ACI L3 Multicast

Roland Ducomble – CCIE 3745
EMEAR ACI solution Tac Team – Technical Leader
28th Feb 2018

Intro

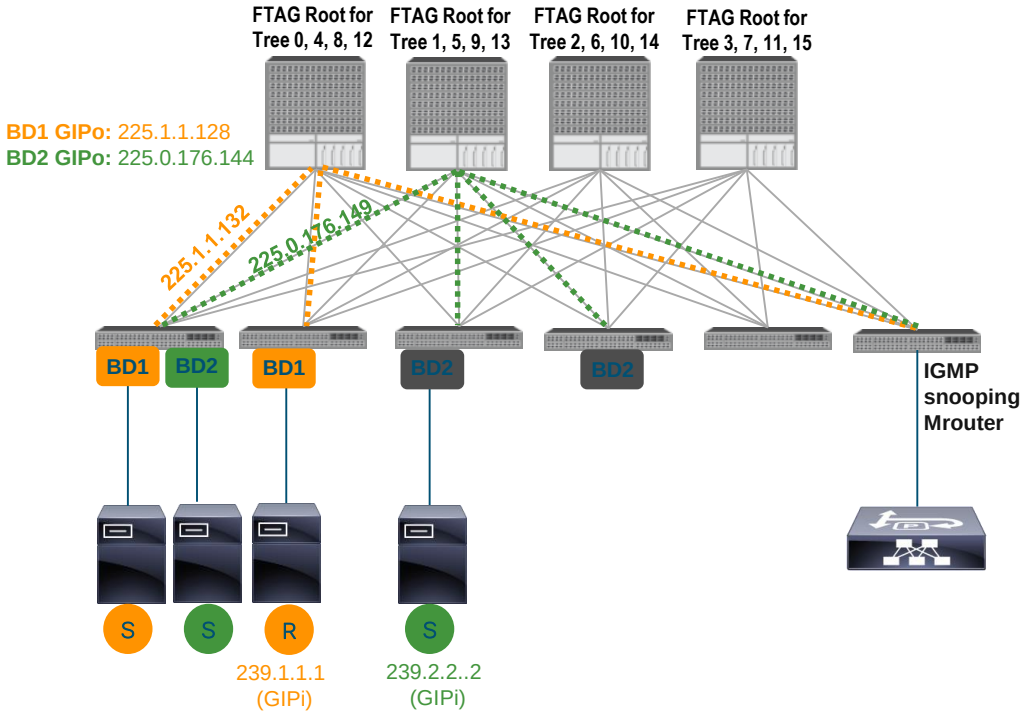
Multicast Routing Support (as of 3.2)

- Multicast routing is supported in ACI release 2.0 (Congo)
- Requires EX switches (i.e N9K-C93180YC-EX)
- ACI release 2.0 supports the following multicast routing features
  - PIM ASM
  - PIM SSM
  - PIM Auto-RP
  - PIM BSR
- The following flows are supported for both ASM and SSM
  - Receiver in fabric with external source
  - Source in fabric with external receiver
  - Source and receiver in fabric
  - External source and external receiver (transit case)
- FEX support (3.1)

Multicast routing limitations (as of 3.2)

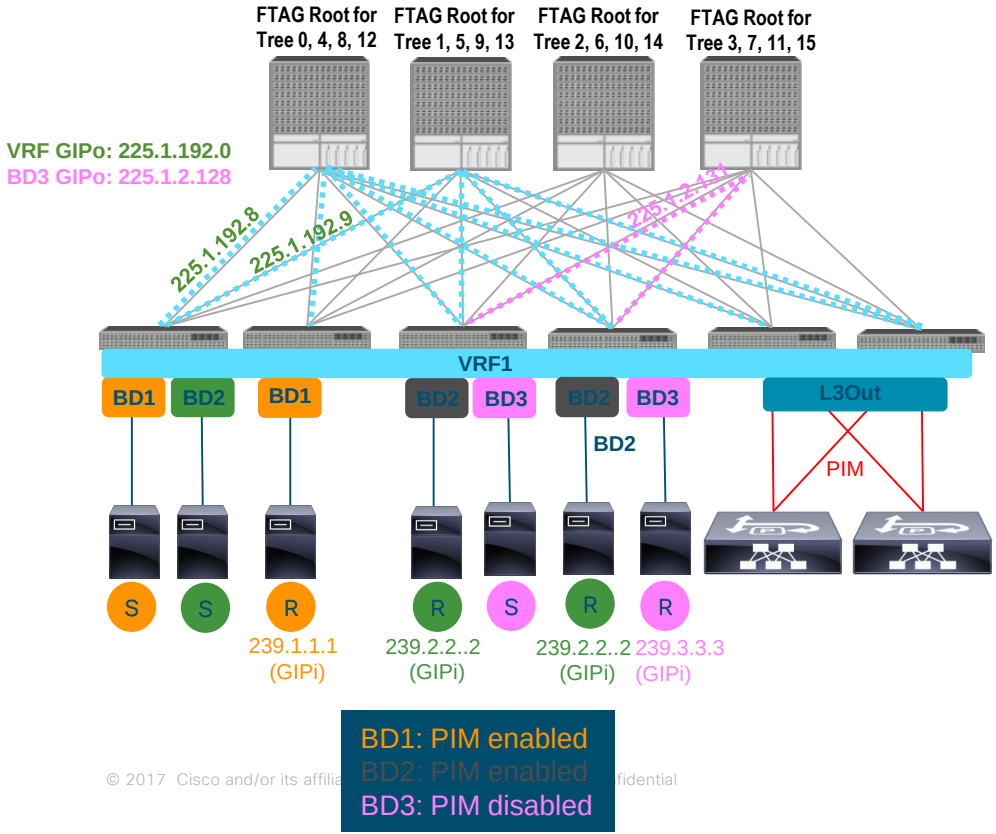
- Does not support BiDir (roadmapped)
- Does not support fabric as RP (Rendezvous Point) – Added in 4.0
- Not supported on 1st Gen leaf switches
- PIM is not supported on L3Out SVIs
- IGMP snooping must be enabled for PIM enabled BDs
- Contracts are not supported for multicast traffic
- No IPv6 multicast routing support
- Multicast routing is not supported over GOLF L3outs
- Not supported across VRFs – Added in 4.0

L2 Multicast Forwarding Review



- Multicast traffic is load balanced across FTAG trees (rooted at the spines).
- There are 16 FTAG trees (12 for user traffic)
- Multicast traffic is scoped to bridge domains
- Traffic is forwarded to all leaf nodes where the bridge domain is deployed
- Traffic is flooded for the BD via a multicast VXLAN group IP address outer address (GIPo)
- The GIPo address is a multicast address with the last four bits set to 0000
- When traffic is forwarded in the fabric the last four bits will be used to select the FTAG tree (selection made by leaf using hash function)
- IGMP snooping runs on leaf switches

L3 Multicast Forwarding Overview

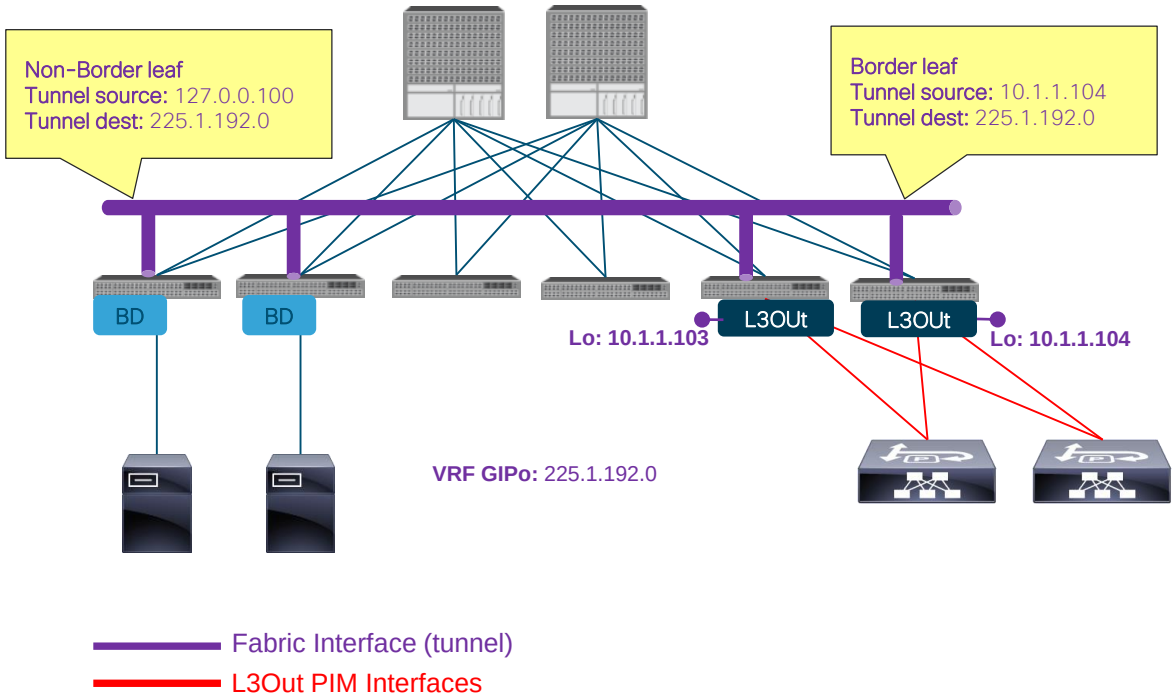


- L3 Multicast is enabled at the **VRF**, **BD**, and **L3Out** level
- Must be enabled at the VRF if there are any multicast enabled BDs or L3Outs in the VRF
- A multicast enabled VRF can have a mix of multicast enabled BDs and multicast disabled BDs
- A single GIPo address will be assigned to the **VRF** when it is enabled for multicast
- **Multicast traffic for all multicast enabled BDs will be forwarded using the VRF GIPo**
- Multicast traffic in multicast disabled BDs will be forwarded in BD GIPo
- A GIPo may be shared (reused) across multiple BDs or VRFs but not a combination of BDs and VRFs

Multicast GIPo Usage

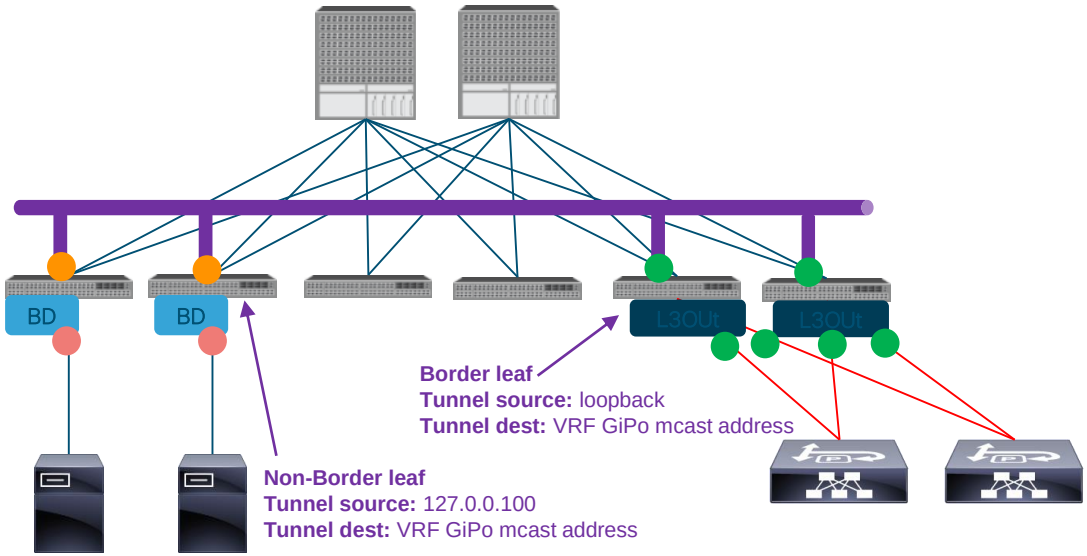
	Non-Multicast Routing Enabled BD	Multicast Routing Enabled BD
Broadcast	BD GIPo	BD GIPo
Unknown Unicast Flood	BD GIPo	BD GIPo
Multicast	BD GIPo	VRF GIPo

Multicast routing – border leaves and non-border leaves



- When the VRF is enabled for multicast routing a fabric interface (tunnel) is created for multicast routing within the fabric.
- The tunnel destination will be the VRF GiPo multicast address
- On border leaf switches the tunnel source will be the loopback interface on the border leaf
- On non-border leaf switches the tunnel source will be a loopback address (127.0.0.100)
- Border leaves send PIM hellos on the fabric interface
- Non-border leaves run in passive mode on the fabric interface, they listen for PIM hellos from the border leaves but do not send PIM hellos. Non-border leaves will not show up in output to “show ip pim neighbor”
- L3 out interfaces run PIM in normal mode (sends and receives hellos, elects DR)

Multicast Routing PIM Interfaces



Fabric Interface (tunnel)

L3Out PIM Interfaces

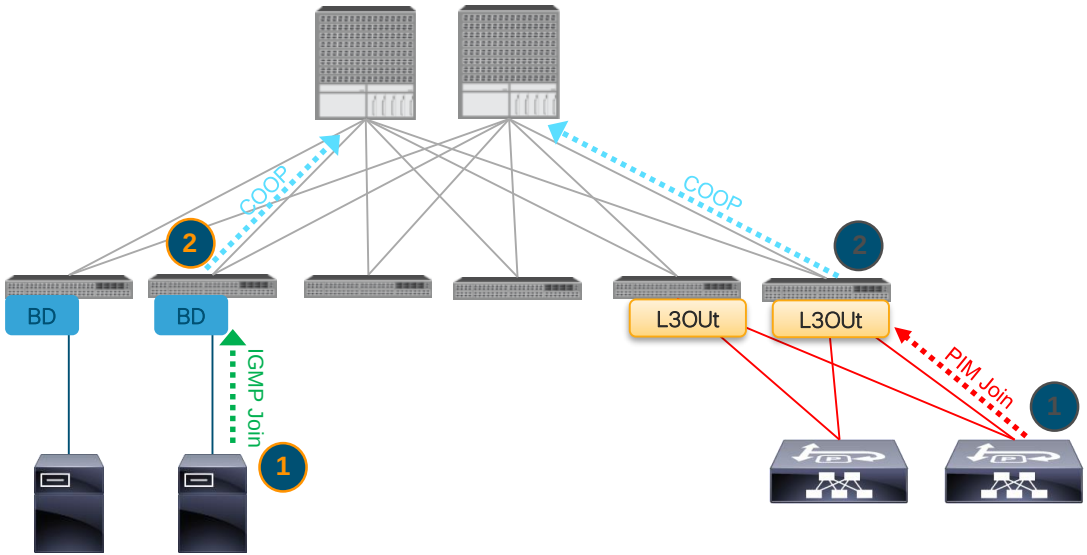
- Multicast routing is enabled in three places in ACI
  - **VRF**
  - **L3Out**
  - **BD**
- When multicast is enabled PIM will run on the following interfaces
  - L3Out interfaces
  - Border leaf tunnel interfaces
  - Non-border leaf tunnel interfaces
  - SVI interfaces
- PIM runs in different modes depending on the interface

Normal PIM interfaces. Sends and receives PIM hellos, follows standard DR election process

PIM Passive mode. Receives PIM hellos from the border leaves. Does not send PIM hellos

PIM Passive Probe Mode. Sends PIM hellos but does not form PIM neighbors

Multicast group interest advertised in COOP



- Multicast group information (*,G), (S,G) is advertised within the fabric using COOP
- IGMP and PIM joins triggers the leaf to publish multicast group interest in COOP (PIM joins/prunes are not used)*

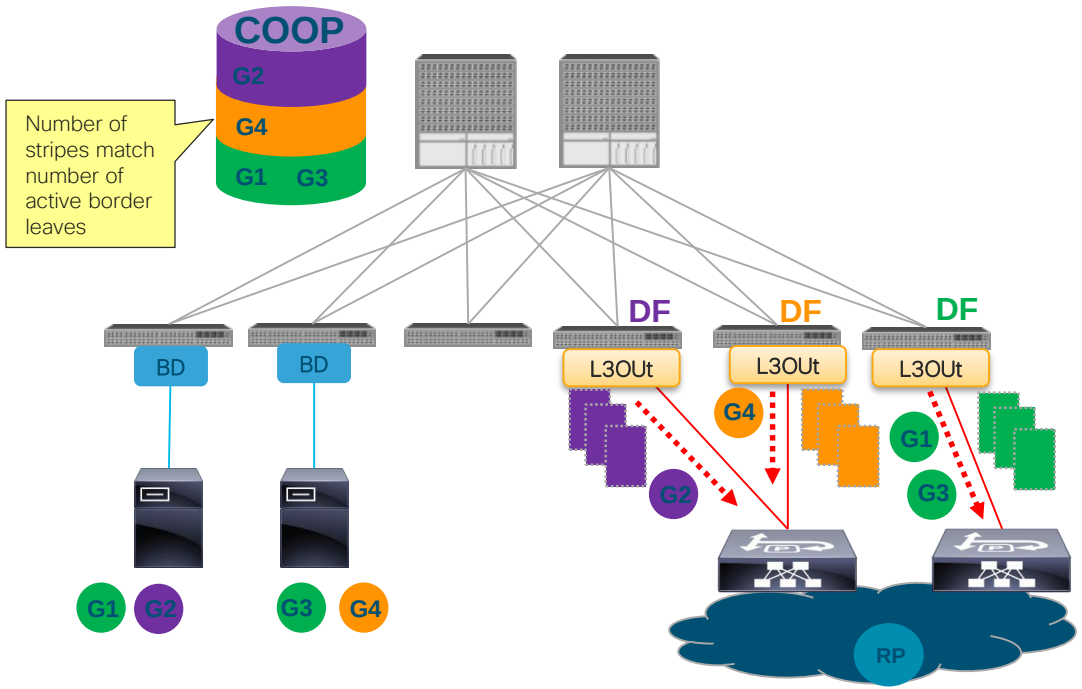
IGMP Join received on pervasive SVI

1. IGMP Join report received
2. Leaf publishes the join information to COOP

PIM Join received on border leaf

1. PIM Join received on border leaf
2. Border leaf publishes the join information to COOP

Multiple Border Leaves and Designated Forwarder



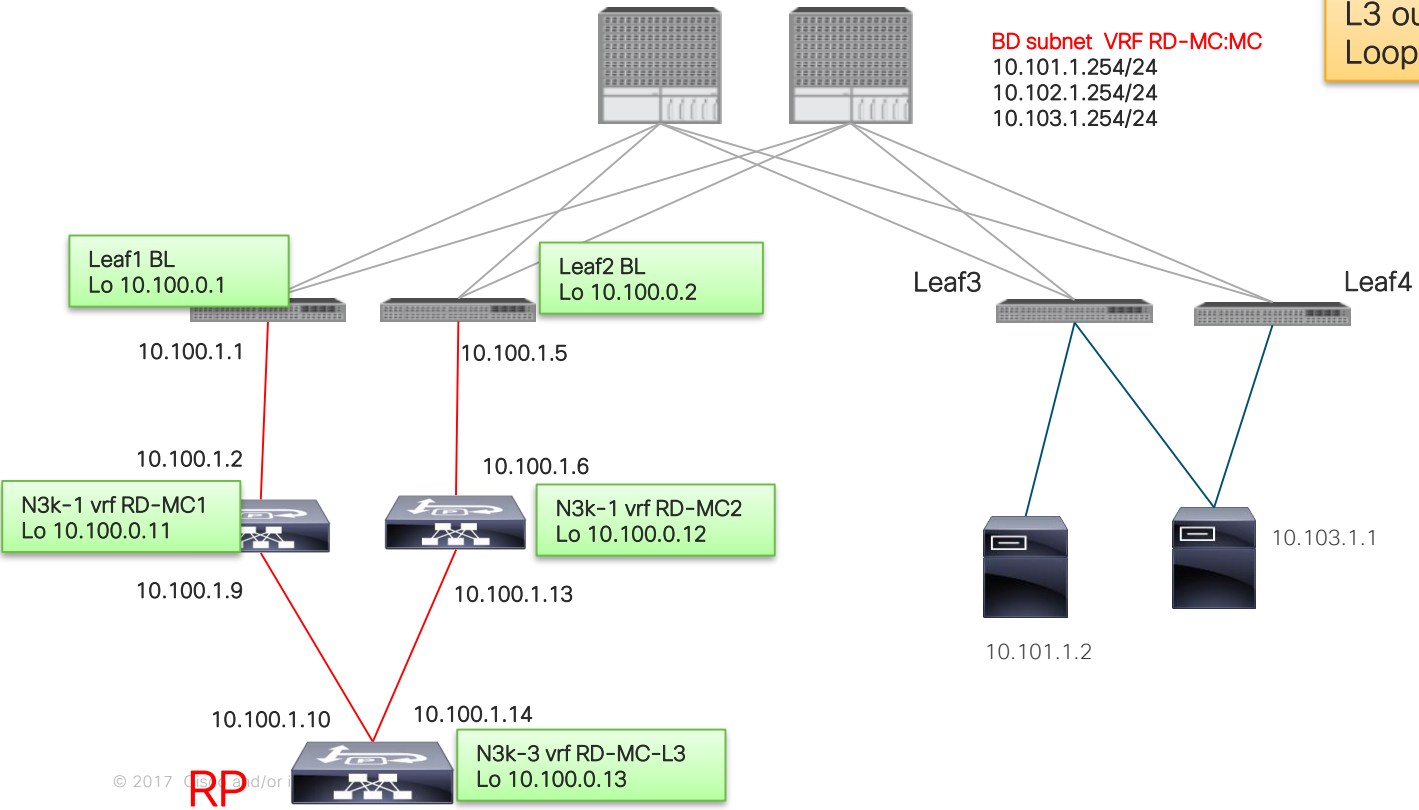
- Multicast routing can be enabled on multiple border leaves
- Only one border leaf will be the designated forwarder for a multicast group to avoid sending duplicate traffic into the fabric
- Load balancing is done by striping group ownership for multicast groups in the COOP database across all active border leaves
- The border leaf that is the stripe winner for a group is responsible for sending PIM joins on behalf of the fabric and forwarding multicast traffic into the fabric
- When a border leaf is removed the groups are restriped across the remaining active border leaves

