IPv4 Multicast VXLAN

SPEAKERS:
• Daryl Wan
• Thomas Samuel
Agenda: IPv4 Multicast VXLAN

1. Overview
2. Use Cases
3. Details/Caveats
4. Configuration and Best Practices
5. Troubleshooting
6. Demo
7. Resources
Overview
IPv4 Multicast VXLAN Overview

• 10.5 did not support IPv4 multicast over VXLAN EVPN networks
• 10.6 CPE adds support for IPv4 multicast over VXLAN EVPN networks
  • On 6300/6400/8325/8360/8400

• What’s new and related to IPv4 Multicast VXLAN in 10.6?
  • L2 IGMP snooping
  • L3 PIM-SM routing
  • IPv4 Centralized L3 Gateways
  • IPv4 Distributed L3 Gateways - recommended for most customers
  • BSR/RPs in the VXLAN overlay
Use Cases
Campus/Edge DC Use Case

- Distributed L3 Gateways (symmetric IRB) recommended for Campus
- Sources are typically located in edge DC, sources sometimes also connected to access VTEPs
- Receivers are typically connected to Campus access VTEPs and require PIM-SM L3 multicast routing to access multicast stream from source in edge DC
DC Use Case

- Distributed L3 Gateways (symmetric IRB) recommended for DC
- Some applications in the DC utilize IPv4 multicast across racks
  - Sources/Receivers could be spread across different racks
  - PIM-SM is required for L3 multicast routing across subnets

- IGMP snooping is used to limit L2 multicast across VTEPs without receivers
  - Also applicable to Campus use case
Details
IGMP snooping

Source and Receivers on VLAN 10 mapped to L2VNI 6500

VLAN 10 : L2_VNI 6500

VTEP-1

VLAN 10 : L2_VNI 6500

VTEP-2

VLAN 10 : L2_VNI 6500

VTEP-3

VLAN 10 : L2_VNI 6500

Receiver-A

Receiver-B

Multicast Source

Overlay tunnels

IGMP Joins

Data

Tunnel source 1.1.1.1

Tunnel source 2.2.2.2

Tunnel source 3.3.3.3

Underlay N/W

VLAN 10 : L2_VNI 6500

Tunnel source
IGMP Snooping – filtered VTEP

VTEP-1# show ip igmp snooping vlan 10 group 239.10.10.1

IGMP ports and group information for group 239.10.10.1

VLAN ID : 10
VLAN Name : VLAN10

Group Address : 239.10.10.1
Last Reporter : 100.3.3.109
Group Type : Filter

<table>
<thead>
<tr>
<th>Port</th>
<th>Vers</th>
<th>Mode</th>
<th>Uptime</th>
<th>Expires</th>
<th>V1 Timer</th>
<th>V2 Timer</th>
<th>Sources</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>vxlan1(3.3.3)</td>
<td>2</td>
<td>EXC</td>
<td>5h 24m</td>
<td>3m 38s</td>
<td>3m 38s</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
PIM Routing use case #1: RP near to Source

- Symmetric Routing with PIM-SM

Source @ VLAN 10
Receiver @ VLAN 30
L3VNI: 6500

VTEP-1# show ip mroute vrf red
IP Multicast Route Entries
VRF : red
Total number of entries : 1
Group Address : 239.1.1.1
Source Address : 200.200.1.10
Neighbor :
Incoming interface : vlan10
Outgoing Interface List :
Interface       State
--------------  --------
vni6500         forwarding

VTEP-3# show ip mroute vrf red
IP Multicast Route Entries
VRF : red
Total number of entries : 1
Group Address : 239.1.1.1
Source Address : 200.200.1.10
Neighbor : 1.1.1.1
Incoming interface : vni6500
Outgoing Interface List :
Interface       State
--------------  --------
vlan30          forwarding
PIM Routing use case #2: RP away from Source

- Symmetric Routing with PIM-SM

- Multicast Source
  - VLAN 10: L3_VNI 6500
  - VTEP-1

- Receiver-B
  - VLAN 30: L3_VNI 6500
  - VTEP-3

- Data
  - Overlay tunnels
  - PIM Joins

- Source @ VLAN 10
  - Receiver @ VLAN 30
  - L3VNI: 6500
VSX Border VTEP with L2 extension (Source case)

