Layer 3 IP Multicast Architecture and Design in Cisco ACI Fabric

What You Will Learn

Many enterprise data center applications require IP multicast support and rely on multicast packet delivery across Layer 3 boundaries to provide necessary services and functions. Cisco® Application Policy Infrastructure Controller (APIC) Release 2.0(1) introduces the capability to support IP multicast routing across bridge domains in tenant Virtual Routing and Forwarding (VRF) instances, using the hardware capabilities of the Cisco Nexus® 9300 EX platform leaf switches that support the Cisco Application Centric Infrastructure (Cisco ACI™) solution.

This document provides an introduction to the multicast routing solution for the Cisco ACI fabric. It discusses best-practice designs and provides a brief technical discussion of the implementation. Configuration and monitoring procedures are provided as well.

Solution Overview

Prior to APIC Release 2.0(1), IP multicast support in the Cisco ACI fabric was limited to constraining IPv4 multicast at Layer 2 within each bridge domain based on the Internet Group Management Protocol (IGMP) snooping state. Any inter–bridge domain multicast routing, as well as multicast routing into or out of the Cisco ACI fabric, required a Protocol-Independent Multicast (PIM) router external to the fabric to perform those functions.

With the introduction of APIC 2.0(1), along with the Cisco Nexus 9300 EX leaf-switch platforms based on the leaf-and-spine engine (LSE) application-specific integrated circuit (ASIC), the Cisco ACI fabric itself provides distributed Layer 3 IP multicast routing between bridge domains, reducing or eliminating the need for external multicast routers.

The following multicast protocols are supported with this release:

- PIM any-source multicast (PIM-ASM)
- PIM source-specific multicast (PIM-SSM)
- Static rendezvous point (RP), Auto-RP, and bootstrap router (BSR) for RP-to-group mapping

Note that bidirectional PIM (PIM-bidir), IPv6 multicast (PIM6 and multicast listener discovery [MLD]), and PIM rendezvous point functions are not supported in the Cisco ACI fabric in APIC 2.0(1). In addition, Layer 3 multicast routing is not supported with fabric extenders or in conjunction with the multi-pod function also introduced in APIC 2.0(1).

Platform Support

Native Layer 3 IP multicast forwarding between bridge domains in the Cisco ACI fabric requires Cisco Nexus 9300 EX platform leaf switches, built with the LSE ASIC. Earlier leaf-switch platforms do not have the hardware capability to perform inter–bridge domain multicast routing and require an external multicast router to perform this function.
The specific leaf switches that support Layer 3 multicast routing in Cisco ACI at the time of this writing include:

- **Cisco Nexus 93180YC-EX Switch**: 48 x 10 and 25-Gbps Enhanced Small Form-Factor Pluggable (SFP+) ports with 6 x 100-Gbps Quad SFP28 (QSFP28) ports
- **Cisco Nexus 93108TC-EX Switch**: 48 x 10GBASE-T ports with 6 x 100-Gbps QSFP28 ports

Future EX platform leaf switches will also support Layer 3 multicast routing; refer to the latest software release notes for current information about platform support for multicast routing in Cisco ACI.

Because IP multicast forwarding is transparent to the spine switches in Cisco ACI, the spine switches have no hardware dependency. Any Cisco Nexus 9300 or 9500 platform with hardware support for Cisco ACI can serve as a spine switch when the leaf switches are performing native Layer 3 multicast routing.

