



#CiscoLive



Deploying IP Multicast in ACI and Multi-Site Fabrics

John Weston, Technical Marketing Engineer

BRKDCN-2623

cisco ile

#CiscoLive

Cisco Webex App

Questions?

Use Cisco Webex App to chat with the speaker after the session

How

- **1** Find this session in the Cisco Live Mobile App
- 2 Click "Join the Discussion"
- 3 Install the Webex App or go directly to the Webex space
- 4 Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 17, 2022.



https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-2623

cisco / ille

Session Objectives



At the end of the session, the participants should be able to:

Understand how Tenant Routed Multicast (TRM) is deployed in ACI and ACI Multi-Site fabrics

Initial assumption:

- The audience already has a good knowledge of ACI main concepts (Tenant, BD, EPG, L3Out, Multi-Pod, Multi-Site, etc)
- The audience already has a good understanding of multicast routing (PIM, IGMP)



- Introduction
- Underlay and Overlay Multicast Control
 Plane
- Bridged IP Multicast
- Tenant Routed Multicast (TRM)
- Multicast in Multi-Pod Fabrics
- Multicast across Multi-Site Fabrics
- Conclusions and Q&A

Overview of ACI Tenant Routed Multicast TRM

- Multi-tenant aware multicast forwarding on the ACI fabric
- VXLAN encapsulated multicast with multicast enabled underlay



cisco ile

Underlay and Overlay Multicast Control Plane

cisco Liver



Multicast Underlay Control Plane



Underlay Control Plane Overview

- ACI reserves a multicast group range (default 225.0.0.0/15) for point to multipoint underlay tunnels
- Multicast addresses from this range are used in the VXLAN packet outer destination IP address called GIPo (Group IP Outer) address
- GIPo address ranges are assigned to BDs and VRFs
- PIM is not used in the underlay



Multicast Underlay Control Plane

Multicast GIPo range selected during initial fabric setup. Cannot be changed later.



cisco / ille

FTAG Distribution Trees



ACI uses Forwarding Tag (FTAG) distribution trees for forwarding multi-destination traffic inside the ACI fabric

- IS-IS is used to build loop free FTAG trees
- Multiple trees achieve load balancing across the fabric for multicast traffic
- FTAGs are assigned by APIC
- FTAGs are always rooted at spines
- FTAGs become inactive if spine is down. FTAGs do not move to other spines

- All tenant multicast is sent to fabric GIPo multicast address
- Fabric GIPo multicast address selects the FTAG

