ACI Multicast

Roland Ducomble - CCIE 3745
EMEAR ACI solution Tac Team – Technical Leader
10th Dec 2019

V2.2

ACI – Tenant L3 Multicast

© 2015 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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)

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
- L3 multicast over multisite – Added in 4.0

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

L3 Multicast Forwarding Overview

BD1: PIM enabled
BD2: PIM enabled
BD3: PIM disabled

- 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

© 2017 Cisco and/or its affilia fidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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)

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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.

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Lab topology

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

© 2017 Cisco and/or i

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 !

System | Tenants | Fabric | VM Networking | L4-L7 Services | Admin | Operations | Apps

ALL TENANTS | Add Tenant | Search: enter name, alias, descr | common | RD-MC | RD | DC | ATenant01

Tenant RD-MC

- Quick Start
- Tenant RD-MC
  - Application Profiles
    - APP
      - Application EPGs
        - EPG1
        - EPG2
        - EPG3
      - uSeg EPGs
      - L4-L7 Service Parameters
  - Networking
    - Bridge Domains
      - BD1
      - BD2
      - BD3
    - VRFs
      - RD
        - Deployed VRFs (Simple Mode)
        - Multicast
        - EPG Collection for VRF
    - External Bridged Networks
    - External Routed Networks
      - Route Maps/Profiles

Multicast

Configuration | Stats

Interfaces | Rendezvous Points | Pattern Policy | PIM Setting

Enable

Bridge Domains

BD	IGMP Policy
RD-MC/BD3	
RD-MC/BD2	
RD-MC/BD1	

Page 1 Of 1 | 10

Interfaces

L3 Out	Interface Group	Interface	IGMP Policy	PIM Policy
MC				
	ospf-if	pod-1/101/[eth1/8] pod-1/102/[eth1/8]		

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

3. Specify a RP

Note ACI fabric can’t be RP so you need an external RP (before 4.0)

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco sco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

IP L3 multicast in ACI
Control plane

Receiver in ACI fabric

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. BL1 sends PIM Join towards RP
6. BL1 receives multicast traffic from external network and sends it to tunnel if
7. BL1 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 ——→

© 2017 Cisco and/or i

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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#
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

PIM Join

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Enabling PIM fast switchover

In VRF/Multicast/PIM setting

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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"
```

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
```

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

```
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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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#
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

PIM Join

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

reserved. Cisco Confidential

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)
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Source in ACI fabric
Receiver outside

Lab topology

© 2017 Cisco and/or

Lab topology

© 2017 Cisco and/or

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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)
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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

© 2017 Cisco and/or

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)
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

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.

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

L2 Multicast with PIM enabled VRF, PIM enabled BD

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confiden

- 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.

PIM RP in the fabric
4.0 feature

RP Behavior on the Fabric (Anycast RP)

- All border leafs are configured to function as the RP projecting the entire fabric as a RP. RP address is a property of a PIM interface on a switch.
- In ACI, upon application of the Fabric RP configuration, a loopback interface will be created in each of the border leafs whose IP address will be configured as the Fabric RP address. This loopback address will be the same across all BLs.
- The RP address will be an anycast address on all BLs and the behavior of Fabric RP is the same as PIM anycast RP with all the BLs in the fabric as the members of the list.
- On NBL, only the RP configuration will be applied. The Loopback interface is not created on NBLs.
- An explicit route policy on the L3out should be applied to advertise this RP address outside of the fabric.

How the anycast RP member list is built:

- When fabric RP is enabled, PIM internally builds an RP list using its neighbor list. This will be the anycast RP list.
- As a result of this, since all the BLs will advertise the hellos, all the BLs will become a part of the anycast RP member list.
- Once the source starts sending traffic, a PIM data register will be sent to the RP address by the router connected to the source.
- By virtue of anycast, this register will be received by one of the BLs. This BL will use the anycast RP member list to send and forward the source register to the members of this list. Hence the source information will be learnt on all the BLs.

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Restrictions

- With this solution, only the ACI fabric can behave as an RP for a given group-range. If there are external RPs for the same group-range, it will not be supported. This means that ACI fabric cannot behave as an anycast RP along with external RPs.
- Fast convergence will not be supported when RP is inside the fabric since all the BLs are RPs and have the capability to reach the source and the source will be registered with all the BLs.

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Lab topology

© 2017 Cisco and/or

Fabric RP config

Multicast

Configuration | Stats | Fa

Interfaces | Rendezvous Points | Pattern Policy | PIM Setting | IGMP Setting | Inter

Static RP

IP	RouteMap
10.100.0.13	239Range

Fabric RP

IP	RouteMap
10.100.0.9	238Range

Auto-RP

RP Updates: ☐ Forward Auto-RP Updates
☐ Listen to Auto-RP Updates

MA Filter: select an option

Bootstrap Router (BSR)

RP Updates: ☐ Forward BSR Updates
☐ Listen to BSR Updates

BSR Filter: select an option

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

RP should be advertised on L3 out

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Checking Fabric RP on each Leaf (here non BL)