- Source connected to VSX border VTEP
- `ip pim-sparse vsx-virtual-neighbor`

- Both Primary and Secondary takes DR role
- VTEP-2 send PIM joins towards the source (via L3VNI interface)
- Either the Primary or Secondary get the PIM join (ECMP Path)
- Whoever get the join process it and send a copy to VSX peer
- Both Primary and Secondary include L3VNI in the Outgoing list
- When ISL is in sync only primary forwards traffic to VTEP-2
- Secondary take the backup role and forwards only - when ISL goes out of sync (Broadcom ASIC’s)
- In Rosewood (Aruba ASIC’s) whoever gets the traffic - will forwards it.
- Note:- Keepalive L3 connectivity is necessary

```
VTEP1# interface vlan20
  vrf attach VRF1
  ip address 10.10.10.2/24
  ip pim-sparse enable
  ip pim-sparse vsx-virtual-neighbor
  ip igmp enable
!
router pim vrf VRF1
  enable
  rp-candidate source-ip=interface loopback3
  rp-candidate group-prefix 224.0.0.0/4
  bsr-candidate source-ip=interface loopback3
```
VSX Border VTEP with L2 extension (Receiver case)

- VSX border VTEP
- Receiver connected to VSX border VTEP
- Both Primary and Secondary takes DR role
- Both primary and secondary sends the PIM joins
- Either one of Primary or Secondary get the traffic (ECMP Path) and route it to the host VLAN

Diagram:
- Multicast Source
- Source @VLAN10
- VTEP-2
- Primary (DR)
  - Tunnel source 1.1.1.1
  - Underlay N/W L3VNI
- Secondary (DR)
  - Tunnel source 2.2.2.2
  - Tunnel source 1.1.1.1
- KA
- ISL
- VSX LAG
- L2 Access
- Receiver @VLAN20
- Host @VLAN20
- Overlay tunnels
- PIM Joins
- Data
L3 extension from non VSX VTEP

- Non VSX VTEPS can be L3 extended
- Can reuse the existing RP from outside the fabric
VSX Border VTEP with L3 extension

- VSX border VTEP can be connected to existing sites/networks
- Support SVI’s and ROP extensions.
- Can reuse the existing RP from outside the fabric.
Active Gateway IP

- VSX VTEPs can be configured with same AG IP
- Same SVI IP on Primary and Secondary
IPv4 Multicast VXLAN Caveats

• AOS-CX doesn’t use EVPN route types 6/7/8
  • It’s not possible to interop with other vendors VXLAN IPv4 multicast (L2 IGMP snooping and L3 PIM-SM routing) implementation

• IPv4 multicast IVRL (Inter VRF Route Leaking) is not supported in 10.6
IPv4 Multicast VSX VXLAN Caveats

- Supported with 10.6 CPE release
  1. VSX switches (Logical VTEPs) for IGMP Snooping and PIM-SM
  2. VSX border VTEP with VSX LAG, L2 extension to sources/receivers in external/non overlay network

- Planned for future 10.6 CPE
  3. VSX border VTEP with VSX LAG, SVI + PIM-SM L3 extension to sources/receivers in external/non overlay network

- Not Supported in 10.6
  4. VSX border VTEP with Routed Only Ports (ROP) + PIM-SM L3 extension to sources/receivers in external/non overlay network
  5. RP on VSX switches
  6. VSX Orphan ports are not supported

The above caveats only apply to VSX switches, they do not apply to standalone or VSF switches.
DC Use Case Caveats

Since RPs cannot be deployed on VSX VTEP switches
• VTEPs can point to external RPs with PIM-SM
• Can’t use VSX border VTEP/L3 PIM until future release
Campus/Edge DC Use Case Caveats

- If source is on external/non overlay network, VSX border VTEP will not be able to support L3 PIM to external/non overlay network with 10.6 initial CPE release (refer to VSX caveats)
- Redundant overlay BSR/RPs can be enabled on VSF switches
Configuration
VXLAN IGMP Snooping