Basic Config

Multicast Routing Configuration

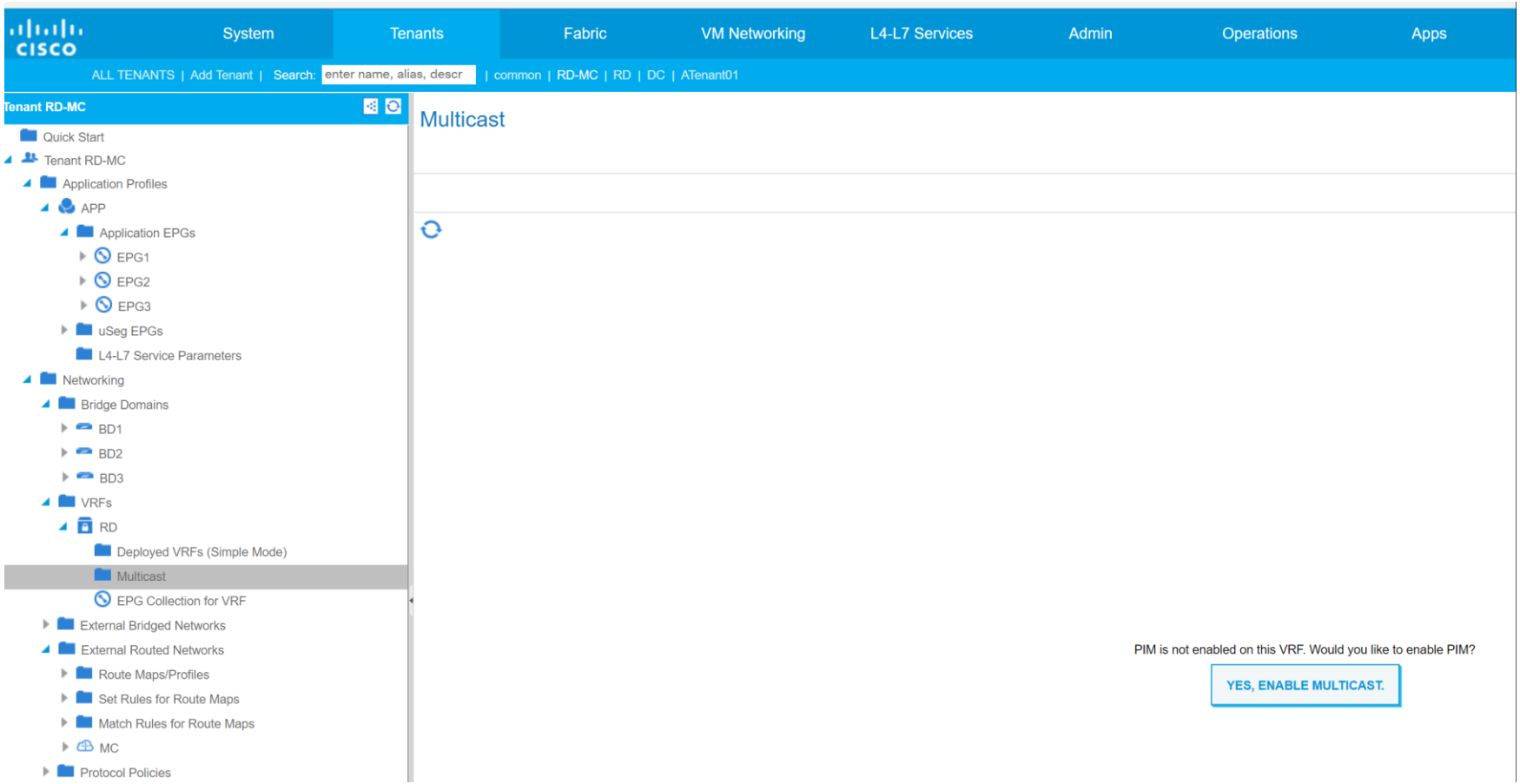
- Multicast routing is enabled at the VRF, BD, and L3out levels.
- Multicast routing must be enabled at the VRF level if there are any multicast enabled BDs or L3outs
- There can be a mix of multicast enabled BDs and multicast disabled BDs in the same VRF
- The VRF configuration in the APIC GUI will have a Multicast folder where multicast routing can be configured
- BDs, L3outs, and multicast IGMP and PIM policies can be configured from the VRF multicast view
- L3Outs on border leaf must be configured with loopback addresses enabled in the Node profile. A loopback is required for multicast routing on the border leaf.

Lab topology



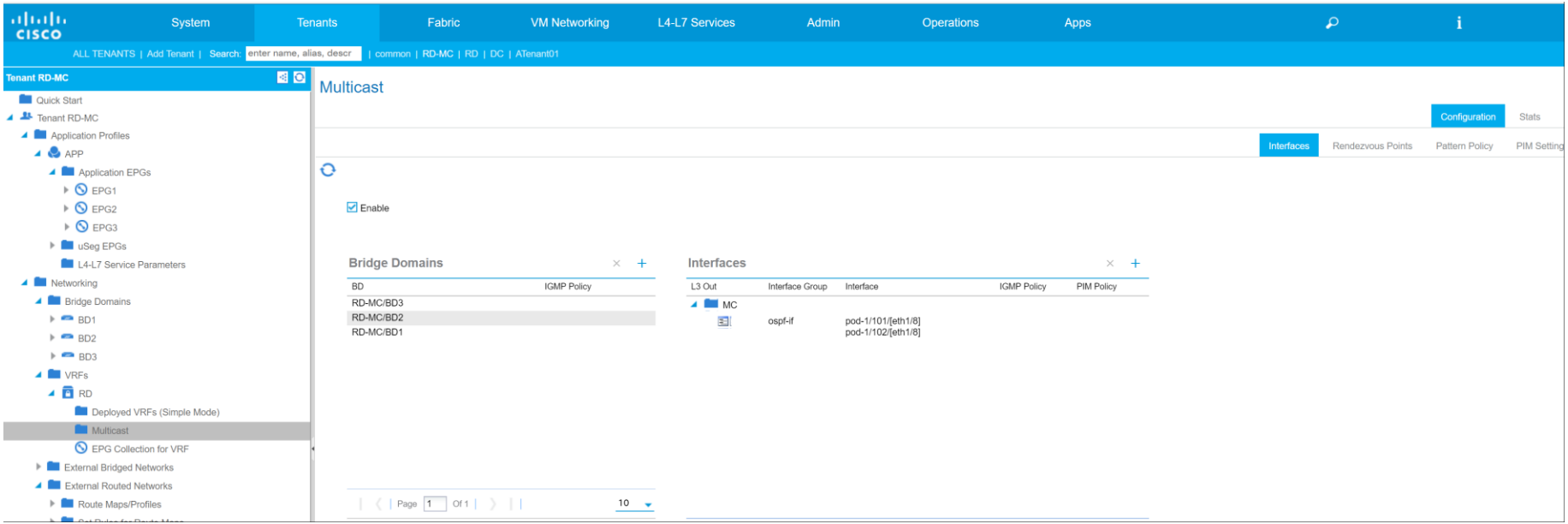
Note:
RP is outside of Fabric
L3 out on subinterface
Loopback mandatory on BL

1. Enable PIM multicast at vrf level



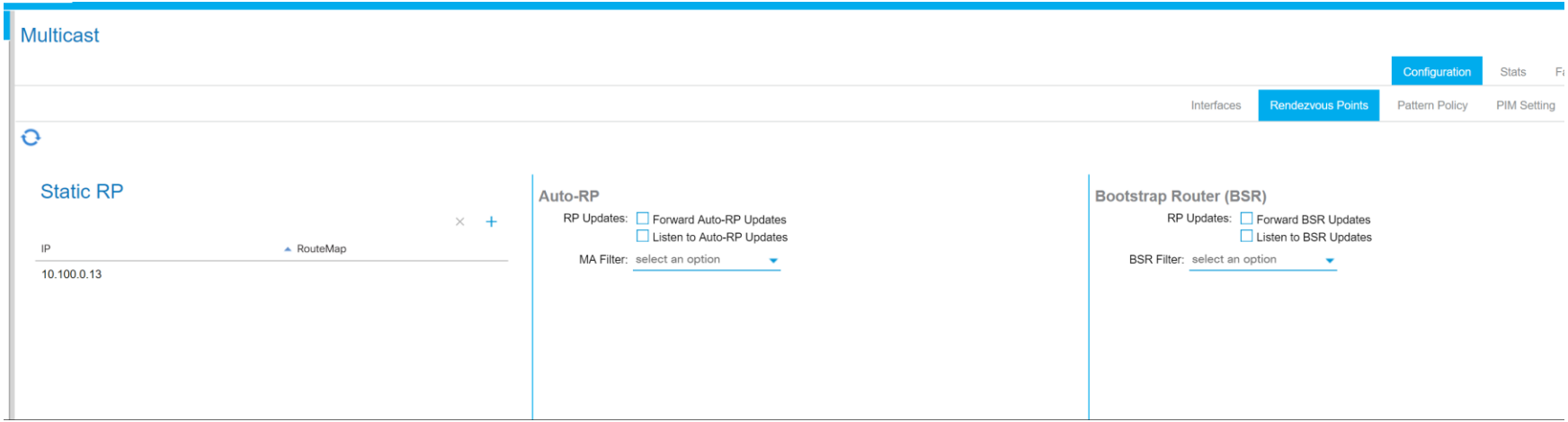
2. Add BD and L3 out where you need PIM

Note : you can as well enable PIM at BD or L3 out level !

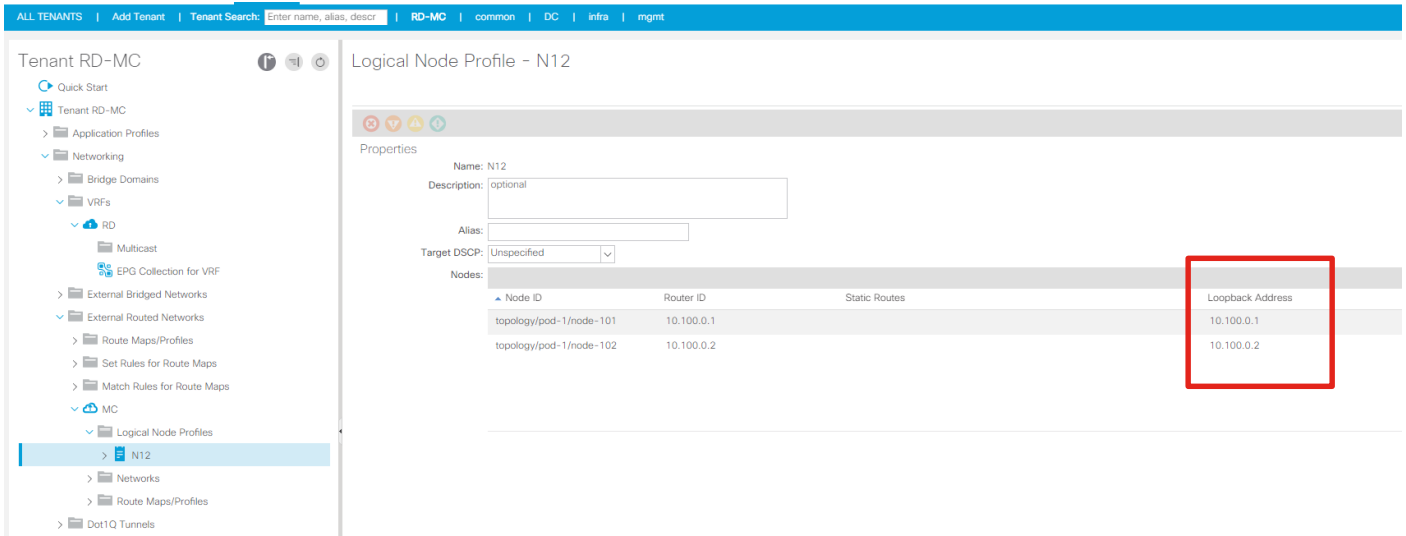


3. Specify a RP

Note ACI fabric can't be RP so you need an external RP



Mandatory Loopback as Router ID on layer 3 multicast enabled Node



Border leaf – PIM if and neighb

BL have PIM neighbor :
- Tunnel to all other BL only if and L3 out If
- No neighbor to Non-BL

BL runs PIM on :
- Rid Lo, L3 out if , Tunnel If, vlan for BD running mcast on that BL
- Tunnel Src is lookpback on BL

```
admin@apic1:pimctxp> moquery -c pim.CtxP -f 'pim.CtxP.vrfGipo
=="225.1.192.48"'
Total Objects shown: 1
# pim.CtxP
..
dn            : uni/tn-RD-MC/ctx-RD/pimctxp
lcOwn         : local
modTs         : 2018-02-19T08:31:13.216+00:00
monPolDn      : uni/tn-common/monepg-default
mtu           : 1500
..
status        :
uid           : 15374
vrfGipo       : 225.1.192.48/32
```

```
bdsol-aci32-leaf1# show ip pim interface brief vrf  RD-MC:RD
PIM Interface Status for VRF "RD-MC:RD"
Interface            IP Address      PIM DR Address  Neighbor  Border
                                                     Count     Interface
Vlan35               10.101.1.254    10.101.1.254    0         no
loopback18           10.100.0.0      10.100.0.0      0         no
Ethernet1/8.30       10.100.1.1      10.100.1.2      1         no
Vlan38               10.102.1.254    10.102.1.254    0         no
Vlan44               10.103.1.254    10.103.1.254    0         no
Tunnel5              10.100.0.0      10.100.0.2      1         no
bdsol-aci32-leaf1# show ip pim neighbor vrf  RD-MC:RD
PIM Neighbor Status for VRF "RD-MC:RD"
Neighbor        Interface       Uptime    Expires   DR       Bidir-  BFD
                                                    Priority Capable State
10.100.1.2      Ethernet1/8.30  00:21:00  00:01:40 1        no    n/a
10.100.0.2      Tunnel5         00:23:34  00:01:26 1        no    n/a

bdsol-aci32-leaf1# show interface tunnel 5
Tunnel5 is up
    MTU 9000 bytes, BW 9 Kbit
    Transport protocol is in VRF "RD-MC:RD"
    Tunnel protocol/transport ivxlan
    Tunnel source 10.100.0.1, destination 225.1.192.48
    Tx
    0 packets output, 1 minute output rate 0 packets/sec
    Rx
    0 packets input, 1 minute input rate 0 packets/sec
```

```
bdsol-aci32-leaf2#  show ip pim interface brief vrf  RD-MC:RD
PIM Interface Status for VRF "RD-MC:RD"
Interface            IP Address      PIM DR Address  Neighbor  Border
                                                     Count     Interfac
Tunnel10             10.100.0.2      10.100.0.2      1         no
loopback7            10.100.0.2      10.100.0.2      0         no
Ethernet1/8.48       10.100.1.5      10.100.1.6      1         no
bdsol-aci32-leaf2#  show ip pim nei  vrf  RD-MC:RD
PIM Neighbor Status for VRF "RD-MC:RD"
Neighbor        Interface       Uptime    Expires   DR       Bidir-  BFD
                                                    Priority Capable Stat
10.100.0.1      Tunnel10        00:25:02  00:01:28  1        no      n/a
10.100.1.6      Ethernet1/8.48  00:20:56  00:01:40  1        no      n/a

bdsol-aci32-leaf2# show interface tunnel 10
Tunnel10 is up
    MTU 9000 bytes, BW 9 Kbit
    Transport protocol is in VRF "RD-MC:RD"
    Tunnel protocol/transport ivxlan
    Tunnel source 10.100.0.2, destination 225.1.192.48
    Tx
    0 packets output, 1 minute output rate 0 packets/sec
    Rx
    0 packets input, 1 minute input rate 0 packets/sec
```

Non BL PIM if and neighbor

PIM runs only on tunnel and
Vlan for mcast BD existing on that leaf
Pim neighbor only through tunnel
Note Tunnel source is a 127.0.0.100

Poim neighbor only on Tunnel interface to BL
No neighbor to other non BL.

```
bdsol-aci32-leaf3# show ip pim interface brief vrf RD-MC:RD
PIM Interface Status for VRF "RD-MC:RD"
Interface            IP Address      PIM DR Address  Neighbor  Border
                                                     Count     Interface
Vlan74               10.101.1.254    10.101.1.254    0         no
Vlan78               10.103.1.254    10.103.1.254    0         no
Vlan76               10.102.1.254    10.102.1.254    0         no
Tunnel61             127.0.0.100     127.0.0.100     2         no
bdsol-aci32-leaf3# show ip pim nei  vrf RD-MC:RD
PIM Neighbor Status for VRF "RD-MC:RD"
Neighbor        Interface            Uptime    Expires   DR       Bidir-  BFD
                                                         Priority Capable State
10.100.0.1      Tunnel61             00:02:23  00:01:17  1        no      n/a
10.100.0.2      Tunnel61             00:02:23  00:01:43  1        no      n/a
bdsol-aci32-leaf3# show interface t
transceiver   trunk         tunnel
bdsol-aci32-leaf3# show interface tunnel 61
Tunnel61 is up
    MTU 9000 bytes, BW 9 Kbit
    Transport protocol is in VRF "RD-MC:RD"
    Tunnel protocol/transport ivxlan
    Tunnel source 127.0.0.100, destination 225.1.192.48
    Tx
    0 packets output, 1 minute output rate 0 packets/sec
    Rx
    0 packets input, 1 minute input rate 0 packets/sec
```

RP

- Both BL and non-BL should know the RP

BL

```
bdsol-aci32-leaf1# show ip pim rp vrf RD-MC:RD
PIM RP Status Information for VRF "RD-MC:RD"
BSR disabled
Auto-RP disabled
BSR RP Candidate policy: None
BSR RP policy: None
Auto-RP Announce policy: None
Auto-RP Discovery policy: None

RP: 10.100.0.13, (0), uptime: 00:12:42, expires: never,
  priority: 0, RP-source: (local), group-map: mcast_rprange_RD-MC:RD_10.100.0.13, group ranges:
      224.0.0.0/4
```

Non-BL

```
bdsol-aci32-leaf3# show ip pim rp vrf RD-MC:RD
PIM RP Status Information for VRF "RD-MC:RD"
BSR disabled
Auto-RP disabled
BSR RP Candidate policy: None
BSR RP policy: None
Auto-RP Announce policy: None
Auto-RP Discovery policy: None

RP: 10.100.0.13, (0), uptime: 00:13:02, expires: never,
  priority: 0, RP-source: (local), group-map: mcast_rprange_RD-MC:RD_10.100.0.13, group ranges:
```

PIM VRF process info

BL

```
bdsol-aci32-leaf1#  show ip pim vrf RD-MC:RD detail
PIM Enabled VRFs
VRF Name              VRF      Table       Interface  BFD
MVPN
                      ID       ID          Count      Enabled
Enabled
RD-MC:RD              21       0x00000014  6          no
no
  State Limit: None
  Register Rate Limit: none
  Register source address : none
  Shared tree ranges: none
  (S,G)-expiry timer: not configured
    (S,G)-list policy: none
    (S,G)-expiry timer config version 0, active version 0

  Pre-build SPT for all (S,G)s in VRF: disabled
  CLI vrf done: TRUE
  PIM cibtype Auto Enabled: yes
  txlist work pending: FALSE
  PIM VxLAN VNI ID: 0
  iVxlan VRF VNID: 3112960
  Fabric IOD: 0x7a
  Fast Convergence: NO
  Num Interfaces: 6
```

Non-BL

```
bdsol-aci32-leaf3# show ip pim vrf RD-MC:RD detail
PIM Enabled VRFs
VRF Name              VRF      Table       Interface  BFD
MVPN
                      ID       ID          Count      Enabled
Enabled
RD-MC:RD              25       0x00000012  4          no          no
  State Limit: None
  Register Rate Limit: none
  Register source address : none
  Shared tree route-map: mcast_permit_all
         route-ranges:
                    224.0.0.0/4 Accept
  (S,G)-expiry timer: not configured
    (S,G)-list policy: none
    (S,G)-expiry timer config version 0, active version 0

  Pre-build SPT for all (S,G)s in VRF: disabled
  CLI vrf done: TRUE
  PIM cibtype Auto Enabled: yes
  txlist work pending: FALSE
  PIM VxLAN VNI ID: 0
  iVxlan VRF VNID: 3112960
  Fabric IOD: 0x77
  Fast Convergence: NO
  Num Interfaces: 4
```

IGMP runs on every BD and L3 out If

```
bdsol-aci32-leaf3# show ip igmp  interface vrf RD-MC:RD  | egrep -A 1 "status"
Vlan74, Interface status: protocol-up/link-up/admin-up
  Active querier: 10.101.1.254, version: 2, next query sent in: 00:00:26
--
Vlan78, Interface status: protocol-up/link-up/admin-up
  Active querier: 10.103.1.254, version: 2, next query sent in: 00:01:17
--
Vlan76, Interface status: protocol-up/link-up/admin-up
  Active querier: 10.102.1.254, version: 2, next query sent in: 00:00:50
--
Tunnel61, Interface status: protocol-up/link-up/admin-up
  Active querier: 0.0.0.0, version: ?, next query sent in: 0.000000
```

MGNVPN General check

Every leaf, BL and non BL should have fabric tunnel and the list
Of active BL in the VRF

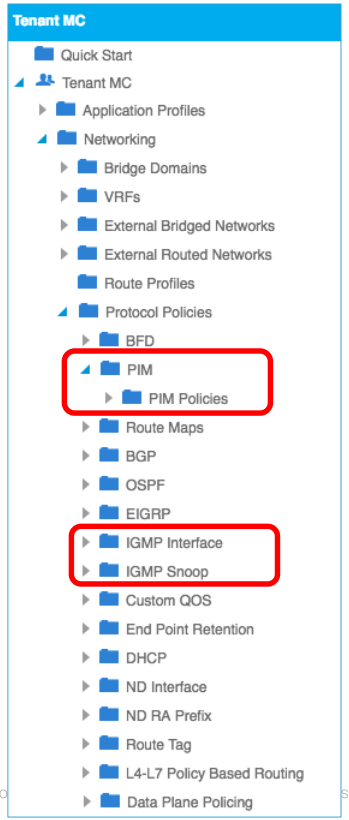
BL

```
bdsol-aci32-leaf1# show fabric multicast vrf RD-MC:RD
Fabric Multicast Enabled VRFs
VRF Name              VRF      Vprime        VN-Seg     VRF    Conv   Tunnel
                      ID       If            ID         Role   Mode   IP
RD-MC:RD              21       Tunnel5       3112960    BL     Reg    10.100.0.1
bdsol-aci32-leaf1# show fabric  multicast internal active-bl-list vrf RD-MC:RD
Active BLs for VRF: RD-MC:RD
    10.100.0.2
    10.100.0.1
```

Non-BL

```
bdsol-aci32-leaf3# show fabric  multicast vrf RD-MC:RD
Fabric Multicast Enabled VRFs
VRF Name              VRF      Vprime        VN-Seg     VRF    Conv   Tunnel
                      ID       If            ID         Role   Mode   IP
RD-MC:RD              25       Tunnel61      3112960    Leaf   Reg    127.0.0.100
bdsol-aci32-leaf3# show fabric  multicast internal active-bl-list vrf RD-MC:RD
Active BLs for VRF: RD-MC:RD
    10.100.0.2
    10.100.0.1
```

