) Click **Update**.

Step 8 Configure inter-VRF multicast:

a) In the **Work** pane, click the **Inter-VRF Multicast** tab.
b) Expand the **Inter-VRF Multicast** table.
c) Enter appropriate value in each field.
d) Click **Update**.

Step 9 When finished, click **Submit**.

Step 10 On the menu bar, navigate to **Tenants** > ***Tenant_name*** > **Networking** > **VRFs** > ***VRF_name*** > **Multicast**, and perform the following actions:

a) In the **Work** pane, **Interfaces** tab, choose the appropriate L3 Out, and from the **PIM Policy** drop-down list, choose the appropriate PIM policy to attach.
b) Click **Submit**.

Step 11 To verify the configuration perform the following actions:

a) In the **Work** pane, click **Interfaces** to display the associated **Bridge Domains**.
b) Click **Interfaces** to display the associated **L3 Out** interfaces.
c) In the **Navigation** pane, navigate to the **BD**.
d) In the **Work** pane, the configured IGMP policy and PIM functionality are displayed as configured earlier.
e) In the **Navigation** pane, the L3 Out interface is displayed.
f) In the **Work** pane, the PIM functionality is displayed as configured earlier.
g) In the **Work** pane, navigate to **Fabric** > **Inventory** > **Protocols** > **IGMP** to view the operational status of the configured IGMP interfaces.
h) In the **Work** pane, navigate to **Fabric** > **Inventory** > ***Pod name*** > ***Leaf_Node*** > **Protocols** > **IGMP** > **IGMP Domains** to view the domain information for multicast enabled/disabled nodes.

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**.
b) Enter the appropriate value in each field.
c) Click **Update**.

Step 5 Configure the pattern policy.

a) From the **Work** pane, click the **Pattern Policy** tab and choose **Any Source Multicast (ASM)**.
b) Enter the appropriate values in each field.

Step 6 Configure the PIM settings.

a) Click the **PIM Setting** tab.
b) Enter the appropriate value in each field.

Step 7 When finished, click **Submit**.

Step 8 On the menu bar, navigate to **Tenants** > ***Tenant_name*** > **Networking** > **VRFs** > ***VRF_name*** > **Multicast IPv6**, and perform the following actions:

a) In the **Work** pane, **Interfaces** tab, choose the appropriate **L3 Out** and from the **PIM Policy** drop-down list, choose the appropriate PIM policy to attach.
b) Click **Submit**.

Step 9 To verify the configuration perform the following actions:

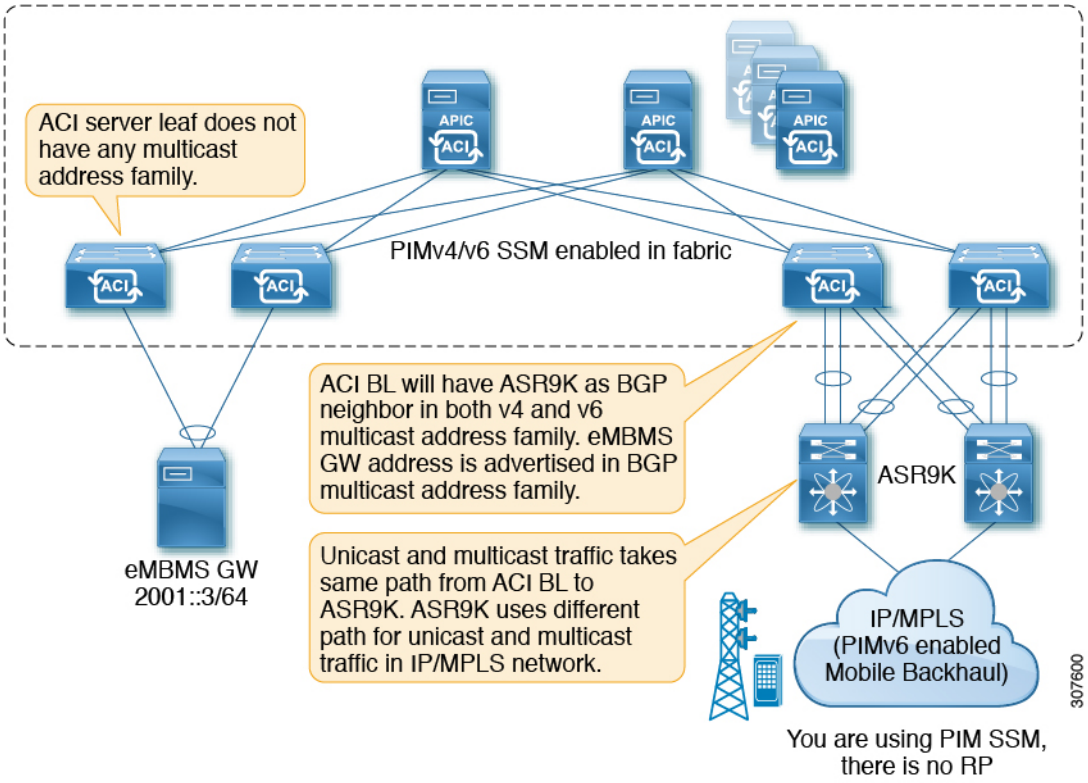
a) In the **Work** pane, click **Interfaces** to display the associated **Bridge Domains**.
b) In the **Navigation** pane, navigate to the associated BD with IPv6 multicast.
In the **Work** pane, the configured PIM functionality is displayed as configured earlier.
c) In the **Navigation** pane, navigate to the associated L3 Out interface.
In the **Work** pane, the PIM6 check box is checked.
d) In the **Work** pane, navigate to **Fabric** > **Inventory** >**Pod NodeProtocols** > **PIM6** and expand PIM.
Under the appropriate PIM6 protocol that was created earlier, you can view information about he associated Neighbors, PIM Interfaces, Routes, Group Ranges, and RPs. You can verify that all these objects are set up.

About BGP IPv4/IPv6 Multicast Address-Family

Note The IPv4 version of the BGP IPv4/IPv6 multicast address-family feature was available as part of Cisco APIC Release 4.1.