Multicast Underlay Distribution Trees FTAG Root: CLI verification



Multicast Underlay Distribution Trees FTAG Root: CLI verification



cisco ile

Multicast Underlay Distribution Trees FTAG Root: UI Verification

System Tenants Fabric		Virtual Networking	Admin Operatio		ons Apps	s Integr	ations			
Inventory Fabric Policies		Access Policies								
✓	IS-IS	-abric Multicast Trees					0.0			
> 🔄 Chassis							0 <u>+</u>			
> 🚞 Interfaces	ID	Root Address	Root Port	Diameter	Origin	Diameter Alert	Operational State			
> 🚞 Control Plane Statistics	0	10.21.184.64	unspecified	4	Static	Normal	Active			
✓	1	10.21.136.64	eth1/49.67	8	Learned through isis	Normal	Active			
> = BGP	2	10.21.136.64	eth1/49.67	8	Learned through isis	Normal	Active			
> = COOP	3	10.21.184.67	eth1/53.69							
> = IPV4 > = IPV6		10.21.184.70	eth1/54.70	Unspe	ecified = no root port.					
> = IPV6	5	10.21.184.64	unspecified							
> <mark>=</mark> IPV6 > = ISIS	6	10.21.184.64	unspecified	Th	is spine	is root.				
> = IPV6 > = ISIS > = ISIS for V/RE-overlay-1	5 6 7	10.21.184.64 10.21.184.64 10.21.184.64	unspecified unspecified	Th	is spine	e is root.				
 PV6 ISIS ISIS for VRF-overlay-1 	5 6 7 8	10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64	unspecified unspecified unspecified unspecified	Th	static	e is root.	Active			
 PV6 ISIS ISIS for VRF-overlay-1 Discovered Tunnel Endpoints Discovered Tunnel Endpoints 	5 6 7 8 9	10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64	unspecified unspecified unspecified unspecified unspecified		iis spine	e is root. Normal Normal	Active Active Active			
 F IPV6 F ISIS F ISIS for VRF-overlay-1 Discovered Tunnel Endpoints LSP Records IO IN Law formation 	5 6 7 8 9 10	10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.240.32	unspecified unspecified unspecified unspecified unspecified eth1/57.68	4 4 4	Static Static Static Learned through isis	e is root. Normal Normal Normal	Active Active Active Active			
 F IPV6 F ISIS F ISIS for VRF-overlay-1 Discovered Tunnel Endpoints LSP Records IS-IS Interfaces 	5 6 7 8 9 10 11 12	10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.240.32 10.21.240.32	unspecified unspecified unspecified unspecified unspecified eth1/57.68 eth1/57.68	4 4 8 8	Static Static Static Learned through isis Learned through isis	e is root. Normal Normal Normal Normal	Active Active Active Active Active			
 PV6 ISIS ISIS for VRF-overlay-1 Discovered Tunnel Endpoints LSP Records IS-IS Interfaces Neighbors 	5 6 7 8 9 10 11 12 12	10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.240.32 10.21.240.32 0.0.00	unspecified unspecified unspecified unspecified unspecified eth1/57.68 eth1/57.68 unspecified	4 4 8 8 0	Static Static Static Learned through isis Learned through isis Static	e is root. Normal Normal Normal Normal Normal	Active Active Active Active Active Active Inactive			
 F IPV6 F ISIS ISIS for VRF-overlay-1 Discovered Tunnel Endpoints LSP Records IS-IS Interfaces Neighbors IS-IS Routes 	5 6 7 8 9 10 11 12 13 14	10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.240.32 10.21.240.32 0.0.00	unspecified unspecified unspecified unspecified unspecified eth1/57.68 eth1/57.68 unspecified unspecified	4 4 4 8 8 0 0	Static Static Static Static Learned through isis Learned through isis Static Static	e is root. Normal Normal Normal Normal Normal Normal	Active Active Active Active Active Inactive			
 PV6 ISIS ISIS for VRF-overlay-1 Discovered Tunnel Endpoints LSP Records IS-IS Interfaces Neighbors IS-IS Routes IS-IS Fabric Multicast Trees 	5 6 7 8 9 10 11 12 13 14 15	10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.184.64 10.21.240.32 10.21.240.32 0.0.00 0.0.00 0.0.00	unspecified unspecified unspecified unspecified unspecified eth1/57.68 eth1/57.68 unspecified unspecified unspecified	4 4 8 8 0 0	Static Static Static Static Learned through isis Static Static Static Static	e is root. Normal Normal Normal Normal Normal Normal Normal	Active Active Active Active Active Inactive Inactive			

cisco

Multicast load balancing in the fabric



- Every bridge domain and PIM enabled VRF is assigned a GIPo address range (/28 range)
- When the leaf switch forwards multicast packets within the fabric it will select an FTAG tree based on a hash of the flow
- The FTAG tree is selected by GIPo address



GIPo: 225.1.192.0 FTAG range: 225.1.192.0-12



Fabric GIPo Usage



- FTAG trees are used for all multi-destination traffic, including broadcast, unknown unicast, and multicast (BUM)
- When PIM is enabled on a bridge domain, all inter and intra BD (same subnet) multicast will be forwarded on the VRF GIPo. This is an always route approach and will follow the rules for multicast routing
- The BD GIPo is used for intra BD multicast when PIM is not enabled. Multicast will not be forwarded outside of the bridge domain.
- The BD GIPo tree is also used for all other multi-destination traffic in the BD (broadcast, unknown unicast, and intra BD non-IP multicast traffic)