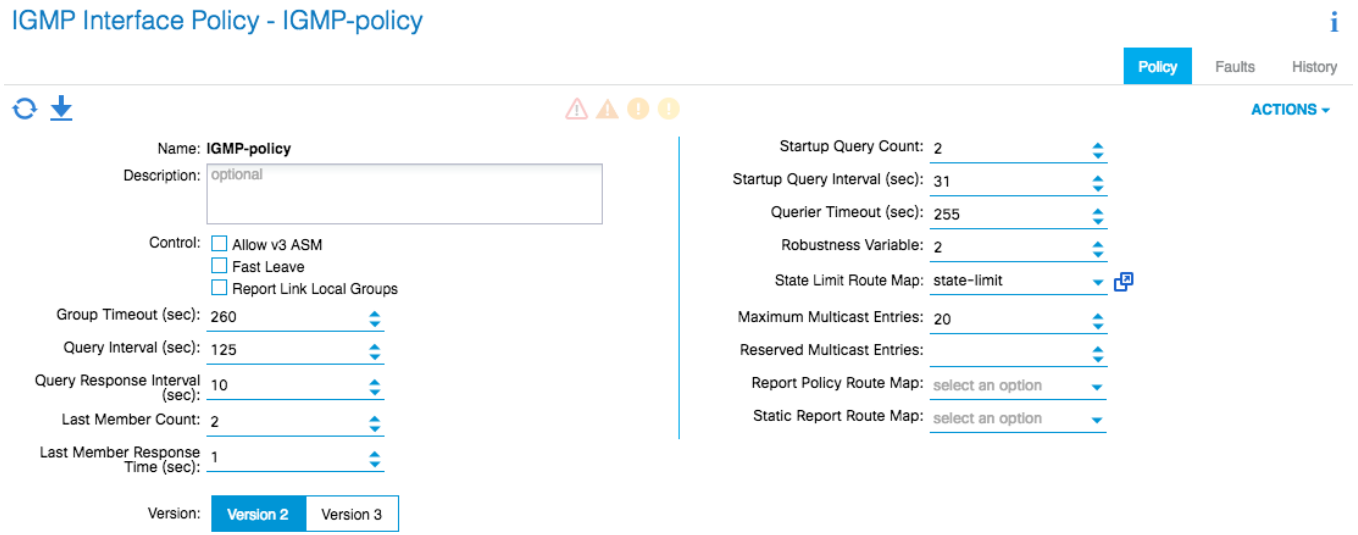
Multicast Policies



- PIM and IGMP policies are configured under tenant **Network→ Protocol Policies**
- These protocol policies can also be configured from the VRF, BD, and L3outs

IGMP Interface Policy

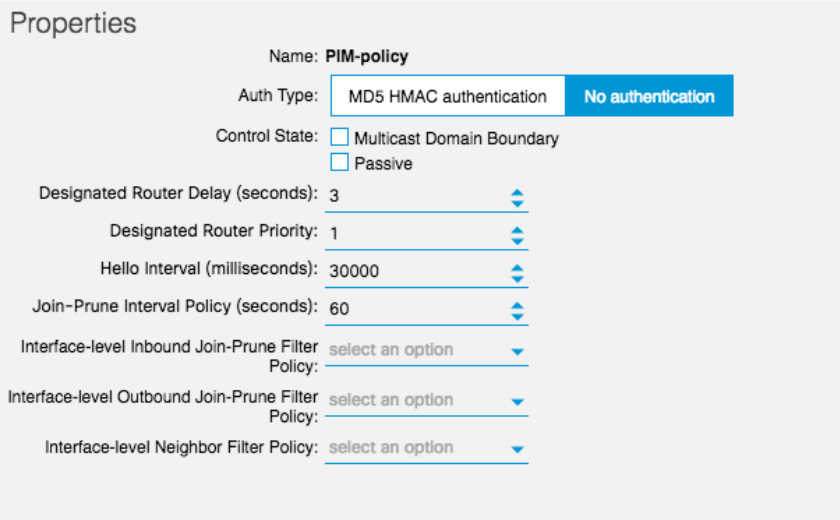
- The IGMP interface policy can be assigned to a BD or an L3out interface profiles
- If no policy is specified the interface will used the default policy.



Default IGMP interface policy settings

PIM Interface Policy

- The PIM interface policy can be assigned to L3out interfaces profiles
- If no policy is specified the interface will used the default policy.



Default PIM interface policy settings

IGMP Features

Feature	Where to configure	Where it is applied	Where to configure it
IGMP Report Policy	IGMP Interface Policy	BD, L3Out interface profile	BD, L3out interface profile or VRF
IGMP static join group	IGMP Interface Policy	BD, L3Out interface profile	BD, L3out interface profile or VRF
IGMP allow v3 ASM	IGMP Interface Policy	BD, L3Out interface profile	BD, L3out interface profile or VRF
IGMP Fast Leaves	IGMP Interface Policy	BD, L3Out interface profile	BD, L3out interface profile or VRF
IGMP state-limit	IGMP Interface Policy	BD, L3Out interface profile	BD, L3out interface profile or VRF
IGMP SSM Translate	VRF	VRF	VRF

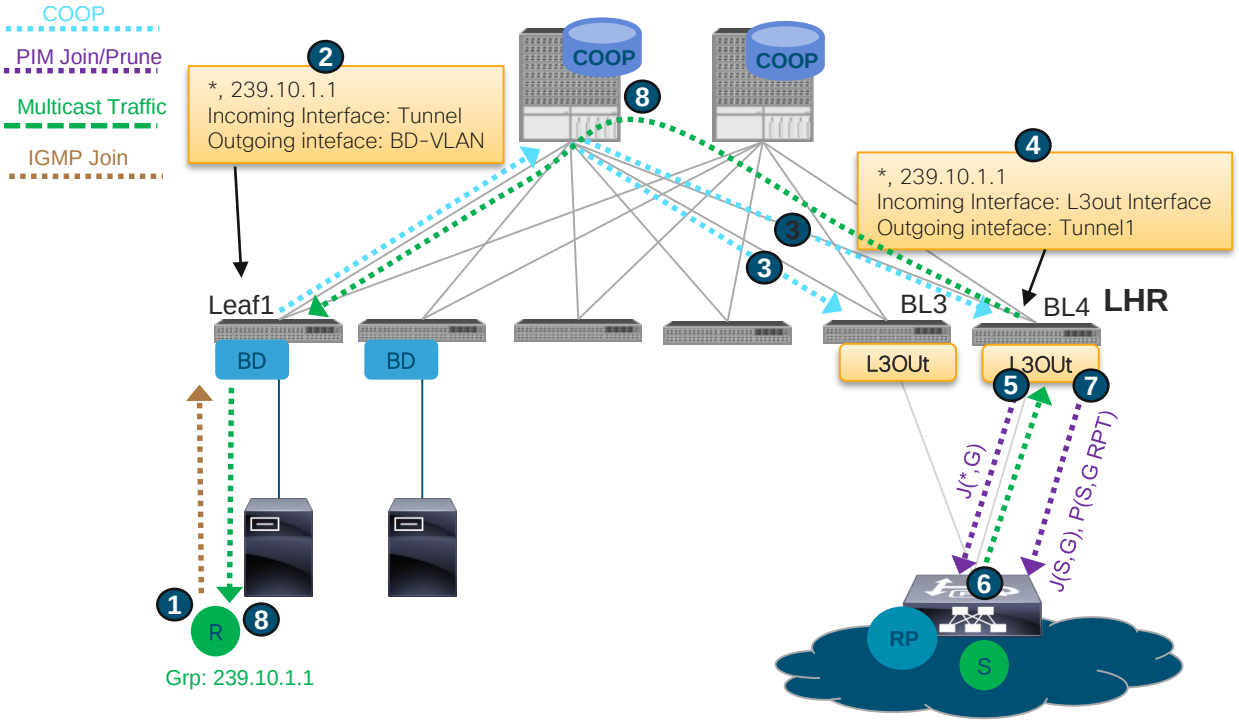
PIM Features

Feature	Where to configure	Where it is applied	Where to configure it
PIM Authentication	PIM interface profile	L3Out interface profile	L3Out interface profile or VRF
PIM timers	PIM interface profile	L3Out interface profile	L3Out interface profile or VRF
PIM Join/Prune Filter (inbound and outbound)	PIM interface profile	L3Out interface profile	L3Out interface profile or VRF
PIM Neighbor Filter	PIM interface profile	L3Out interface profile	L3Out interface profile or VRF
PIM multicast domain boundary	PIM Interface profile	L3Out interface profile	L3Out interface profile or VRF
Fast Convergence	VRF	VRF	VRF
Resource Policy	VRF	VRF	VRF
Shared tree only	VRF	VRF	VRF
expiry timer	VRF	VRF	VRF
Auto-RP	VRF	VRF	VRF
BSR	VRF	VRF	VRF

IP L3 multicast in ACI
Control plane

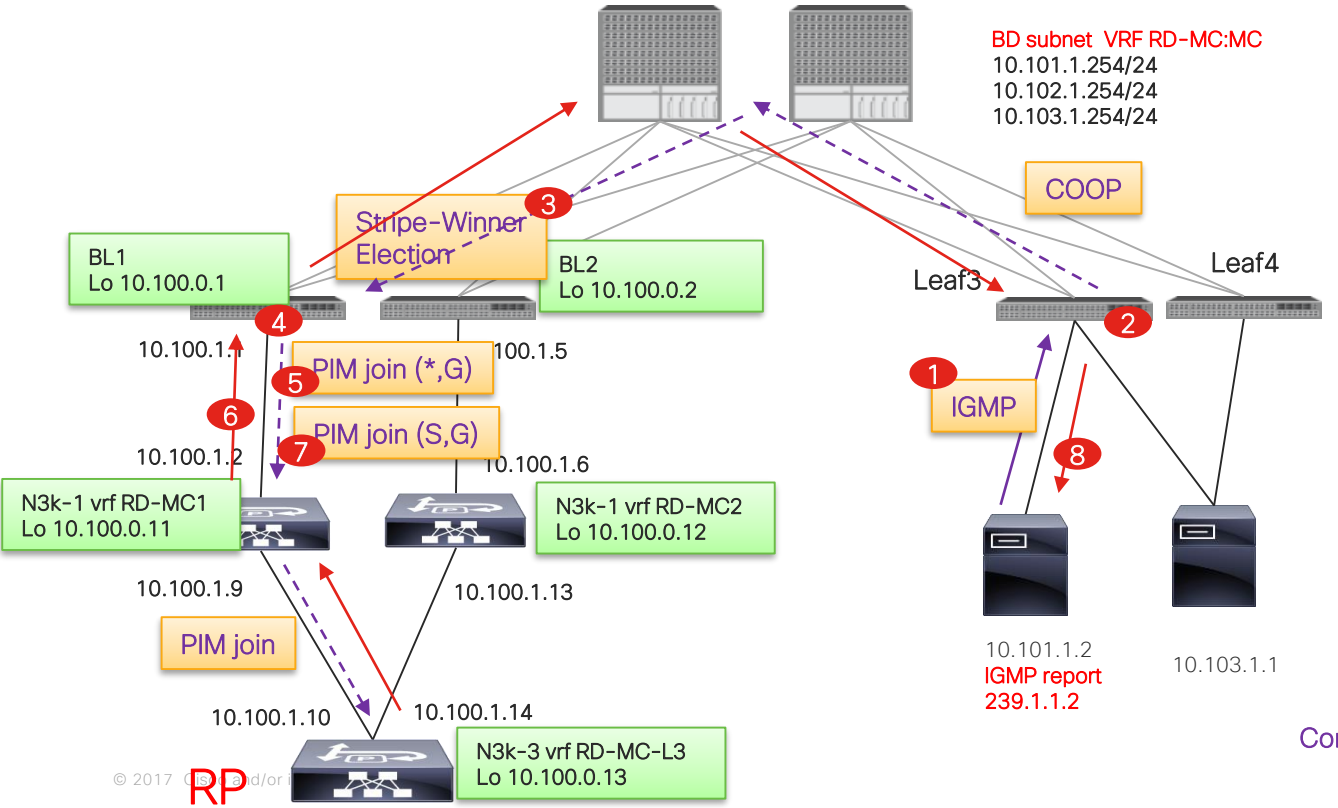
Receiver in ACI fabric

ASM L3 Multicast Receiver inside fabric Default Mode



1. Receiver sends IGMP Join report
2. Leaf 1 installs (*,G)in MRIB and publishes group interest to COOP
3. BL3 and BL4 receive group interest from COOP
4. BL4 is the stripe winner for the group. Mroute will be installed in the MRIB. BL3 is the stripe loser for the group. Mroute will not be installed in MRIB
5. BL4 sends PIM Join towards RP
6. BL4 receives multicast traffic from external network
7. BL4 sends join towards source and prune towards RP. **Border leaves perform the LHR function for the fabric.** Non-border leaves have SPT threshold set to infinity and will not have S,G routes in MRIB
8. Leaf1 receives traffic from source

Receiver in ACI – Shared Tree build up



1. Receiver sends IGMP Join report
2. Leaf 3 installs (*,G)in MRIB and publishes group interest to COOP
3. BL1 and BL2 receive group interest from COOP
4. BL1 is the stripe winner for the group. Mroute will be installed in the MRIB. BL2 is the stripe loser for the group. Mroute will not be installed in MRIB
5. BL4 sends PIM Join towards RP
6. BL4 receives multicast traffic from external network and sends it to tunnel if
7. BL4 sends join towards source and prune towards RP. **Border leaves perform the LHR function for the fabric.** Non-border leaves have SPT threshold set to infinity and will not have S,G routes in MRIB
8. Leaf1 receives traffic from source

Control plane flow (igmp-pim-coop) - - - - - - →

Dataplane Flow ——→

Server Leaf – Gets IGMP report

```
bdsol-aci32-leaf3# show ip igmp groups  vrf RD-MC:RD
IGMP Connected Group Membership for VRF "RD-MC:RD" - 1 total entries
Type: S - Static, D - Dynamic, L - Local, T - SSM Translated
Group Address      Type Interface           Uptime    Expires   Last Reporter
239.1.1.2           D    Vlan74              00:01:08  00:04:05  10.101.1.2

bdsol-aci32-leaf3# show ip mroute 239.1.1.2 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 00:01:39, igmp ip pim
  Incoming interface: Tunnel61, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 1)
    Vlan74, uptime: 00:01:39, igmp
```

Spine COOP Multicast DB

```
bdsol-aci32-spine2# show coop   internal info repo mgroup

------------------------------------------
Repo Hdr Checksum : 8317
Repo Hdr record timestamp : 02 23 2018 07:54:37 124288068
Repo Hdr last pub timestamp : 02 23 2018 07:54:37 124433199
Repo Hdr last dampen timestamp : 01 01 1970 00:00:00 0
Repo Hdr dampen penalty : 0
Repo Hdr flags : IN_OBJ EXPORT
VRF Vnid : 3112960
mgroup src ip : 0.0.0.0
mgroup group ip : 239.1.1.2
Flags : 0x0x1 afi 0
Local leafs 1 (active: 1 deleted: 0)
Leaf 0 Info :
Leaf Repo Hdr Checksum : 0
Leaf Repo Hdr record timestamp : 02 23 2018 07:54:37 124288068
Leaf Repo Hdr last pub timestamp : 02 23 2018 07:54:37 124433199
Leaf Repo Hdr last dampen timestamp : 01 01 1970 00:00:00 0
Leaf Repo Hdr dampen penalty : 0
Leaf Repo Hdr flags : IN_OBJ
Leaf tep ip : 10.0.88.91
oldest local publish timestamp: 02 23 2018 07:53:28 624297072
MPOD Pub id : 0.0.0.0
MPOD leafs 0 (active: 0 deleted: 0)
MSITE Pub id : 0.0.0.0
MSITE leafs 0 (active: 0 deleted: 0)
Hash: 3038634040 owner: 10.0.128.64
```

```
bdsol-aci32-spine2# acidiag fnvread | egrep "10.0.88.91"
     103        1     bdsol-aci32-leaf3      FDO20230USW      10.0.88.91/32    leaf         active   0
```

Spine – coop trace-detail-mc

```
Show coop internal trace-detail-mc


24) 2018 Feb 23 14:52:17.523975 TID 22:coop_oracle_process_publish_tep_add:305: (1519372497:700748648) STORE
message with trans_id 3132: Get TEP record <10.0.96.66> 0x1b303874 from hash table
25) 2018 Feb 23 14:52:17.523973 TID 22:coop_oracle_publish_tep_rmt_preprocess:236: TEP record <10.0.96.66>
Op:Refresh received: return offset:36 ACTION:Forward
26) 2018 Feb 23 14:52:17.523040 TID 18:coop_oracle_process_publish_tep_add:305: (1519372497:700748648) PUBLISH
message with trans_id 3132: Get TEP record <10.0.88.90> 0x1b0a3cf4 from hash table
27) 2018 Feb 23 14:52:17.523036 TID 18:coop_oracle_publish_tep_rmt_preprocess:236: TEP record <10.0.88.90>
Op:Refresh received: return offset:0 ACTION:Local Process
28) 2018 Feb 23 14:52:15.332808 TID 05:coop_repo_tep_aging_handler:983: TEP Aging Callback.
29) 2018 Feb 23 14:51:56.946393 TID 33:coop_oracle_process_publish_mgroup_add:1061: <0x2f8000, V0, 0.0.0.0,
239.1.1.2> (1519372477:124433199) STORE message with trans_id 2954: Get MGROUP record 0x1b607694 from hash table
(hash=b51de038) publish from 10.0.88.91 nleafs=1 flags=1 rep
o_hdr.flags=22
30) 2018 Feb 23 14:51:56.946390 TID 33:coop_oracle_publish_mgroup_rmt_preprocess:890: <0x2f8000, V0, 0.0.0.0,
239.1.1.2> MGROUP record Op:Add received: return offset:76 ACTION:Forward
```

Stripe winner - loser

Non BL have mroute per igmp
Stripe winner have mroute (mgmvpn)
Stripe loser has no states

BL – Stripe winner

```
bdsol-aci32-leaf1# show ip pim internal stripe-winner
239.1.1.2 vrf RD-MC:RD
PIM Stripe Winner info for VRF "RD-MC:RD" (BL count:
2)
(*, 239.1.1.2)
BLs: 10.100.0.1  hash: 1742568489 (local)
      10.100.0.2  hash: 758146928
Winner: 10.100.0.1 best_hash: 1742568489

bdsol-aci32-leaf1# show ip mroute 239.1.1.2 vrf RD-
MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 00:10:37, ngmvpn ip pim
  Incoming interface: Ethernet1/8.30, RPF nbr:
10.100.1.2
  Outgoing interface list: (count: 1) (Fabric OIF)
    Tunnel5, uptime: 00:10:37, ngmvpn
```

Non BL – Stripe loser

```
bdsol-aci32-leaf3# show ip pim  internal stripe-winner
239.1.1.2 vrf RD-MC:RD
PIM Stripe Winner info for VRF "RD-MC:RD" (BL count:
3)
(*, 239.1.1.2)
BLs: 127.0.0.100  hash: 905105898 (local)
      10.100.0.1  hash: 1742568489
      10.100.0.2  hash: 758146928
Winner: 10.100.0.1 best_hash: 1742568489
bdsol-aci32-leaf3# show ip mroute 239.1.1.2 vrf RD-
MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 00:09:21, igmp ip pim
  Incoming interface: Tunnel61, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 1)
    Vlan74, uptime: 00:09:21, igmp
```

BL – Stripe loser

```
bdsol-aci32-leaf2# show ip pim internal stripe-winner
239.1.1.2 vrf RD-MC:RD
PIM Stripe Winner info for VRF "RD-MC:RD" (BL count:
2)
(*, 239.1.1.2)
BLs: 10.100.0.2  hash: 758146928 (local)
      10.100.0.1  hash: 1742568489
Winner: 10.100.0.1 best_hash: 1742568489
bdsol-aci32-leaf2# show ip mroute 239.1.1.2 vrf RD-
MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

Group not found
```

Stripe winner sends PIM join

```
bdsol-aci32-leaf1# show ip pim event-history join-prune | egrep " Feb 23 15:10:52.2"
2018 Feb 23 15:10:52.239138 pim [14079]: TID 14120:pim_process_periodic_for_context:7392:(RD-MC:RD-base) -----
2018 Feb 23 15:10:52.238975 pim [14079]: TID 14120:pim_send_jp:584:(RD-MC:RD-base) Send Join-Prune on
Ethernet1/8.30, length: 62 in context 21
2018 Feb 23 15:10:52.238730 pim [14079]: TID 14120:pim_merge_and_send_jp_msg:3634:(RD-MC:RD-base) Merging curr-
len 62, curr-num-grps 1
2018 Feb 23 15:10:52.238726 pim [14079]: TID 14120:pim_merge_and_send_jp_msg:3611:(RD-MC:RD-base) curr_grp
239.1.1.2, next_grp NULL, size 20
2018 Feb 23 15:10:52.238722 pim [14079]: TID 14120:pim_merge_and_send_jp_msg:3578:(RD-MC:RD-base) curr_grp_nbr
10.100.1.2, next_grp_nbr NULL, temp_glob_nbr 10.100.1.2
2018 Feb 23 15:10:52.238707 pim [14079]: TID 14120:pim_store_in_list:2914:(RD-MC:RD-base) Put (*, 239.1.1.2/32),
WRS in join-list for nbr 10.100.1.2
2018 Feb 23 15:10:52.238700 pim [14079]: TID 14120:pim_store_in_list:2909:(RD-MC:RD-base) wc_bit = TRUE, rp_bit
= TRUE

bdsol-aci32-leaf1# show ip pim neighbor vrf RD-MC:RD
PIM Neighbor Status for VRF "RD-MC:RD"
Neighbor        Interface            Uptime    Expires   DR       Bidir-  BFD
                                                         Priority Capable State
10.100.1.2      Ethernet1/8.30       3d23h     00:01:34  1        no      n/a
10.100.0.2      Tunnel5              3d23h     00:01:27  1        no      n/a
```

MGMVPN – Fabric mroute

Leaf 1- BL stripe winner

```
bdsol-aci32-leaf1# show fabric multicast ipv4 mroute
239.1.1.2 vrf RD-MC:RD
VRF "RD-MC:RD" Fabric mroute Database VNI: 3112960

Fabric Mroute: (*, 239.1.1.2/32)
  COOP related flags:  remote stripe-winner
  MRIB related flags: installed-to-mrib
```

Leaf 2- BL stripe loser

```
bdsol-aci32-leaf2# show fabric multicast ipv4 mroute
239.1.1.2 vrf RD-MC:RD
VRF "RD-MC:RD" Fabric mroute Database VNI: 3112960

Fabric Mroute: (*, 239.1.1.2/32)
  COOP related flags:  remote
  MRIB related flags:
```