Beginning with Cisco APIC release 4.2(1), the BGP multicast address-family feature adds support for IPv6 for BGP peers towards external routers in the tenant VRF on the border leaf switch. You can specify if the peer will also be used separately to carry multicast routes in the IPv4/IPv6 multicast address-family.

The following figure shows how this feature might be implemented.



Guidelines and Limitations for BGP IPv4/IPv6 Multicast Address-Family

Guidelines and Restrictions for the BGP Multicast Address-Family Feature for IPv6

- The Rendezvous Point (RP) is an IP address that is external the Cisco ACI fabric. Fabric RP is not supported for IPv6 multicast.
- The multicast source is within the Cisco ACI fabric, and the receivers are outside of the fabric.
- Transit L3Out is not supported for BGPv4/v6 address-family.

Guidelines and Restrictions for the BGP Multicast Address-Family Feature for Both IPv4 and IPv6

- There is no support for BGPv4/v6 multicast address-family within the Cisco ACI fabric.
- RP reachability should be present in the unicast address-family, if that is being used. For PIM Source-Specific Multicast (SSM), there is no need for RP.

Configuring BGP IPv4/IPv6 Multicast Address-Family Using the GUI

The following procedure describes how to configure the BGP IPv4/IPv6 multicast address-family feature using the GUI.

Before you begin

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

- Configure the tenant, 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 Locate the VRF that you will be using with the L3Out, or create the VRF, if necessary.

Tenants > ***tenant*** > **Networking** > **VRFs**

Step 2 Enable PIMv4 or PIMv6 under the VRF.

- To enable PIMv4 under the VRF, on the menu bar, navigate to **Tenants** > ***Tenant_name*** > **Networking** > **VRFs** > ***VRF_name*** > **Multicast**.
  - If you see the message **PIM is not enabled on this VRF. Would you like to enable PIM?**, then click **Yes, enable Multicast**.
  - If you see the main **Multicast** window, check the **Enable** box, if it is not already checked.
- To enable PIMv6 under the VRF, on the menu bar, navigate to **Tenants** > ***Tenant_name*** > **Networking** > **VRFs** > ***VRF_name*** > **Multicast IPv6**.
  - If you see the message **PIMv6 is not enabled on this VRF. Would you like to enable PIMv6?**, then click **Yes, enable multicast IPv6**.
  - If you see the main **Multicast IPv6** window, check the **Enable** box, if it is not already checked.

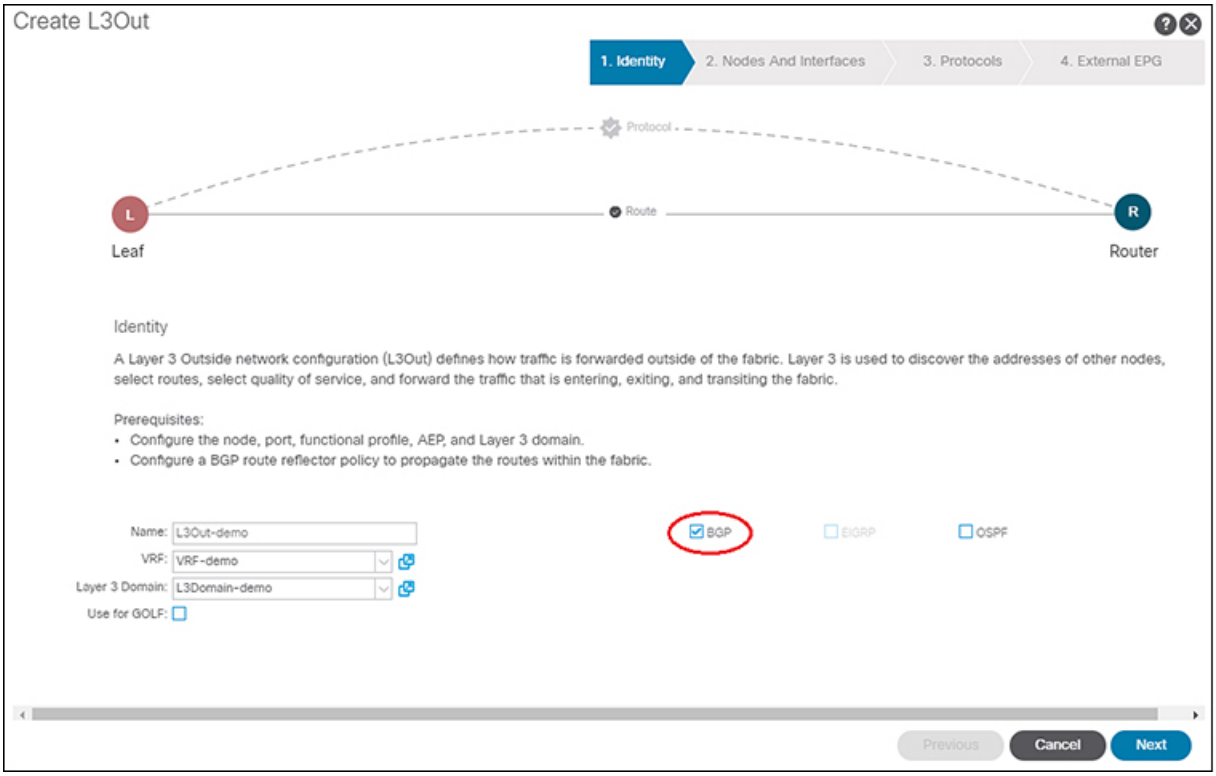
Step 3 Create the L3Out and configure the BGP for the L3Out:

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.

In the **Identity** page:

- Select the VRF that you configured in the previous step.

- 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.