Traffic Type	BD without PIM enabled	BD with PIM enabled
Broadcast	BD GIPo	BD GIPo
Unknown Unicast Flood	BD GIPo	BD GIPo
Multicast	BD GIPo	VRF GIPo

Multicast Overlay Control Plane



Overlay Control Plane

- Receivers connected to the fabric will send IGMPv2/IGMPv3 join reports
- The leaf switch receiving the IGMP join reports will send a COOP message to the spine announcing group interest
- Spines maintain a database of all multicast interest (pod-wide IGMP snooping table)
- Border leafs subscribe to spines to receive
 multicast interest information
- Border leafs send joins to RP/Source. Border leafs perform the LHR function for non-border leafs with connected receivers



Multicast Overlay Control Plane

Check fabric IGMP groups - UI

- Pod operational tab displays IGMP group information for pod
- Shows leaf, VLAN, EPG, port information for each group

System	Tenants	Fabric	Virtual Networking	Admin	Operations	Apps	Integrations							
Inve	ntory Fab	ric Policies	Access Policies											
Inventory			\odot	Pod -	1									Q
> 🕩 Quick St	tart V								Summary	Dashboard	Topology	Interface Operational	Stats Faults	History
~ 🖨 Pod 1					Se	earchab	le by					BGP IGMP Multicas	st MLD Multicast	Routes
> ::::: F1-P	1-Leaf-101 (№	Node-101)			m	ulticast	group					Crown	Search Croups	Deutere
> 📰 F1-P	91-Leaf-102 (N	lode-102)			_	\sim						Group	Search Groups	Routers
> F1-P	91-Leaf-103 (N	lode-103)		MCast	Group Address: e	nter the addres	s Q							0 % .
> F1-P	91-Leaf-104 (N	Node-104)		Name	е	Node	Encap	EPG Name	Gq Misse	d Expires	Group Vers	ion Latest Reporter	No. Of Sources Joined	d Source IF
> 📰 F1-P	91-Leaf-105 (N	lode-105)		✓ 8	239.10.1.1	Node-102					v2	10.90.1.2	0	
> 🗐 F1-P	1-Spine-1001	(Node-1001)		~	82		vlan-2006	MC:AP:S-MCR-EPG1						
> 🗐 F1-P	1-Spine-1002	? (Node-1002)			v a wildcard									wildcard
🚞 Pod Fab	ric Setup Polic				po5				0	1969-12-3	i1			
E Fabric N	lembership			> 2	239.10.1.1	Node-101					v2	10.90.1.2	0	
🚞 Disabled	d Interfaces an	d Decommissio	ned Switches	>	239.10.1.1	Node-103					v2	10.90.1.1	0	
🚞 Duplicat	e IP Usage			> 8	239.10.1.1	Node-103					v2	10.90.2.1	0	
				>	239.10.1.1	Node-104					v2	10.90.1.1	0	
				> É	239.10.1.1	Node-104					v2	10.90.2.1	0	
				4										

Multicast Overlay Control Plane PIM Modes



L3Out Loopbacks

- PIM enabled L3Outs must be configured with loopback interfaces
- BLs form PIM adjacencies across the fabric with other BLs (this is used by BLs to create BL list for stripe winner selection)

PIM normal mode

- L3Out interfaces run standard PIM
- Forms adjacency with neighbor devices
- Configurable PIM settings



Multicast Overlay Control Plane PIM Neighbor verification - UI



cisco ile

Multicast Overlay Control Plane

PIM Neighbor verification – CLI verification

• Standard NXOS CLI commands can be used on leaf to view PIM information

Leaf-303# show ip pim neighbor vrf MC2:vrf									
PIM Neighbor information for Dom:MC2:vrf									
Neighbor	Interface	Uptime	Expires	DRPrior	rity Bidir				
BFDState	/								
10.214.2.1/32	eth1/16.80	7d5h	00:01:26	1	yes				
$\frac{1}{1}$ 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	tunnel19	00.04.18	00.01.41	1	no				
n/a	cumery	00.01.10	00.01.11	±	110				

cisco / il

Bridged IP Multicast

۲

۲

cisco live!