- Using example above for VLAN 10
  - Leaf1a# vlan 10
  - ip igmp snooping enable

- It’s not possible to use CLI to prove IGMP snooping on VTEP1 prevents multicast stream on uplinks
  - Wireshark packet capture can be used on VTEP uplinks if really required

- Only active joins are shown

Leaf1a# sh ip igmp snooping vlan 10
IGMP Snooping Protocol Info

Total VLANs with IGMP enabled : 1
Current count of multicast groups joined : 1

IGMP Drop Unknown Multicast : Global
VLAN ID : 10
VLAN Name : VLAN10
IGMP Configured Version : 3
IGMP Operating Version : 3
Querier Address [this switch] : 10.1.10.2
Querier Port :
Querier Uptime : 2d 23h 11m
Querier Expiration Time : 1m 0s

Active Group Address Tracking Vers Mode Uptime Expires
------------------------------------------ ---- ---- ---------- -------
239.255.255.250 Filter 3 EXC 1d 6h 36m 3m 17s
VXLAN PIM-SM – External Anycast RP

Can't use VSX border VTEP/L3 PIM until future release

RP1#
interface loopback 10
  ip address 192.168.200.4/32
  ip ospf 15 area 0.0.0.0
interface loopback 11
  description MSDP Anycast
  ip address 192.168.200.8/32
  ip ospf 15 area 0.0.0.0
!
router pim
  enable
  rp-address 192.168.200.8
!
router msdp
  enable
  ip msdp peer 192.168.200.5
    connect-source loopback10
    enable
    mesh-group m1

RP2#
interface loopback 10
  ip address 192.168.200.5/32
  ip ospf 15 area 0.0.0.0
interface loopback 11
  description MSDP Anycast
  ip address 192.168.200.8/32
  ip ospf 15 area 0.0.0.0
!
router pim
  enable
  rp-address 192.168.200.8
!
router msdp
  enable
  ip msdp peer 192.168.200.4
    connect-source loopback10
    enable
    mesh-group m1
VXLAN PIM-SM – External Anycast RP

- Using example above for VLAN 10

```
Leaf1a#
interface vlan10
  vrf attach VRF1
  ip address 10.10.10.2/24
  active-gateway ip 12:00:00:00:01:00
  ip pim-sparse enable
  ip igmp enable

  router pim vrf VRF1
    enable
    rp-address 192.168.200.8
```

Anycast RP on external/non overlay network has to be reachable via VRF1

- Verification

```
Leaf1a# sh ip pim neighbor vrf VRF1
PIM Neighbor
VRF : VRF1
Total number of neighbors : 2

IP Address : 10.10.10.3
Interface : vlan10
Up Time (HH:MM:SS) : 3 days 02:49:07
DR Priority : 1
Hold Time (HH:MM:SS) : 00:01:45

IP Address : 192.168.11.5
Interface : vni100001
Up Time (HH:MM:SS) : 3 days 02:49:43
DR Priority : 1
Hold Time (HH:MM:SS) : 00:01:45
```
VXLAN PIM-SM – External Anycast RP

- Verification

Leaf1a# sh ip mroute vrf VRF1
IP Multicast Route Entries
VRF : VRF1
Total number of entries : 1

<table>
<thead>
<tr>
<th>Group Address</th>
<th>Source Address</th>
<th>Neighbor</th>
<th>Incoming interface</th>
<th>Outgoing Interface List</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>239.10.10.10</td>
<td>10.10.10.12</td>
<td></td>
<td>vlan10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Leaf1a# sh ip pim rp-set vrf VRF1
VRF: VRF1
Status and Counters - PIM-SM Static RP-Set Information

<table>
<thead>
<tr>
<th>Group Address</th>
<th>Group Mask</th>
<th>RP Address</th>
<th>Override</th>
</tr>
</thead>
<tbody>
<tr>
<td>224.0.0.0</td>
<td>240.0.0.0</td>
<td>192.168.200.8</td>
<td>No</td>
</tr>
</tbody>
</table>
**VXLAN PIM-SM – External Anycast RP**