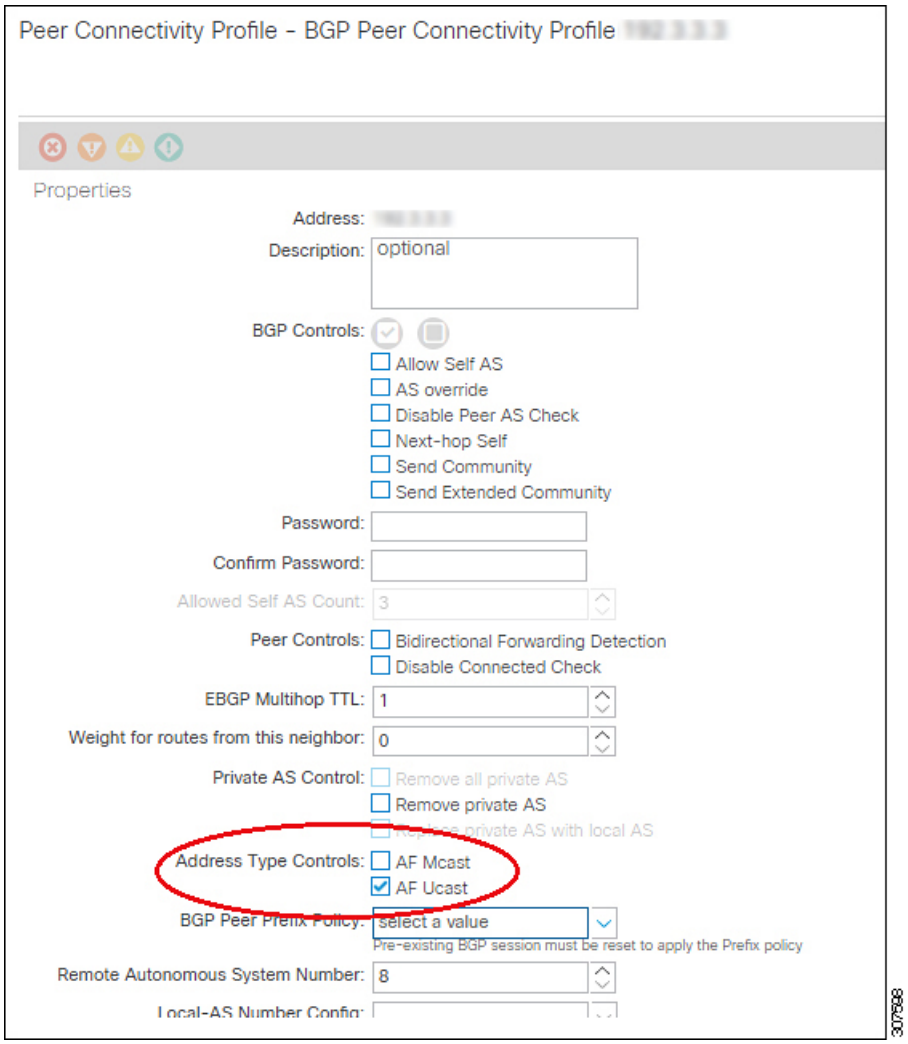
Step 4 After you have completed the L3Out configuration, configure the BGP IPv4/IPv6 multicast address-family feature:

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 **Address Type Controls** field and make the following selections:

- Select **AF Mcast**.
- Leave **AF Ucast** selected, if it is already selected.



c) Click **Submit**.

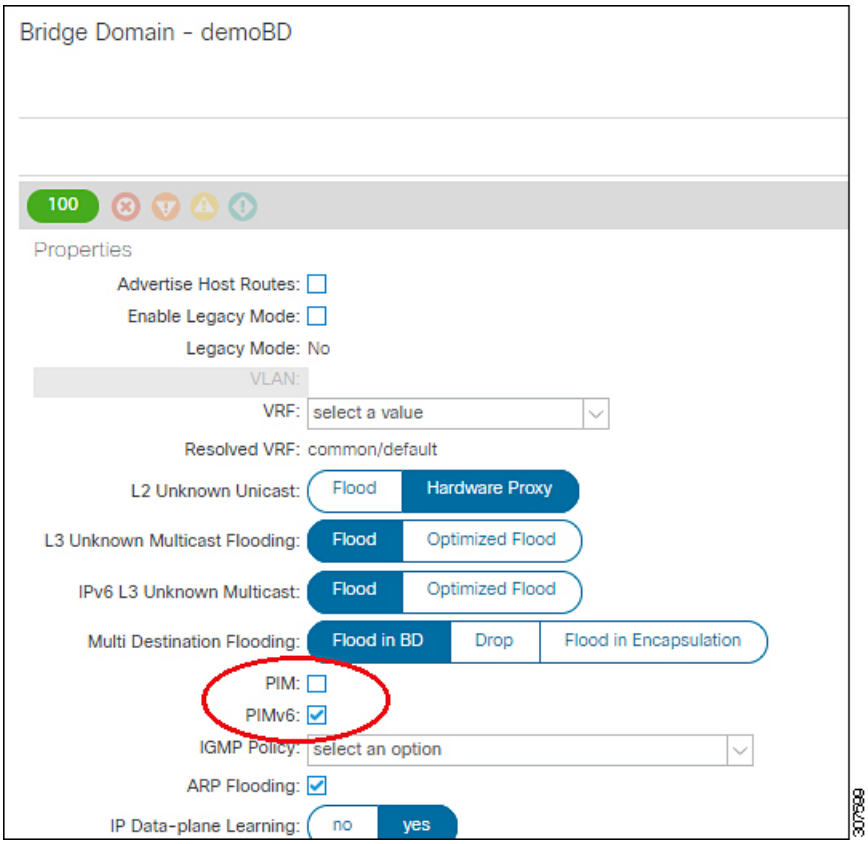
d) Navigate to the bridge domain with the subnet that needs to be redistributed to the peer's IPv4 or IPv6 multicast address-family:

Tenants > ***tenant*** > **Networking** > **Bridge Domains** > ***bridge_domain-name***

e) In the main pane, click the **Policy/General** tabs.

f) Enable PIMv4 or PIMv6 on the bridge domain.

- To enable PIMv4 on the bridge domain, scroll down to the **PIM** field and check the box next to that field to enable it.
- To enable PIMv6 on the bridge domain, scroll down to the **PIMv6** field and check the box next to that field to enable it.



g) Click **Submit**.

About Multicast Filtering

ACI supports control plane configurations that can be used to control who can receive multicast feeds and from which sources. The filtering options can be IGMP report filters, PIM join or prune filters, PIM neighbor filters, and Rendezvous Point (RP) filters. These options rely on control plane protocols, namely IGMP and PIM.