Leaf 4- non BL local receiver

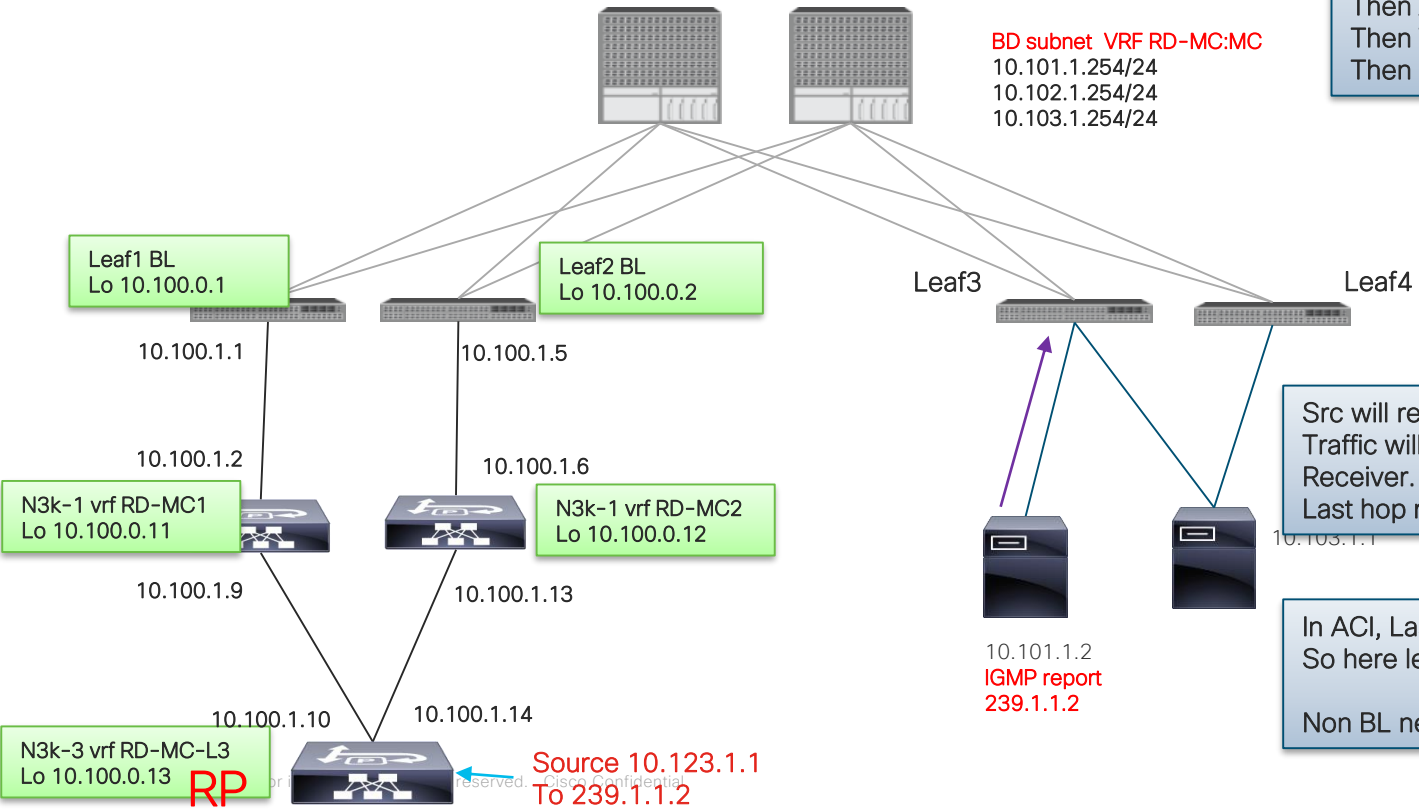
```
bdsol-aci32-leaf3# show fabric multicast ipv4 mroute
239.1.1.2 vrf RD-MC:RD
VRF "RD-MC:RD" Fabric mroute Database VNI: 3112960

Fabric Mroute: (*, 239.1.1.2/32)
  COOP related flags: local
  MRIB related flags:
```

Leaf 4- no info

```
bdsol-aci32-leaf4# show fabric multicast ipv4 mroute
239.1.1.2 vrf RD-MC:RD
VRF "RD-MC:RD" Fabric mroute Database VNI: 3112960

bdsol-aci32-leaf4#
```

Receiver in ACI fabric + source external

Receiver in ACI – Adding Source



Share tree already build
From RP to n3k1- VRF RD-MC1 (left Router)
Then ACI leaf 1 (stripe winner)
Then Tunnel int to distribute to Fabric
Then Leaf3 IGMP receiver

Src will register to RP (outside of ACI)
Traffic will start to flow down the share tree to Receiver.
Last hop router should join Source tree

In ACI, Last hop router role is only the BL.
So here leaf 1 will join SRc Tree,

Non BL never install (s,g) but only keep (*,G)

BL joining source tree

- We can see BL is sending Pim (s,g) join to source tree

```
bdsol-aci32-leaf1# show ip pim internal  event-history join-prune

2018 Feb 27 14:38:38.684902 pim [30562]: TID 30611:pim_store_in_list:2929:(RD-MC:RD-base) Put (10.123.1.1/32,
239.1.1.2/32), S in join-list for nbr 10.100.1.2
2018 Feb 27 14:38:38.684893 pim [30562]: TID 30611:pim_store_in_list:2918:(RD-MC:RD-base) wc_bit = FALSE,
rp_bit = FALSE
2018 Feb 27 14:38:38.684755 pim [30562]: TID 30611:pim_store_in_list:2923:(RD-MC:RD-base) Put (*,
239.1.1.2/32), WRS in join-list for nbr 10.100.1.2
2018 Feb 27 14:38:38.684747 pim [30562]: TID 30611:pim_store_in_list:2918:(RD-MC:RD-base) wc_bit = TRUE,
rp_bit = TRUE
```

Resulting mroute

BL

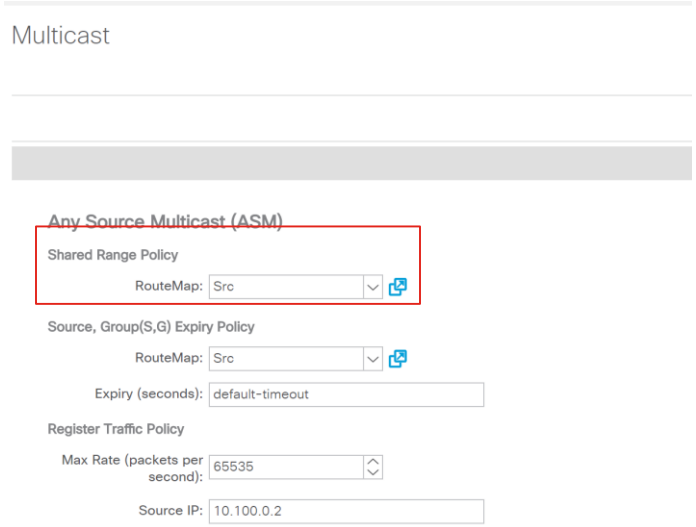
```
bdsol-aci32-leaf1# show ip mroute 239.1.1.2 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 20:02:37, ngmvpn ip pim igmp
  Incoming interface: Ethernet1/8.13, RPF nbr: 10.100.1.2
  Outgoing interface list: (count: 2) (Fabric OIF)
    Vlan29, uptime: 00:03:53, igmp
    Tunnel16, uptime: 20:02:37, ngmvpn

(10.123.1.1/32, 239.1.1.2/32), uptime: 00:00:21, ip mrib pim
  Incoming interface: Ethernet1/8.13, RPF nbr: 10.100.1.2
  Outgoing interface list: (count: 2)
    Vlan29, uptime: 00:00:21, mrib
    Tunnel16, uptime: 00:00:21, mrib


bdsol-aci32-leaf2# show ip mroute 239.1.1.2 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

Group not found

bdsol-aci32-leaf2#
```

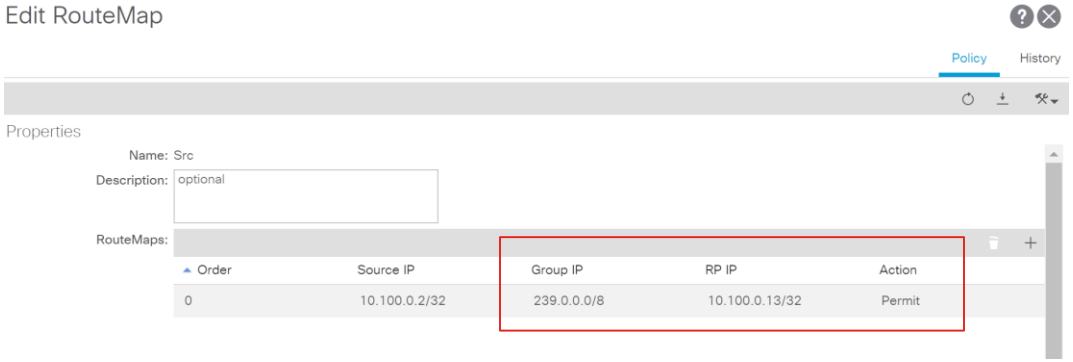
Non-BL

```
bdsol-aci32-leaf3# show ip mroute 239.1.1.2 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 21:53:44, igmp ip pim
  Incoming interface: Tunnel22, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 1)
    Vlan62, uptime: 21:53:44, igmp
```

Setting STP threshold to infinity

- We can prevent BL to join source-tree by setting STP threshold to infinity.
- This is done in vrf/Multicast config/pattern policy/shared Range policy

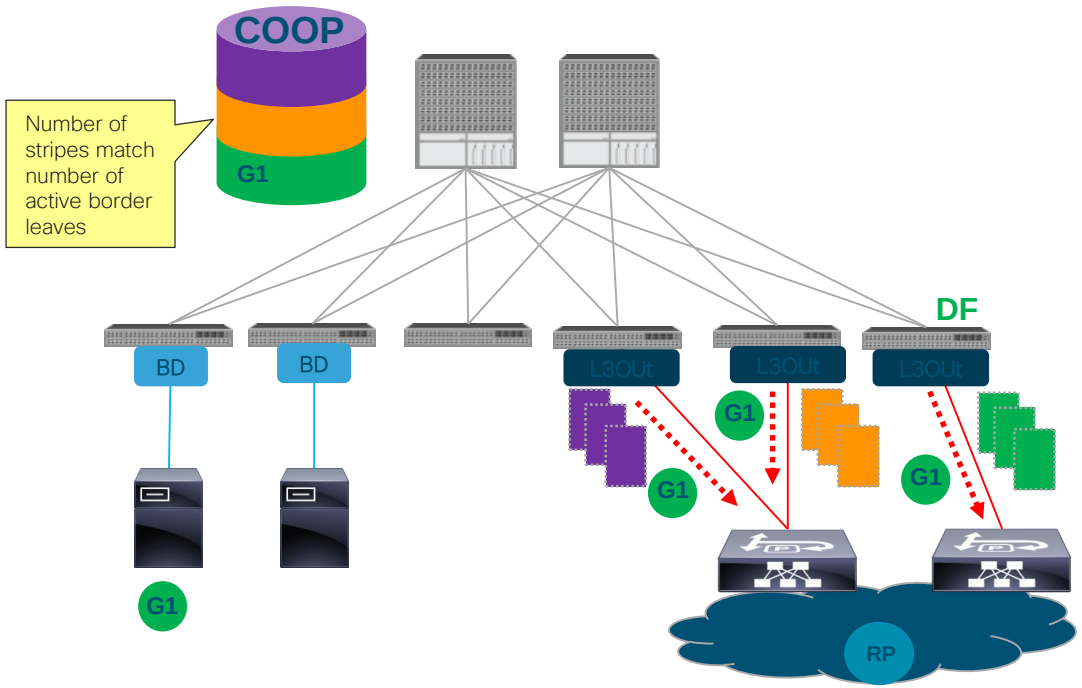




No More Switchover – (s,g) not created on BL

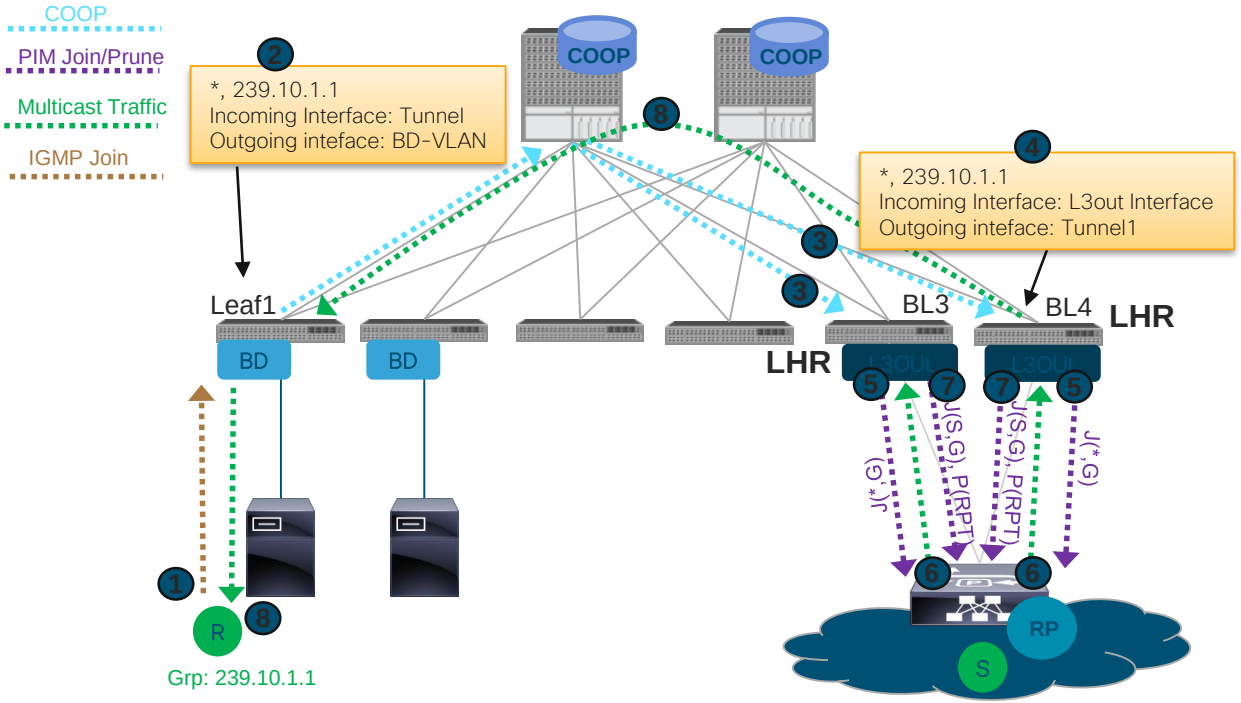
```
bdsol-aci32-leaf1#  show ip pim group-range vrf RD-MC:RD

PIM Group-Range Configuration for VRF "RD-MC:RD"
Group-range        Action    Mode      RP-address       Shrd-tree-range     Origin
232.0.0.0/8        Accept    SSM       -                -                   Local
224.0.0.0/4        Accept    ASM       10.100.0.13      -                   -
239.0.0.0/8        Accept    -         10.100.0.13      Yes                 -

bdsol-aci32-leaf1# show ip mroute 239.1.1.2 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 22:00:34, ngmvpn ip pim igmp
  Incoming interface: Ethernet1/8.13, RPF nbr: 10.100.1.2
  Outgoing interface list: (count: 2) (Fabric OIF)
    Vlan29, uptime: 02:01:51, igmp
    Tunnel16, uptime: 22:00:34, ngmvpn
```

Fast Convergence Mode



- When **Fast Convergence** mode is enabled all border leaves will send joins towards the external network.
- To prevent duplicates, only one border leaf will forward traffic onto the fabric
- The border leaf that forwards traffic for the group is the **designated forwarder** for the group
- The stripe winner decides which BL will be the designated forwarder
- If the stripe winner has a route to the source/RP it will be the designated forwarder

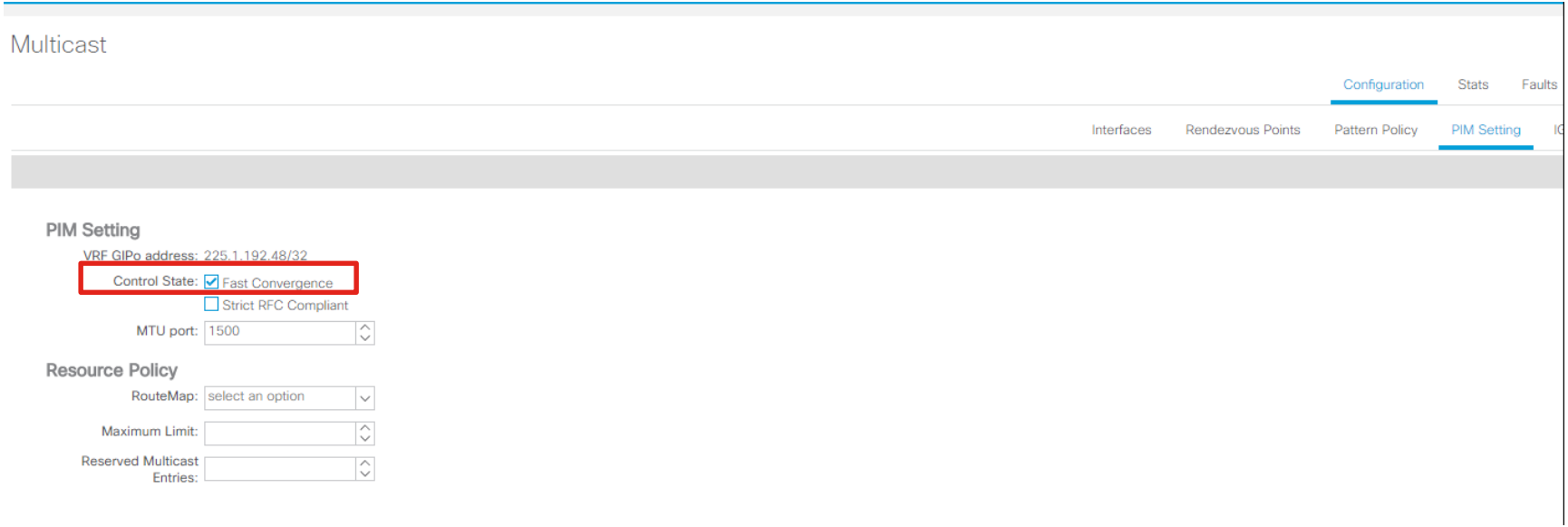
ASM L3 Multicast Receiver inside fabric Fast Convergence Mode



1. Receiver sends IGMP Join report
2. Leaf 1 installs (*,G) in MRIB and publishes group interest to COOP
3. BL3 and BL4 receive group interest from COOP
4. Mroute will be installed in MRIB on both L3 and L4.
5. Both BL3 and BL4 send PIM Joins for (*,G) towards RP
6. Both BL3 and BL4 receive multicast traffic from external network
7. Both BL3 and BL4 sends (S,G) join and (*,G) prune. **Border leaves perform the LHR for the fabric.** Non-border leaves have SPT threshold set to infinity and will not have S,G routes in MRIB
8. BL4 is stripe winner and designated forwarder. It will forward multicast traffic over the fabric interface. BL3 is the stripe loser and drops multicast traffic

Enabling PIM fast switchover

In VRF/Multicast/PIM setting



Fast Convergence check on leaf

```
bdsol-aci32-leaf2# show  ip pim vrf RD-MC:RD detail
PIM Enabled VRFs
VRF Name              VRF      Table       Interface  BFD        MVPN
                      ID       ID          Count      Enabled    Enabled
RD-MC:RD              5        0x00000005  3          no          no
  State Limit: 4294967295,  Available States: 4294967295
  Register Rate Limit: 65535 pps
  Register source address : 10.100.0.2
  Shared tree ranges: none
  (S,G)-expiry timer: not configured
    (S,G)-list policy: none
    (S,G)-expiry timer config version 0, active version 0

  Pre-build SPT for all (S,G)s in VRF: disabled
  CLI vrf done: TRUE
  PIM cibtype Auto Enabled: yes
  txlist work pending: FALSE
  PIM VxLAN VNI ID: 0
  iVxlan VRF VNID: 3112960
  Fabric IOD: 0x53
  Fast Convergence: YES
  Num Interfaces: 3
```

Non Stripe Winner routing-table

Before – no fast failover

```
bdsol-aci32-leaf2# show ip mroute vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"
```

Upstream router of Leaf2
Now have states and
Send Mcast to leaf2.
Leaf2 is not stripe winner
So do not actually forward
To tunnel 16

After – with fast failover

```
bdsol-aci32-leaf2# show ip mroute 239.1.1.2 vrf RD-MC:RD

(*, 239.1.1.2/32), uptime: 00:00:03, ngmvpn ip pim
  Incoming interface: Ethernet1/8.12, RPF nbr: 10.100.1.6
  Outgoing interface list: (count: 1) (Fabric OIF) (Fabric
Forwarding Loser)
    Tunnel16, uptime: 00:00:03, ngmvpn

(10.123.1.1/32, 239.1.1.2/32), uptime: 00:00:01, ip mrib pim
  Incoming interface: Ethernet1/8.12, RPF nbr: 10.100.1.6
  Outgoing interface list: (count: 1) (Fabric Forwarding Loser)
    Tunnel16, uptime: 00:00:01, mrib
```

```
bdsol-aci32-n3k-1# show ip mroute vrf RD-MC2
IP Multicast Routing Table for VRF "RD-MC2"

(*, 232.0.0.0/8), uptime: 1w1d, pim ip
  Incoming interface: Null, RPF nbr: 0.0.0.0, uptime: 1w1d
  Outgoing interface list: (count: 0)

(10.123.1.1/32, 239.1.1.2/32), uptime: 00:00:27, pim ip
  Incoming interface: Vlan11, RPF nbr: 10.100.1.14, uptime: 00:00:27
  Outgoing interface list: (count: 1)
    Ethernet1/10.10, uptime: 00:00:27, pim
```