**Best Practice Recommendations**

This section describes recommended best practices for three possible Cisco ACI fabric deployment scenarios. The scenarios differ in the capabilities of the leaf-switch platforms (Figure 1):

- **All leaf switches are first-generation switches that do not use the Cisco Nexus EX platform.** They are based on the application leaf engine (ALE) ASICs and require external multicast routers to perform inter-bridge domain and entry and exit multicast routing.
- **All leaf switches are second-generation Cisco Nexus EX platform switches.** They are based on the LSE ASIC and support native inter-bridge domain Layer 3 multicast routing as well as entry and exit multicast routing at the border leaf.
- **The leaf switches are a hybrid of some Cisco Nexus EX platform leaf switches and some leaf switches that do not use the EX platform.**

**Figure 1.** Three Possible Multicast Deployment Scenarios

**Scenario 1: Leaf Switches Not Based on Cisco Nexus EX Platform**

There are no changes to the best-practices recommendations for scenario 1 (Figure 2). Refer to the user documentation for information about the integration of external multicast routers with the Cisco ACI fabric to support inter-bridge domain and entry and exit IP multicast routing.
Scenario 2: Leaf Switches Based on Cisco Nexus EX Platform

For Cisco ACI fabrics in which all leaf switches are based on the EX platform (Figure 3), the best-practices recommendation is to enable native IP multicast routing in the Cisco ACI fabric. This configuration uses the latest technology generation, simplifies the network design, and simplifies IP multicast routing configuration and management. This document describes how to configure Layer 3 multicast in the Cisco ACI fabric and discusses the technical details of the solution.

Figure 2. Scenario 1: External PIM Router

Figure 3. Scenario 2: Native Layer 3 Multicast
Scenario 3: Hybrid Fabric with Leaf Switches both Based on and Not Based on Cisco Nexus EX Platform

In a hybrid environment (Figure 4), in which some of the leaf switches are not based on the EX platform and others are based on the EX platform, the best-practice recommendation is to continue to use an external router to perform multicast routing. Although it is technically possible to combine native multicast routing on EX platform leaf switches for some bridge domains with external multicast routing for other bridge domains, design, configuration, and management becomes increasingly complex and error prone. Although this type of design is discouraged, some details of this approach are presented in the appendix.

![Figure 4. Scenario 3: Hybrid Leaf Capability](image)

Layer 3 Multicast Routing Configuration

This document does not cover the process of configuring tenants, VRF instances, and bridge domains within the APIC. Only the process of enabling Layer 3 multicast is discussed. Refer to the user documentation for general Cisco ACI configuration procedures.

When you enable multicast routing in the APIC, you enable it at the tenant VRF level and then, optionally, at the bridge domain level. For example, if you have a tenant VRF instance with multiple bridge domains, you can enable Layer 3 multicast on all those bridge domains or only on a subset. In either case, you must first enable multicast at the VRF level in order to enable multicast routing on one or more bridge domains within that VRF instance (Figure 5).

![Figure 5. Layer 2 Versus Layer 3 Multicast for Tenant VRF Instances and Bridge Domains](image)
As shown in Figure 5, Tenant VRF1 has Layer 3 multicast enabled for the VRF instance and for all the bridge domains in that VRF instance. Leaf switches can route multicast traffic between any of those bridge domains, and border leaf switches can route traffic into and out of the Cisco ACI fabric for those bridge domains.

Tenant VRF2 has Layer 3 multicast enabled for the VRF instance, but not all the bridge domains have Layer 3 multicast enabled. Leaf switches can route multicast traffic between BD1 and BD2, but not into BD3. BD3 may or may not have Layer 2 multicast enabled (Layer 2 multicast with IGMP snooping in the bridge domain is enabled by default but can be disabled). If it does, IP multicast traffic can be constrained within the bridge domain, but it cannot be routed to other bridge domains or into and out of the fabric.

Tenant VRF3 does not have Layer 3 multicast enabled, but may have Layer 2 multicast enabled for some or all the bridge domains. The leaf switches perform no inter-bridge domain routing in this case. An external PIM router must provide any inter-bridge domain multicast routing.

Assuming that you have created a tenant with VRF instances and bridge domains on which you want to enable Layer 3 multicast, navigate to the Tenants main tab and select the desired tenant from the navigation pane at the left. Expand Networking > VRFs, expand the target VRF instance, and select Multicast (Figure 6).