In some deployments, it may be desirable to constrain the sending and/or receiving of multicast streams at the data plane level. For example, you may want to allow multicast senders in a LAN to only send to specific multicast groups or to allow receivers in a LAN to only receive specific multicast groups originated from all the possible sources or from specific sources.

Beginning with Cisco APIC Release 5.0(1), the multicast filtering feature is now available, which allows you to filter multicast traffic from two directions:

- Configuring Multicast Filtering: Source Filtering at First-Hop Router, on page 27
- Configuring Multicast Filtering: Receiver Filtering at Last-Hop Router, on page 27
- Combined Source and Receiver Filtering on Same Bridge Domain, on page 27

Configuring Multicast Filtering: Source Filtering at First-Hop Router

For any sources that are sending traffic on a bridge domain, if you have configured a multicast source filter for that bridge domain, then the source and group will be matched against one of the entries in the source filter route map, where one of the following actions will take place, depending on the action that is associated with that entry:

- If the source and group is matched against an entry with a **Permit** action in the route map, then the bridge domain will allow traffic to be sent out from that source to that group.
- If the source and group is matched against an entry with a **Deny** action in the route map, then the bridge domain will block traffic from being sent out from that source to that group.
- If there is no match with any entries in the route map, then the bridge domain will block traffic from being sent out from that source to that group as the default option. This means that once the route map is applied, there is always an implicit "deny all" statement in effect at the end.

You can configure multiple entries in a single route map, where some entries can be configured with a **Permit** action and other entries can be configured with a **Deny** action, all within the same route map.

Note When a source filter is applied to a bridge domain, it will filter multicast traffic at the source. The filter will prevent multicast from being received by receivers in different bridge domains, the same bridge domain, and external receivers.

Configuring Multicast Filtering: Receiver Filtering at Last-Hop Router

Multicast receiver filtering is used to restrict from which sources receivers in a bridge domain can receive multicast for a particular group. This feature provides source or group data plane filtering functionality for IGMPv2 hosts, similar to what IGMPv3 provides at the control plane.

For any receivers sending joins on a bridge domain, if you have configured a multicast receiver filter for that bridge domain, then the source and group will be matched against one of the entries in the receiver filter route map, where one of the following actions will take place, depending on the action that is associated with that entry:

- If the source and group is matched against an entry with a **Permit** action in the route map, then the bridge domain will allow traffic to be received from that source for that group.
- If the source and group is matched against an entry with a **Deny** action in the route map, then the bridge domain will block traffic from being received from that source for that group.
- If there is no match with any entries in the route map, then the bridge domain will block traffic from being received from that source for that group as the default option. This means that once the route map is applied, there is always an implicit "deny all" statement in effect at the end.

You can configure multiple entries in a single route map, where some entries can be configured with a **Permit** action and other entries can be configured with a **Deny** action, all within the same route map.

Combined Source and Receiver Filtering on Same Bridge Domain

You can also enable both multicast source filtering and multicast receiver filtering on the same bridge domain, where one bridge domain can perform blocking or can permit sources to filtering when sending traffic to a

group range, and can also perform restricting or can allow restricting to filtering when receiving traffic from sources to a group range.

Guidelines and Restrictions for Multicast Filtering

Following are the guidelines and restrictions for the multicast filtering feature:

- While you can enable either the multicast source filtering or the receiver filtering on a bridge domain, you can also have both multicast source filtering and receiver filtering enabled on the same bridge domain.
- Multicast filtering is supported only for IPv4.
- If you do not want to have multicast filters on a bridge domain, then do not configure a source filter or destination filter route maps on that bridge domain. By default, no route maps are associated with a bridge domain, which means that all sources and groups are allowed. If a route map with source filters or destination filters is associated with a bridge domain, only the permit entries in that route map will be allowed, and all deny entries will be blocked (including the implicit "deny-all" statement always present at the end).
- If you attach an empty route map to a bridge domain, route maps assume a deny-all by default, so all sources and groups will be blocked on that bridge domain.
- The multicast filtering feature is applied at the bridge domain level. ACI supports configuration of multiple EPGs in a single bridge domain. When this configuration is used with the bridge domain filtering features, the filter will be applied across all EPGs in the bridge domain as it is a bridge domain level setting.
- The multicast filtering feature is intended to be used for Any-Source Multicast (ASM) ranges only. If, however, you have support for Source-Specific Multicast (SSM) ranges, then we recommend that sources and joins be filtered in the SSM join itself using IGMPv3.

  If you configure SSM ranges for the multicast filtering feature, the following restrictions apply:

  - **Bridge domain source filtering with SSM**: Source filtering is not supported with SSM.
  - **Bridge domain receiver filtering with SSM**: Receiver filtering can be used with SSM group ranges. One of the main use cases for receiver filtering is to filter multicast streams from specific sources. In most cases, receiver filtering is not needed with SSM as this functionality is already provided by the SSM protocol.

- Source and receiver filtering use an ordered list of route-map entries. Route-map entries are executed with the lowest number first until there is a match. If there is a match, even if it is not the longest match in the list, it will exit the program and will not consider the rest of the entries.

  For example, assume that you have the following route map for a specific source (192.0.3.1/32), with these entries:

Table 2: Route Map

Order	Source IP	Action
1	192.0.0.0/16	Permit
2	192.0.3.0/24	Deny

The route map is evaluated based on the order number. Therefore, even though the second entry (192.0.3.0/24) is a longer match for the source IP, the first entry (192.0.0.0/16) will be matched because of the earlier order number.

Configuring Multicast Filtering Using the GUI

You will be configuring multicast filtering at the bridge domain level. Use the procedures in this topic to configure either source filtering or receiver filtering, or both, at the bridge domain level.

Before you begin

- The bridge domain where you will be configuring multicast filtering is already created.
- The bridge domain is a PIM-enabled bridge domain.
- Layer 3 multicast is enabled at the VRF level.

Procedure

Step 1 Navigate to the bridge domain where you want to configure multicast filtering.