```
bdsol-aci32-leaf3# show ip pim rp 238.0.0.0 vrf RD-MC:RD

PIM RP Status Information for VRF:"RD-MC:RD"
PIM RP Information for group 238.0.0.0 in VRF "RD-MC:RD"
BSR: Not Operational
Auto-RP: Not Operational
RP: 10.100.0.9, uptime: 00:05:45, expires: never, FabricRP
  priority: 0, RP-source:  (local), group-map: mcast_rprange_RD-MC:RD_10.100.0.9, group ranges:
    238.0.0.0/8
Fabric RP members: 10.100.0.2 10.100.0.1

bdsol-aci32-leaf3# show ip pim neighbor vrf RD-MC:RD

PIM Neighbor information for Dom:RD-MC:RD
Neighbor            Interface           Uptime              Expires             DRPriority Bidir
BFDState
10.100.0.2/32       tunnel19            25d22h              00:01:25            1          no         n/a
10.100.0.1/32       tunnel19            25d22h              00:01:44            1          no         n/a
```

See each leaf sees who are the Fabric RP member
Should be all BL.
Fabric RP members is build of PIM neighbor on tunnel interface

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Checking loopback on BL

- Once fabric RP is configured, each BL should have the fabric RP as anycast loopback and as PIM interface

```
bdsol-aci32-leaf2# show ip interface vrf RD-MC:RD
IP Interface Status for VRF "RD-MC:RD"(5)
loopback2, Interface status: protocol-up/link-up/admin-up, iod: 103, mode: unspecified, vrf_vnid: 3112960
  IP address: 10.100.0.2, IP subnet: 10.100.0.2/32
  IP primary address route-preference: 1, tag: 0
loopback5, Interface status: protocol-up/link-up/admin-up, iod: 121, mode: unspecified, vrf_vnid: 3112960
  IP address: 10.100.0.9, IP subnet: 10.100.0.9/32
  IP primary address route-preference: 1, tag: 0
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     Interface
Tunnel16             10.100.0.2      10.100.0.2      1         no
loopback2            10.100.0.2      10.100.0.2      0         no
Ethernet1/8.41       10.100.1.5      10.100.1.6      1         no
Ethernet1/47.40      10.100.1.17     10.100.1.18     1         no
loopback5            10.100.0.9      10.100.0.9      0         no
```

Source Inside Receiver Inside (SIRI)

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Debugging States and Traffic Flows: Source and receivers inside the fabric - SIRI

1. IGMP join received on NBL-1 will be installed to MRIB in BL-4 because BL-4 is the stripe winner.
2. Since fabric is RP, RPF on BL-4 will point to itself which is the loopback interface.
3. Once Source become active, (S,G) will be created on NBL-2 by virtue of FHR functionality. States will be the same as in 3.x.
4. NBL-2 also builds a PIM data register packet and sends it to the anycast RP address.
5. Since fabric is RP, we will use unicast Routing (likely ECMP) to reach RP. Here we assume it reaches BL-1, BL-1 will receive the register packet and install the (S,G).
6. Since BL-1 is an anycast RP, it will forward the register to all the anycast RP members and S,G will be created in each BL.
7. Only the BL who is the stripe-winner for the group will inherit the OIF in MRIB. (aka forward it to tunnel

NBL-1	NBL-2	BL-1	BL-2	BL-3	BL-4 (SW)
(*,G) RPF: Tunnel OIF: Vlan10 (igmp)					(*,G) RPF: Loopback OIF: Tunnel (ngmvpn)
	(S,G) data-created RPF: Tunnel OIF: Null (traffic is put on VRF GIPO)	(S,G) reg RPF: Tunnel OIF: Null	(S,G) reg RPF: Tunnel OIF: Null	(S,G) reg RPF: Tunnel OIF: Null	(S,G) reg RPF: Tunnel OIF: Tunnel (mrib)

Lab topology

© 2017 Cisco and/or

Receiver joined – (no src yet)

Leaf 1- BL stripe winner