**Figure 6.** Navigate to the Tenant VRF Multicast Screen

Click the button in the work area to enable multicast for the VRF instance (Figure 7).

**Figure 7.** Enable Layer 3 Multicast in the Tenant VRF Instance

PIM is not enabled on this VRF. Would you like to enable PIM?

YES, ENABLE MULTICAST.
After you enable multicast, five Configuration subtabs are displayed, where you can adjust the multicast parameters for the VRF instance (Figure 8).

**Figure 8.** Multicast Configuration Options

Use these subtabs to modify the various multicast parameters for the tenant VRF instance:

- **Interfaces:** Select the bridge domains and Layer 3 outside (L3out) interfaces on which multicast will run in the VRF instance.
- **Rendezvous Points:** Define the static rendezvous point, Auto-RP, and bootstrap router (BSR) parameters.
- **Pattern Policy:** Define various PIM-ASM policies (expiration timers, register policy, etc.) as well as the PIM source-specific multicast (PIM-SSM) group-range policy (the default SSM group range is 232.0.0.0/8).
- **PIM Settings:** View the VRF GiPo address (see the Technical Details section for more information about the GiPo address) and enable Fast Convergence. Also limit the number of multicast state entries that the VRF instance can create.
- **IGMP Settings:** Define SSM translation entries.

**Minimum Multicast Configuration: PIM-ASM**

The minimum configuration for basic Layer 3 PIM-ASM requires you to enable multicast for the VRF instance, add one or more bridge domains on the Interfaces configuration subtab, and define a static rendezvous point address in the Rendezvous Points subtab. The PIM rendezvous point must be located outside the Cisco ACI fabric in APIC 2.0(1). Verify that the rendezvous point IP address is reachable from inside the fabric.

This section describes the steps to follow to enable Layer 3 multicast in a tenant VRF instance using PIM-ASM. These steps assume that unicast forwarding with an L3Out interface is already configured and working.

1. In the navigation pane, expand Networking > VRFs, expand the target VRF instance, and select Multicast.
2. Click the button to enable multicast for the VRF instance.
3. On the Interfaces subtab in the work area, add one or more bridge domains on which to enable multicast and add an associated L3Out interface. This process enables PIM on the border leaf switches and on the target bridge domains (Figure 9).

**Figure 9.** Tenant VRF Instance with Two Bridge Domains Enabled for Multicast and One Associated L3Out Interface
4. Assuming that you are enabling PIM-ASM with a static rendezvous point, select the Rendezvous Points subtab in the work area and enter the IP address of the PIM rendezvous point. The PIM rendezvous point must be located outside the Cisco ACI fabric and must be reachable in unicast routing (Figure 10).

**Figure 10.** Tenant VRF Instance with Static Rendezvous Point IP Address Defined

Note: For a multicast source inside the Cisco ACI fabric, you must verify unicast reachability for the source’s pervasive subnet inside the fabric. Reachability requires one of the following:

- Add an explicit contract between multicast sources (as the provider endpoint group [EPG]) and receivers (as the consumer EPG), if such a contract does not already exist for unicast flows. This contract is used only to ensure unicast reachability. Contracts are not enforced for multicast traffic.
- Use unenforced contracts in the VRF instance, so that all pervasive subnets in that VRF instance are pushed to all leaf switches that have the VRF instance deployed.

At this point, any combination of the following traffic patterns should be operational for PIM-ASM:

- Inter–bridge domain multicast routing for sources and receivers in either of the bridge domains you specified in step 3
- Multicast routing for sources inside the Cisco ACI fabric to receivers outside the fabric
- Multicast routing for sources outside the Cisco ACI fabric to receivers inside the fabric

**Minimum Multicast Configuration: PIM-SSM**

The minimum configuration for basic Layer 3 PIM-SSM requires you to enable multicast for the VRF instance, add one or more bridge domains on the Interfaces configuration subtab, and enable IGMPv3 processing on those bridge domains (PIM-SSM does not require a rendezvous point).