Tenant > ***tenant-name*** > **Networking** > **Bridge Domains** > ***bridge-domain-name***

The Summary page for this bridge domain appears.

Step 2 Select the **Policy** tab, then select the **General** subtab.

Step 3 In the **General** window, locate the **PIM** field and verify that PIM is enabled (that there is a check in the box next to the **PIM** field).

If PIM is not enabled, put a check in the box next to the **PIM** field to enable that now. The **Source Filter** and **Destination Filter** fields become available.

Note Multicast filtering is supported only for IPv4 (PIM), and is not supported for IPv6 (PIM6) at this time.

Step 4 Determine whether you want to enable multicast *source* or *receiver* filtering.

Note You can also enable both source and receiver filtering on the same bridge domain.

- If you want to enable multicast *source* filtering at the first-hop router, in the **Source Filter** field, make one of the following selections:
  - **Existing route map policy**: Select an existing route map policy for multicast for the source filtering, then go to Step 7, on page 31.
  - **New route map policy**: Select **Create Route Map Policy for Multicast**, then proceed to Step 5, on page 30.
- If you want to enable multicast *receiver* filtering at the last-hop router, in the **Destination Filter** field, make one of the following selections:
  - **Existing route map policy**: Select an existing route map policy for multicast for the receiver filtering, then go to Step 7, on page 31.

- **New route map policy**: Select **Create Route Map Policy for Multicast**, then proceed to Step 6, on page 31.

Step 5 If you selected the **Create Route Map Policy for Multicast** option to enable multicast **source** filtering at the first-hop router, the **Create Route Map Policy for Multicast** window appears. Enter the following information in this window:

a) In the **Name** field, enter a name for this route map, and enter a description in the **Description** field, if desired.

b) In the **Route Maps** area, click **+**.

The **Create Route Map Entry** window appears.

c) In the **Order** field, if multiple access groups are being configured for this interface, select a number that reflects the order in which this access group will be permitted or denied access to the multicast traffic on this interface.

Lower-numbered entries are ordered before higher-numbered entries. The range is from 0 to 65535.

d) Determine how you want to allow or deny traffic to be sent for multicast source filtering.

- If you want to allow or deny multicast traffic to be sent from **a specific source to any group**, in the **Source IP** field, enter the IP address of the specific source from which the traffic is sent, and leave the **Group IP** field empty.
- If you want to allow or deny multicast traffic to be sent from **any source to a specific group**, in the **Group IP** field, enter the multicast IP address to which the traffic is sent, and leave the **Source IP** field empty.
- If you want to allow or deny multicast traffic to be sent from **a specific source to a specific group**, enter the necessary information in both the **Group IP** and the **Source IP** fields.

Note The **RP IP** field is not applicable for multicast source filtering or multicast receiver filtering. Any entry in this field will be ignored for multicast filtering, so do not enter a value in this field for this feature.

e) In the **Action** field, choose **Deny** to deny access or **Permit** to allow access for the target source.

f) Click **OK**.

The **Create Route Map Policy for Multicast** window appears again, with the route map entry that you configured displayed in the **Route Maps** table.

g) Determine if you want to create additional route map entries for this route map.

You can create multiple route map entries for a route map, each with their own IP addresses and related actions. For example, you might want to have one set of IP addresses with a **Permit** action applied, and another set of IP addresses with a **Deny** action applied, all within the same route map.

If you want to create additional route map entries for this route map, click **+** in the **Route Maps** area again, then go to 5.c, on page 30 to repeat the steps for filling in the necessary information in the **Create Route Map Entry** window for the additional route map entries for this route map.

h) When you have completed all of the route map entries for this route map, click **Submit**. Go to Step 7, on page 31.

Step 6 If you selected the **Create Route Map Policy for Multicast** option to enable multicast **destination** (receiver) filtering at the last-hop router, the **Create Route Map Policy for Multicast** window appears. Enter the following information in this window:

a) In the **Name** field, enter a name for this route map, and enter a description in the **Description** field, if desired.

b) In the **Route Maps** area, click **+**.

The **Create Route Map Entry** window appears.

c) In the **Order** field, if multiple access groups are being configured for this interface, select a number that reflects the order in which this access group will be permitted or denied access to the multicast traffic on this interface.

Lower-numbered entries are ordered before higher-numbered entries. The range is from 0 to 65535.

d) Determine if you want to allow or deny traffic to be received for multicast receiver filtering.

- If you want to allow or deny traffic from being received from **any source to a specific group**, in the **Group IP** field, enter the multicast IP address to which the traffic is sent, and leave the **Source IP** field empty.
- If you want to allow or deny traffic from being received from **a specific source to any group**, in the **Source IP** field, enter the IP address of the specific source from which the traffic is sent, and leave the **Group IP** field empty.
- If you want to allow or deny traffic from being received from **a specific source to a specific group**, enter the necessary information in both the **Group IP** and the **Source IP** fields.

Note The **RP IP** field is not applicable for multicast source filtering or multicast receiver filtering. Any entry in this field will be ignored for multicast filtering, so do not enter a value in this field for this feature.

e) In the **Action** field, choose **Deny** to deny access or **Permit** to allow access for the target group.

f) Click **OK**.

The **Create Route Map Policy for Multicast** window appears again, with the route map entry that you configured displayed in the **Route Maps** table.

g) Determine if you want to create additional route map entries for this route map.

You can create multiple route map entries for a route map, each with their own IP addresses and related actions. For example, you might want to have one set of IP addresses with a **Permit** action applied, and another set of IP addresses with a **Deny** action applied, all within the same route map.

If you want to create additional route map entries for this route map, click **+** in the **Route Maps** area again, then go to 6.c, on page 31 to repeat the steps for filling in the necessary information in the **Create Route Map Entry** window for the additional route map entries for this route map.

h) When you have completed all of the route map entries for this route map, click **Submit**. Go to Step 7, on page 31.

Step 7 At the bottom righthand corner of the Policy/General page, click **Submit**.