Forwarding software info

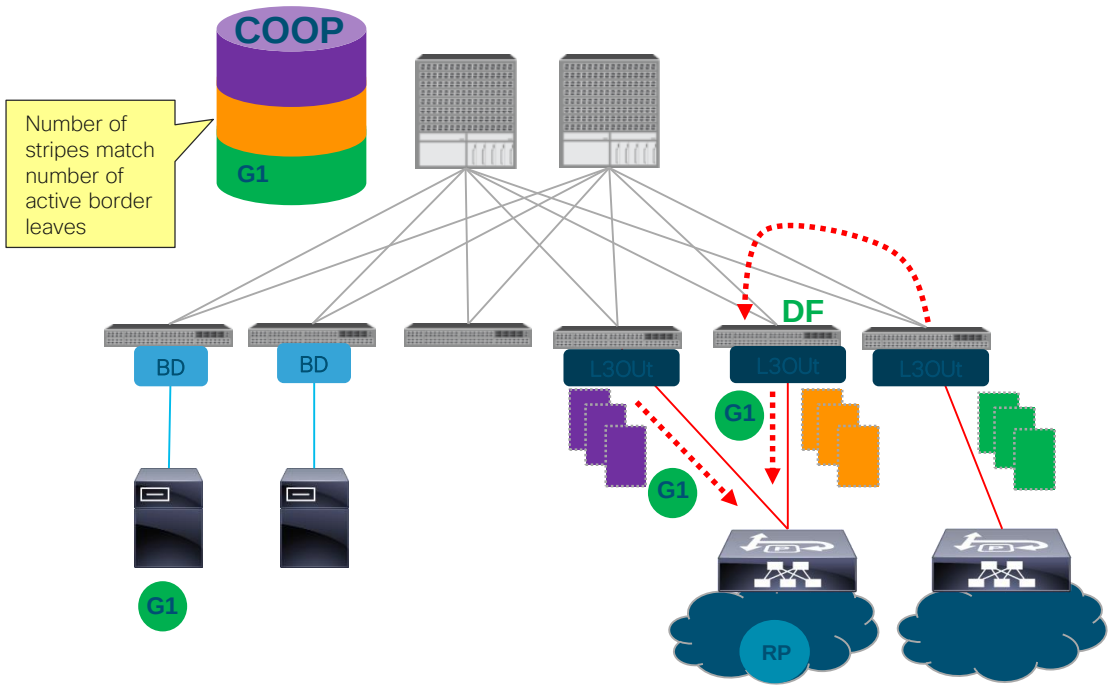
Stripe winner BL does have OIL
Matching mroute

```
bdsol-aci32-leaf1# show forwarding distribution
multicast route vrf RD-MC:RD group 239.1.1.2

  (*, 239.1.1.2/32), RPF Interface: Ethernet1/8.13,
flags: G
    Received Packets: 0 Bytes: 0
    Number of Outgoing Interfaces: 2
    Outgoing Interface List Index: 8205
      Vlan29
      Tunnel16

  (10.123.1.1/32, 239.1.1.2/32), RPF Interface:
Ethernet1/8.13, flags:
    Received Packets: 13 Bytes: 845
    Number of Outgoing Interfaces: 2
    Outgoing Interface List Index: 8205
      Vlan29
      Tunnel16
```

Stripe loser BL do not have OIL
Matching mroute (mroute oil just there
To pull traffic with pim join)

```
bdsol-aci32-leaf2# show forwarding distribution
multicast route vrf RD-MC:RD group 239.1.1.2

  (*, 239.1.1.2/32), RPF Interface: Tunnel16, flags: G
    Received Packets: 0 Bytes: 0
    Number of Outgoing Interfaces: 0
    Null Outgoing Interface List

  (10.123.1.1/32, 239.1.1.2/32), RPF Interface:
Tunnel16, flags:
    Received Packets: 12 Bytes: 780
    Number of Outgoing Interfaces: 0
    Null Outgoing Interface List
bdsol-aci32-leaf2#
```

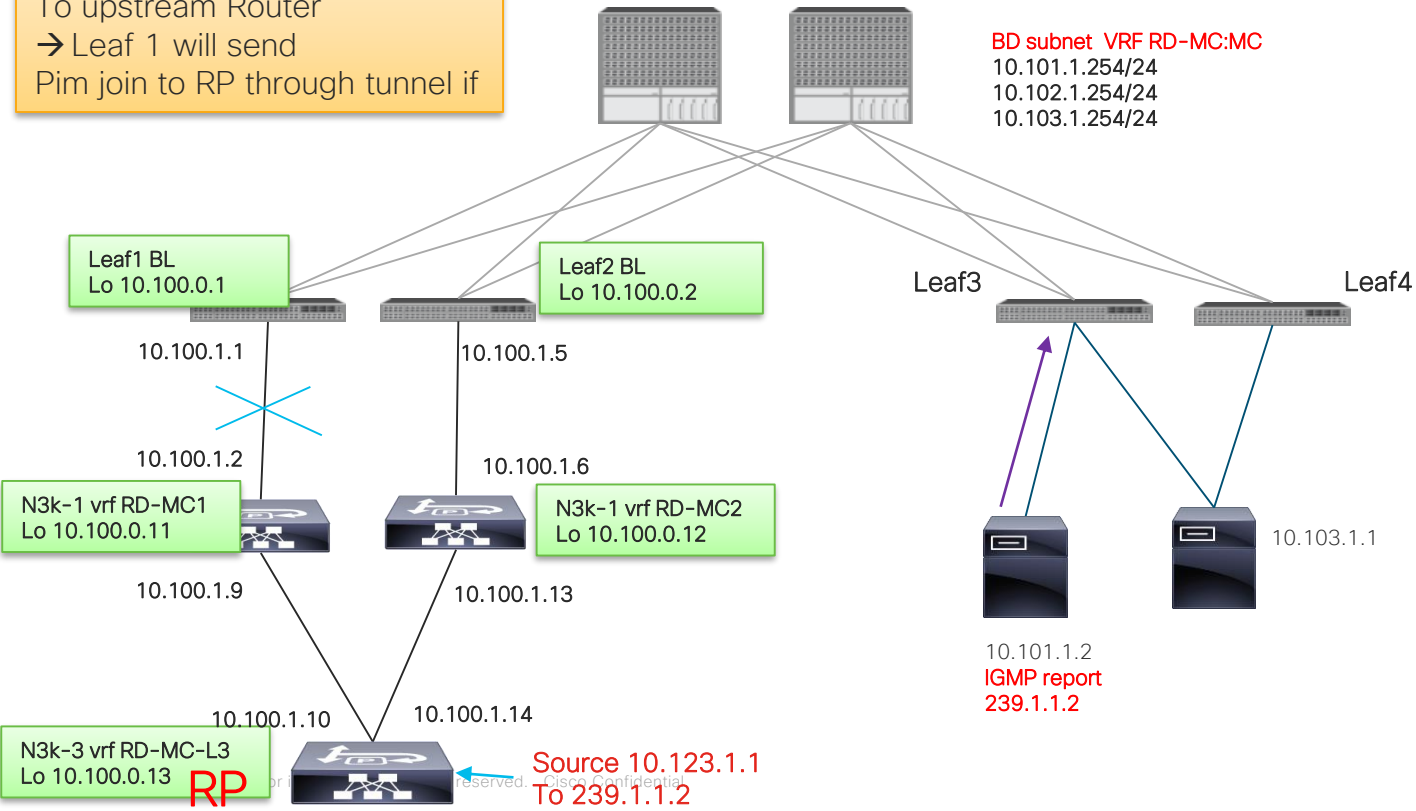
Stripe winner does not have route to source



- If the stripe winner does not have a route to the root (RP/source) it will select another BL as the designated forwarder by sending a PIM join over the fabric
- This is the only time PIM joins are sent within the fabric
- All border leaves with a route to the root will send PIM joins
- The selected DF will forward traffic onto the fabric

Receiver in ACI – BL route to RP through ACI

Leaf1 is stripe winner
We break link from leaf 1
To upstream Router
→ Leaf 1 will send
Pim join to RP through tunnel if



Leaf1 mroute

Before – Inc IF is Eth1/8

```
bdsol-aci32-leaf1# show ip mroute 239.1.1.2 vrf RD-
MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 1d20h, ngmvpn ip pim igmp
  Incoming interface: Ethernet1/8.13, RPF nbr:
10.100.1.2
  Outgoing interface list: (count: 2) (Fabric OIF)
    Vlan29, uptime: 1d00h, igmp
    Tunnel16, uptime: 1d20h, ngmvpn

(10.123.1.1/32, 239.1.1.2/32), uptime: 00:27:22, ip
mrib pim
  Incoming interface: Ethernet1/8.13, RPF nbr:
10.100.1.2
  Outgoing interface list: (count: 2)
    Vlan29, uptime: 00:27:22, mrib
    Tunnel16, uptime: 00:27:22, mrib
```

After – Incoming if is Tunnel -> we send pim join

```
bdsol-aci32-leaf1# show ip mroute 239.1.1.2 vrf RD-
MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.2/32), uptime: 1d20h, ngmvpn ip pim igmp
  Incoming interface: Tunnel16, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 2) (Fabric OIF)
    Vlan29, uptime: 1d00h, igmp
    Tunnel16, uptime: 1d20h, ngmvpn, (RPF)

(10.123.1.1/32, 239.1.1.2/32), uptime: 00:27:38, ip
mrib pim
  Incoming interface: Tunnel16, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 2)
    Vlan29, uptime: 00:27:38, mrib
    Tunnel16, uptime: 00:27:38, mrib, (RPF)
```

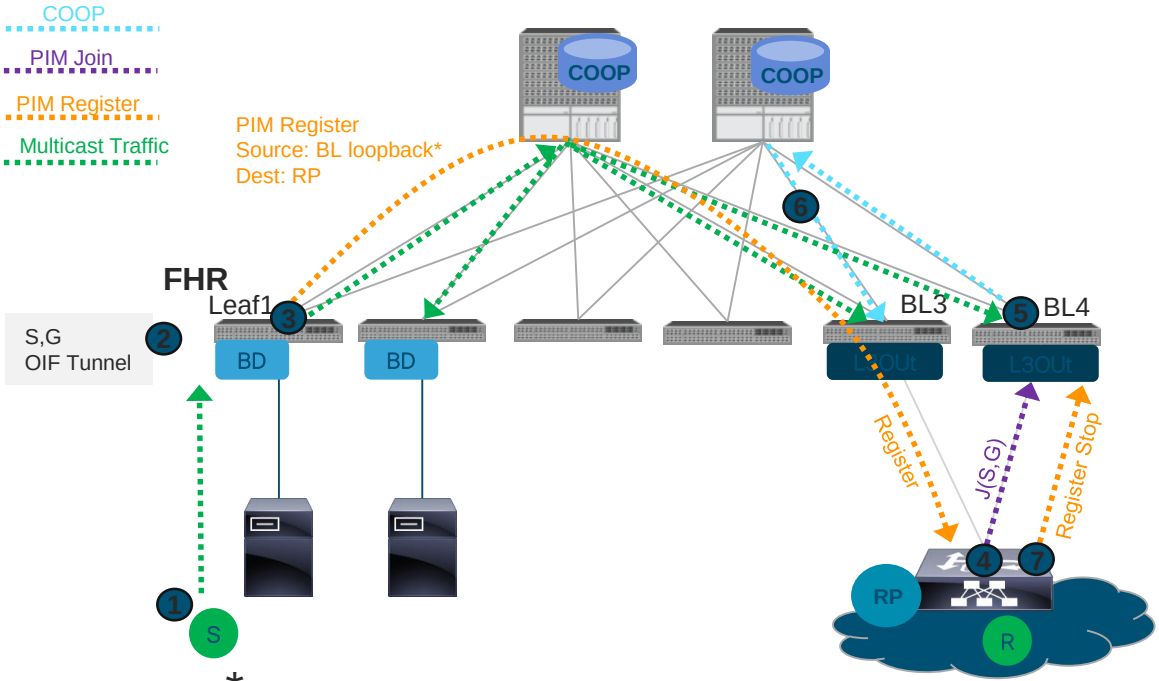
PIM join across the fabric

- Leaf does sen a PIM join to new RPF interface (tunnel 16) Leaf2 PIM neighbor 10.100.0.2

```
Show ip pim event-history join-prun

2018 Feb 28 13:16:20.484957 pim [30562]: TID 30611:pim_store_in_list:2929:(RD-MC:RD-base) Put (10.123.1.1/32,
239.1.1.2/32), S in join-list for nbr 10.100.0.2
2018 Feb 28 13:16:20.484948 pim [30562]: TID 30611:pim_store_in_list:2918:(RD-MC:RD-base) wc_bit = FALSE,
rp_bit = FALSE
2018 Feb 28 13:16:20.484847 pim [30562]: TID 30611:pim_age_route:6489:(RD-MC:RD-base) (10.123.1.1/32,
239.1.1.2/32) expiration timer updated due to data activity
2018 Feb 28 13:16:20.484798 pim [30562]: TID 30611:pim_store_in_list:2923:(RD-MC:RD-base) Put (*,
239.1.1.2/32), WRS in join-list for nbr 10.100.0.2
```

Source in ACI fabric
Receiver outside

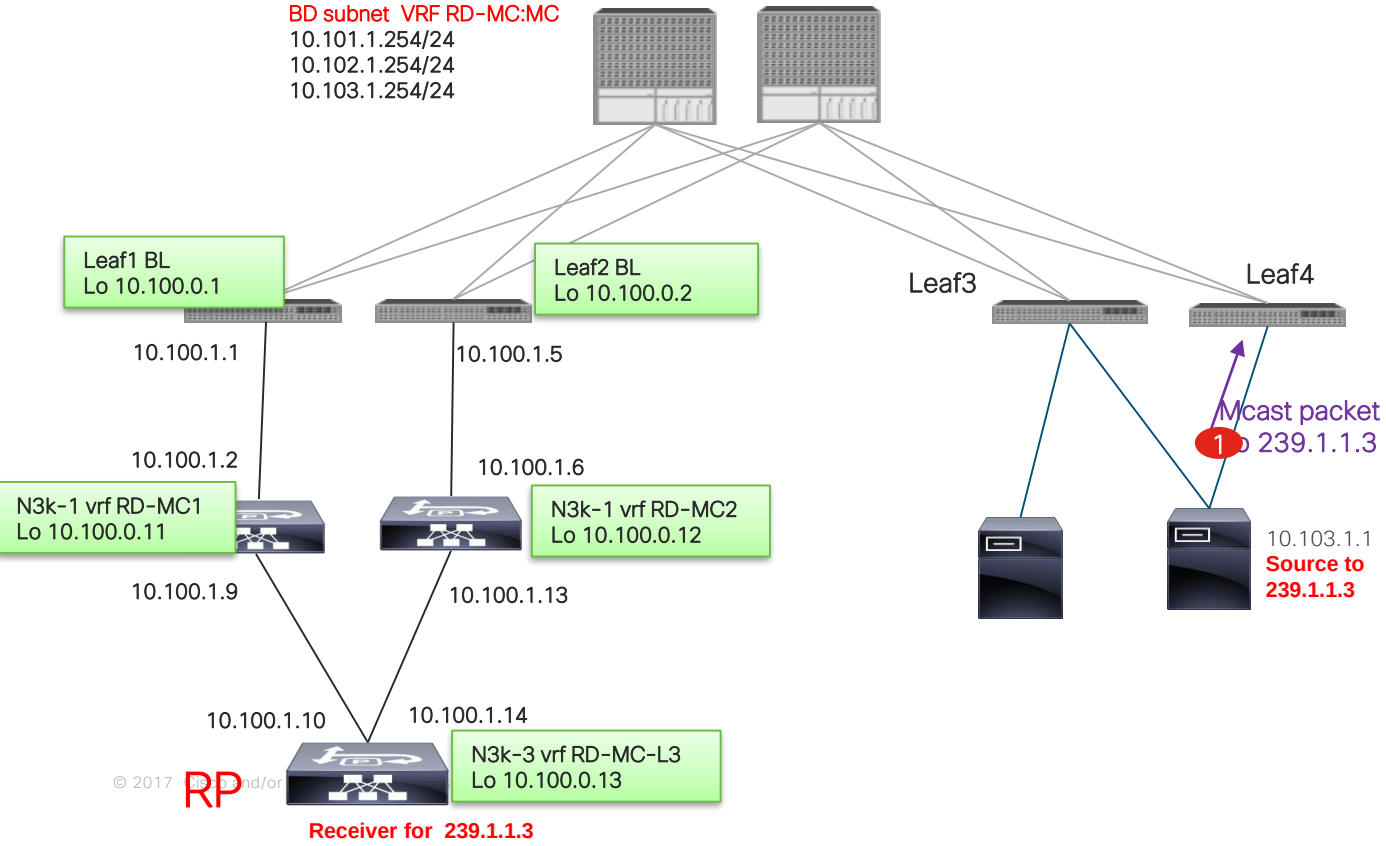
ASM L3 Multicast Source inside fabric



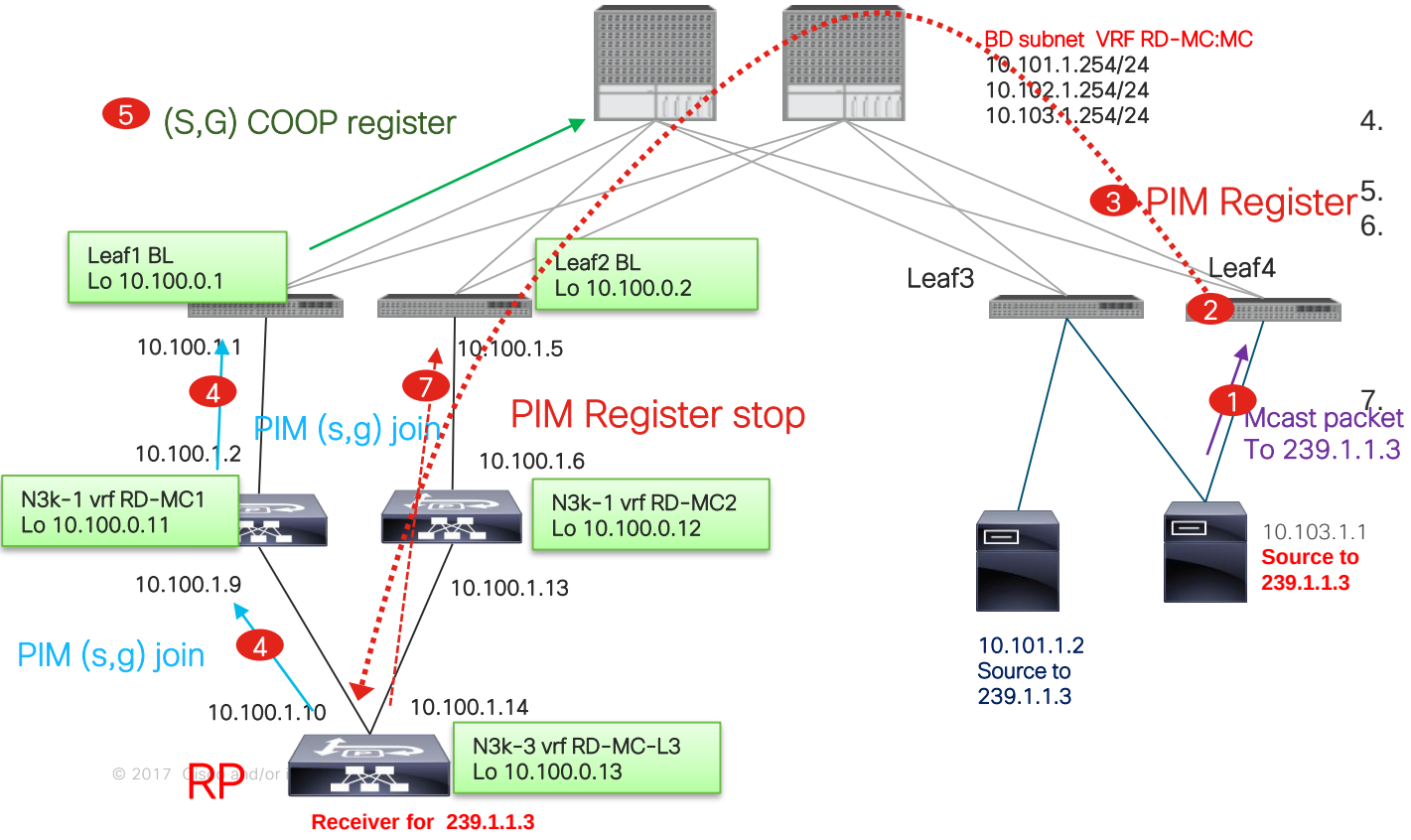
1. Source starts sending multicast traffic
2. Leaf 1 performs FHR function, installs S,G with fabric interface in Inc Inteface
3. Leaf 1 sends PIM register towards RP. Leaf 1 selects an active border leaf (BL3) using a hash. Forwards traffic over fabric to all leaves in VRF
4. RP receiving register joins the SPT. Sends (S,G) join to BL4
5. BL4 publishes join to COOP
6. BL3 receives COOP update for S,G. Source is local. S,G is installed on border leaf. RPF is fabric tunnel interface, L3OUt interface is OIF.
7. Register stop sent by RP

* By default the fabric will select a BD subnet IP address as the source of the register message. This may be a different BD from where the source is connected. If the BD subnet is not advertised externally the RP will not be able to send the register stop.