Bridge IP Multicast with IGMP Snooping

Same subnet IP multicast "L2 multicast"



- Multicast will be forwarded to all leaf switches where the BD is deployed
- Each Leaf switch will receive a copy of the multicast stream and forward out the front panel ports based on IGMP snooping and the bridge domain's Unknown Multicast Flooding configuration
 - Flood
 - Optimized Multicast Flood •

Spines forward based on BD GIPo FTAG tree. Multicast stream will be sent to all leaf switches where BD is deployed

IGMP snooping will forward multicast only out port with interested receivers

cisco /

BD1

BD2

Bridged IP multicast L3 Unknown Multicast Flooding: Flood



#CiscoLive BRKDCN-2623 © 2022 Cisco and/or its affiliates. All rights reserved. Cisco Public

23

Bridged IP Multicast

L3 Unknown Multicast Flooding: Optimized Flood



Bridged IP Multicast

L3 Unknown Multicast Flooding: Which mode to use?

L3 Unknown Multicast Flooding: Flood Optimized Flood

Optimized Flood: IGMP snooping behavior across the BD

- Optimized behavior, does not forward out ports with no interested receivers
- Recommended when all switches are 2nd generation*
- Flood: Flood on switches where no interested receivers connected
 - Non-optimized behavior
 - May result in congestion on switches with no receivers

*Optimized flood limited to 50 BDs for 1st generation switches

Tenant Routed Multicast (TRM)

cisco live!

Tenant Routed Multicast

Guidelines and Limitations

- Hardware Requirements
 - Leaf switch: Cisco Nexus 9300-EX, 9300-FX, FX2, FX3, 9300-GX, 9300-GX2
 - Spine switch: all models
- Supports PIM ASM and PIM SSM only, PIM BiDir is not supported
- Fabric RP and external RP supported
- Intra-VRF and Inter-VRF (extranet) supported
- Multicast traffic not filtered with contracts
- Control plane and data plane filtering options supported

L3 Unknown Multicast Flooding with TRM

- IGMP snooping must be enabled for PIM enabled BDs
- Multicast packets will only be forwarded out BD ports if there is an IGMP snooping entry
- The Bridge Domain L3 Unknown Multicast Flooding setting does not apply to PIM enabled BDs. Forwarding is always optimized



It is not required to change setting for PIM enabled BDs Always Optimized for PIM enabled BDs

TRM Multicast Forwarding

Different and same subnet - Always Route



cisco live!



Tenant Routed Multicast

Always Route Approach

Tenant Routed Multicast follows an "Always Route" approach

- TRM is enabled at the VRF and BD levels. Enabling PIM at the BD level enables TRM for that BD
- When PIM is enabled for the BD, all IP multicast will be routed across the fabric. Routed multicast has the following behavior
 - Sent in the VRF GIPo, VRF VNID
 - TTL is decremented (the TTL will be decremented twice when sent over the fabric. Once on the sending leaf and once on the receiving leaf)
 - Source MAC is rewritten
 - RP is required for ASM flows
- Does not apply to the link local multicast range 224.0.0.0/8. This ranges is always bridged

Always route forwarding behavior TTL Decrement Behavior



#CiscoLive BRKDCN-2623 © 2022 Cisco and/or its affiliates. All rights reserved. Cisco Public 32