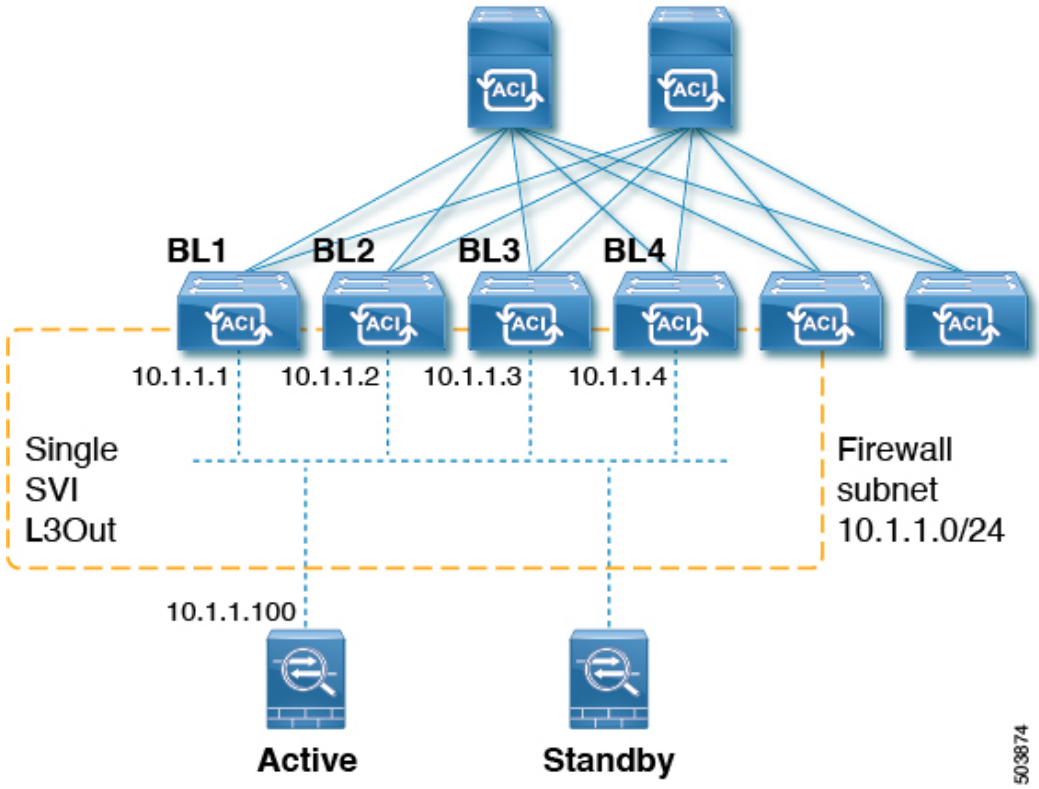
The **Policy Usage Warning** window appears.

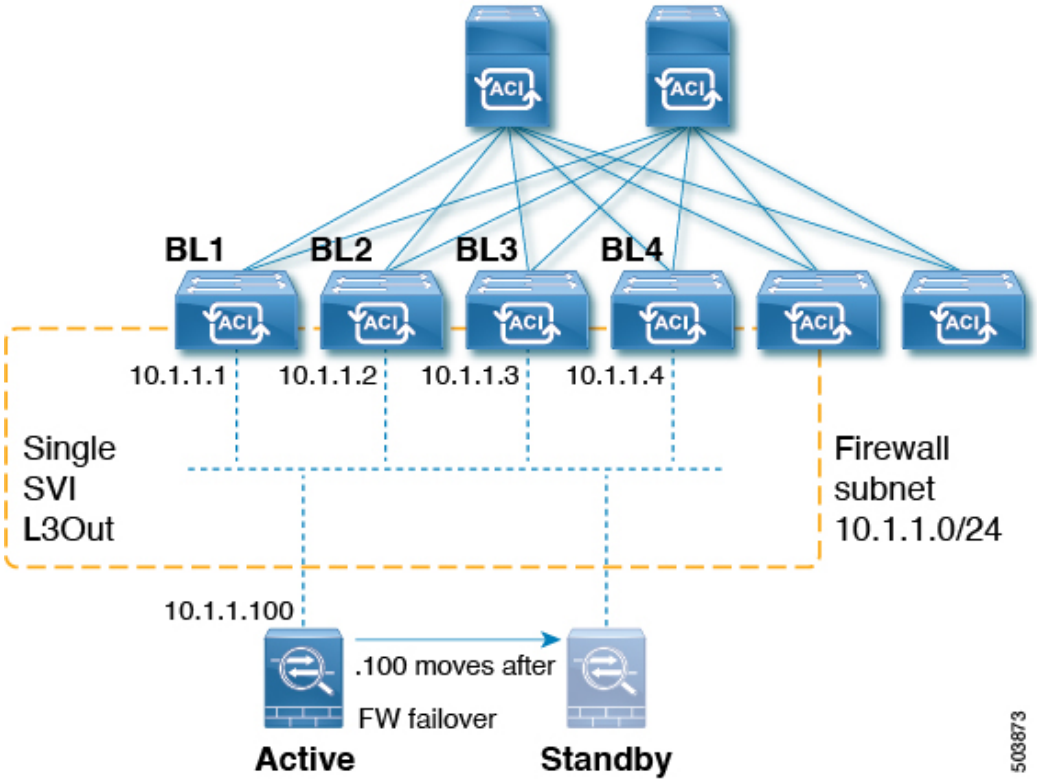
Step 8 Verify that it is acceptable that the nodes and policies displayed in the table in the Policy Usage Warning window will be affected by this policy change to enable multicast source and/or destination filtering, then click **Submit Changes**.

About Layer 3 Multicast on an SVI L3Out

Layer 3 multicast on an L3Out SVI adds support for enabling PIM on L3Out SVIs. This allows the ACI border leaf switch configured with an L3Out SVI to establish PIM adjacencies with an external multicast router or firewall.

Firewalls are usually deployed in active/standby pairs, where both firewalls connect to the fabric on the same VLAN and subnet, as shown below.



