AOS-CX 10.10 Update
June 2022

FIB optimization

aruba
a Hewlett Packard Enterprise company

Vincent Giles & Daryl Wan
Technical Marketing Engineering

Agenda

1. Overview
2. Use Cases
3. Details and Caveats
4. Configuration
5. Best Practices
6. Troubleshooting
7. Demo
8. Additional Resources

Definitions

Acronyms

- VXLAN — **V**irtual e**X**tensible **LAN**
- VTEP — **V**XLAN **T**unnel **E**nd **P**oint
- VNI — **V**XLAN **N**etwork **I**dentifier
- L2VNI — **L**ayer2 **V**XLAN **N**etwork **I**dentifier (to extend L2 traffic)
- L3VNI — **L**ayer3 **V**XLAN **N**etwork **I**dentifier (to send routed traffic)
- EVPN — **E**thernet **V**irtual **P**rivate **N**etwork
- MP-BGP — **M**ulti-**P**rotocol **B**order **G**ateway **P**rotocol
- AF — **A**ddress **F**amily (Ex: IPv4, IPv6 or EVPN address families used in MP-BGP)
- MP-BGP EVPN — Refers to the EVPN AF in MP-BGP
- RT — Refers to EVPN **R**oute-**T**ype or Type of Route: (AOS-CX supports RT2, RT3, RT5)
- VRF — **V**irtual **R**outing and **F**orwarding
- IRB — **I**ntegrated **R**outing and **B**ridging (symmetric or asymmetric IRB used in VXLAN overlay)
- VSX — **V**irtual **S**witching e**X**tension
- ISL — **I**nter **S**witch **L**ink (link between VSX peers)
- AG — **A**ctive **G**ateway (anycast IP address used for default-gateway)
- VSX VTEP — VTEP function hosted on a VSX cluster for dual-homing capability

- NHS — **N**ext-**H**op-**S**elf
- NHU — **N**ext-**H**op-**U**nchanged
- Border VTEP — VTEP acting as boundary for the Fabric
- Border-Leader — Border VTEP hosting BGP sessions with other Fabrics
- Fabric — Set of fully-meshed VTEPs for the VXLAN dataplane
- Local Fabric — internal Fabric (iBGP)
- Remote Fabric — external Fabric (eBGP)
- iBGP — internal BGP
- eBGP — external BGP
- ASN — Autonomous System Number (used in BGP)
- DCI — Data-Center-Interconnect
- POD — Point Of Delivery
- Routing table — Valid routing entries selected from each active routing protocols based on the administrative distance
- FIB — **F**orwarding **I**nformation **B**ase, active forwarding entries programmed into ASIC based on the routing table
- RIB — **R**outing **I**nformation **B**ase, selected and non-selected candidate routes per routing protocol

Overview

FIB optimization

Overview of EVPN host routes FIB optimization

- All VTEP access switches learn host routes (/32) from all connected devices per shared VRF.
  This overloads ASIC host-table on access switches which have typically smaller hardware-tables.
  (A separate lookup is performed into the ASIC FIB for hosts, using the IP Host table, 49k size for 6300).

- The proposed optimization allows to:
  - program in ASIC only active routed destination (mechanism called conversational host routes learning)
  - age out non-active EVPN host route entries after configured ageout timer.
  - deploy large scale EVPN Campus with switches having small/medium hw-tables.

