Hewlett Packard Enterprise

HPE aruba networking

AOS-CX 10.13 Update:

Multicast Update: PIM Bidirectional (PIM-Bidir) & Multicast Optimizations

November 2023

Agenda

1	Overview
2	Feature details and configuration

Overview

Overview Multicast Terminology

Term	Definition
PIM-SM	Protocol Independent Multicast Sparse Mode
PIM-Bidir	Protocol Independent Multicast Bidirectional
IGMP	Internet Group Management Protocol –three versions, used buy receivers who want to acquire a multicast stream
RP	Rendezvous Point –builds a multicast map of sources and receivers and other RPs, and is root of the shared tree
SPT	Shortest Path Tree AKA Source Tree
RPT	Rendezvous Point Tree AKA Shared Tree
(S) DR	Designated Router - send PIM join/prune messages, register source to RP and forward multicast, sometimes called the Source DR see also FHR
FHR	First Hop Router - closest to source interchangeably referred to as DR or Source DR in some cases
LHR	Last Hop Router - connected to receivers, IGMP information is collected on receivers and forwarded upstream [hop by hop] to RP. The LHR is AKA Receiver DR
(R) DR	Receiver DR, see LHR
Querier	Querier, communicates with receivers using IGMP to determine multicast group membership [IGMPv2 Querier is Separate to RDR , IGMPv1 DR and Querier are same device
RPF	Reverse Path Forwarding –Forwarding multicast traffic away from source toward receiver. Check multicast source with unicast routing table if it's a match accept only that traffic flow.
RPF (check)	Reverse Path Forwarding –multicast loop /duplication prevention for PIM SM. Forwarding multicast traffic away from source toward receiver. Check multicast source with unicast routing table if it's a match accept only that traffic flow.
OIL	Outgoing Interface List - Interface facing downstream towards receivers build the OIL after RPF check
DF	Designated Forwarder used in PIM-Bidir to forwards multicast sources towards the RP
DF (E)	Designated Forwarder (Election) used in PIM-Bidir each segment or link becomes DF based on best unicast routing metric to RP
RPTF	RP Tree Forwarding in PIM-Bidir the RP forwards multicast traffic to the receivers along the receiver side RP/Shared Tree

Overview

- Part 1 of 2
  - PIM Bidirectional for IPv4

- **Part 2 of 2**
  - **Multicast Optimizations**
    - **Querier wait time**
    - **Handling of new and old joins**
    - **Preprograming (*,G) bridge entries**