- **Verification**

  Leaf2a# sh ip mroute vrf VRF1
  IP Multicast Route Entries
  
  VRF : VRF1
  Total number of entries : 1
  
  Group Address : 239.10.10.10
  Source Address : 10.1.10.12
  Neighbor : 192.168.11.3
  Incoming interface : vni100001
  Outgoing Interface List :
  Interface State
  ----------------- ------------
  vlan12 forwarding

  Leaf2a# sh ip pim rp-set vrf VRF1
  
  VRF: VRF1
  
  Status and Counters - PIM-SM Static RP-Set Information
  
  Group Address    Group Mask       RP Address       Override
  ----------------- --------------- --------------- --------
  224.0.0.0        240.0.0.0        192.168.200.8    No
**VXLAN PIM-SM – Overlay BSR/RP**

**VTEP1#**
```
interface loopback 3
  vrf attach VRF1
  ip address 60.1.1.1/32
  ip pim-sparse enable
!
interface vlan10
  vrf attach VRF1
  ip address 10.10.10.2/24
  active-gateway ip mac 12:00:00:00:01:00
  active-gateway ip 10.10.10.1
  ip pim-sparse enable
  ip igmp enable
!
router pim vrf VRF1
  enable
  rp-candidate source-ip-interface loopback3
  rp-candidate group-prefix 224.0.0.0/4
  bsr-candidate source-ip-interface loopback3
```

**VTEP2#**
```
interface loopback 3
  vrf attach VRF1
  ip address 60.2.1.1/32
  ip pim-sparse enable
!
interface vlan10
  vrf attach VRF1
  ip address 10.10.10.3/24
  active-gateway ip mac 12:00:00:00:01:00
  active-gateway ip 10.10.10.1
  ip pim-sparse enable
  ip igmp enable
!
router pim vrf VRF1
  enable
  rp-candidate source-ip-interface loopback3
  rp-candidate group-prefix 224.0.0.0/4
  bsr-candidate source-ip-interface loopback3
VTEP1# show ip pim bsr elected all-vrfs
Status and Counters - PIM-SM Elected Bootstrap Router Information

VRF : VRF1
E-BSR Address : 60.1.1.1
E-BSR Priority : 0
E-BSR Hash Mask Length : 30
E-BSR Up Time : 49 mins 58 secs
Next Bootstrap Message : 1 mins 47 secs

VTEP1# show ip pim rp-set vrf VRF1
VRF: VRF1
Status and Counters - PIM-SM Learned RP-Set Information

<table>
<thead>
<tr>
<th>Group Address</th>
<th>Group Mask</th>
<th>RP Address</th>
<th>Hold Time</th>
<th>Expire Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>224.0.0.0</td>
<td>240.0.0.0</td>
<td>60.2.1.1</td>
<td>150</td>
<td>133</td>
</tr>
<tr>
<td>224.0.0.0</td>
<td>240.0.0.0</td>
<td>60.1.1.1</td>
<td>150</td>
<td>133</td>
</tr>
</tbody>
</table>
Best Practices
IPv4 Multicast VXLAN – IGMP Snooping Best Practices

• Full mesh tunnels connections are a must

• “redistribute local-svi” should be enabled on all VSX VTEP’s
  Required to learn the querier mac address.

  8325-1(config)# evpn
  8325-1(config-evpn)# redistribute
  local-svi  Redistribute local SVI IP address via EVPN

• On VSX cases set the robustness to a higher value
  Useful to retain the joins during device reboots events
IPv4 Multicast VXLAN – PIM-SM Best Practices

- Full L3 mesh tunnels connections are a must
- For VSX leaf routers need to configure dedicated Keepalive interface and need to be kept UP
- PIM should not be enabled on the Keepalive interface
- PIM Active-Active feature is not supported with VXLAN
- `vsx-virtual-neighbor` command must be configured on the VSX border VTEP SVI’s
Troubleshooting
IPv4 Multicast VXLAN Troubleshooting Details
– IGMP Snooping

• Tunnels are operational and unicast reachability
• Querier elections are proper. All router should know about the querier
• Check the joined VTEP tunnels (show ip igmp snooping vlan x group y)
• Show ip igmp interface vlanX (when snooping not in use)
• Look for Querier detected port
• Look for join uptime and see any fluctuations
• Look for igmp copp drops
IPv4 Multicast VXLAN Troubleshooting Details – PIM-SM

- Verify the Tunnels are created and operational