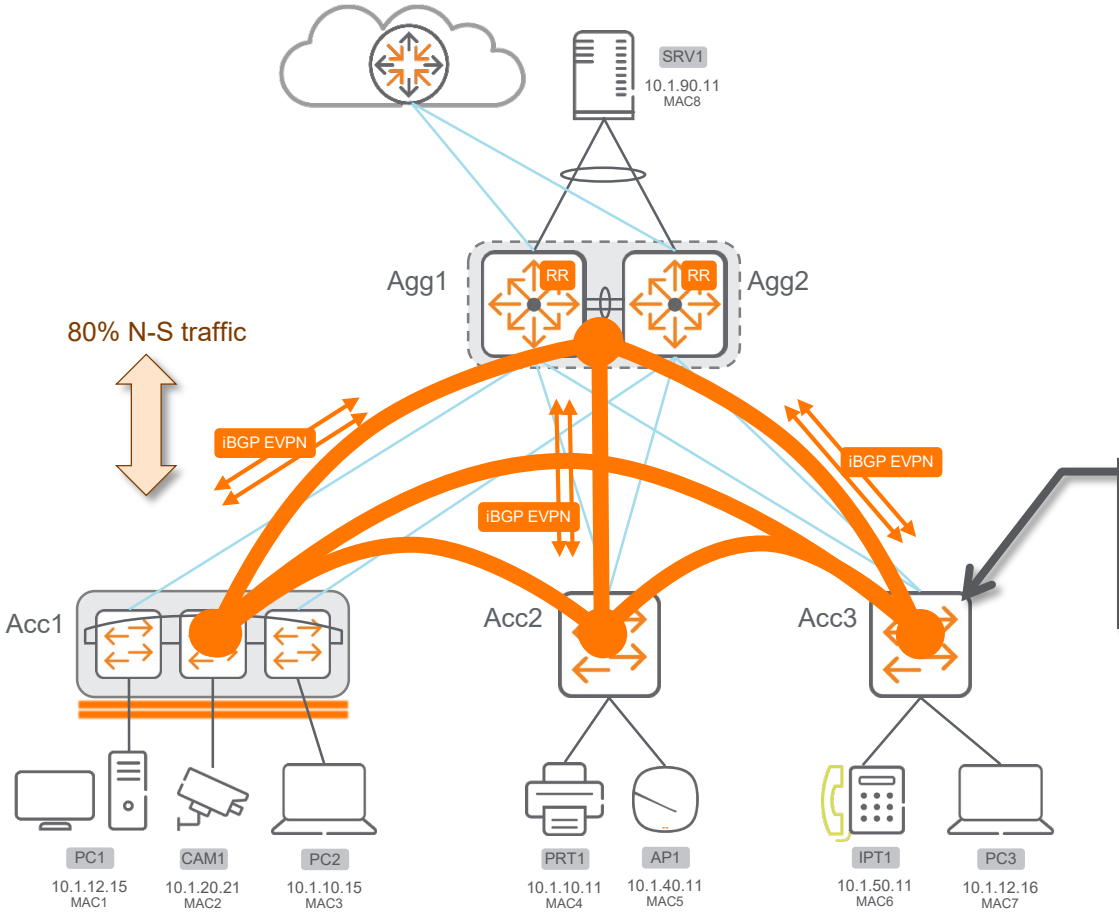
- Remote MACs are still programmed in ASIC for stretched VLANs.
  (i.e. there is no conversational MAC learning mechanism yet).
  If VLAN is not shared, remote MACs for non-stretched VLANs are not programmed in ASIC.

- FIB Optimization is also called: **A**ruba **I**ntelligent **F**orwarding (AIF)

Use Cases

Campus Use-case

Distributed GW – 80% North-South traffic



- Each VTEP learns all EVPN host routes even if many hosts never communicate together.
- ASIC host-table is unnecessary programmed with all host entries.

2 local host routes: IPT1 and PC3
5 remote host routes: PC1, CAM1, PC2, PRT1, AP1, SRV1

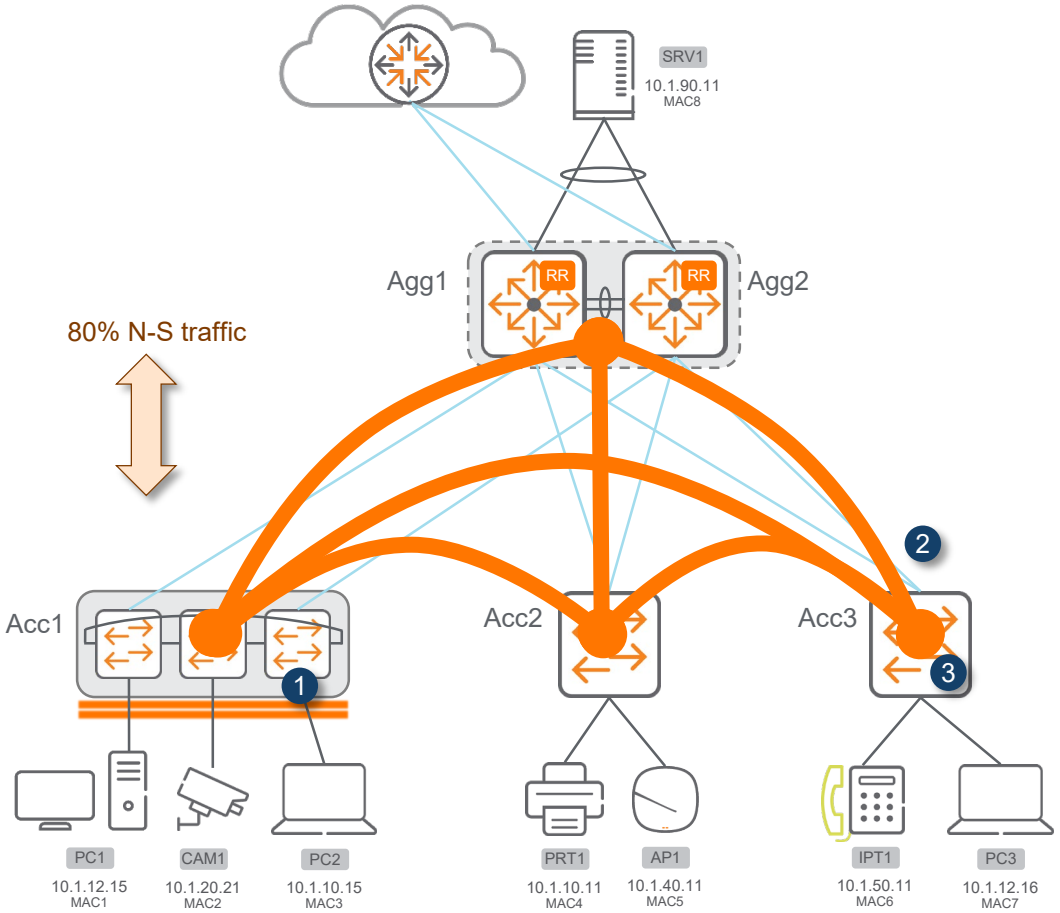
Acc3 VTEP has ~80% unnecessary host routes: CAM1, AP1, PC1…