```
bdsol-aci32-leaf1#  show ip mroute 238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 238.1.1.1/32), uptime: 00:07:18, ngmvpn ip pim
  Incoming interface: loopback6, RPF nbr: 10.100.0.9
  Outgoing interface list: (count: 1) (Fabric OIF)
    Tunnel16, uptime: 00:07:18, ngmvpn
```

Leaf 2- BL stripe loser

```
bdsol-aci32-leaf2#  show ip mroute 238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

Group not found
```

Leaf 3 - non BL local receiver

```
bdsol-aci32-leaf3# show ip mroute 238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 238.1.1.1/32), uptime: 00:07:41, igmp ip pim
  Incoming interface: Tunnel19, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 1)
    Vlan83, uptime: 00:07:41, igmp
bdsol-aci32-leaf3# show ip igmp  snooping groups 238.1.1.1
Type: S - Static, D - Dynamic, R - Router port, F -
Fabricpath core port

Vlan  Group Address      Ver  Type  Port list
83    238.1.1.1          v2   D     Eth1/9
```

Leaf 4- non-BL - no info

```
bdsol-aci32-leaf4#  show ip mroute 238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

Group not found
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Src is starting on leaf 4 – Flow of register

Leaf 1- receive Register, send register to other BL
And send reg stop

```
11:31:30.452009 IP (tos 0xc0, ttl 32, id 34164, offset 0, flags [none], proto PIM (103),
length 1526)
    10.101.1.254 > 10.100.0.9: PIMv2, length 1506
        Register, cksum 0xdeff (incorrect), Flags [ none ]
        IP (tos 0x0, ttl 32, id 65194, offset 512, flags [none], proto UDP (17), length
1498)
    10.101.1.23 > 238.1.1.1: udp
11:40:04.616583 IP (tos 0xc0, ttl 33, id 36332, offset 0, flags [none], proto PIM (103),
length 1526)
    10.100.0.1 > 10.100.0.2: PIMv2, length 1506
        Register, cksum 0xdeff (incorrect), Flags [ none ]
        IP (tos 0x0, ttl 32, id 35849, offset 512, flags [none], proto UDP (17), length
1498)
    10.101.1.23 > 238.1.1.1: udp

11:31:30.456224 IP (tos 0xc0, ttl 33, id 46454, offset 0, flags [none], proto PIM (103),
length 38)
    10.100.0.9 > 10.101.1.254: PIMv2, length 18
        Register Stop, cksum 0xe160 (correct) group=238.1.1.1 source=10.101.1.23
```

1

2

3

Leaf 3 – No register received

```
bdsol-aci32-leaf3# tcpdump -xxvvi kpm_inb host 10.100.0.9
tcpdump: listening on kpm_inb, link-type EN10MB (Ethernet),
capture size 65535 bytes
```

Leaf 4- non-BL – Send pim register to the anycast addres

```
bdsol-aci32-leaf4# tcpdump -vvni  kpm_inb host 10.100.0.9
tcpdump: listening on kpm_inb, link-type EN10MB (Ethernet), capture
size 65535 bytes
11:31:30.452039 IP (tos 0xc0, ttl 33, id 34164, offset 0, flags
[none], proto PIM (103), length 1526)
    10.101.1.254 > 10.100.0.9: PIMv2, length 1506
        Register, cksum 0xdeff (incorrect), Flags [ none ]
        IP (tos 0x0, ttl 32, id 65194, offset 512, flags [none],
proto UDP (17), length 1498)
    10.101.1.23 > 238.1.1.1: udp
```