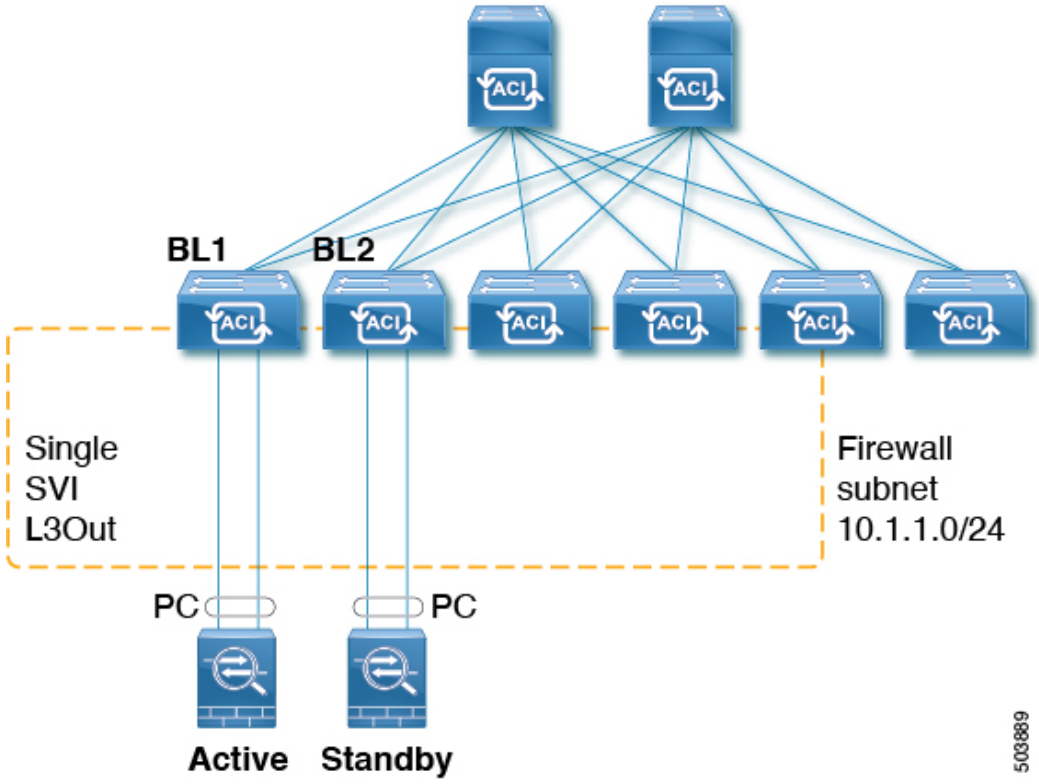


Because this is a LAN-like topology, it requires an SVI L3Out on the fabric side. Beginning with release 5.2(3), support is available for Layer 3 multicast on an SVI L3Out.

An L3Out SVI is an interface type where a Layer 3 SVI interface is configured on every border leaf switch where the SVI is deployed. When PIM is enabled on an L3Out that is configured with an SVI, the PIM protocol will be enabled on the border leaf switch that is part of the SVI. All SVIs will then form PIM adjacencies with each other and any external PIM-enabled devices.

L3Out to Firewall Example Topology

The following figure shows an example topology for an L3Out to firewalls.

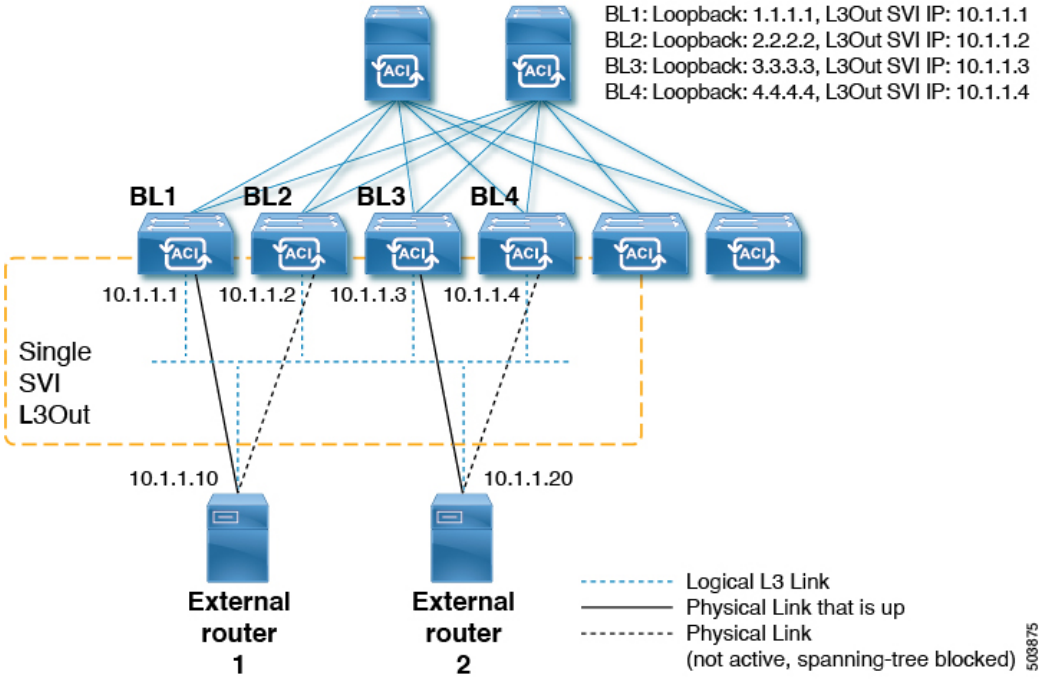


In this example, BL1 and BL2 are the border leaf switches on the fabric. Both border leaf switches are on the same SVI L3Out that connects to the external firewalls. Each firewall is connected to one of the two border leaf switches over a port-channel (non-vPC).

- Each border leaf switch will form a PIM neighbor adjacency to the active firewall.
- BL2 in the example will peer to the active firewall over the fabric tunnel for the L3Out external bridge domain.
- The active firewall can send PIM joins/prunes to both BL1 and BL2.
- One of the two border leaf switches will send the PIM joins towards the firewall. The border leaf switch that sends the PIM join towards the firewall is determined by the stripe winner selection for the multicast group (group and source for SSM).
- BL2 can be selected as the stripe winner for a multicast group. BL2 in the example topology is not directly connected to the active firewall. BL1 will notify BL2 that it is the directly connected reverse path forwarding (RPF) to the source. BL2 can send the PIM via BL1. BL2 must be able to perform a recursive lookup for the IP address of the firewall. This functionality is provided by the attached-host redistribution feature. A route-map matching the firewall subnet must be configured for attached-host redistribution on the L3Out.