20% E-W traffic

VXLAN Overlay Tunnel
L2 link
L3 link

Campus Use-case

Without FIB optimization



1. PC2 initiates traffic to upstream server.
   10.1.10.15/32 and MAC3 are dynamically learnt on Acc1 and populated in EVPN routes as RT-2.

2. PC2 10.1.10.15/32 EVPN RT-2 is received on Acc3

3. On Acc3, 10.1.10.15/32 is programmed in the ASIC host-table regardless it is the destination for an existing traffic flow or not.

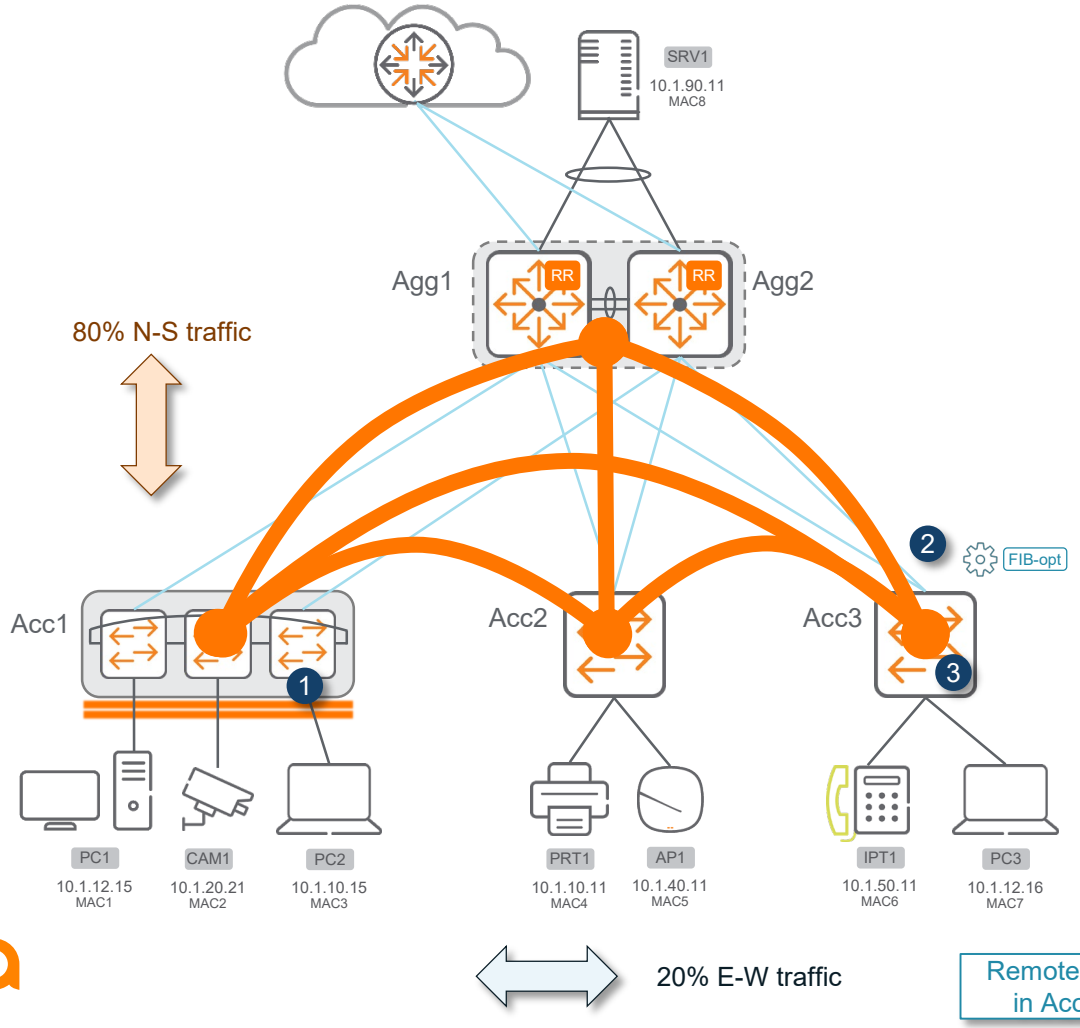
```
Acc3# show capacities-status l3-resources

System Capacities Status: Filter L3 Resources
Capacities Status Name                                                                    Value Maximum
-------------------------------------------------------------------------------------------------
Number of IP neighbor (IPv4+IPv6) entries                                                    43   49152
Number of IP Directed Broadcast neighbor entries                                              0    4096
Number of IPv4 neighbor(ARP) entries                                                         43   49152
Number of IPv6 neighbor(ND) entries                                                           0   49152
...
```

```
Acc3:/home/admin# ovs-appctl l3pd/show route -a
l3pd_lib_route_s information
  Total Cache Entries: 60, Cache Mask: 3F
  Host_routes Cache Entries: 40
  Route Cache Entries: 20
  * Indicates KEY field.

VRF*  IP-SUBNET*                MASK  L3-DEST-ID  ECMP-GROUP-ID  IPv6  DP-STATE           HA-STATE  SERVICE-LABEL-IDX  COMP-GRP-ID
-------------------------------------------------------------------------------------------------------------------------------
1     192.168.2.5               32    N/A         0x0000001D     NO    HOST_ROUTE_ECMP    NORMAL    0                  0
...
1     192.168.40.1              32    0x00000002  N/A            NO    HOST_ROUTE_SINGLE  NORMAL    0                  0
3     10.1.10.0                 24    0x00000065  N/A            NO    SINGLE             NORMAL    0                  0