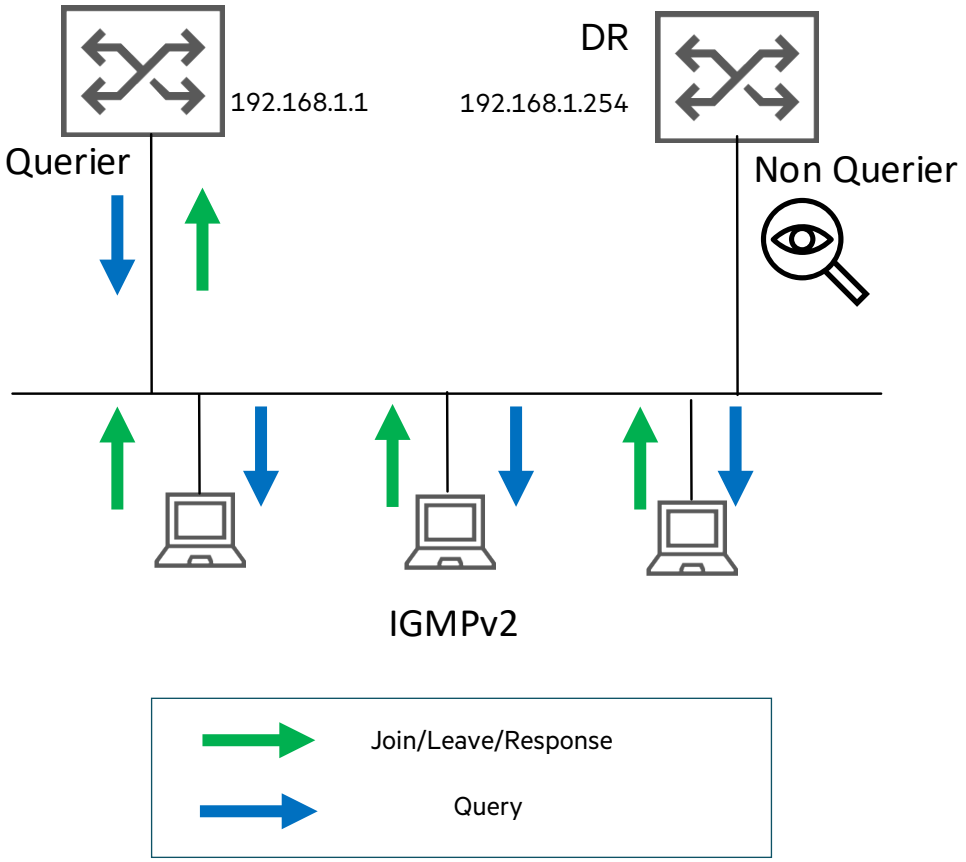
Overview Multicast Optimizations

- The optimization to follow detail providing reduced times/quicker access to multicast streams often termed “Latency”
- Optimize performance
- Predictable and deterministic use
- Latency in this context is **NOT** packet switching throughput

- Why is this important ?
- Improved user experience i.e video streams accessed with reduced delay, channel “changing”
- Monitoring i.e availability in distributed command and control CCTV monitoring
- Critical time dependent data i.e market data ticker

IGMP Overview

- IGMPv2 Has Querier and Non Querier
  - Querier is **elected** and is the lowest IP address
  - Non Querier is passive and listens and records host responses
  - Querier is active and exchanges IGMP/MLDv1 messages
- IGMPv2 has four types of message exchange between Querier and hosts
  - membership query
    - general query, check if multicast is still required on segment
    - group specific query, IGMPv2 check if multicast is still required
  - v1 membership report AKA Join
  - v2 membership report AKA join
  - v2 leave group
- NOTE: Hosts can also send unsolicited join
- NOTE: IGMPv2 and MLDv1 operation similar



Querier Wait Timer

- Querier wait time. What is it ?
  - Time to wait before the Querier election process begins
  - The default Querier wait time ~250 seconds, and non-configurable prior to 10.13
- Querier wait time only occurs on:
  - Reboot i.e power outage, software upgrade
  - protocol restart i.e crash or line card (OIR) on a vlan
  - Initial configuration i.e turn up and test
- **NOTE:** The Querier wait time is overridden if PIM is active on a segment, as PIM starts the Querier election process
- Querier wait time use case
  - Initial faster join of groups on boot or applying configuration in Layer 2 environment only
  - L2VNI segment with IGMP/MLD only
  - Not applicable in PIM environments as wait timer is overridden

```
show ip igmp interface vlan 100

IGMP Configured Version  : 3
IGMP Operating Version   : 3
Querier State              : Initial Wait
Querier IP                 :
Querier Uptime             :
Querier Expiration Time    :
IGMP Snoop Enabled on VLAN : True
```

```
show ip igmp int vlan 100

IGMP Configured Version  : 3
IGMP Operating Version   : 3
Querier State              : Querier-election
Querier IP                 : 100.100.100.1
Querier Uptime             : 0m 29s
Querier Expiration Time    : 0m 3s
IGMP Snoop Enabled on VLAN : True
```

Optimization 1:- Querier Wait Timer

- Configurable querier wait timer
- Supported on all platforms
- Supported for both IGMP and MLDP
- No restrictions i.e VXLAN/Non VXLAN support
- Querier wait timer can be adjusted between 1-300 secs

```
switch(config)# interface vlan 100
switch(config-if-vlan)# ip igmp querier-wait-time <1 - 300>

switch(config)# interface vlan 100
switch(config-if-vlan)# ipv6 mld querier-wait-time <1 - 300>
```