L3Out SVI to External Switch/Router Example Topology

The following figure shows an example topology for an L3Out SVI to external switches or routers.



With respect to the Layer 3 multicast states and multicast data traffic, the components in the figure above are affected in the following manner:

- BL1, BL2, BL3, and BL4 are the border leaf switches on the fabric. All of these border leaf switches are on the same SVI L3Out that connect to the external boxes, where the external boxes could be any external switch or router.
- Logically, the Layer 3 link is up between the border leaf switches and the external routers. So a full mesh adjacency exists with regards to unicast routing protocol(s) or PIM across the border leaf switches and the external switches/routers on the SVI L3Out.
- Since the SVI L3Out is a bridge domain, even if there are multiple physical connections from border leaf switches to the external switches/routers, only one link among them will be up at the Layer 2 level to each external switch/router. All of the other links will be blocked by STP.

  For example, in the figure above, only the following links at the Layer 2 level are up:

  - The link between BL1 and external router 1
  - The link between BL3 and external router 2

  So for all of the other border leaf switches, this makes the IP addresses 10.1.1.10 reachable only through BL1 and 10.1.1.20 reachable only through BL3.

Guidelines and Limitations

- An attached-host route-map must be configured for the PIM-enabled SVI L3Out. This route-map should match all directly-connected external PIM neighbors. The 0.0.0.0/0 subnet can be used.
- For the Layer 3 multicast on an SVI L3Out feature, the following areas are supported or unsupported:
  - **Supported**:

- Protocol Independent Multicast (PIM) Any Source Multicast (ASM) and Source-Specific Multicast (SSM)
- SVI with physical interfaces
- SVI with direct port-channels (non-vPC)
- All topology combinations:
  - Source inside receiver inside (SIRI)
  - Source inside receiver outside (SIRO)
  - Source outside receiver inside (SORI)
  - Source outside receiver outside (SORO)

- **Unsupported**:
  - Layer 3 multicast with VPC over an SVI L3Out
  - Source or receiver hosts connected directly on the SVI subnet (source or receiver hosts must be connected behind a router on the SVI L3Out)
  - Stretched SVI L3Out between local leaf switches (ACI main data center switches) and remote leaf switches
  - Stretched SVI L3Out across sites (Cisco ACI Multi-Site)
  - SVI L3Out for PIMv6
  - Secondary IP addresses. PIM joins/prunes will not be processed if sent to the secondary IP address of the border leaf switch. Secondary IP address are typically used for configuring a shared (virtual) IP address across border leaf switches for static routing. We recommend that you use dynamic routing when configuring PIM over SVIs or create static routes to each border leaf switch primary address.

Configuring Layer 3 Multicast on an SVI L3Out Using the GUI

Procedure

Step 1 Configure a standard L3Out using the Create L3Out wizard with `SVI` set as the Layer 3 interface type.

a) In the GUI **Navigation** pane, under the Tenant Example, navigate to **Networking** > **L3Outs**.
b) Right-click and choose **Create L3Out**.
c) In the **Create L3Out** screen, in the **Identity** window, enter a name for the L3Out and select a VRF and L3 domain to associate with this L3Out.
d) Click Next when you have entered the necessary information in the **Identity** window.
The **Nodes and Interfaces** window appears.
e) In the **Nodes and Interfaces** window, in the **Interface Types: Layer 3** field, choose `SVI` as the Layer 3 interface type.

f) Continue configuring the individual fields through the Create L3Out wizard until you have completed the L3Out configuration.

Step 2 Navigate to the configured L3Out:

Tenants > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name***

The **Summary** page for the configured L3Out is displayed.

Step 3 Click on the **Policy** tab, then the **Main** subtab.

The **Properties** page for the configured L3Out is displayed.

Step 4 In the **Route Profile for Redistribution** field, click **+** to configure a route profile for redistribution.

Step 5 In the **Source** field, choose **attached-host**.

Step 6 In the **Route Map** field, configure a route map that permits all.

a) Click **Create Route Maps for Route Control**.

The **Create Route Maps for Route Control** window is displayed.

b) Enter a name and description for this route map, then click **+** in the **Contexts** area.

The **Create Route Control Context** window is displayed.

c) Configure the necessary parameters in the **Create Route Control Context** window, with the value in the **Action** field set to **Permit**.

d) Click **+** in the **Associated Match Rules** area, then choose **Create Match Rule for a Route Map** to configure the match rules for this route control context.

The **Create Match Rule** window is displayed.

e) Click **+** in the **Match Prefix** area.

The **Create Match Route Destination Rule** window is displayed.

f) In the **Create Match Route Destination Rule** window, enter the following values in these fields to configure a match rule with an aggregate route matching the subnet or 0.0.0.0/0 route and aggregate setting:

- **IP**: 0.0.0.0/0
- **Aggregate**: Check the box in this field. The **Greater Than Mask** and **Less Than Mask** fields appear.
- **Greater Than Mask**: 0
- **Less Than Mask**: 0

g) Click **Submit** to configure this match route destination rule.

Step 7 Once you have configured a route map that permits all, configure an external EPG with an export route control subnet that does an aggregate export of the aggregate route or the 0.0.0.0/0 route.

a) Navigate to the configured external EPG:

Tenants > ***tenant_name*** > **Networking** > **L3Outs** > ***L3Out_name*** > **External EPGs** > ***external_EPG_name***

The **Properties** page for the configured L3Out is displayed. You should be in the **Policy/General** page by default.

b) In the **Subnets** area, double-click on the 0.0.0.0/0 entry that you just configured.

The **Properties** window for this configured subnet is displayed.

c) In the **Route Control** area, make the following selections:

- Check the box next to the **Export Route Control Subnet** field.
- In the **Aggregate** area, check the box next to the **Aggregate Export** field.

d) Click **Submit**.