TRM: Layer-3 Outside Role of L3Out with TRM



- PIM enabled L3Outs are always required for PIM ASM (not required for SSM if there are no external sources or receivers)
- Provides path to external RP
- Takes role of anycast RP for fabric RP
- PIM enabled BLs perform the LHR function for non-BLs with connected receivers
- LHR function distributed across all active BLs using hash function (stripe winner)
- PIM enabled L3Outs must be configured with a loopback interface
- L3Outs support routed, routed subinterfaces, L3 port-channels, and SVI interfaces (non-VPC)

cisco /

TRM L3Outs

Stripe winner function





All PIM enabled BLs learn of other BLs from

PIM hellos over the fabric

TRM L3Outs Fast Convergence



- When fast convergence is configured, all PIM enabled BLs will send PIM joins
- All BLs will receive multicast streams from external sources
- Only the stripe winner BL forwards multicast onto the fabric
- Multicast traffic can converge faster during certain failure scenarios because state is created in the eternal multicast network
- Tradeoff is that multiple copies of the streams are sent in the external network

TRM L3Outs

Stripe winner function: CLI verification


TRM L3Outs

Stripe Winner: Failure Scenarios, no route to source/RP



- If the stripe winner for a group does not have a route to the source/RP, it will send a PIM join to another BL (ECMP if multiple BLs have route to source/RP)
- Does not change stripe winner selection. The BL list does not change if PIM is up within the fabric

TRM L3Outs

Stripe Winner: Failure Scenarios, BL failure



L3Out PIM support on SVI interfaces

Guidelines and limitations

- Main use case is for connectivity to active/standby firewalls
- Supports PIM ASM and SSM
- Supported on physical interfaces and port-channel interfaces (not supported on VPC interfaces)
- L3Out SVI cannot be stretched to remote leaf switches
- PIM6 not supported
- Requires attached-host route-map distribution configuration

L3Out PIM support on SVI interfaces Firewall use case



- Both active and standby firewalls need to connect to the same subnet
- L3Out SVIs create a bridge domain (external bridge domain)
- L3Out SVIs support BGP, OSPF, EIGRP, and static routing
- L3Out SVIs can also form PIM peering adjacencies to PIM enabled routers/firewalls

L3Out PIM support on SVI interfaces PIM Neighborship



PIM hellos/JP messages are sent between BLs on the fabric VRF GIPo tunnel. This is existing behavior. The purpose of PIM hellos on the fabric VRF GIPo tunnel is to discover other BLs for stripe winner calculation

PIM on SVI feature uses the fabric External BD GIPo tunnel for forwarding PIM Hello/JP messages sent by devices on the external BD (BLs, and external devices)



L3Out PIM support on SVI interfaces Source inside fabric



- When PIM join/prunes from the external device are sent to the BL the join target may be any BL where the SVI is deployed (all devices are PIM neighbors)
- The PIM join may arrive on the BL that is not the join/prune target
- The PIM join will be forwarded ٠ across the fabric to the BL that is the PIM join target
- The BL that is directly connected • to the neighbor will implicitly install the multicast state (mroute), even though it is not the target

L3Out PIM support on SVI interfaces

Source outside fabric, receiver inside: Non-working case

Requires feature Redistribute Attached host route-map



- BLs with PIM enabled SVIs can be selected as stripe winners (no change to stripe winner calculation)
- A BL selected as the stripe winner may be the BL connected to the standby FW
- The PIM join sent to the external neighbor will go over the Ext-BD tunnel
- The directly connected BL (BL 103) will forward the join in the external BD but will not create the multicast state
- The BL will drop the multicast traffic received from the FW (as there is no (S,G) for group)

cisco / ila

Redistribute Attached-host route-map



Redistribute Attached-host route-map



L3Out PIM support on SVI interfaces

Attached host route-map redistribution



- PIM over SVI support requires requires "Attached host route-map" redistribution when there are external sources.
- "Attached host route-map" is a feature when configured will redistribute host routes into BGP for devices that are directly connected to the leaf in the L3Out SVI.
- The route-map should be configured to match the SVI subnet
- With this feature enabled, the stripe winner BL will learn the host route of the external PIM neighbor over MP-BGP
- The stripe winner BL will then use the fabric tunnel to join the source tree. This will create state on the directly connected BL