Lab topology



Lab topology



1. Source starts sending multicast traffic
2. Leaf 4 performs FHR function, installs (S,G) with fabric interface in inc Interface
3. Leaf 4 sends PIM register towards RP. Leaf 4 selects an active border leaf (BL1 or BL2) using a hash. Forwards traffic over fabric to all leaves in VRF
4. RP receiving register joins the SPT. Sends (S,G) join to BL1 (RPF tree)
5. BL1 publishes join to COOP
6. BL1 receives COOP update for (S,G). Source is local. (S,G) is installed on border leaf. RPF is fabric tunnel interface, L3OUT interface is OIF.
7. Register stop sent by RP

Server leaf

- (s,g) entry created on server leaf where source is connected
- RPF is set to Fabric interface (route to Src is BD subnet)
- RPF nbr is the Anycast v4 spine address in overlay-1

```
bdsol-aci32-leaf4# show ip mroute 239.1.1.3 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(10.101.1.1/32, 239.1.1.3/32), uptime: 00:01:46, ip pim
  Incoming interface: Tunnel59, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 0)

bdsol-aci32-leaf4# show ip pim  route  239.1.1.3 vrf RD-MC:RD
PIM Routing Table for VRF "RD-MC:RD" - 2 entries

(10.101.1.1/32, 239.1.1.3/32), expires 00:02:43
  Incoming interface: Tunnel59, RPF nbr 0.0.0.0
  Oif-list:       (0) 00000000, timeout-list: (0) 00000000
  Immediate-list: (0) 00000000, timeout-list: (0) 00000000
  Sgr-prune-list: (0) 00000000  Assert-win-oif-list: (0) 00000000
  Timeout-interval: 3, JP-holdtime round-up: 3
```

Leaf 4 leaks new source traffic to CPU to send register

Mcast leak

```
bdsol-aci32-leaf4# tcpdump -i kpm_inb host 239.1.1.2
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on kpm_inb, link-type EN10MB (Ethernet), capture size 65535 bytes
16:43:13.819251 IP hsrp-10-101-1-1.cisco.com > 239.1.1.2: ICMP echo request, id 1452, seq 104, length 64
16:43:14.819230 IP hsrp-10-101-1-1.cisco.com > 239.1.1.2: ICMP echo request, id 1452, seq 105, length 64
16:43:15.819209 IP hsrp-10-101-1-1.cisco.com > 239.1.1.2: ICMP echo request, id 1452, seq 106, length 64
```

PIM register send

```
bdsol-aci32-leaf4# tcpdump -xxvvi kpm_inb pim
tcpdump: verbose output suppressed, use -v or -vv for full protocol decode
listening on kpm_inb, link-type EN10MB (Ethernet), capture size 65535 bytes
16:44:57.515872 IP 10.101.1.254 > 10.100.0.13: PIMv2, Register, length 92
16:45:12.544750 IP 10.101.1.254 > 10.100.0.13: PIMv2, Register, length 28

16:47:55.421896 IP (tos 0xc0, ttl 2, id 45672, offset 0, flags [none], proto PIM (103), length 112)
    10.101.1.254 > 10.100.0.13: PIMv2, length 92
        Register, cksum 0xdeff (correct), Flags [ none ]
        IP (tos 0x0, ttl 1, id 0, offset 512, flags [none], proto ICMP (1), length 84)
    hsrp-10-101-1-1.cisco.com > 239.1.1.2: icmp
        0x0000:  0000 0000 0000 0000 0000 0000 0800 45c0
        0x0010:  0070 b268 0000 0267 ee2b 0a65 01fe 0a64
        0x0020:  000d 2100 deff 0000 0000 4500 0054 0000
        0x0030:  0040 0101 bdfb 0a65 0101 ef01 0102 0800    10.101.1.1 → 239.1.1.2
        0x0040:  1d29 05af 0001 a545 905a 0000 0000 dcb3
        0x0050:  0400 0000 0000 1011 1213 1415 1617 1819
        0x0060:  1a1b 1c1d 1e1f 2021 2223 2425 2627 2829
        0x0070:  2a2b 2c2d 2e2f 3031 3233 3435 3637
```

Leaf4 pim register event-history

```
bdsol-aci32-leaf4# show ip pim event-history null-register

 null-register events for PIM process


2018 Feb 23 16:48:17.398369 pim [14254]: TID 14282:pim_receive_register_stop:1660:(RD-MC:RD-base) Received
Register-Stop from 10.100.0.13 for (10.101.1.1/32, 239.1.1.2/32)
2018 Feb 23 16:48:17.396737 pim [14254]: TID 14301:pim_send_null_register:3258:(RD-MC:RD-base) Send Null
Register to RP 10.100.0.13 for (10.101.1.1/32, 239.1.1.2/32)
2018 Feb 23 16:47:16.701007 pim [14254]: TID 14301:pim_send_null_register:3258:(RD-MC:RD-base) Send Null
Register to RP 10.100.0.13 for (10.101.1.1/32, 239.1.1.2/32)
```

Note the register stop is first receive by BL (owning BD subnet as well)

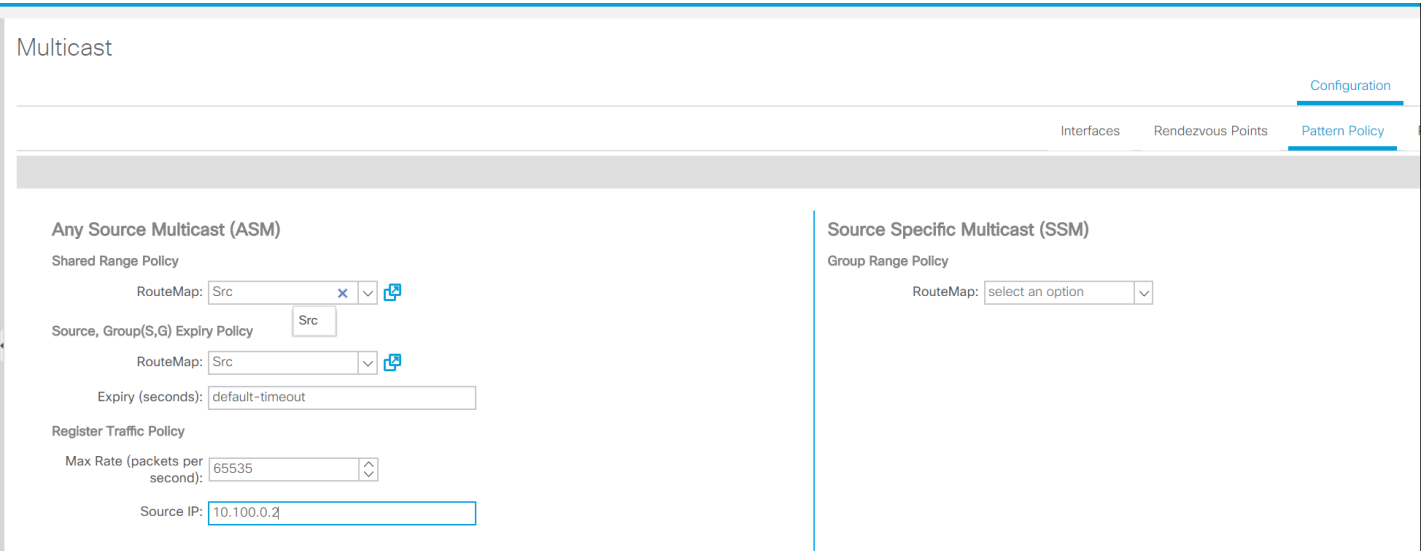
```
bdsol-aci32-leaf1# show ip pim  event-history null-register

 null-register events for PIM process

2018 Feb 23 16:45:12.545676 pim [14079]: TID 14092:pim_receive_register_stop:1660:(RD-MC:RD-base) Received
Register-Stop from 10.100.0.13 for (10.101.1.1/32, 239.1.1.2/32)
```

Debug PIM register on RP

Register is send from BD subnet pervasive GW !

```
2018 Feb 23 16:28:18.269380 pim: [3758] (RD-MC-L3-base) Received Register from 10.101.1.254 for
(10.101.1.1/32, 239.1.1.2/32), border_bit FALSE
2018 Feb 23 16:28:18.269893 pim: [3758] (RD-MC-L3-base) Create route for (10.101.1.1/32, 239.1.1.2/32)
2018 Feb 23 16:28:18.270350 pim: [3758] (RD-MC-L3-base) Add route (10.101.1.1/32, 239.1.1.2/32) to MRIB,
multi-route TRUE
2018 Feb 23 16:28:18.270652 pim: [3758] (RD-MC-L3-base) Send Register-Stop to 10.101.1.254 for (10.101.1.1/32,
239.1.1.2/32)
```

Changing PIM register source to a BL Loopback



Source Register with BD subnet is not the best ,
It is recommended to change the src of register to be a BL loopback
In Tn/VRF/Multicast/ Pattern policy

PIM register source address check

```
bdsol-aci32-leaf3# show ip pim vrf RD-MC:RD detail
PIM Enabled VRFs
VRF Name              VRF      Table       Interface  BFD        MVPN
                      ID       ID          Count      Enabled    Enabled
RD-MC:RD              19       0x00000013  4          no          no
  State Limit: 4294967295,  Available States: 4294967294
  Register Rate Limit: 65535 pps
  Register source address : 10.100.0.2
  Shared tree route-map: mcast_permit_all
         route-ranges:
                    224.0.0.0/4 Accept
  (S,G)-expiry timer: not configured
    (S,G)-list policy: none
    (S,G)-expiry timer config version 0, active version 0

  Pre-build SPT for all (S,G)s in VRF: disabled
  CLI vrf done: TRUE
  PIM cibtype Auto Enabled: yes
  txlist work pending: FALSE
  PIM VxLAN VNI ID: 0
  iVxlan VRF VNID: 3112960
  Fabric IOD: 0x79
  Fast Convergence: NO
  Num Interfaces: 4
```

register debug

- RP debug : Now we can see Source of Register is a loopback of BL

```
bdsol-aci32-n3k-3(config-line)# 2018 Feb 26 14:50:18.087246 pim: [3758] (RD-MC-L3-base) Received Register from
10.100.0.2 for (10.101.1.1/32, 239.1.1.3/32), border_bit FALSE
2018 Feb 26 14:50:18.087459 pim: [3758] (RD-MC-L3-base) Send Register-Stop to 10.100.0.2 for (10.101.1.1/32,
239.1.1.3/32)
```

- Tcpdump on server leaf – Src is loopback of BL2

```
bdsol-aci32-leaf4# tcpdump -i kpm_inb pim
15:41:04.965586 IP 10.100.0.2 > 10.100.0.13: PIMv2, Register, length 28
```

RP sends PIM join on its RPF interface to Source

- Here RP sends PIM join vlan 10 to neighbor 10.100.1.9 (N3k1- VRF RD-MC1)

```
bdsol-aci32-n3k-3(config-if)# show ip mroute 239.1.1.3 vrf RD-MC-L3
IP Multicast Routing Table for VRF "RD-MC-L3"

(*, 239.1.1.3/32), uptime: 2d20h, igmp pim ip
  Incoming interface: loopback80, RPF nbr: 10.100.0.13
  Outgoing interface list: (count: 1)
    loopback80, uptime: 2d20h, igmp, (RPF)

(10.101.1.1/32, 239.1.1.3/32), uptime: 00:02:59, pim mrib ip
  Incoming interface: Vlan10, RPF nbr: 10.100.1.9, internal
  Outgoing interface list: (count: 1)
    loopback80, uptime: 00:02:59, mrib

bdsol-aci32-n3k-3(config-if)# show ip route 10.101.1.1 vrf RD-MC-L3

10.101.1.0/24, ubest/mbest: 2/0
    *via 10.100.1.9, Vlan10, [110/20], 6d21h, ospf-666, type-2
    *via 10.100.1.13, Vlan11, [110/20], 1w2d, ospf-666, type-2
```

N3k-1 propagates (s,g) Join

- N3k-a sends (s,g) join on 1/8.10 towards ACI BL leaf 1

```
bdsol-aci32-n3k-1# show ip mroute 239.1.1.3 vrf RD-MC1
IP Multicast Routing Table for VRF "RD-MC1"

(*, 239.1.1.3/32), uptime: 00:02:58, pim ip
  Incoming interface: Vlan10, RPF nbr: 10.100.1.10, uptime: 00:02:58
  Outgoing interface list: (count: 0)

(10.101.1.1/32, 239.1.1.3/32), uptime: 00:03:18, pim ip
  Incoming interface: Ethernet1/8.10, RPF nbr: 10.100.1.1, uptime: 00:03:18
  Outgoing interface list: (count: 1)
    Vlan10, uptime: 00:03:18, pim
bdsol-aci32-n3k-1# show lldp neighbors interface ethernet 1/8
Capability codes:
  (R) Router, (B) Bridge, (T) Telephone, (C) DOCSIS Cable Device
  (W) WLAN Access Point, (P) Repeater, (S) Station, (O) Other
Device ID           Local Intf      Hold-time  Capability  Port ID
bdsol-aci32-leaf1.cisco.com
                    Eth1/8          120        BR          Eth1/8
```

COOP citizen on leaf 1

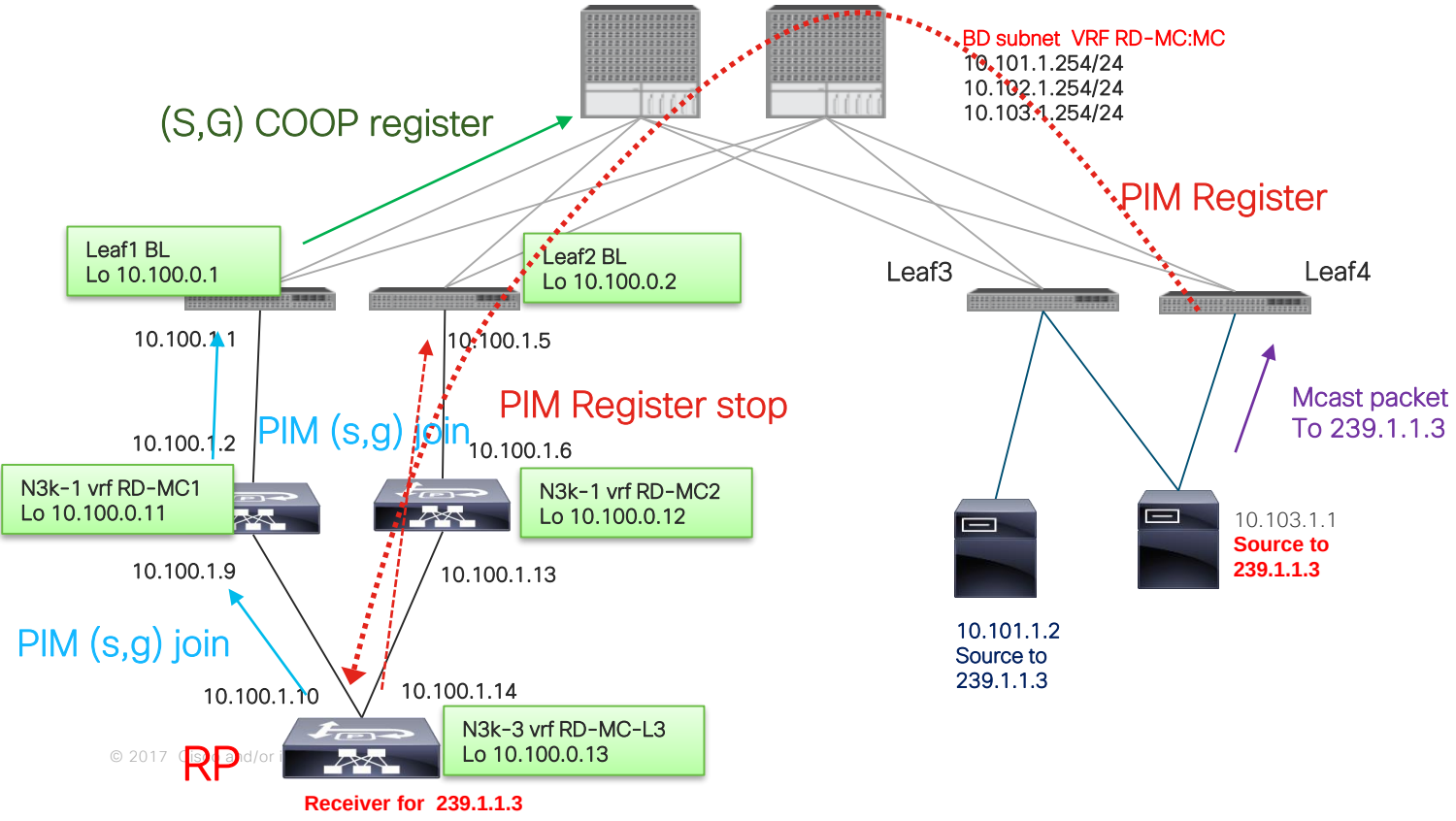
- BL that receive the PIM join from RP publish group to COOP

```
bdsol-aci32-leaf1# show coop internal  event-history trace-detail-mc

2) 2018 Feb 26 13:20:59.144537 TID 01:coop_igmp_process_one_mgroup_common:711: Received ADD L3 MGROUP response
<0, 10.101.1.1, 239.1.1.3, 2f8000> (seqno 1)

5) 2018 Feb 26 13:20:59.141918 TID 01:coop_citizen_publish_mgroup_common:891: Received ADD MGROUP msg <0,
10.101.1.1, 239.1.1.3, 2f8000>  with trans_id 5078 rec_ts: 1519626219:326068515 pub_ts: 1519626219:326231665
flags=0
```

Spine COOP entry

```
bdsol-aci32-spine2# show coop internal  info  repo mgroup

----------------------------------------
Repo Hdr Checksum : 46449
Repo Hdr record timestamp : 02 26 2018 06:23:39 326068515
Repo Hdr last pub timestamp : 02 26 2018 06:23:39 326231665
Repo Hdr last dampen timestamp : 01 01 1970 00:00:00 0
Repo Hdr dampen penalty : 0
Repo Hdr flags : IN_OBJ EXPORT
VRF Vnid : 3112960
mgroup src ip : 10.101.1.1
mgroup group ip : 239.1.1.3
Flags : 0x0x1 afi 0
Local leafs 1 (active: 1 deleted: 0)
Leaf 0 Info :
Leaf Repo Hdr Checksum : 0
Leaf Repo Hdr record timestamp : 02 26 2018 06:23:39 326068515
Leaf Repo Hdr last pub timestamp : 02 26 2018 06:23:39 326231665
Leaf Repo Hdr last dampen timestamp : 01 01 1970 00:00:00 0
Leaf Repo Hdr dampen penalty : 0
Leaf Repo Hdr flags : IN_OBJ
Leaf tep ip : 10.0.88.95                                LEAF 1 (BL TEP )
oldest local publish timestamp: 02 26 2018 06:12:22 31282186
MPOD Pub id : 0.0.0.0
MPOD leafs 0 (active: 0 deleted: 0)
MSITE Pub id : 0.0.0.0
MSITE leafs 0 (active: 0 deleted: 0)
Hash: 3546171375 owner: 10.0.88.65
```

Lab topology



Resulting mroute

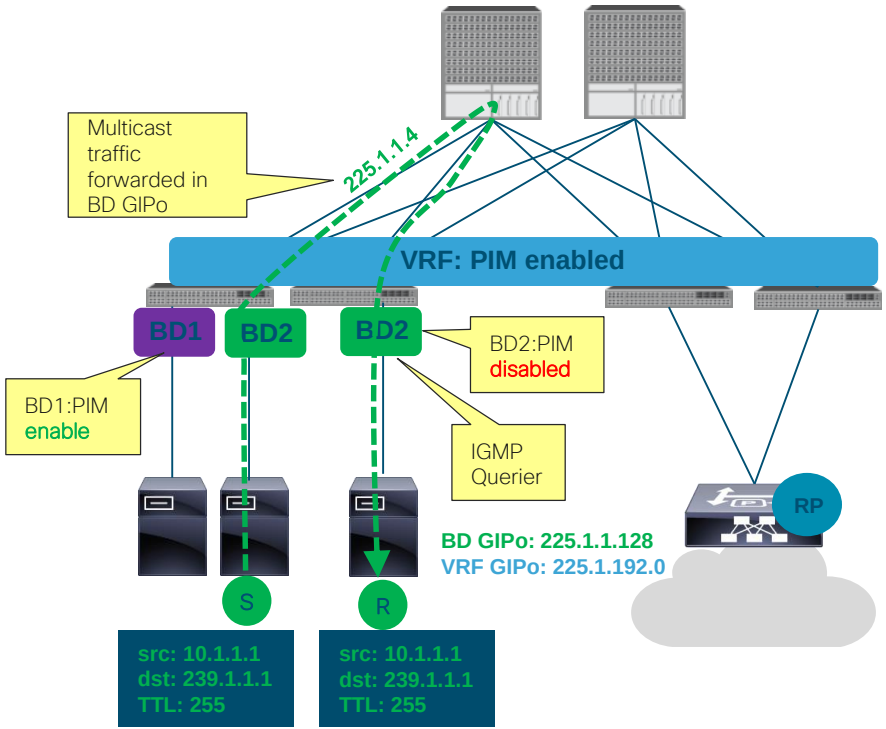
```
bdsol-aci32-leaf1# show ip mroute 239.1.1.3 vrf RD-
MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(10.101.1.1/32, 239.1.1.3/32), uptime: 01:05:54, pim
ip ngmvpn
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65
(pervasive)
  Outgoing interface list: (count: 2) (Fabric OIF)
    Ethernet1/8.13, uptime: 00:00:01, pim
    Tunnel16, uptime: 00:17:39, ngmvpn, (RPF)
```

```
bdsol-aci32-leaf2# show ip mroute 239.1.1.3 vrf RD-
MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(10.101.1.1/32, 239.1.1.3/32), uptime: 00:17:51, pim
ip
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65
(pervasive)
  Outgoing interface list: (count: 0)
```

```
bdsol-aci32-leaf4# show ip mroute 239.1.1.3 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(10.101.1.1/32, 239.1.1.3/32), uptime: 00:39:20, ip pim
  Incoming interface: Tunnel21, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 0)
```

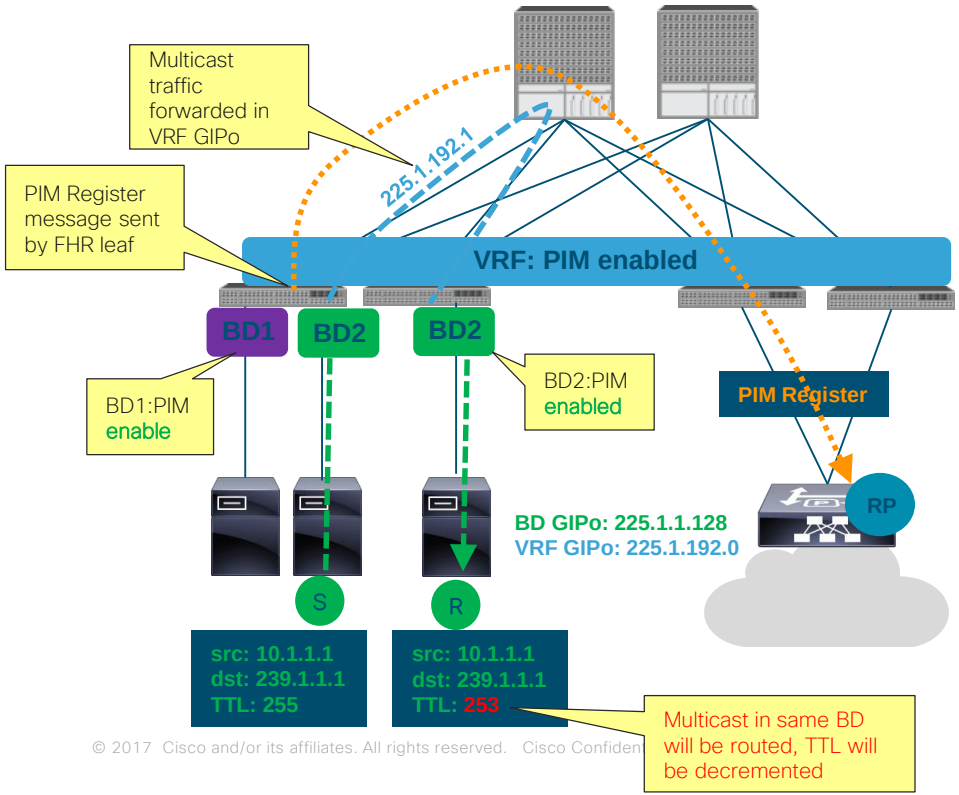
Other Scenario

L2 Multicast with PIM enabled VRF, PIM disabled BD



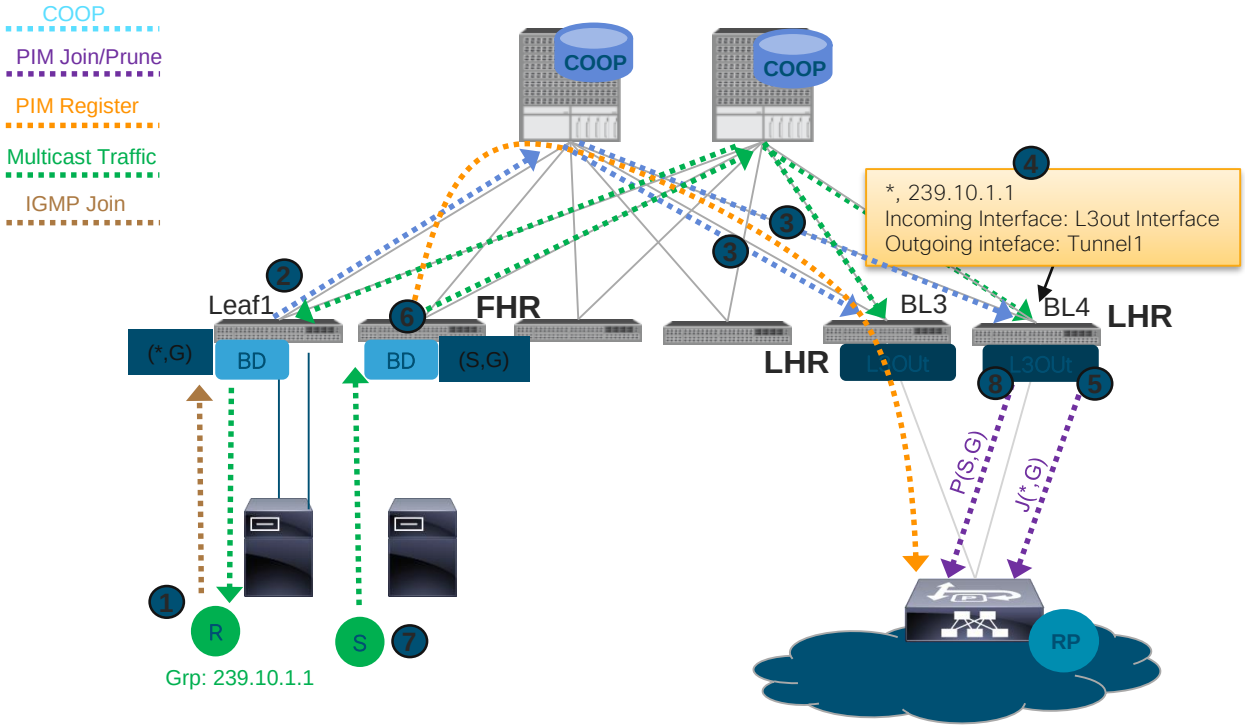
- Multicast enabled VRFs can have both PIM enabled BDs and PIM disabled BDs
- Multicast traffic for BDs with PIM disabled will be forwarded with BD GIPo
- Multicast traffic will be bridged in the BD (TTL unchanged)
- PIM disabled BDs used for L2 multicast should be configured with an IGMP snooping querier, use an external querier, or have IGMP snooping disabled.