3     10.1.10.15                32    0x00000065  N/A            NO    HOST_ROUTE_SINGLE  NORMAL    0                  0
```

Campus Use-case

With FIB optimization

Benefit: save space in ASIC host-table



1. PC2 initiates traffic to upstream server.
   10.1.10.15/32 and MAC3 are dynamically learnt on Acc1 and populated in EVPN routes as RT-2.
2. PC2 10.1.10.15/32 EVPN RT-2 is received on Acc3
3. On Acc3, with FIB-optimization enabled, 10.1.10.15/32 is no longer programmed in the ASIC host-table.

```
Acc3# show capacities-status l3-resources

System Capacities Status: Filter L3 Resources
Capacities Status Name                                                                    Value Maximum
-------------------------------------------------------------------------------------------------
Number of IP neighbor (IPv4+IPv6) entries                                                    21   49152
Number of IP Directed Broadcast neighbor entries                                              0    4096
Number of IPv4 neighbor(ARP) entries                                                         21   49152
Number of IPv6 neighbor(ND) entries                                                           0   49152
...
```

```
Acc3:/home/admin# ovs-appctl l3pd/show route -a
l3pd_lib_route_s information
  Total Cache Entries: 60, Cache Mask: 3F
  Host_routes Cache Entries: 40
  Route Cache Entries: 20
  * Indicates KEY field.

VRF*  IP-SUBNET*                MASK  L3-DEST-ID  ECMP-GROUP-ID  IPv6  DP-STATE           HA-STATE  SERVICE-LABEL-IDX  COMP-GRP-ID
-------------------------------------------------------------------------------------------------------------------------------
1     192.168.2.5               32    N/A         0x0000001D     NO    HOST_ROUTE_ECMP    NORMAL    0                  0
...
1     192.168.40.1              32    0x00000002  N/A            NO    HOST_ROUTE_SINGLE  NORMAL    0                  0
3     10.1.10.0                 24    0x00000065  N/A            NO    SINGLE             NORMAL    0                  0
```