This section describes the steps to follow to enable Layer 3 multicast in a tenant VRF instance using PIM-SSM. These steps assume that unicast forwarding with an L3Out interface is already configured and working.

1. In the navigation pane, expand Networking > VRFs, expand the target VRF instance, and select Multicast.
2. Click the button to enable multicast for the VRF instance.
3. On the interfaces subtab in the work area, add one or more bridge domains on which to enable multicast. This process enables PIM on the target bridge domains. If desired, add an associated L3Out interface for the VRF instance. This step is optional with PIM-SSM and is required only if you want to route PIM-SSM traffic into or out of the Cisco ACI fabric (PIM-SSM routing does not require a rendezvous point, and therefore an L3Out interface is not required to perform inter–bridge domain SSM routing inside the Cisco ACI fabric).
4. In the navigation pane, expand Networking > Bridge Domains and select the target bridge domain (Figure 11).

**Figure 11.** Navigate to and Select the Bridge Domain to Enable IGMPv3

5. In the work area, create a new IGMP policy, or apply an existing one, for the bridge domain (Figure 12).

**Figure 12.** Create (or Apply an Existing) IGMP Policy for the Bridge Domain
6. Verify that the IGMP policy enables IGMPv3. This setting is required to process IGMPv3 (S, G) reports and to create PIM-SSM state on the leaf switches. Repeat steps 5 and 6 for each bridge domain in which you want to enable PIM-SSM routing (Figure 13).

Figure 13. Enable IGMPv3 in the IGMP Policy Applied to the Bridge Domain

At this point, any combination of the following traffic patterns should be operational for PIM-SSM:

- Inter-bridge domain multicast routing for sources and receivers in either of the bridge domains you specified in step 3.
- Multicast routing for sources inside the Cisco ACI fabric to receivers outside the fabric, assuming that you associated an L3Out interface in step 3.
- Multicast routing for sources outside the Cisco ACI fabric to receivers inside the fabric, assuming that you associated an L3Out interface in step 3.
Monitoring Layer 3 Multicast

This section describes how to perform basic multicast monitoring tasks using the APIC GUI. Two main multicast monitoring tasks are discussed:

- How to monitor multicast state on fabric nodes
- How to enable and monitor multicast statistics on leaf-switch interfaces

Monitoring Multicast State

You can monitor a variety of multicast state entries on the leaf and spine switches. Follow these steps to view the multicast protocol state in the fabric:

1. Navigate to the Fabric main tab.
2. In the navigation pane at the left, expand the desired leaf or spine switch, expand Protocols, and select a multicast protocol to view the state. PIM is used in this example; other relevant protocol items for multicast are IGMP and IGMP snooping.
3. Expand the protocol item (in this case, PIM), expand the desired VRF instance, and select the item you want to monitor (Figure 14).

Figure 14. Select and Expand the Target Protocol to Monitor State
You can monitor a variety of states for PIM, IGMP, and IGMP snooping:

- **PIM (Figure 15):**
  - Neighbors
  - Interfaces
  - Routes
  - Group ranges
  - Rendezvous points
  - BSRs
  - Auto-RPs

- **IGMP (Figure 16):**
  - Groups
  - Interfaces

- **IGMP snooping (Figure 17):**
  - Groups
  - Routers
  - Queriers
  - Statistics

*Figure 15. PIM Routes on Leaf Switch*
Monitoring Multicast Interface Statistics

By default, the APIC does not collect multicast interface statistics. Follow these steps to enable multicast interface statistics on one or more interfaces:

1. Navigate to the Fabric main tab.
2. In the navigation pane at the left, expand the desired leaf or spine switch, expand Interfaces > Physical Interfaces, and select the desired physical interface (Figure 18).
3. In the work area, select the Stats subtab and then click the Select Stats icon (the checkmark icon; Figure 19).