1

Leaf 2- Receiver register from BL-1
(the one that got the original register

```
11:39:09.326510 IP (tos 0xc0, ttl 254, id 33668, offset 0, flags
[none], proto PIM (103), length 48)
    10.100.0.1 > 10.100.0.2: PIMv2, length 28
        Register, cksum 0x9eff (correct), Flags [ Null ]
        IP (tos 0x0, ttl 255, id 0, offset 0, flags [none], proto
PIM (103), length 20)
    10.101.1.23 > 238.1.1.1:
```

2

SIRI – final mroute

Leaf 1- BL stripe winner
(*,g) to tunnel because of Inside Rcv and stripe winner
(S,g) because inside src and stripe winner

```
(*, 238.1.1.1/32), uptime: 01:16:10, ngmvpn ip pim
  Incoming interface: loopback6, RPF nbr: 10.100.0.9
  Outgoing interface list: (count: 1) (Fabric OIF)
    Tunnel16, uptime: 01:16:10, ngmvpn

(10.101.1.23/32, 238.1.1.1/32), uptime: 00:28:39, pim mrib ip
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 1)
    Tunnel16, uptime: 00:28:39, mrib, (RPF)
```

Leaf 2- BL stripe loser

```
(10.101.1.23/32, 238.1.1.1/32), uptime: 00:29:29, pim ip
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65
(pervasive)
  Outgoing interface list: (count: 0)
```

Leaf 3 - non BL local receiver
Only (*,G)

```
(*, 238.1.1.1/32), uptime: 01:17:33, igmp ip pim
  Incoming interface: Tunnel19, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 1)
    Vlan83, uptime: 01:17:33, igmp
```

Leaf 4- non-BL – only (s,g)
Note though OIL if emptry, forwarding does really
Send it to tunnel interface as expected

```
(10.101.1.23/32, 238.1.1.1/32), uptime: 00:07:35, ip pim
  Incoming interface: Tunnel26, RPF nbr: 10.0.8.65
(pervasive)
  Outgoing interface list: (count: 0)

bdsol-aci32-leaf4# show forwarding distribution multicast
route vrf RD-MC:RD group 238.1.1.1

  (10.101.1.23/32, 238.1.1.1/32), RPF Interface: Tunnel26,
flags: O
    Received Packets: 1 Bytes: 65
    Number of Outgoing Interfaces: 1
    Outgoing Interface List Index: 8205
      Tunnel26
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Source Inside Receiver Outside (SIRO)

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Debugging traffic Flows: Source inside and receivers outside the fabric: SIRO

NBL-1	NBL-2	BL-1	BL-2	BL-3	BL-4 (SW)
			*,G RPF:Loopback OIF: L3out (pim – Ext join)		(*,G) RPF: Loopback OIF: Tunnel (ngmvpn)
	(S,G) data-created RPF: Tunnel OIF: Null (traffic is put on VRF GIPO)	(S,G) reg RPF: Tunnel OIF: Null	(S,G) reg RPF: Tunnel OIF: L3out (pim – Ext join)	(S,G) reg RPF: Tunnel OIF: Null	(S,G) reg RPF: Tunnel OIF: Tunnel (mrib)

1. *,G join from the external receiver is directed towards the anycast RP. The join can land on any BL.
   Assuming that the *,G join lands on BL-2, a (*,G) state will be created and synced via COOP to all BLs.
1. Upon stripe-winning, BL-4 is be chosen as forwarder and hence BL-4 will install to MRIB.
2. Once source becomes active, NBL-2 will originate a source-register packet towards the anycast RP address.
3. Since the register is received on an anycast RP address, the register will be forwarded to all the members in the anycast RP list and
   hence all the BLs will receive the source-register.
4. S,G will created on all the RP BLs.
5. On BL-2 which received the external PIM join, OIF will be added by PIM by virtue of register receipt and
   also because (S,G) join that is sent from receiver.
6. On BL-4 (S,G) state will be created because it's a stripe winner for this group.

Lab topology – adding receiver outside

```
bdsol-aci32-leaf2# show ip pim internal stripe-winner 238.1.1.1 vrf RD-MC:RD
PIM Stripe Winner info for VRF "RD-MC:RD" (BL count: 2)
(*, 238.1.1.1)
BLs: 10.100.0.2  hash: 173930397 (local)
      10.100.0.1  hash: 1336992484
Winner: 10.100.0.1 best_hash: 1336992484
```

Stripe winner is Leaf1
However Pim Join are received
On Leaf2
→ Leaf 2 forwards traffic to external receiver

© 2017 Cisco and/or

SIRO – final mroute on both BL

Leaf 1- BL stripe winner