Redistribute Attached-host route-map

L3 Outside - FW-L3Out					00
	Summary	Policy	Stats	Faults	History
		Main	Node Profiles	s Exter	rnal EPGs
8 👽 🛆 📀				Ċ	<u>+</u> ***
Properties PIM: P					
Route Profile for Redistribution:					+
▲ Source attached-host	Route Map SVI-20-RM				



L3 Multicast Rendezvous Point (RP) Options

External RP Fabric RP Fabric RP peering with Static external RP Can be RP for external network Auto-RP (listen/forward) (static only) BSR (listen/forward) Uses PIM Anycast-RP RFC 4610 • 10.20.0.1 10.20.0.1 Anycast RP Anycast RP RP RP https://datatracker.ietf.org/doc/html/rfc4610 RP 10.10.0.1 Not Supported ip pim rp-address 10.20.0.1 External PIM network can use fabric RP with static configuration cisco / ille #CiscoLive BRKDCN-2623 48 © 2022 Cisco and/or its affiliates. All rights reserved. Cisco Public

Fabric RP Operation RFC 4160

- 1) FHR sends PIM register to anycast RP 10.1.1.101 from source of FHR 10.2.2.2
- 2) BL1 receives register from 10.2.2.2. Source of register is not on anycast BL list.
- 3) BL1 sends PIM register to all anycast RP members on anycast RP list (BL2 and BL3)
- 4) BL2 and BL3 receive PIM register from source of BL1 (1.1.1.101). Source of PIM register is on anycast BL list. BL2 and BL3 do not send PIM register packets.



PIM-ASM Rendezvous Point (RP) RP Filtering options

- RP configuration supports group ranges with route-map configuration (static and fabric RP)
- Auto-RP and BSR support filtering Auto-RP/BSR messages by source (MA filter/BSR filter)

Static RP: 10.10.0.1 Group Range 239.0.00/8 Fabric RP: 10.20.0.1 Group Range 238.0.0.0/8



PIM ASM Rendezvous Point (RP)

RP Verification: CLI Commands

RPs with different group ranges

Static RP: 10.10.0.1 Group Range 239.0.00/8 Fabric RP: 10.20.0.1 Group Range 238.0.0.0/8



L3 Multicast Rendezvous Point (RP)

Multiple RPs for same group range?

What happens if multiple RPs are configured for the same range?

- For static and fabric RP: RP with the highest IP is used
- Auto-RP or BSR RP is always preferred over static or fabric RP. RP override not supported.





PIM-ASM Rendezvous Point (RP) Multiple RPs for same group range CLI verification



PIM-ASM Rendezvous Point (RP)

Multiple RPs for same group range CLI verification





PIM ASM: Source starts sending multicast



VRF deployed on all leafs

#CiscoLive

PIM ASM: Source starts sending multicast CLI verification



PIM ASM: PIM register CLI verification

Leaf-101# show ip pim internal event-history null-register 2022-04-28T17:02:38.662590000-07:00 pim [24171]: TID 24206:pim_send_null_register:1295:(MC3:vrf-base) Send Null Register to RP 10.20.0.1 for (10.1.1.10/32, 239.1.1.1/32)

PIM debug logs stored on leaf. Use event-history to check PIM null-register events. FHR sends periodic null registers for active sources.



cisco /

VRF deployed on all leafs

PIM ASM: Receiver joins multicast group

- FHR for ASM will only have (S,G) entry
- IIF and OIF will show tunnel interface
- (*,G) on leaf where receiver is connected •
- (*,G) on stripe winner BL



cisco ila

VRF deployed on all leafs

#CiscoLive

PIM ASM: Receiver joins multicast group, CLI verification



cisco live!