L2 Multicast with PIM enabled VRF, PIM enabled BD



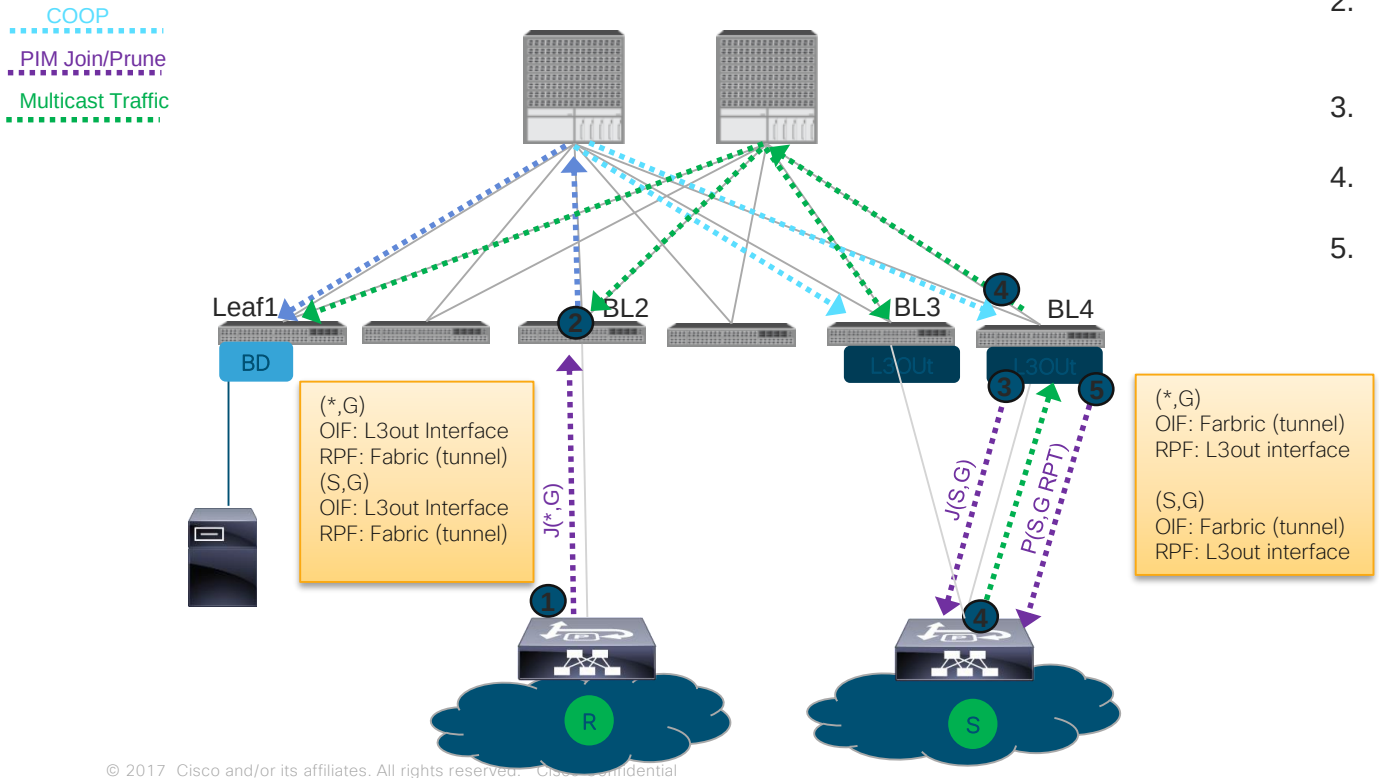
- When PIM is enabled for a BD L2 multicast traffic between leaves will be forwarded in VRF GIPo
- All multicast traffic (L2 and L3) forwarded in a multicast enabled BD will be routed across the fabric (**TTL will be decremented by 2**)
- A reachable RP must be defined for the VRF.

ASM L3 Multicast Source and Receiver inside fabric



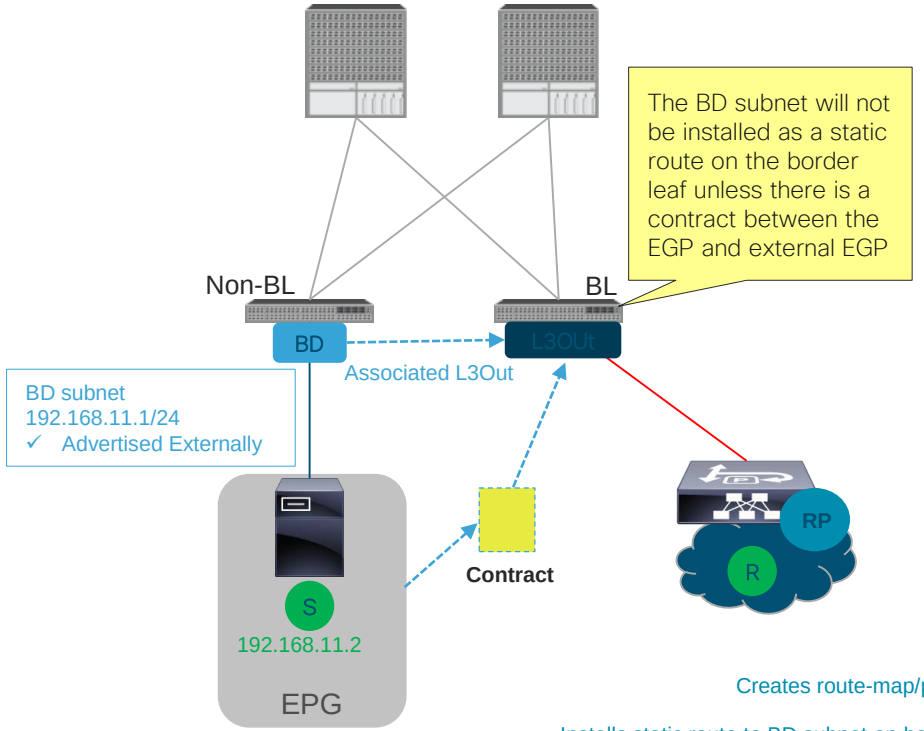
1. Receiver sends IGMP Join report
2. Leaf 1 installs (*,G) in MRIB and publishes group interest to COOP
3. BL3 and BL4 receive group interest from COOP
4. BL4 is stripe winner. Will install (*,G) in MRIB
5. BL4 sends (*,G) PIM join towards RP
6. Source on Leaf2 starts sending multicast traffic. Leaf 2 (FHR) sends register towards RP
7. Multicast traffic is sent to all leaves in VRF.
8. BL4 performs LHR function for fabric and sends (S,G) prune towards RP

ASM Transit multicast



1. BL2 receives PIM (*,G) join for group
2. BL2 publishes join into COOP. All leaves receive group interest from COOP
3. BL4 is the stripe winner and sends PIM (*,G) join to RP
4. RP forwards traffic to BL4 on the (*,G) tree
5. BL 4 receives traffic and sends a prune message towards the RP for the (*,G) tree

Contract considerations for multicast



- ACI security policies (Contracts) are not enforced for multicast traffic. Multicast traffic between different EPGs will be permitted without a contract
- A contract is required for one specific use case
  - Source and Receiver inside fabric (contract not needed)
  - Receiver inside fabric, External source (contract not needed)
  - **Source inside fabric, External receiver (contract required)***
- The contact association is used for policy enforcement but is also used for installing static routes to BD subnets that are remote from the leaf where the contract is configured
- In the case of external multicast receivers the source subnet must be advertised outside of the fabric so that the external network can join the source tree.
- The following configuration is required to allow a BD subnet to be advertised out of an L3Out

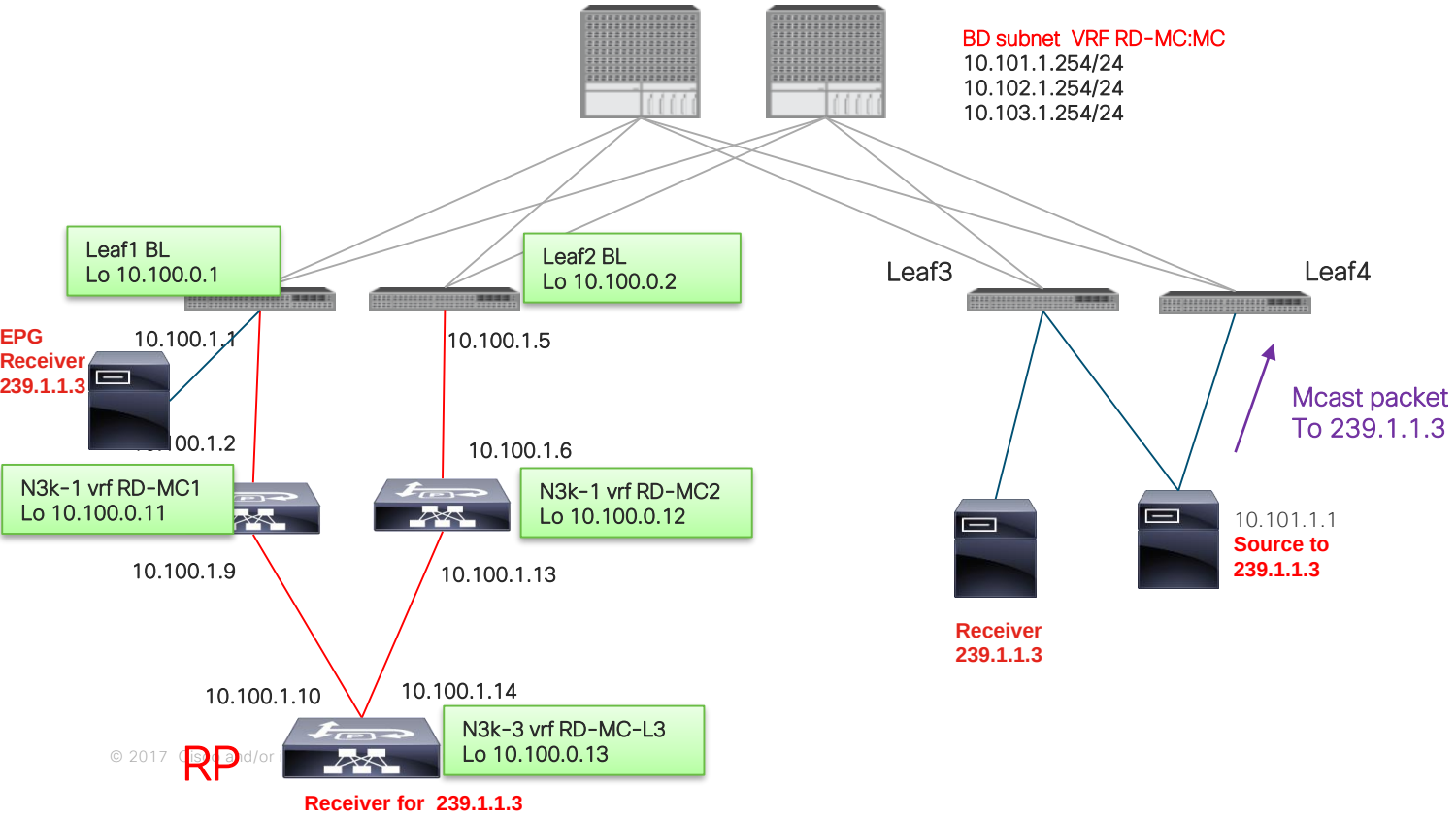
Creates route-map/prefix-list
- ✓ Mark subnet as Advertised Externally
- ✓ Associate BD with L3out

Installs static route to BD subnet on border leaf
- ✓ Create a contract between EPG and L3Out external EPG

*The contract is not required if the BD is deployed on the border leaf

Data plane check

Lab topology



Mroute

BL stripe winner + local receiver

```
bdsol-aci32-leaf1# show ip mroute 239.1.1.3  vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.3/32), uptime: 00:03:03, ngmvpn ip pim igmp
  Incoming interface: Ethernet1/8.13, RPF nbr: 10.100.1.2
  Outgoing interface list: (count: 2) (Fabric OIF)
    Vlan29, uptime: 00:02:25, igmp
    Tunnel16, uptime: 00:03:03, ngmvpn

(10.101.1.1/32, 239.1.1.3/32), uptime: 1d22h, pim ip ngmvpn mrib
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 3) (Fabric OIF)
    Vlan29, uptime: 00:02:25, mrib
    Ethernet1/8.13, uptime: 00:23:24, pim
    Tunnel16, uptime: 1d22h, ngmvpn, mrib, (RPF)
```

Vlan29 → local Rcv in EPG/BD
Eth 1/8 → l3 out Interface to RP
Tunnel 16 – Fabric Interface (inc IF for (s,g))

Server leaf with receiver

```
bdsol-aci32-leaf3# show ip mroute 239.1.1.3 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 239.1.1.3/32), uptime: 00:04:42, igmp ip pim
  Incoming interface: Tunnel22, RPF nbr: 10.0.88.95
  Outgoing interface list: (count: 1)
    Vlan62, uptime: 00:04:42, igmp
```

Leaf where src is connected Should send to Fab tunnel

```
bdsol-aci32-leaf4#  show ip mroute 239.1.1.3 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(10.101.1.1/32, 239.1.1.3/32), uptime: 1d22h, ip pim
  Incoming interface: Tunnel21, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 0)
```

Show ip Mroute det vrf XXX

Server leaf with receiver

```
(*, 239.1.1.3/32), uptime: 00:15:56, igmp(1) ip(0) pim(0)
  Data Created: No
  Stats: 0/0 [Packets/Bytes], 0.000   bps
  Incoming interface: Tunnel22, RPF nbr: 10.0.88.95
  Outgoing interface list: (count: 1)
    Vlan62, uptime: 00:15:56, igmp (vpc-svi)
```

BL stripe winner + local receiver

```
(*, 239.1.1.3/32), uptime: 00:16:24, ngmvpn(1) ip(0) pim(0) igmp(1)
  Data Created: No
  Fabric dont age route
  Stats: 0/0 [Packets/Bytes], 0.000   bps
  Incoming interface: Ethernet1/8.13, RPF nbr: 10.100.1.2
  Outgoing interface list: (count: 2) (Fabric OIF)
    Vlan29, uptime: 00:15:46, igmp (vpc-svi)
    Tunnel16, uptime: 00:16:24, ngmvpn

(10.101.1.1/32, 239.1.1.3/32), uptime: 1d23h, pim(1) ip(0)
ngmvpn(1) mrib(2)
  Data Created: Yes
  Fabric dont age route
  Pervasive
  External PIM Interest
  VPC Flags
    RPF-Source Forwarder
  Stats: 0/0 [Packets/Bytes], 0.000   bps
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 3) (Fabric OIF)
    Vlan29, uptime: 00:15:46, mrib (vpc-svi)
    Ethernet1/8.13, uptime: 00:36:45, pim
    Tunnel16, uptime: 1d22h, ngmvpn, mrib, (RPF)
```

Leaf where src is connected Should send to Fab tunnel

```
(10.101.1.1/32, 239.1.1.3/32), uptime: 1d22h, ip(0) pim(0)
  Data Created: Yes
  Pervasive
  VPC Flags
    RPF-Source Forwarder
  Stats: 2802/142902 [Packets/Bytes], 6.800   bps
  Incoming interface: Tunnel21, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 0)
```

Leaf 4 Detail Check

Ingress leaf 4

MFDM in vsh

```
bdsol-aci32-leaf4# show forwarding distribution
multicast route  vrf RD-MC:RD group 239.1.1.3

  (10.101.1.1/32, 239.1.1.3/32), RPF Interface:
Tunnel21, flags: O
    Received Packets: 2832 Bytes: 184080
    Number of Outgoing Interfaces: 1
    Outgoing Interface List Index: 8200
      Tunnel21
```

MFIB in vsh_lc

```
module-1# show forwarding  multicast route  group 239.1.1.3 vrf RD-MC:RD
  (*, 239.1.1.3/32), RPF Interface: Tunnel21, flags:
    Received Packets: 0 Bytes: 0
    Number of Outgoing Interfaces: 0
    Outgoing Interface List Index: 8191

          Platform Route OIFL Information
            Parent oiflist index: 8191
            Refcount : 1
            Hash : 0x0
            Repl List id : 0xc0000006

          Platform Route RPF Information
            Rpf Type  :0
            Rpf Id    :18010015
            Rpf Index :18010015
            Ref Count :2

  (10.101.1.1/32, 239.1.1.3/32), RPF Interface: Tunnel21, flags: O
    Received Packets: 1369094286720630784 Bytes: 15204152342002794496
    Number of Outgoing Interfaces: 1
    Outgoing Interface List Index: 8200
      Tunnel21 Outgoing Packets:0 Bytes:0

          Platform Route OIFL Information
            Parent oiflist index: 8200
            Refcount : 1
            Hash : 0x96486d70
            Repl List id : 0xc0000007
              Ifindex:18010015
                  Vrf id:9
                  RID:0x0
                  Type:GPIC
                    Repl id: 80000052

          Platform Route RPF Information
            Rpf Type  :0
            Rpf Id    :18010015
            Rpf Index :18010015
            Ref Count :2
```

Note here the Repl List id
Is directly the HAL object

HAL mcast routes

```
module-1(DBG-elam-insel6)# show platform internal hal l3 mcast routes vrf 0x9
=============================================================================================================================================================================================================
                                          | ------------------------------------------------------------ Mcast Replication List  ------------------------------------------------------------
                                          |              | ------------------------------------------------- Mcast Replication Entry -------------------------------------------------
                                          | Mcast    No. | Mc    Rpl                                                                  | ---------------------------------- IFs ----------------------------------
Vrf                                       | Rpl      Rpl | Repl  Ent Vrf           L3           Mc           Alt  Num | L2
Name          Group           Source      | Lst-Id   Ent | Id    Typ Name          IfName       Idx    T Dvif Dvif Ifs | Ifs
=============================================================================================================================================================================================================
RD-MC:RD      239.1.1.3       0.0.0.0       c0000006 -     -     -   -             -            -      - -    -    -    -
RD-MC:RD      239.1.1.3       10.101.1.1    c0000007 1     80000052STI 40000051OvF -    - *  overlay-1    225.1.192.48  F 1ffc  - 0    0    4     Eth1/49(1a030000) Eth1/50(1a031000) Eth1/51(1a0
00) Eth1/52(1a033000)

RD-MC:RD      232.0.0.0       0.0.0.0       c0000005 1     80000052STI 40000051OvF -    - *  overlay-1    225.1.192.48  F 1ffc  - 0    0    4     Eth1/49(1a030000) Eth1/50(1a031000) Eth1/51(1a0
00) Eth1/52(1a033000)
```

Tough hard to read this is the easiest way to see if mroute OIL have been pushed to HAL

Ingress leaf 4 – HAL object

```
module-1# show platform internal hal  objects mcast l3mcastroute groupaddr
239.1.1.3/32 extensions

  OBJECT 1:
Handle                                                  : 78725
groupaddr                                               : 239.1.1.3/32
grpprefixlen                                            : 0x20
sourceaddr                                              : 10.101.1.1/32
ispimbidir                                              : Disabled
ctrlflags                                               : UseMetFlag,
rtflags                                                 : UseMetEntry,
NoDcSupRedirect, ForceRpfPass, UseExplRPFBD,
acirtpolicy                                             : none
aciepgid                                                : 0x0
aciclass                                                : 0x4
aciageinterval                                          : 0x37
minmtu                                                  : 0x5dc
l3iif                                                   : 0x0
rpfbd                                                   : 0x120a
id                                                      : 0x9
Relation Object repllistnextobj :
  rel-repllistnextobj-mcast-mcast_mcast_repl_list-handle  : 78722
  rel-repllistnextobj-mcast-mcast_mcast_repl_list-id      : 0xc0000007
Relation Object mtu :
  rel-mtu-l3-l3_ip_mtu-handle                           : 25188
  rel-mtu-l3-l3_ip_mtu-mtuidx                           : 0x4
```

```
GPD OBJECT MCAST GPDL3VRFSOURCEGROUP

        APD OBJECT MCAST SUGASICPDL3VRFSOURCEGROUP
srcrtidx                                                : 0x603f
issrcrttcam                                             : Disabled
grprtidx                                                : 0x6065
isgrprttcam                                             : Disabled
l2ptridx                                                : 0xb
primarysgidx                                            : id 5000 type 2
shadowsgidx                                             : id -1 type 0
sgoffset                                                : 0x0
overlaysrcgrptblindex                                   : id -1 type 0
sgphysidx                                               : 0x4bc
Executing Custom Apd Private Handler function

                VRF HWId    : 0
               Route Phytype     : 5         HitIndex    : 0x4bc
                   tile-entry-1212-tile-1-subtile-0-fp-id-6

               L2Ptr Phytype     : 5         HitIndex    : 0xa00b
                     tile-entry-11-tile-2-subtile-5-fp-id-9

          AsicId      : 0         SliceVector: 0x3       MetPTR      : 0x5f

          Hit Status: no hit

          sw_mtu_idx : 0x4

          hw_mtu_idx : 0x4
```