```
bdsol-aci32-leaf1# show ip mroute 238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 238.1.1.1/32), uptime: 03:33:41, ngmvpn ip pim
  Incoming interface: loopback6, RPF nbr: 10.100.0.9
  Outgoing interface list: (count: 1) (Fabric OIF)
    Tunnel16, uptime: 03:33:41, ngmvpn

(10.101.1.23/32, 238.1.1.1/32), uptime: 01:09:19, pim mrib ip
ngmvpn
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 1) (Fabric OIF)
    Tunnel16, uptime: 01:09:19, mrib, ngmvpn, (RPF)
```

Leaf 2 – Stripe loser but receive PIM join (*,G) and (S,g)

```
bdsol-aci32-leaf2# show ip mroute 238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 238.1.1.1/32), uptime: 00:02:53, pim ip
  Incoming interface: loopback5, RPF nbr: 10.100.0.9
  Outgoing interface list: (count: 1)
    Ethernet1/8.41, uptime: 00:02:53, pim

(10.101.1.23/32, 238.1.1.1/32), uptime: 01:10:16, pim ip
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65
(pervasive)
  Outgoing interface list: (count: 1)
    Ethernet1/47.40, uptime: 00:02:53, pim
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Source outside Receiver inside (SORI)

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Debugging traffic Flows: Source outside and receivers inside the fabric: SORI

1. IGMP join received on NBL-1 will be installed to MRIB in BL-1 because BL-1 is the stripe winner.
2. Upon the receipt of register on BL-4 from the external source, S,G entry will be created on BL-4.
3. Since the register is received on an anycast RP address, the register will be forwarded to all the members in the anycast RP list
   and hence all the BLs will receive the source-register.
4. (S,G) will created on all the RP BLs.
5. On the stripe-winner BL, the (S,G) will be programmed with the inherited OIF from the *,G. and SW sends (S,G) join

Even with the case of receiver being on one of the border-leafs, the states would look similar.

NBL-1	NBL-2	BL-1 (SW)	BL-2	BL-3	BL-4
(*,G) RPF: Tunnel OIF: Vlan10(igmp)		(*,G) RPF: Loopback OIF: Tunnel (ngmvpn)			
		(S,G) reg RPF: L3out OIF: Tunnel (mrib)	(S,G) reg RPF: L3out OIF: Null	(S,G) reg RPF: L3out OIF: Null	(S,G) reg RPF: L3out OIF: Null

Lab topology – adding a source outside

PIM register from External nexus will reach one of the BL (anycast),
BL sync Register together.
Only Stripe winner (SW) does joint pim (here BL leaf1) and populate OIL to tunnel Interface

© 2017 Cisco and/or

SORI – final mroute on both BL –new source 10.124.1.3

Leaf 1- BL stripe winner

```
bdsol-aci32-leaf1# show ip mroute  238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 238.1.1.1/32), uptime: 06:06:27, ngmvpn ip pim
  Incoming interface: loopback6, RPF nbr: 10.100.0.9
  Outgoing interface list: (count: 1) (Fabric OIF)
    Tunnel16, uptime: 06:06:27, ngmvpn

(10.101.1.23/32, 238.1.1.1/32), uptime: 03:42:05, pim mrib ip
ngmvpn
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 1) (Fabric OIF)
    Tunnel16, uptime: 03:42:05, mrib, ngmvpn, (RPF)

(10.124.1.3/32, 238.1.1.1/32), uptime: 00:01:41, pim mrib ip
  Incoming interface: Ethernet1/47.66, RPF nbr: 10.100.1.22,
internal
  Outgoing interface list: (count: 1)
    Tunnel16, uptime: 00:01:41, mrib
```