**Figure 19.** Select the Stats Subtab and Click to Select Stats

4. In the Select Stats window, select the desired multicast-related statistics from the Available pane and click the right arrow button to move them to the Selected pane; then click Submit (Figure 20).
Figure 20. Move the Multicast Statistics Items to the “Selected” Pane

At this point, the interface statistics graph should show the multicast-related statistics. You can roll the mouse over the items in the key to highlight just that item, or click an item to toggle between hiding and showing that item (Figure 21).
Technical Details

This section describes in more detail the technical implementation and operation of Layer 3 multicast routing in the Cisco ACI fabric when you use Cisco Nexus EX platform leaf switches. The role that each leaf switch plays in multicast routing depends on whether that leaf functions as a border leaf (BL) switch or a non–border leaf (non-BL) switch (Figure 22):

- **Border leaf switch**: Responsible for routing multicast traffic into and out of the Cisco ACI fabric, and for performing inter–bridge domain multicast routing for any devices directly connected to the border leaf.
- **Normal leaf switch**: Responsible for performing inter–bridge domain multicast routing for multicast traffic entering the fabric at that leaf, and for routing traffic toward the border leaf switches, if necessary.
When you enable multicast routing in the APIC, you enable it at the VRF level and then, optionally, at the bridge domain level. For example, if you have a tenant VRF instance with multiple bridge domains, you can enable Layer 3 multicast on all those bridge domains or only on a subset. Note that the VRF instance must have multicast enabled for Layer 3 multicast routing on any bridge domains in that VRF instance to operate.

**GiPo Addresses**

Prior to APIC 2.0(1), software assigned a group IP address called bridge domain group IP–outer (BD GiPo) for each bridge domain on which Layer 2 multicast is enabled. This is the destination multicast group address used in the outer IP header whenever a leaf switch encapsulates Layer 2 broadcast, multicast, and unknown unicast traffic in the Cisco ACI fabric for transport to other leaf switches.

Beginning with APIC 2.0(1), an additional group IP address—the VRF GiPo—is assigned for each VRF instance on which you enable Layer 3 multicast. The VRF GiPo becomes the outer destination IP address for routed multicast traffic carried over the fabric.

APIC 2.0(1) also introduces a new virtual interface for multicast routing on the leaf switches: the underlay multicast tunnel interface. When you enable Layer 3 multicast routing for a VRF instance, the underlay multicast tunnel interface becomes the single outgoing interface (OIF) for the VRF GiPo on leaf switches for any multicast routes (mroutes) learned in the fabric. Multicast traffic routed within the Cisco ACI fabric traverses the underlay multicast tunnel, with each leaf switch load-sharing the traffic over all the available multi-destination trees (MDTs). All leaf switches on which a Layer 3 multicast-enabled VRF instance is active become receivers for the VRF GiPo address and receive multicast traffic from the underlay multicast tunnel interface. Each leaf switch is responsible for constraining multicast traffic received from the fabric to just those bridge domains and edge ports that have interest in the multicast stream.
Border Leaf Designated Forwarder

For each (*, G) and (S, G) mroute, software selects a single border leaf switch to be the designated forwarder (DF) if the VRF instance has an associated L3Out interface (Figure 23). The designated forwarder is responsible for sending PIM joins upstream to the rendezvous point or the source to attract the traffic to the Cisco ACI fabric. When a fabric has multiple border leaf switches, each border leaf switch is aware of all other border leaf switches in the fabric, and all the switches use a deterministic method to select the border leaf that will be designated forwarder for each (*, G) and (S, G) mroute. Only one designated forwarder can be active at one time for a given mroute to prevent duplicate multicast packets from being sent to the Cisco ACI fabric.