```
leaf1# show interface vxlan vtep
Source Destination Origin Status VNI Routing VLAN VRF
----------------- --------------- ------ --------------------- -------- ------ ------ ----
3.3.3.3 1.1.1.1 evpn operational 100 enabled -- red
```

- Verify that PIM neighbors exist on L3VNI from all the remote VTEP’s.

  *Mirror traffic in the underlay interface and look for VXLAN encapsulated PIM hello’s once every 30 second. Unicast reachability issues in the underlay can lead to neighbors getting lost.*

```
leaf1# show ip pim neighbor vrf red

PIM Neighbor
VRF : red
Total number of neighbors : 1

IP Address : 3.3.3.3
Interface : vni6500
Up Time (HH:MM:SS) : 05:32:17
Expires Time (HH:MM:SS) : 00:01:33
DR Priority : 1
Hold Time (HH:MM:SS) : 00:01:45
```
IPv4 Multicast VXLAN Troubleshooting Details – PIM-SM

• Check BSR and RP-set informations are learned on the tenant’s VRF’s.

```plaintext
leaf2# show ip pim bsr vrf red
Status and Counters- PIM-SM Bootstrap Router Information

VRF: red
E-BSR Address : 200.200.1.2
E-BSR Priority : 0
E-BSR Hash Mask Length : 30
E-BSR Up Time : 3 mins 24 secs
Next Bootstrap Message : 1 mins 51 secs

C-BSR Admin Status : This system is not a Candidate-BSR
C-RP Admin Status : This system is a Candidate-RP
C-RP Address : 100.100.1.2
C-RP Hold Time : 150
C-RP Advertise Period : 60
C-RP Priority : 192
C-RP Source IP Interface : vlan10

Group Address    Group Mask
---------------    -----------
224.0.0.0        240.0.0.0

leaf2# show ip pim rp-set vrf red
VRF: red

Status and Counters - PIM-SM Learned RP-Set Information

Group Address    Group Mask       RP Address       Hold Time  Expire Time
---------------    -----------       -----------       ---------  ----------
224.0.0.0        240.0.0.0       100.100.1.2      150        118
```

---

36
IPv4 Multicast VXLAN Troubleshooting Details – PIM-SM

- One of the basic criteria for PIM to work on overlay is the unicast reachability.
- Ensure the interested IP in question (say EBSR, RP, Mcast source, DR etc) are reachable via overlay.
- Check their respective unicast routes.

```plaintext
leaf1(config)# show ip route bgp vrf red
Displaying ipv4 routes selected for forwarding
' [x/y]' denotes [distance/metric]
100.100.1.0/24, vrf red
via 3.3.3.3[vrf default], [200/0], bgp, encap vxlan, l3vni 100

leaf1(config)# show ip route vrf red
Displaying ipv4 routes selected for forwarding
' [x/y]' denotes [distance/metric]
20.20.20.0/24, vrf default
via 10.10.10.2, [110/200], ospf
100.100.1.0/24, vrf red
via 3.3.3.3[vrf default], [200/0], bgp, encap vxlan, l3vni 100
200.200.1.0/24, vrf red
via vlan20, [0/0], connected
200.200.1.2/32, vrf red
via vlan20, [0/0], local
```
Resources
Feature/Solution References

– RFCs
– IEEE
– Standards
– Industry Standard References
– Public Blogs

– Anything you’ve found useful that would be interesting to someone talking to customers or supporting your feature/solution.
Demo
Demo#1 – IPv4 Multicast VXLAN with Distributed L3 Gateways

- Show IGMP snooping working, VTEP1 should not forward multicast traffic to VTEP2 (without a receiver) on VLAN 10
- Show PIM neighbors, show L3 multicast traffic and mroute between VTEP1/VTEP2
Demo#2 – Troubleshooting IPv4 Multicast VXLAN

- Walkthrough a troubleshooting example, what to look out for on both VTEPs
Thank you

daryl.wan@hpe.com
thomas.samuel@hpe.com