Leaf 2 – Stripe loser but receive PIM join (*,G) and (S,g)

```
bdsol-aci32-leaf2#  show ip mroute  238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 238.1.1.1/32), uptime: 02:34:51, pim ip
  Incoming interface: loopback5, RPF nbr: 10.100.0.9
  Outgoing interface list: (count: 1)
    Ethernet1/8.41, uptime: 02:34:51, pim

(10.101.1.23/32, 238.1.1.1/32), uptime: 03:42:14, pim ip
  Incoming interface: Tunnel16, RPF nbr: 10.0.8.65 (pervasive)
  Outgoing interface list: (count: 1)
    Ethernet1/47.40, uptime: 02:34:51, pim

(10.124.1.3/32, 238.1.1.1/32), uptime: 00:01:49, pim mrib ip
  Incoming interface: Ethernet1/47.40, RPF nbr: 10.100.1.18,
internal
  Outgoing interface list: (count: 0)
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Source outside Receiver outside (SORO)

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Debugging traffic Flows: Source outside and receivers inside the fabric: SORO

1. *,G join from the external receiver is directed towards the anycast RP. The join can land on any BL. Assuming that the *,G join lands on BL-2, a *,G state will be created and synced via COOP to all BLs.
2. Upon stripe-winning, BL-4 is be chosen as forwarder and hence BL-4 will install to MRIB.
3. Once source becomes active, FHR will originate a source-register packet towards the anycast RP address.
4. Since the register is received on an anycast RP address, the register will be forwarded to all the members in the anycast RP list and hence all the BLs will receive the source-register.
5. S,G will created on all the RP BLs.
6. On BL-2 which received the external PIM join, OIF will be added by PIM by virtue of register receipt and also because S,G join that is sent from receiver.
7. On BL-4 S,G state will be created because it's a stripe winner for this group.

NBL-1	NBL-2	BL-1	BL-2	BL-3	BL-4 (SW)
			*,G RPF:Loopback OIF: L3out (pim - Ext join)		(*,G) RPF: Loopback OIF: Tunnel (ngmvpn)
		(S,G) data-created RPF: L3out OIF: Tunnel	(S,G) reg RPF: Tunnel OIF: L3out (pim – Ext join)	(S,G) reg RPF: Tunnel OIF: Null	(S,G) reg RPF: Tunnel OIF: Tunnel (mrib)

Inter VRF Multicast 4.0

Use case for Inter VRF multicast

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Inter VRF multicast high level

- Allows the receiver VRF to perform the reverse path forwarding (RPF) lookup for the multicast route in the source VRF.
- When a valid RPF interface is formed in the source VRF, this enables an outgoing interface (OIF) in the receiver VRF.
- All inter-VRF multicast traffic will be forwarded within the fabric in the source VRF.
- The inter-VRF forwarding and translation is performed on the leaf switch where the receivers are connected.
- For any-source multicast, the RP used must be in the same VRF as the source.
- Source and receiver VRFs can be in an EPG or connected behind an L3 Out.

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

VRF deployement

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

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Inter VRF requirement

- All sources for a particular group must be in the same VRF (the source VRF).
- Source VRF and source EPGs need to be present on all leafs where there are receiver VRFs.
- For ASM (Sparse or BiDir) :
  - The RP must be in the same VRF as the sources (the source VRF).
  - The source VRF must be using fabric RP.
  - The same RP address configuration must be applied under the source and all receiver VRFs for the given group-range.

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Mo for pim InterVRF

- MO on each leaf in dom of rcv vrf/.

```
apic1# moquery -d topology/pod-1/node-103/sys/pim/inst/dom-RD-MC:RD2/intervrf/intervrf-mcast_intervrf_RD-
MC:RD2_tn-RD-MC_ctx-RD
Total Objects shown: 1

# pim.InterVRF
rtMap        : mcast_intervrf_RD-MC:RD2_tn-RD-MC_ctx-RD
adminSt      : disabled
childAction  :
ctrl         :
dn           : topology/pod-1/node-103/sys/pim/inst/dom-RD-MC:RD2/intervrf/intervrf-mcast_intervrf_RD-
MC:RD2_tn-RD-MC_ctx-RD
lcOwn        : local
modTs        : 2018-12-28T04:12:58.376+00:00
mtu          : 1500
name         :
rn           : intervrf-mcast_intervrf_RD-MC:RD2_tn-RD-MC_ctx-RD
srcVrfName   : tn-RD-MC:ctx-RD
status       :
```

Leaf RPF check

- We can check on leaf that Rcv VRF do RPF for the shared range in vrf of src (RD-MC:RD)