10.1.10.15 /32 is not longer listed in this table

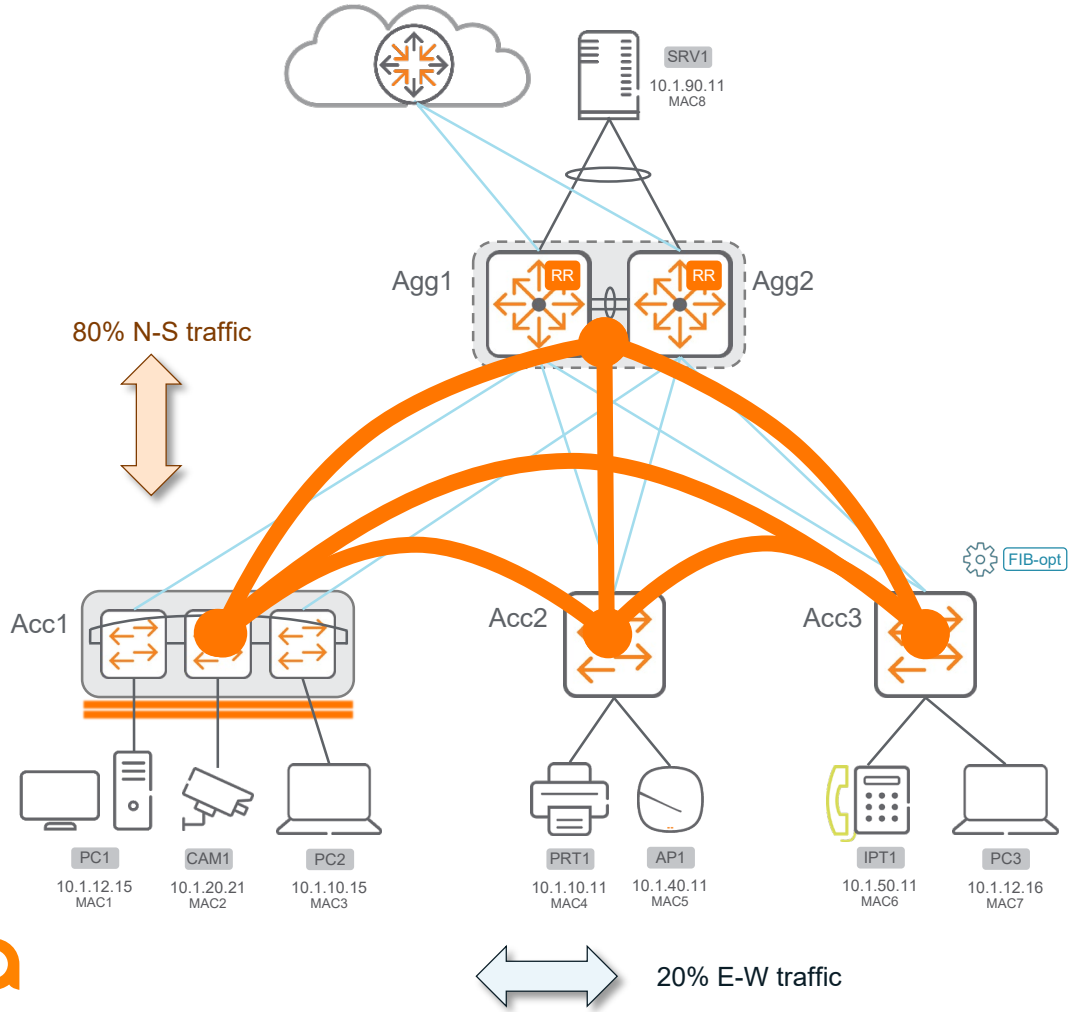
```
Acc3# show mac-address-table
MAC age-time                : 300 seconds
Number of MAC addresses : 2

MAC Address          VLAN     Type                      Port
-----------------------------------------------------------------
00:50:56:8e:43:7f    12       evpn                      vxlan1(192.168.2.5)
00:50:56:8e:fc:39    12       dynamic                   1/1/1
```

Remote MAC1 in same VLAN than VLAN12 present in Acc3 are learnt and still programmed in ASIC

Campus Use-case

FIB optimization – HW savings

80-90% of space in ASIC host-table is saved



- **Without FIB-optimization**:

Programmed host entries in ASIC =
number of hosts on average per VRF
* number of VRFs

- **With FIB-optimization**:

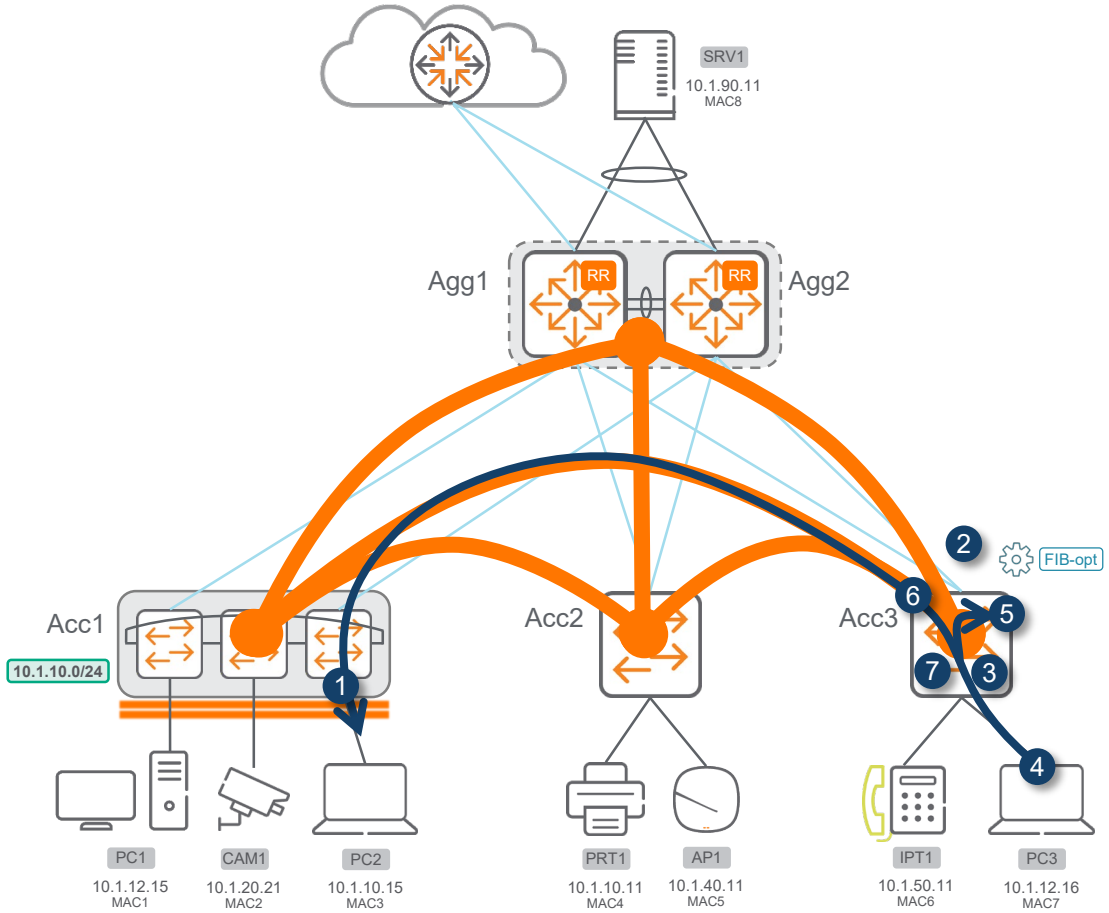
Programmed host entries in ASIC =
number of hosts on average per VRF
* number of VRFs * 0.2 (or even less)

E-W traffic might be 20% in volume and even less in term of number of involved hosts.

Details / Caveats

FIB-optimization process

Packet forwarding and conversational host-route learning



1. PC2 initiates traffic to upstream server.
   10.1.10.15/32 and MAC3 are dynamically learnt on Acc1 and populated in EVPN routes as RT-2.

2. PC2 10.1.10.15/32 EVPN RT-2 is received on Acc3

3. On Acc3, with FIB-optimization enabled, 10.1.10.15/32 is no longer programmed in the ASIC host-table.

4. PC3 initiates routed traffic to PC2. The packet is routed based on RT-5 information for destination subnet 10.1.10.0/24. If multiple VTEPs host this subnet, one is preferred (likely lowest router-id).

5. A **copy of the packet** is sent to Acc3 switch CPU until the host route entry is programmed in HW. This copy to CPU is only used to inform the CPU that, if this is an EVPN host route, the entry must be programmed in the ASIC host-table. There is no further SW forwarding and the copied packet is simply dropped after CPU took actions for ASIC programming. This packet drop after the CPU processing does not increment any drop counter as this is the expected behavior, and “packet passed” counter will increment in COPP policy. However, if too many packets get copied to the CPU, the COPP policy (default 100pps rate) would drop packets before copy, and such drops before the CPU can process the information will be reported as COPP drops.