r uepioyeu on all lears

#CiscoLive

PIM SSM: Source starts sending multicast



cisco /

VRF deployed on all leafs

#Ciscol ive

PIM SSM: Source starts sending multicast, CLI verification



cisco

VRF deployed on all leafs

PIM SSM: Receiver joins source and group

- Receiver sends IGMPv3 join for group and source
- (S,G) state created on leaf connected to receiver



cisco / ille

VRF deployed on all leafs

#CiscoLive BRKDCN-2623

PIM SSM: Receiver joins source and group, CLI verification



#CiscoLive BRKDCN-2623

Multicast with Multi-Pod Fabrics





ACI Multi-Pod

Review



- Multiple ACI Pods connected by an IP Inter-Pod L3 Network
- Managed by a single APIC cluster
- Forwarding control plane fault isolation (IS-IS, COOP)
- VXLAN data plane between pods

Multicast over Multi-Pod Fabrics

- An ACI Multi-Pod Fabric operates as a single ACI Fabric
- All policies (VRFs, Bridge Domains, EPGs, etc) are available across all pods
- This applies to both bridge domain multicast (L2) and TRM
- Bidirectional PIM (BiDir) is used in the Inter-Pod (IPN) network to extend the multicast underlay across pods

ACI Multi-Pod Multicast Data Plane: Inter-Pod Network



- PIM BiDir in the IPN is used to extend multicast distribution trees across pods
- Spines send IGMP joins on IPN links for BD and VRF GIPos (Spines do not run PIM)
- One spine link per pod is selected as the designated forwarder (DF) for each BD and VRF GIPo (distributed across available spine links)

ACI Multi-Pod GIPo DF CLI verification



- Spines send IGMP joins on IPN links for BD and VRF GIPos (Spines do not run PIM)
- One spine link per pod is selected as the designated forwarder (DF) for each BD and VRF GIPo (distributed across available spine links)

ACI Multi-Pod

IPN GIPo Verification



- PIM BiDir in the IPN is used to extend multicast distribution trees across pods
- Spines send IGMP joins on IPN links for BD and VRF GIPos (Spines do not run PIM)
- One spine link per pod is selected as the designated forwarder (DF) for each BD and VRF GIPo (distributed across available spine links)

ACI Multi-Pod Multicast Control Plane: Pod-to-pod



cisco live

ACI Multi-Pod

Multicast Control Plane



cisco live!

ACI Multi-Pod

Multicast Control Plane: Stripe winner CLI verification



cisco
ACI Multi-Pod

IPN Multicast Configuration

Bidir PIM configuration on all IPN devices

vrf context IPN-1
ip pim rp-address 192.168.100.1 group-list 225.0.0.0/15 bidir
ip pim rp-address 192.168.100.1 group-list 239.255.255.240/28 bidir



- Phantom RP configuration is applied to multiple IPN devices (provides RP redundancy)
- Each phantom RP is configured with a different mask
- RP with the longest mask will be the root of the shared tree
- Loopback is configured with an IP in the same subnet as the RP but not the RP address

Multicast across Multi-Site Fabrics



Tenant Route Multicast with Multi-Site

• Support for PIM-ASM and PIM-SSM Tenant routed multicast

For PIM-ASM external RP or RP in the fabric are possible options (Fabric RP with Multi-Site from ACI 5.0(1)) When using external RP, all sites must have reachability to the RP via a local L3Out

- Each BL node runs PIM in active mode and forms neighborship with other BLs in the same site and with the external router(s)
- BLs do not form neighborship with BLs in other sites
- Supports sources attached to the fabric and external sources (reachable via local L3Out) BDs with L3 Multicast sources or receivers may or may not be stretched across the sites External sources must be reachable independently from each site via a local L3Out
- Supports receivers attached to the fabric and external receivers (reachable via local or remote L3Out)

BDs with L3 Multicast receivers or receivers may or may not be stretched across the site