```
bdsol-aci32-leaf3#  show ip multicast vrf RD-MC:RD2
Multicast Routing VRFs (21 VRFs)
VRF Name                VRF      Table       Route   Group   Source  (*,G)   State
                        ID       ID          Count   Count   Count   Count
RD-MC:RD2               24       0x00000018  1       0       0       0       Up
    Multipath configuration: legacy
    Resilient configuration: Disabled
    MFDM outstanding ack count: 0
    Fabric Interface:  0x74
    RPF Select configuration:
      238.0.0.0/8 in vrf RD-MC:RD
          owners: mcast_intervrf_RD-MC:RD2_tn-RD-MC_ctx-RD
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Adding a receiver on leaf3 in Rcv VRF

- We have a receiver in leaf3 for group 238.1.1.1 in rcv vrf
- Note the RPF interface is the mcast tunnel in src VRF !!!

```
bdsol-aci32-leaf3# show ip  igmp groups 238.1.1.1 vrf RD-MC:RD2
Type: S - Static, D - Dynamic, L - Local, T - SSM Translated
Displaying Groups for vrf:RD-MC:RD2
Group Address         Type   Interface    Uptime               Expires              Last Reporter
238.1.1.1             D      vlan35       00:09:44             00:02:31             10.110.1.3
bdsol-aci32-leaf3#

bdsol-aci32-leaf3# show ip mroute 238.1.1.1 vrf RD-MC:RD2
IP Multicast Routing Table for VRF "RD-MC:RD2"

(*, 238.1.1.1/32), uptime: 00:09:04, igmp ip pim
  Incoming interface: Tunnel19, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 1)
    Vlan35, uptime: 00:09:04, igmp

bdsol-aci32-leaf3# show fabric multicast vrf all
Fabric Multicast Enabled VRFs
VRF Name              VRF      Vprime        VN-Seg      VRF    Conv   Tunnel
                      ID       If            ID          Role   Mode   IP
RD-MC:RD2             24       Tunnel21      2129922     Leaf   Reg    127.0.0.100
RD-MC:RD              13       Tunnel19      3112960     Leaf   Reg    127.0.0.100
bdsol-aci32-leaf3#
```

Lab topology – with receive in leaf3 in Rcv VRF

© 2017 Cisco and/or

Mroute in src VRF in rcv leaf

- Note vrf RD-MC:RD2 is extranet VRF in the OIL

```
bdsol-aci32-leaf3# show ip mroute 238.1.1.1 vrf RD-MC:RD
IP Multicast Routing Table for VRF "RD-MC:RD"

(*, 238.1.1.1/32), uptime: 1d01h, igmp ip pim
  Incoming interface: Tunnel19, RPF nbr: 10.0.80.94
  Outgoing interface list: (count: 1)
    Vlan83, uptime: 1d01h, igmp
  Extranet receiver list: (vrf count: 1, OIF count: 1)
  Extranet receiver in vrf RD-MC:RD2:
  (*, 238.1.1.1/32) OIF count: 1
```

Multicast forwarding info

- See here MFDM info in vrf Src has Vlan 35 as OIL.
- Vlan 35 is part of vrf RD-MC:RD2 (rcv VRF)

```
bdsol-aci32-leaf3# show forwarding  distribution multicast route vrf RD-MC:RD group 238.1.1.1

  (*, 238.1.1.1/32), RPF Interface: Tunnel19, flags: Es
    Received Packets: 0 Bytes: 0
    Number of Outgoing Interfaces: 2
    Outgoing Interface List Index: 8205
      Vlan35
      Vlan83
bdsol-aci32-leaf3# show ip interface vrf RD-MC:RD2
IP Interface Status for VRF "RD-MC:RD2"(24)
Vlan35, Interface status: protocol-up/link-up/admin-up, iod: 138, mode: pervasive, vrf_vnid: 2129922
  IP address: 10.110.1.254, IP subnet: 10.110.1.0/24
  IP primary address route-preference: 1, tag: 0
```

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential

Inter VRF event trace – look for mrib_extnet

```
bdsol-aci32-leaf3# show routing ip multicast event-history rib | egrep mrib_extent
```