   Outcome: EVPN host routes are programmed in hardware only when there is active traffic.

6. Daemon on Acc3 will keep polling the hit bit status from the ASIC for traffic to PC2 destination, to check if PC2 is still an active destination.

7. If there is no active traffic any longer, PC2 host route entry is removed from the ASIC host-table and marked as inactive (i.e. aged-out).

   Outcome: This conversational mechanism for L3VNI host routes preserves HW-table to get filled unnecessarily and allows large scale Campus deployment.

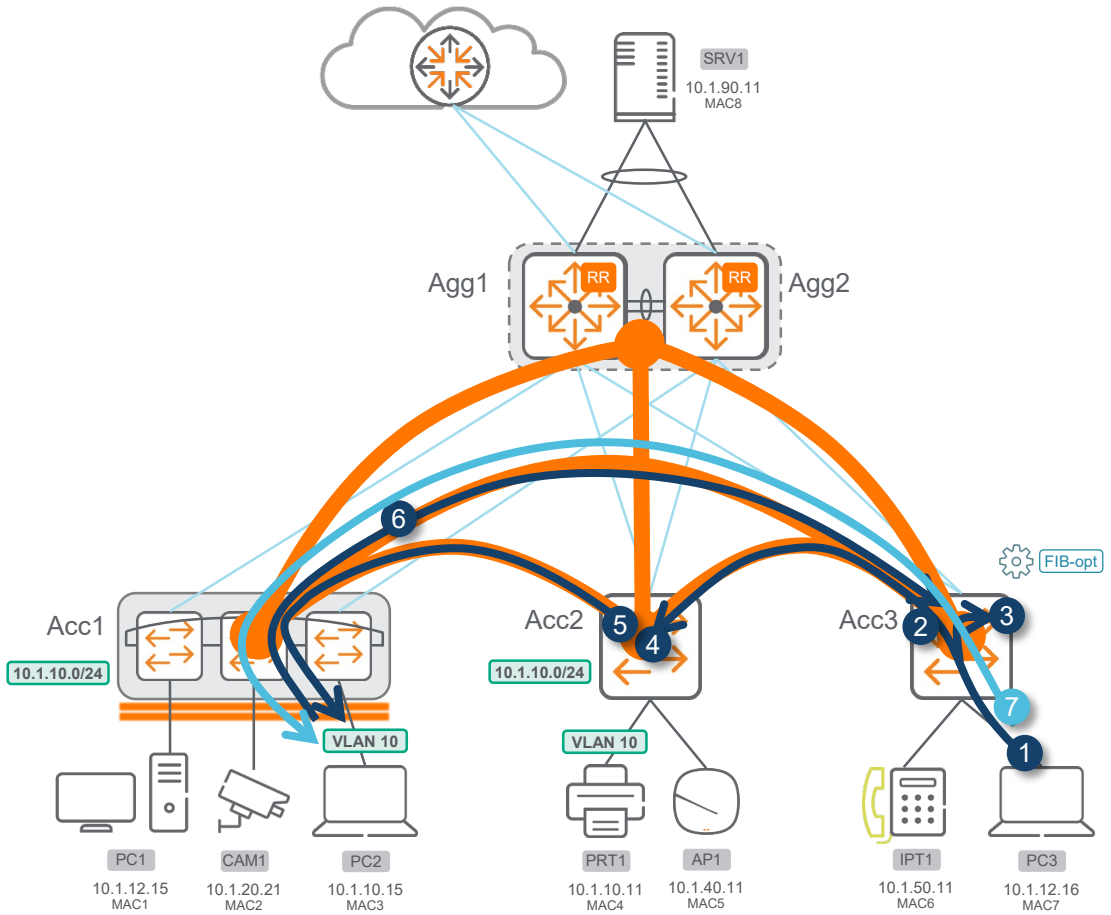
FIB optimization

Caveats

- IPv4 only
- First packet and potential sub-optimum path
- Stretched VLAN use-case
- Aggregation border VTEP: not supported
- COPP, border VTEP exclusion, cascaded L3 SW
- VSX