Figure 23. Designated Forwarder

Optionally, you can enable the Fast Convergence option so that all multicast-enabled border leaf switches with direct reachability to the rendezvous point or source send PIM joins to attract the traffic. (Fast Convergence is enabled on the PIM Setting subtab of the Multicast configuration screen.) Only the designated forwarder will add the fabric as an OIF; the other border leaf switches simply discard the traffic. This option can accelerate convergence in the event of a border-leaf designated forwarder failure, because the new designated forwarder only needs to add the fabric as an OIF instead of having to send PIM joins to the rendezvous point or source to build the multicast distribution tree (Figure 24). However, the trade-off is that multicast streams are pulled in parallel to all of the border leaf switches, increasing network bandwidth utilization and potentially replication load on upstream devices.

Figure 24. Accelerating Convergence after Designated Forwarder Failure
First-Hop Router
A leaf switch is considered a first-hop router (FHR) if a directly connected multicast source is attached to a segment on that switch. The basic mechanics of an FHR in the Cisco ACI fabric are similar to those of an FHR in a traditional Ethernet/IP network. For PIM-ASM, the FHR must detect packets from new sources, encapsulate those packets in PIM register packets, and send those packets as unicast traffic to the PIM rendezvous point.

However, in the Cisco ACI fabric, an FHR can source PIM register packets from the IP address of any of the bridge domains in the tenant VRF instance active on that leaf switch, which is not necessarily the IP address of the bridge domain receiving new multicast source traffic (Figure 25).

Figure 25. First-Hop Router

Last-Hop Router
A leaf switch is considered a last-hop router (LHR) if a directly connected multicast receiver is attached to a segment on that switch. However, the behavior of a non-border leaf LHR in the Cisco ACI fabric varies slightly from that of a traditional LHR. Like any LHR, both border leaf and non-border leaf LHRs periodically send PIM Hellos and IGMP queries on multicast-enabled bridge domains and process IGMPv2 and v3 reports and leaves from hosts.

Upon receipt of an IGMPv2 or v3 report, the LHR takes one or more of several actions, depending on the scenario:

- If the IGMP report was the first join for a particular (*, G) or (S, G), the LHR switch:
  - Creates a new Layer 2 IGMP snooping table entry in the receiving bridge domain
  - Creates a new Layer 3 IGMP table entry in the receiving VRF instance
  - Creates a new Layer 3 mroute with the receiving bridge domain as an OIF
  - Updates the Cisco ACI Council of Oracles Protocol (COOP) with both the bridge domain VXLAN network identifier (VNID) and the VRF VNID
- If the IGMP report was not the first join, the LHR switch simply updates the various tables based on the receiving interface, bridge domain, and VRF instance.

Upon receiving interest for a new (*, G) or (S, G), the COOP oracle publishes an update to the border leaf switches active for the target bridge domain and VRF VNIDs. If the border leaf switch is the designated forwarder for the mroute, a PIM join is sent upstream if necessary to attract the traffic to the Cisco ACI fabric (Figure 26).

**Figure 26.** Last-Hop Router: IGMP Reports

For IGMP leaves, the LHR processing is similar to that for the IGMP report case:

- If the IGMP leave was the last leave for a particular (*, G) or (S, G), the LHR switch:
  - Removes the Layer 2 IGMP snooping table entry
  - Removes the Layer 3 IGMP table entry
  - Removes the Layer 3 mroute
  - Withdraws the bridge domain VNID and the VRF VNID from COOP

- If the IGMP leave was not the last leave, the LHR switch simply updates the various tables based on the receiving interface, bridge domain, and VRF instance.

If the leaf was the last switch in the fabric with interest in the (*, G) or (S, G), the COOP oracle publishes an update to the border leaf switches active for the target bridge domain and VRF VNIDs. If the border leaf is the designated forwarder for the mroute, a PIM prune is sent upstream if necessary (Figure 27).
The border leaf switches in Cisco ACI maintain the traditional LHR behavior of setting the shortest path tree (SPT) threshold to zero, but the SPT threshold on non-border leaf switches is always set to infinity (Figure 28). This configuration prevents PIM-ASM multicast traffic from being punted to the CPU at a non-border leaf LHR, and also avoids unnecessary (S, G) state in the non-border leaf LHRs. Therefore, only the border leaf that is the designated forwarder for a particular (*, G) mroute will punt the packets arriving on the shared tree to the CPU to create (S, G) state and trigger a PIM (S, G) join to the source (assuming that the source is located outside the Cisco ACI fabric).