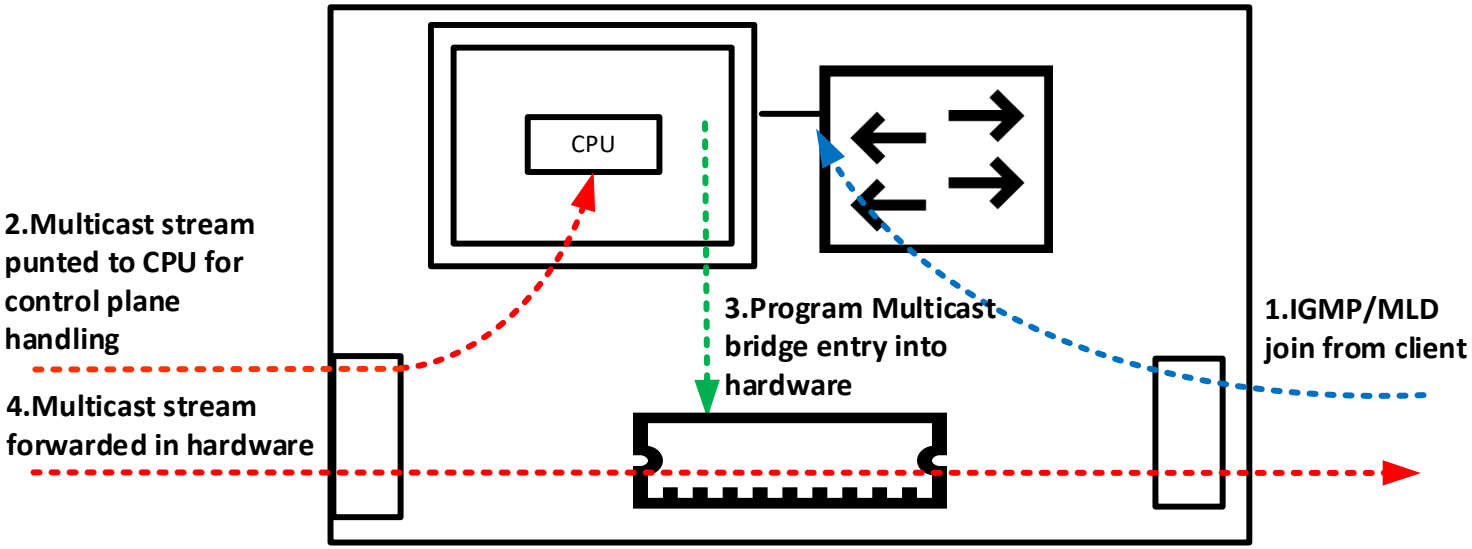
Platform/Feature	4100i	6000	6100	6200	6300	6400 (v1/V2)	8100	8320	8325	8360 (v1/v2)	8400	9300	10000
Querier Wait timer	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Optimization 2 :-Handling of old and new join