Start by hal object mcast l3mcastroute object and follow the Handle chain
Here it has Handle 78725 which rel handle 78722

Next Handle (mcast rep list)

First MET ptr is 95 = 0x5f
92 = 0x5c
The met0 first pointer is what you shall see in ELAM !!!

```
module-1# show platform internal hal  objects mcast mcastrepllist id  0xc0000007 extensions
## Get Extended Objects for mcast mcastrepllist for Asic 0

  OBJECT 0:
Handle                                                  : 78722
id                                                      : 0xc0000007
rsvdmetptr                                              : 0x0
ctrlflags                                               :
Relation Object mcastreplentry :
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle  : 78708
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id      : 0x80000052

    GPD OBJECT MCAST GPDMCASTREPLLIST

        APD OBJECT MCAST ASICPDMCASTREPLLIST
replindex                                               : 0xc0000007
Executing Custom Apd Private Handler function

Repl-List Asicpd Debug :
Entry-Num 0
Repl Entry Id:          0 Hw Epg Id:       0 Hw Bd Id:      0
Mc Id:               8191 Met Id:         95 Encap Id:     -1
Sh Grp:                 0 Next Met Id:    92
Entry-Num 1
Repl Entry Id:   0x40000051 Hw Epg Id:   65533 Hw Bd Id:  65534
Mc Id:               8188 Met Id:         92 Encap Id:   6173
Sh Grp:                11 Next Met Id:     0
```

Now we can follow next handle which is 78708

Next Handle – 2 ways to find the next HAL handle

```
module-1(DBG-elam-insel6)# show platform internal hal objects mcast mcaststitchreplentry handle 78708
No sandboxes exist
## Get Objects for mcast mcaststitchreplentry for Asic 0

  OBJECT 0:
Handle                                                  : 78708
id                                                      : 0x80000052
entrytype                                               : StichReplEntry
ctrlflags                                               :
dirtyflags                                              :
Relation Object ovlystitchreplentrytooverlayrepllist :
  rel-ovlystitchreplentrytooverlayrepllist-mcast-mcast_mcast_repl_list-handle : 78705
  rel-ovlystitchreplentrytooverlayrepllist-mcast-mcast_mcast_repl_list-id : 0x40000051
```

If you know the table

```
module-1(DBG-elam-insel6)# show platform internal hal objects all | egrep -A 25 -B 3 "Handle.*78708"


  OBJECT 28:
Handle                                                  : 78708
id                                                      : 0x80000052
entrytype                                               : StichReplEntry
ctrlflags                                               :
dirtyflags                                              :
Relation Object ovlystitchreplentrytooverlayrepllist :
  rel-ovlystitchreplentrytooverlayrepllist-mcast-mcast_mcast_repl_list-handle : 78705
  rel-ovlystitchreplentrytooverlayrepllist-mcast-mcast_mcast_repl_list-id : 0x40000051
```

If you do not know the table

Next Handle again

```
module-1(DBG-elam-insel6)# show platform internal hal objects all extension | egrep -A 25 -B 5 "Handle.*78705"

  OBJECT 67:
Handle                                                    : 78705
id                                                        : 0x40000051
rsvdmetptr                                                : 0x0
ctrlflags                                                 :
Relation Object mcastreplentry :
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle  : 78704
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id      : 0x40000051
GPD OBJECT MCAST GPDMCASTREPLLIST

        APD OBJECT MCAST ASICPDMCASTREPLLIST
replindex                                                 : 0x40000051
Executing Custom Apd Private Handler function

Repl-List Asicpd Debug :
Entry-Num 0
Repl Entry Id:   0x40000051 Hw Epg Id:   65533 Hw Bd Id:  65534
Mc Id:               8188 Met Id:         92 Encap Id:   6173
Sh Grp:                11 Next Met Id:     0
```

Next Handle ptr to overlay encap

```
module-1(DBG-elam-insel6)# show platform internal hal objects all | egrep -A 25 -B 5 "Handle.*78704"

  OBJECT 21:
Handle                                                  : 78704
isvrfmode                                               : 0x1
vrfid                                                   : 0x9
id                                                      : 0x40000051
entrytype                                               : OvlyFabricReplEntry
ctrlflags                                               : UseMetFlag,
dirtyflags                                              :
Relation Object ovlyfabricreplentrytoportsfanout :
  rel-ovlyfabricreplentrytoportsfanout-mcast-mcast_ports_fanout-handle : 6915
  rel-ovlyfabricreplentrytoportsfanout-mcast-mcast_ports_fanout-mc_idx : 0x1ffc
Relation Object ovlyfabricreplentrytoovlyencapentry :
  rel-ovlyfabricreplentrytoovlyencapentry-mcast-mcast_ovly_encap_entry-handle : 78701
  rel-ovlyfabricreplentrytoovlyencapentry-mcast-mcast_ovly_encap_entry-groupaddr : 225.1.192.48/0
  rel-ovlyfabricreplentrytoovlyencapentry-mcast-mcast_ovly_encap_entry-ovlyencaptype : ivxlan_l3
```

Next handle Port fanout

```
module-1# show platform internal  hal l2 port gpd  | egrep "1a03"
1a030000 Eth1/49      0 1     42     0  41 1  18 30 b0   1   0 0 0 0 0 0  0  0
1a031000 Eth1/50      0 2     44     0  49 1  20 38 b8   1   0 0 0 0 0 0  0  0
1a032000 Eth1/51      0 3     46     0  19 0  18 30 30   1   0 0 0 0 0 0  0  0
1a033000 Eth1/52      0 4     48     0  21 0  20 38 38   1   0 0 0 0 0 0  0  0
module-1#
```

```
  OBJECT 32:
Handle                                                : 6915
mc_idx                                                : 0x1ffc
ismctransitsuported                                   : Enabled
ctrls                                                 :
spltype                                               : none
destvif                                               : 0x0
alternatedestvif                                      : 0x0
Relation Object mcastportsfanout2brportatt :
  rel-mcastportsfanout2brportatt-l2-l2_br_if-handle   : 26639
  rel-mcastportsfanout2brportatt-l2-l2_br_if-id       : 0x1a030000
Relation Object mcastportsfanout2brportatt :
  rel-mcastportsfanout2brportatt-l2-l2_br_if-handle   : 86253
  rel-mcastportsfanout2brportatt-l2-l2_br_if-id       : 0x1a031000
Relation Object mcastportsfanout2brportatt :
  rel-mcastportsfanout2brportatt-l2-l2_br_if-handle   : 28055
  rel-mcastportsfanout2brportatt-l2-l2_br_if-id       : 0x1a032000
Relation Object mcastportsfanout2brportatt :
  rel-mcastportsfanout2brportatt-l2-l2_br_if-handle   : 83796
  rel-mcastportsfanout2brportatt-l2-l2_br_if-id       : 0x1a033000
 GPD OBJECT MCAST GPDMCASTPORTSFANOUT

        APD OBJECT MCAST SUGASICPDMCASTPORTSFANOUT
mc_idx                                                : 0x1ffc
ovmcsg_tblidx                                         : id -1 type 0
rwniv_dstvif_tblidx                                   : id -1 type 0
rwniv_altvif_tblidx                                   : id -1 type 0
Executing Custom Apd Private Handler function
```

Next Handle – Overlay encap

```
  OBJECT 19:
Handle                                                    : 78701
groupaddr                                                 : 225.1.192.48/0
ovlyencaptype                                             : ivxlan_l3
overlayencapmode                                          : mcast_ftag
overlayouterbd                                            : 0x0
ovlyencapentryctrlflags                                   :
id                                                        : 0x4
Relation Object portgroup :
  rel-portgroup-l2-l2_portgroup-handle                    : 6814
  rel-portgroup-l2-l2_portgroup-id                        : 0x2
```

ELAM what to check

- Ovector is irrelevant (will be 0x0) as elam comes before MET replica
- Ftag is usefull to know which one we will follow in fabric
- Met pointer can be correlated with previous HAL info
- Make sure to check TTL (iperf for example send mcast with ttl 1 by default

```
module-1(DBG-elam-insel6)# report | egrep
"pad.ftag|met|ip.ttl:|ip.da"
    sug_pr_lu_vec_l3v.ip.ttl: 0xFF
    sug_pr_lu_vec_l3v.ip.da: 0x00000000000000000EF010103
  sug_lurw_vec.ol_fb_metric: 0x0
      sug_lu2ba_sb_info.mc_info.mc_info_nopad.ftag: 0x9
      sug_lu2ba_sb_info.mc_info.mc_info_nopad.met0_v: 0x0
      sug_lu2ba_sb_info.mc_info.mc_info_nopad.met0_idx: 0x5f
      sug_lu2ba_sb_info.mc_info.mc_info_nopad.met1_v: 0x0
      sug_lu2ba_sb_info.mc_info.mc_info_nopad.met1_idx: 0x0
```

Ftag 0x9 – VRF GIPo being 225.1.192.48
We will send to 225.1.192.57

Leaf 1 Detail Check

Ingress leaf 4

MFDM in vsh

```
bdsol-aci32-leaf1# show forwarding distribution
multicast route vrf RD-MC:RD group 239.1.1.3 source
10.103.1.1

  (10.103.1.1/32, 239.1.1.3/32), RPF Interface:
Tunnel16, flags:
    Received Packets: 3 Bytes: 195
    Number of Outgoing Interfaces: 3
    Outgoing Interface List Index: 8214
      Vlan29
      Tunnel16
      Ethernet1/8.13
```

Note here the Repl List id
Is directly the HAL object

MFIB in vsh_lc

```
module-1# show forwarding  multicast route  group 239.1.1.3 vrf RD-MC:RD

.103.1.1/32, 239.1.1.3/32), RPF Interface: Tunnel16, flags:
 Received Packets: 216172782113783808 Bytes: 14051230837395947520
 Number of Outgoing Interfaces: 3
 Outgoing Interface List Index: 8214
   Vlan29 Outgoing Packets:0 Bytes:0
   Tunnel16 Outgoing Packets:0 Bytes:0
   Ethernet1/8.13 Outgoing Packets:0 Bytes:0

        Platform Route OIFL Information
          Parent oiflist index: 8214
          Refcount : 2
          Hash : 0x9cf56a39
          Repl List id : 0xc0000004
            Ifindex:901001d
                Vrf id:6
                RID:0x3
                Type:L2
                Num Repl: 1
                Repl id:
                   0x60
            Ifindex:18010010
                Vrf id:6
                RID:0x0
                Type:GPIC
                  Repl id: 80000057
            Ifindex:1a00700d
                Vrf id:6
                RID:0x0
                Type:L3
                MCIDX: 1f81
                  Repl id: 60000000

        Platform Route RPF Information
          Rpf Type  :0
          Rpf Id    :18010010
          Rpf Index :18010010
          Ref Count :2
```

Leaf1 HAL routes

- We see 3 copy, one to local EPG receiver (1/11), one to L3 out (1/8) and one to Stitching overlay (here it is rpf)

```
module-1# show platform internal  hal l3  mcast routes  vrf 0x6


RD-MC:RD      239.1.1.3       10.101.1.1        c0000004 3  60        EPG -    -   1e   0 -  -            -            - 4     - 0    800  1     Eth1/11(1a00a000)
                                                           60000000  L3  RD-MC:RD      Eth1/8.13    1f81  - 0    0    1     Eth1/8(1a007000)
                                                           80000057  STI 40000056OvF -    - *  overlay-1     225.1.192.48  F 1ffc  - 0    0    4
Eth1/49(1a030000) Eth1/50(1a031000) Eth1/51(1a032000) Eth1/52(1a033000)
```

Egress leaf 1 – HAL object

```
module-1# show platform internal   hal objects mcast l3mcastroute
groupaddr 239.1.1.3/32 sourceaddr 10.103.1.1/32 extensions
## Get Extended Objects for mcast l3mcastroute for Asic 0

  OBJECT 0:
Handle                                                    : 219170
groupaddr                                                 : 239.1.1.3/32
grpprefixlen                                              : 0x20
sourceaddr                                                : 10.103.1.1/32
ispimbidir                                                : Disabled
ctrlflags                                                 : UseMetFlag,
rtflags                                                   : UseMetEntry,
NoDcSupRedirect, ForceRpfPass, RpfFailureSupRedirect, UseExplRPFBD,
acirtpolicy                                               : none
aciepgid                                                  : 0x0
aciclass                                                  : 0x4
aciageinterval                                            : 0x37
minmtu                                                    : 0x5dc
l3iif                                                     : 0x0
rpfbd                                                     : 0x1203
id                                                        : 0x6
Relation Object repllistnextobj :
  rel-repllistnextobj-mcast-mcast_mcast_repl_list-handle  : 219164
  rel-repllistnextobj-mcast-mcast_mcast_repl_list-id      : 0xc0000009
Relation Object mtu :
  rel-mtu-l3-l3_ip_mtu-handle                             : 44407
  rel-mtu-l3-l3_ip_mtu-mtuidx                             : 0x6
```

```
GPD OBJECT MCAST GPDL3VRFSOURCEGROUP

        APD OBJECT MCAST SUGASICPDL3VRFSOURCEGROUP
srcrtidx                                                  : 0xffffffff
issrcrttcam                                               : Disabled
grprtidx                                                  : 0x6035
isgrprttcam                                               : Disabled
l2ptridx                                                  : 0x0
primarysgidx                                              : id -1 type 0
shadowsgidx                                               : id -1 type 0
sgoffset                                                  : 0x0
overlaysrcgrptblindex                                     : id -1 type 0
sgphysidx                                                 : 0x0
Executing Custom Apd Private Handler function

                VRF HWId    : 0

          AsicId      : 0          SliceVector: 0x3       MetPTR      : 0x85

          Hit Status: no hit

          sw_mtu_idx : 0x6

          hw_mtu_idx : 0x8
```

Start by hal object mcast l3mcastroute object and follow the Handle chain
Here it has Handle 219170 which rel handle 219164

Next Handle (mcast rep list)

First MET ptr is 133 = 0x85
The met0 first pointer is what you shall see in ELAM !!!

```
module-1# show platform internal   hal objects all extensions | egrep -B 4 -A 50 "^Han

  OBJECT 37:
Handle                                                  : 219164
id                                                      : 0xc0000009
rsvdmetptr                                              : 0x0
ctrlflags                                               :
Relation Object mcastreplentry :
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle  : 219163
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id      : 0x60
Relation Object mcastreplentry :
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle  : 219162
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id      : 0x60000000
Relation Object mcastreplentry :
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle  : 81300
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id      : 0x80000057

    GPD OBJECT MCAST GPDMCASTREPLLIST

        APD OBJECT MCAST ASICPDMCASTREPLLIST
replindex                                               : 0xc0000009
Executing Custom Apd Private Handler function

Repl-List Asicpd Debug :
Entry-Num 0
Repl Entry Id:       0x60 Hw Epg Id:   11285 Hw Bd Id:    517
Mc Id:                  4 Met Id:        133 Encap Id:     -1
Sh Grp:                 0 Next Met Id:   134
Entry-Num 1
Repl Entry Id:   0x60000000 Hw Epg Id:   11301 Hw Bd Id:   4639
Mc Id:               8065 Met Id:        134 Encap Id:     -1
Sh Grp:                 0 Next Met Id:    90
Entry-Num 2
Repl Entry Id:   0x40000056 Hw Epg Id:   65533 Hw Bd Id:  65534
Mc Id:               8188 Met Id:         90 Encap Id:   6174
Sh Grp:                11 Next Met Id:     0
```

Now we can follow next handle which is 219163, 219162 and 8130

Next Handle – 1st MET entry (HAL chain

```
module-1# show platform internal   hal objects all extensions | egrep -B 4 -A 50 "^Handle.*219163"
  OBJECT 3:
Handle                                                  : 219163
epgval                                                  : 0x1e
mtu_vld                                                 : 0x0
mtu                                                     : 0x0
ovlyencapvalid                                          : Disabled
id                                                      : 0x60
entrytype                                               : EPGReplEntry
ctrlflags                                               : UseMetFlag,
dirtyflags                                              :
Relation Object epgreplentrytoportsfanout :
  rel-epgreplentrytoportsfanout-mcast-mcast_ports_fanout-handle : 79558
  rel-epgreplentrytoportsfanout-mcast-mcast_ports_fanout-mc_idx : 0x4
```

1st rel from the
Mcast rep list
And the epg id 0x1e
Which translate to access encap
Vlan 1006

```
module-1# show platform internal   hal objects all extensions | egrep -B 4 -A 50 "^Handle.*79558"
          Slice-Id    : 1          IfMap[63..0]    : 0


  OBJECT 30:
Handle                                                  : 79558
mc_idx                                                  : 0x4
ismctransitsuported                                     : Disabled
ctrls                                                   :
spltype                                                 : none
destvif                                                 : 0x0
alternatedestvif                                        : 0x800
Relation Object mcastportsfanout2brportatt :
  rel-mcastportsfanout2brportatt-l2-l2_br_if-handle     : 76942
  rel-mcastportsfanout2brportatt-l2-l2_br_if-id         : 0x1a00a000
```

Gives us an l2 ptr to 1/11

```
module-1# show platform internal hal l2  port gpd port_id 0x1a00a000 | egrep 1a00
1a00a000 Eth1/11      0 6e   52    0  b  0  a  14 14   1   0 0 0 0 0 0  0  0 0 0   0 0   0   0  0  D-260  -      0    0      1 0   c    0
module-1#
module-1# show platform internal hal l2  fd pi 0x1e | egrep 1e
1e    FD-30     V  2c15  1d   BD-29  V  3ee    IV 2069   d     0       1 1   0 1 0 0   0 0 1 0
module-1# dec 0x3ee
1006
```

Next Handle – 2nd MET entry (HAL chain)

```
module-1# show platform internal   hal objects all extensions | egrep -B 4 -A 50 "^Handle.*219162"
 OBJECT 3:
Handle                                                  : 219162
vrf                                                     : 0x6
l3ifindex                                               : 0x1a00700d
l3hwbdid                                                : 0x121f
l3hwepgid                                               : 0x2c25
isfabricintf                                            : Disabled
isipn                                                   : Disabled
isdci                                                   : Disabled
tunnelifindex                                           : 0x0
id                                                      : 0x60000000
entrytype                                               : L3ReplEntry
ctrlflags                                               : UseMetFlag,
dirtyflags                                              :
Relation Object l3replentrytoportsfanout :
  rel-l3replentrytoportsfanout-mcast-mcast_ports_fanout-handle : 220971
  rel-l3replentrytoportsfanout-mcast-mcast_ports_fanout-mc_idx : 0x1f81
```

2nd rel from the
Mcast rep list gives l3 out subif

```
module-1# show platform internal   hal objects all extensions | egrep -B 4 -A 50 "^Handle.*220971"

  OBJECT 16:
Handle                                                  : 220971
mc_idx                                                  : 0x1f81
ismctransitsuported                                     : Disabled
ctrls                                                   :
spltype                                                 : none
destvif                                                 : 0x0
alternatedestvif                                        : 0x0
Relation Object mcastportsfanout2brportatt :
  rel-mcastportsfanout2brportatt-l2-l2_br_if-handle     : 197537
  rel-mcastportsfanout2brportatt-l2-l2_br_if-id         : 0x1a007000
```

```
module-1# show platform internal hal l2  port gpd port_id 0x1a007000 | egrep 1a00
1a007000 Eth1/8      0 7b   54    0  10 0  f  1e 1e   1   0 0 0 0 0 0  0  0 0 0   1 f   0   2  0  D-2ad  -      0    0      1 0   22   0
module-1#
module-1# show platform internal hal l3 subif gpd  | egrep 1a00700d
1a00700d Eth1/8.13    6      RD-MC:RD      200 200 D-7f7  D-363  -      -      -      -      -      -      -      -      -       0 0 0
```

Next Handle – 3rd MET entry (HAL chain)

```
module-1# show platform internal   hal objects all extensions | egrep -B 2 -A 50 "^Handle.*81300"
  OBJECT 22:
Handle                                                  : 81300
id                                                      : 0x80000057
entrytype                                               : StichReplEntry
ctrlflags                                               :
dirtyflags                                              :
Relation Object ovlystitchreplentrytooverlayrepllist :
  rel-ovlystitchreplentrytooverlayrepllist-mcast-mcast_mcast_repl_list-handle : 81297
  rel-ovlystitchreplentrytooverlayrepllist-mcast-mcast_mcast_repl_list-id : 0x40000056
```

3rd rel from mcast repl list
Gives the overlay if like on leaf 1

```
module-1# show platform internal   hal objects all extensions | egrep -B 2 -A 50 "^Handle.*81297"
  OBJECT 23:
Handle                                                  : 81297
id                                                      : 0x40000056
rsvdmetptr                                              : 0x0
ctrlflags                                               :
Relation Object mcastreplentry :
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-handle  : 81296
  rel-mcastreplentry-mcast-mcast_mcast_repl_entry-id      : 0x40000056
```

```
module-1# show platform internal   hal objects all extensions | egrep -B 2 -A 50 "^Handle.*81296"

  OBJECT 19:
Handle                                                  : 81296
isvrfmode                                               : 0x1
vrfid                                                   : 0x6
id                                                      : 0x40000056
entrytype                                               : OvlyFabricReplEntry
ctrlflags                                               : UseMetFlag,
dirtyflags                                              :
Relation Object ovlyfabricreplentrytoportsfanout :
  rel-ovlyfabricreplentrytoportsfanout-mcast-mcast_ports_fanout-handle : 8017
  rel-ovlyfabricreplentrytoportsfanout-mcast-mcast_ports_fanout-mc_idx : 0x1ffc
Relation Object ovlyfabricreplentrytoovlyencapentry :
  rel-ovlyfabricreplentrytoovlyencapentry-mcast-mcast_ovly_encap_entry-handle : 81293
  rel-ovlyfabricreplentrytoovlyencapentry-mcast-mcast_ovly_encap_entry-groupaddr : 225.1.192.48/0
  rel-ovlyfabricreplentrytoovlyencapentry-mcast-mcast_ovly_encap_entry-ovlyencaptype : ivxlan_l3
```

Would give list of Fab upling