Rendezvous Point
The industry best-practice for rendezvous point configuration is AnycastRP using MSDP, with static rendezvous point address configuration. In APIC 2.0(1), the Cisco ACI fabric itself cannot function as the PIM rendezvous point for PIM-ASM; the rendezvous point must exist outside the fabric. The Layer 3 multicast configuration in the Cisco ACI fabric provides support for specifying a static rendezvous point address for PIM-ASM, as well as dynamic options for disseminating rendezvous point information such as BSR and Auto-RP.
Spine Functions
Multicast functions in the spine switches are limited to packet replication to the necessary leaf switches. The fanout is determined by the output interface list (OIL) for the VRF GIPo mroute that the packet matches. The GIPO OIL always consists of all the leaf switches on which the VRF instance exists. No pruning occurs in the spine switch based on receiver interest. Each leaf switch receiving packets from the fabric determines whether any local egress replications need to occur for those packets. If one or more bridge domains have interested receivers attached, the leaf transmits a copy of the packets on the relevant ports in each bridge domain (Figure 29).

Figure 29. Multicast Fan-Out at the Spine

Multicast Scale
Layer 3 multicast scale parameters are documented in the Cisco ACI Verified Scale Guide for APIC Release 2.0(1).

Conclusion
Cisco APIC 2.0(1) delivers new capabilities for data center applications requiring Layer 3 IP multicast forwarding, using the hardware capabilities of the Cisco Nexus 9300 EX leaf switches that support Cisco ACI. As shown in this document, you can now use the latest technology to simplify the network design as well as streamline configuration and management of IP multicast routing, reducing or eliminating the need to have an external PIM router connected to the Cisco ACI fabric.
Appendix: Hybrid Fabric Details

This section discusses the hybrid leaf scenario, in which some leaf switches are not based on the Cisco Nexus EX platform (and therefore are not capable of performing Layer 3 multicast routing), and other leaf switches are based on the EX platform (and therefore are capable of Layer 3 multicast routing). The best practice recommendation for this type of deployment is to use an external PIM router to perform multicast routing for all VRF instances and bridge domains, to keep the design straightforward and to simplify operation and troubleshooting.

However, technically it is possible to combine multicast routing using external PIM routers for some VRF instances and bridge domains that are deployed on leaf switches that are not EX based (or on a combination of leaf switches that are and are not EX based), while enabling native IP multicast routing for VRF instances and bridge domains that are deployed only on leaf switches that are EX based (Figure 30).

**Figure 30.** Combination of External PIM Routing with Native IP Routing in Cisco ACI Fabric
In designs combining external multicast routing for some VRF instances and bridge domains with native multicast routing for other VRF instances and bridge domains, be sure that you do **not** deploy VRF instances or bridge domains with Layer 3 multicast enabled on leaf switches that are not EX based. APIC will raise a fault if you do so, and multicast will not function correctly for sources or receivers connected to the VRF instances and bridge domains on those switches.

Also note the following additional points about the hybrid leaf design:

- If the border leaf switches are EX based, they can route multicast traffic in and out of the Cisco ACI fabric for those bridge domains deployed only on EX leaf switches, while simultaneously carrying traffic for externally routed bridge domains.
- Alternatively, you can specify different border leaf switches for native multicast routing and for external multicast routing.
- A given EX leaf switch can simultaneously have some bridge domains that use external multicast routing while also having some bridge domains that use native multicast routing.
- EX leaf switches (whether border leaf switches or non-border leaf switches) cannot perform “proxy” multicast routing for switches that are not EX based. You must use a multicast router external to the Cisco ACI fabric to perform this function.