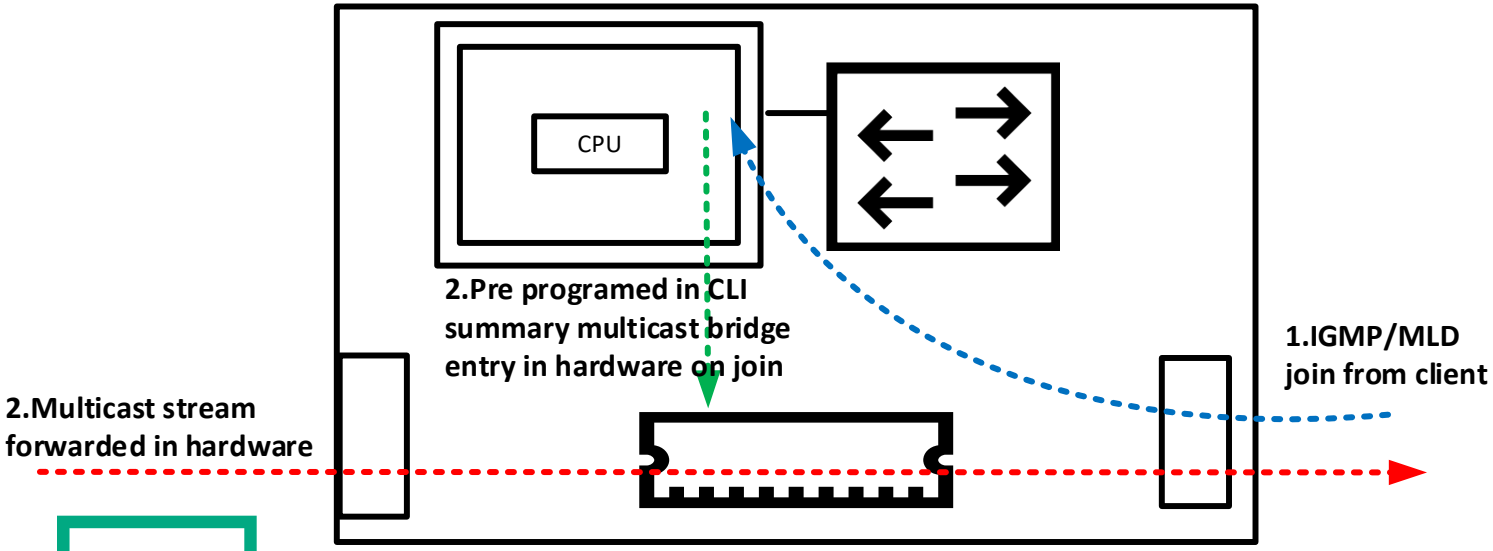
- New joins are prioritized over existing/older joins, in effect this is new groups
- Goal to improve join latency
- Supported on all platforms
- Supported for both IGMP and MLD
- No configuration required, enabled by default from 10.13

Platform/Feature	4100i	6000	6100	6200	6300	6400 (v1)	6400 (v2)	8100	8320	8325	8360 (v1/v2)	8400	9300	10000
Old and new multicast join handling	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Optimization 3 :-Preprogram (*,G) bridge entries overview



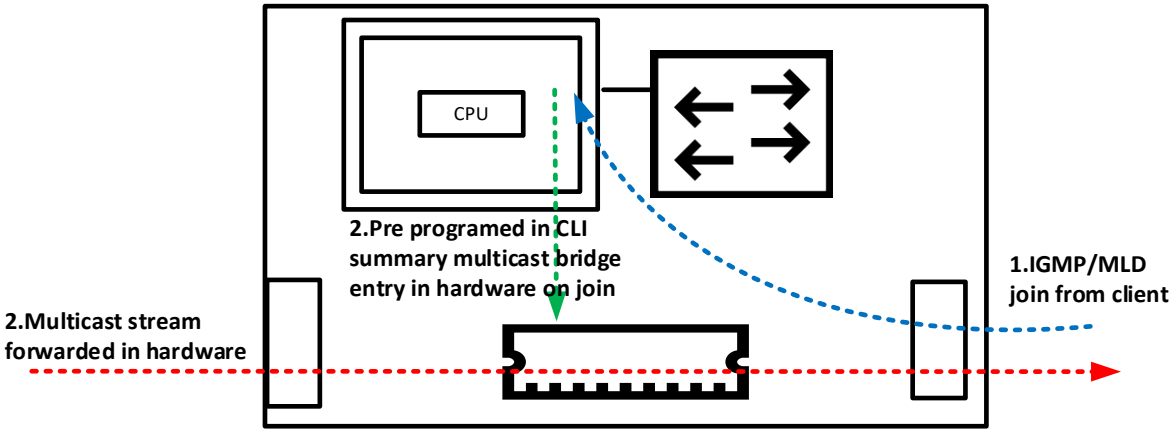
- IGMP/MLD creates multicast bridge entries when L2 device snooping
- Default operation requires CPU intervention and processing to program L2 entries into hardware before multicast can be forwarded



- 10.13 allows IGMP/MLD entries to be preprogramed when L2 device snooping
- Summary (*,G) entries are created and allowed to be programmed in hardware prior to multicast source arrival
  - The removal programming by CPU improves join latency
  - The multicast stream can be accessed faster
  - Less initial multicast packets are dropped
  - CPU overhead may also be reduced

Confidential | Authorized

Optimization 3 :-Preprogram (*,G) bridge entries configuration




```
vlan <ID>
    ip igmp snooping enable
    ip igmp snooping preprogram-starg-flow enable
```