Review



- Multiple ACI sites connected by an L3 Inter-Site (ISN) network
- Each site managed by independent APIC clusters
- Orchestrated by Nexus Dashboard Orchestrator (NDO)
- MP-BGP EVPN control plane between sites
- VXLAN data plane between sites

cisco / ile

Multicast Data Plane



- ACI Multi-Site does not require multicast in the ISN network
- BD and VRF multicast traffic is forwarded over head-end replication (HREP) tunnels.
- Each ACI site will be assigned an Overlay Multicast TEP (O-TEP) address. This is an anycast address programmed on all spines in the site (across all pods in case of Multi-Pod sites)

Multicast Data Plane CLI verification





Multicast Control Plane



- Multicast group interest is not shared across sites (no EVPN Type-6)
- Each site must have its own L3Outs for external sources

cisco / ile !

TRM: Source inside fabric



cisco live!



- Receivers in the fabric must use local L3Outs to receive multicast from external sources
- Multicast traffic from external sources will not be sent over the ISN (blocked by ACL)

ACI Multi-Site Fabric RP across Multi-Site



- Fabric RP support across Multi-Site fabrics supported from ACI 5.2(1)
- Fabric RP BLs send registers to All PIM routers multicast group 224.0.0.13

```
cisco live!
```

TRM with Remote Leaf

cisco live!

TRM with Remote Leaf

- Support for BUM traffic to/from remote leaf switches has been available since the initial release of remote leaf switches.
- Forwarding of BD GIPo traffic to/from remote leaf switches uses HREP tunnels similar to Multi-Site
- TRM support for remote leaf switches was added in the ACI 5.1(3) release.
- No configuration changes are required to support TRM. Change in 5.1(3) was to add support for VRF GIPo HREP tunnels
- All current layer-3 multicast features are supported with remote leaf

ACI Remote Leaf TEP address configuration



- VXLAN tunnels between RLs and RLs and the main pod use externally routable addresses (different from pod TEP range)
- Spine TEPs (anycast TEPs) are assigned from the External TEP range
- In Multi-Pod fabric, each pod will be assigned a unique anycast TEP
- RLs are assigned TEPs from the Remote Pool. Each RL will have a unique TEP.



ACI Remote Leaf Head-End Replication Data Plane: Local Leaf (LL) to Remote Leaf (RL)



- The main pod will send TRM multicast over a head-end replication (HER) tunnel to each RL
- This is the same forwarding behavior for BD GIPo tunnels (BUM)
- RLs will forward multicast locally based on IGMP membership (local receivers) or PIM joins (L3Outs)



ACI Remote Leaf Head-End Replication

Data Plane: Local Leaf (LL) to Remote Leaf (RL) CLI verification



- This is the same forwarding behavior for BD GIPo tunnels (BUM)
- RLs will forward multicast locally based on IGMP membership (local receivers) or PIM joins (L3Outs)

ACI Remote Leaf Head-End Replication Data Plane: RL to LL and RL to RL



 If there is more than one pod, the RL will forward directly to each pods anycast TEP (HREP tunnels go directly from RL to all pods)



Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!
- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.
- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.



Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

E Learn

Cisco U.

IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning

Subscription-based product, technology, and certification training

Cisco Modeling Labs

Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network

Resource community portal for certifications and learning

En Train

Cisco Training Bootcamps Intensive team & individual automation and technology training programs

Cisco Learning Partner Program

Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training

Accelerated curriculum of product, technology, and certification courses

E Certify

Cisco Certifications and Specialist Certifications

Award-winning certification program empowers students and IT Professionals to advance their technical careers

Cisco Guided Study Groups

180-day certification prep program with learning and support

Cisco Continuing Education Program

Recertification training options for Cisco certified individuals

Here at the event? Visit us at The Learning and Certifications lounge at the World of Solutions



Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at <u>www.CiscoLive.com/on-demand</u>



CISCO The bridge to possible

Thank you



#CiscoLive





#CiscoLive