When mrib creates a route, it will look for InterVRF configuration. If found, the match will be shown like below. RPF lookups will then be done in the selected VRF

```
2018 Dec 28 12:22:13.097261 mrib [13359]: TID 13428:mrib_extnet_debug:2528: Extranet: (VRF RD-MC:RD2) Find a match as [238.0.0.0/255.0.0.0, select VRF RD-MC:RD]
2018 Dec 28 12:22:13.096613 mrib [13359]: TID 13442:mrib_extnet_src_route_locate:2492: Extranet: found (*, 238.1.1.1/32) 1/0 in source (VRF RD-MC:RD) succeeded
2018 Dec 28 12:22:13.096470 mrib [13359]: TID 13442:mrib_extnet_src_route_locate:2462: Extranet: locating (*, 238.1.1.1/32) 1/0 in source (VRF RD-MC:RD), triggered by
receiver (VRF RD-MC:RD2)
2018 Dec 28 12:22:13.096179 mrib [13359]: TID 13442:mrib_extnet_debug:2588: Extranet: refreshing (*, 238.1.1.1/32) in source (VRF RD-MC:RD) and receiver (VRF RD-MC:RD2)
2018 Dec 28 12:22:13.095838 mrib [13359]: TID 13442:mrib_extnet_debug:2606: Extranet: (VRF RD-MC:RD2) (*, 238.1.1.1/32) new RPF update, RPF interface Tunnel19 (VRF RD-
MC:RD)->(VRF RD-MC:RD)
```

When mcast route is created in src vrf, we attach receiver vrf

```
2018 Dec 28 12:22:13.095338 mrib [13359]: TID 13442:mrib_extnet_debug:2568: Extranet: attaching (*, 238.1.1.1/32) 1/0 receiver (VRF RD-MC:RD2) to source (VRF RD-MC:RD),
total receiver VRFs (0->1)
2018 Dec 28 12:22:13.095177 mrib [13359]: TID 13442:mrib_extnet_src_route_locate:2492: Extranet: found (*, 238.1.1.1/32) 1/0 in source (VRF RD-MC:RD) succeeded
2018 Dec 28 12:22:13.094794 mrib [13359]: TID 13442:mrib_extnet_src_route_locate:2462: Extranet: locating (*, 238.1.1.1/32) 1/0 in source (VRF RD-MC:RD), triggered by
receiver (VRF RD-MC:RD2)
```

When RPF lookup succeeds with incoming interface in the Source VRF, we find/create the corresponding route in the Source VRF and link the Source and Receiver VRF routes.

```
2018 Dec 28 12:22:13.094519 mrib [13359]: TID 13442:mrib_extnet_debug:2578: Extranet: enqueuing (*, 238.1.1.1/32) receiver (VRF RD-MC:RD2) into source (VRF RD-MC:RD)
2018 Dec 28 12:22:13.094243 mrib [13359]: TID 13442:mrib_extnet_debug:2606: Extranet: (VRF RD-MC:RD2) (*, 238.1.1.1/32) RPF update, RPF interface Tunnel19 (VRF
<unknown>)->(VRF RD-MC:RD)Enqueueing the
extranet route
2018 Dec 28 12:22:13.093951 mrib [13359]: TID 13442:mrib_extnet_debug:2606: Extranet: (VRF RD-MC:RD2) (*, 238.1.1.1/32) RPF update, RPF interface Tunnel19 (VRF
<unknown>)->(VRF RD-MC:RD)
```

Troubleshooting Steps

- Debug on leaf where a receiver is connected, or a BL receiving external joins
- Is (*,G) created ((S,G) for SSM)? This should be there regardless of feature.
  - If not, non-intervrf issue
- Does the above mroute have an RPF interface in the expected Source VRF (slide 14)?
  - If in receiver VRF, check that interVRF is configured for the group (slides 13)
  - If NULL, there may be no route in the Source VRF
  - Check logs (slide 17)
- If yes, look for route in Source VRF with extranet info (slide 15)
  - If not, check logs (slide 17)
- Is MFDM state correct (slide 16)?

- If all the above looks fine, debug multicast in the Source VRF as if you have local receivers, and check on other routers towards the source

© 2017 Cisco and/or its affiliates. All rights reserved. Cisco Confidential