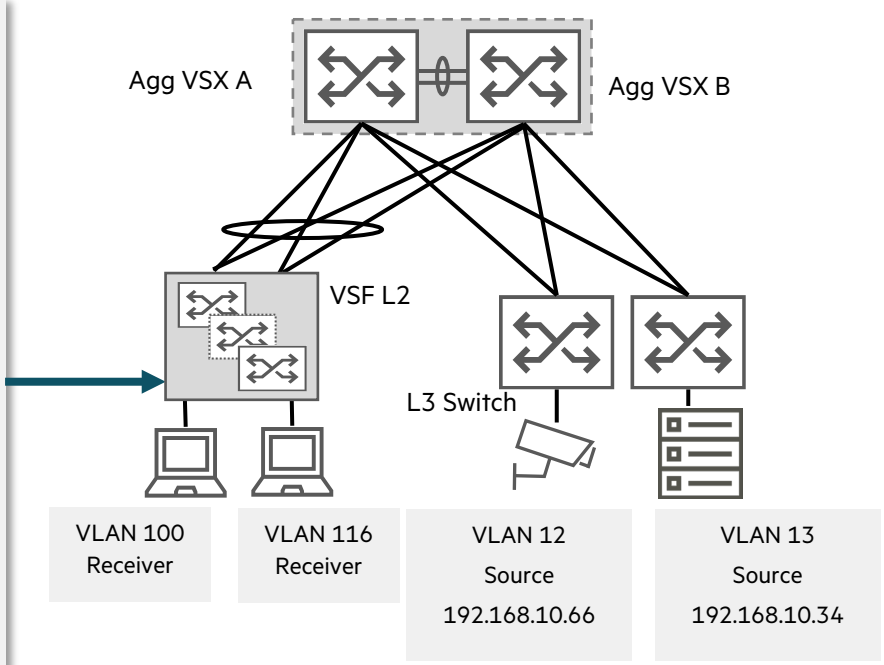
Note: For VSX pair configuration consistency is mandated for this feature, configure preprograming on the VSX pair devices. From a functionality/design perspective, this feature doesn't need any special functionality/design perspective

Platform/Feature	4100i	6000	6100	6200	6300	6400 (v1)	6400 (v2)	8100	8320	8325	8360 (v1/v2)	8400	9300	10000
Preprogramed (*,G) entries	Y	Y	Y	Y	Y	Y	Y	Y	N/A	N/A	Y	N/A	N/A	N/A

(* , G) Preprograming bridge entries- configuration and trouble shooting

1. igmp snooping configured on VLANs

```
hostname VSF-L2
!
vlan 100
    ip igmp snooping enable

vlan 116
    ip igmp snooping enable
!
interface 1/1/3
    no shutdown
    no routing
    vlan access 100
!
interface 1/1/4
    no shutdown
    no routing
    vlan access 116
```



3. preprogram VLAN 116 on VSF-L2

```
hostname VSF-L2
!
vlan 100
    ip igmp snooping enable

vlan 116
    ip igmp snooping enable
    ip igmp snooping preprogram-starg-flow enable
!
interface 1/1/3
    no shutdown
    no routing
    vlan access 100
!
interface 1/1/4
    no shutdown
    no routing
    vlan access 116
```

2. L2 multicast bridge entries hit CPU for both VLANs with (S,G)

```
show ip igmp snooping packet-exceptions
List of L2 Multicast Bridge entries for which data packets are hitting CPU

VRF: default

Vlan   Group Address          Source-Address          Packet Count   Last Seen Time
----   --------------------   ---------------------   ------------   ---------------
100    239.1.1.1/32           192.168.10.66/32        403            00h:07m:30s
116    239.2.2.1/32           192.168.10.34/32        17             00h:07m:10s
116    239.2.2.2/32           192.168.10.34/32        6              00h:07m:10s
116    239.2.2.3/32           192.168.10.34/32        5              00h:07m:10s
116    239.2.2.4/32           192.168.10.34/32        14             00h:07m:10s
116    239.2.2.5/32           192.168.10.34/32        408            00h:07m:10s
```

4. L2 multicast entries (*,G) for VLAN 116 no CPU hit

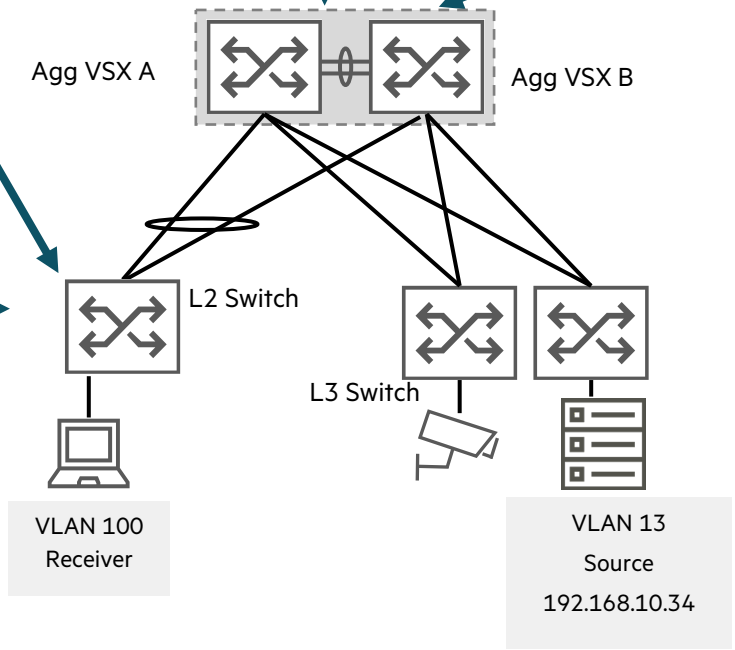
```
show ip igmp snooping packet-exceptions
List of L2 Multicast Bridge entries for which data packets are hitting CPU

VRF: default

Vlan   Group Address          Source-Address          Packet Count   Last Seen Time
----   --------------------   ---------------------   ------------   ---------------
100    239.1.1.1/32           192.168.10.66/32        404            00h:00m:20s
116    239.2.2.1/32           any                     0              00h:00m:00s
116    239.2.2.2/32           any                     0              00h:00m:00s
116    239.2.2.3/32           any                     0              00h:00m:00s
116    239.2.2.4/32           any                     0              00h:00m:00s
116    239.2.2.5/32           any                     0              00h:00m:00s
```

Optimization 3:– PIM and Preprograming (*,G) configuration

This optimization is only applicable for L2 switches that have IGMP snooping only i.e we have bridge entries

Where you have IGMP/PIM on L3 switch this optimization is not applicable i.e we have route entries

```
hostname L2-Switch
!
vlan 100
    ip igmp snooping enable
    ip igmp snooping preprogram-starg-flow enable
!
interface 1/1/3
    no shutdown
    no routing
    vlan access 100
!
```




```
hostname Agg-VSX
!
vlan 100
    ip igmp snooping enable
!
!
interface vlan 100
    ip address 192.168.10.34/28
    active-gateway ip mac 12:00:00:00:19:69
    active-gateway ip 192.168.10.34
    ip igmp enable
    ip pim-sm
```

```
show ip mroute 239.1.1.1

IP Multicast Route Entries

VRF : default
Total number of entries : 1

Group Address                  : 239.1.1.1
Source Address                 : any
Neighbor                       :
Incoming interface             : 1/1/3
Multicast Routing Protocol     : PIM-BIDIR
Unicast Routing Protocol       : OSPF
Uptime (HH:MM:SS)              : 00:01:11
Downstream Interface
Interface        State          VSX Role
-----------      ----------     --------
vlan100          forwarding     DF
```

```
show ip igmp snooping packet-exceptions
List of L2 Multicast Bridge entries for which data packets are hitting CPU

VRF: default

Vlan   Group Address          Source-Address          Packet Count   Last Seen Time
----   --------------------   ---------------------   ------------   ---------------
100    239.1.1.1/32           192.168.10.34/32        403            00h:07m:30s
```

Thank you

© 2023 Hewlett Packard Enterprise Development LP