Caveat #1: potential sub-optimum path for first packet

Same RT-5 from different VTEPs

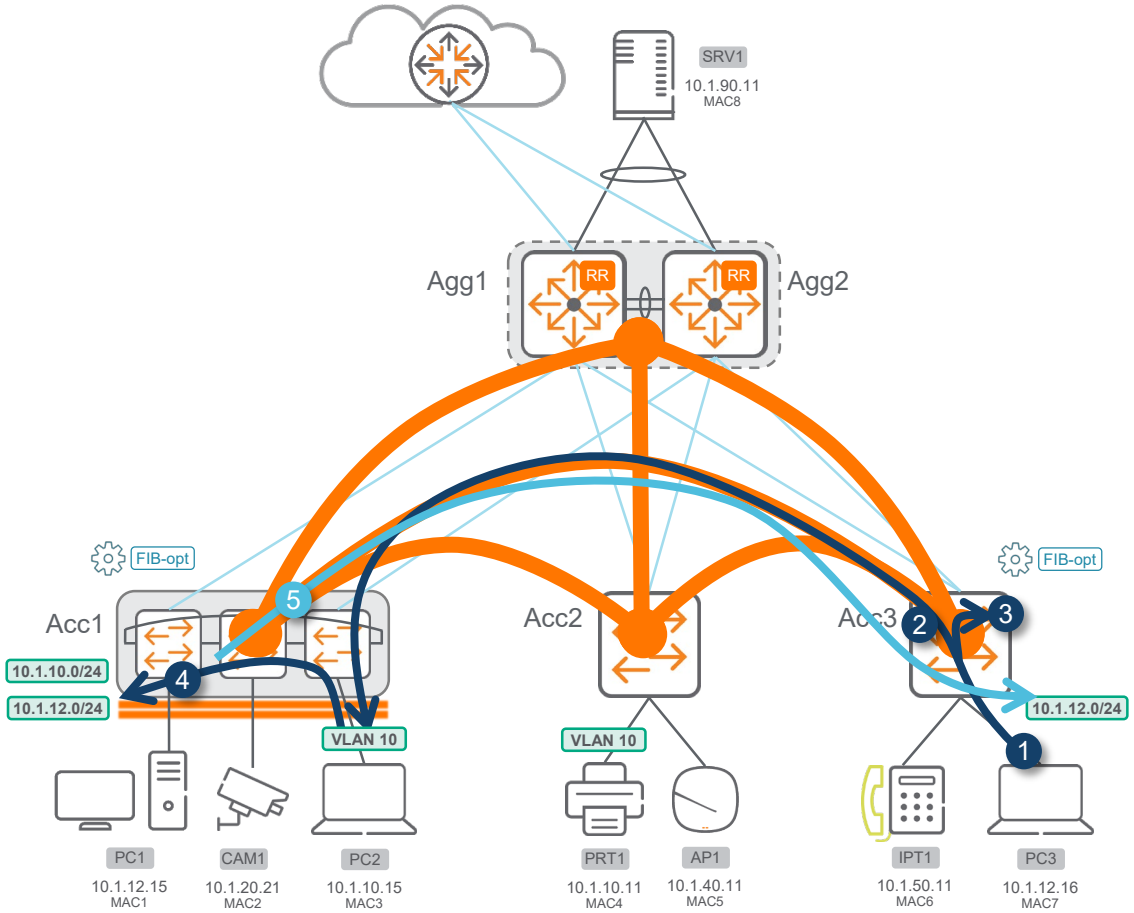


Assumption: FIB-optimization enabled **only on Acc3**.
10.1.10.15/32 is not programmed in the neighbor hw-table.

1. PC3 initiates routed traffic to PC2. The packet is routed based on RT-5 information for destination subnet 10.1.10.0/24.
   2 VTEPs own this 10.1.10.0/24 subnet, one is preferred (likely lowest router-id). Let's assume Acc2 is the preferred next-hop due to its lowest router-id.
2. Traffic is sent to Acc2 (say lowest router-id) over L3VNI.
3. A copy of the packet is send to Acc3 switch CPU. The PC2 entry is programmed in the ASIC host-table of Acc3.
4. Packet is received on Acc2 VTEP from L3VNI, and get pushed to L2VNI (VLAN10) as Acc2 has the PC2 entry already programmed in Acc2 host-table.
5. Acc2 VTEP sends packet to Acc1 over L2VNI for VLAN10. If PC2 is a silent host and its ARP is unknown, packet is forwarded as unknown-unicast packet over L2VNI.
6. PC2 response packet to PC3 over L3VNI directly to Acc3.
7. PC3 sends second packet to PC2 over L3VNI directly to Acc1.

Caveat #2: fib-optimization and stretched VLAN

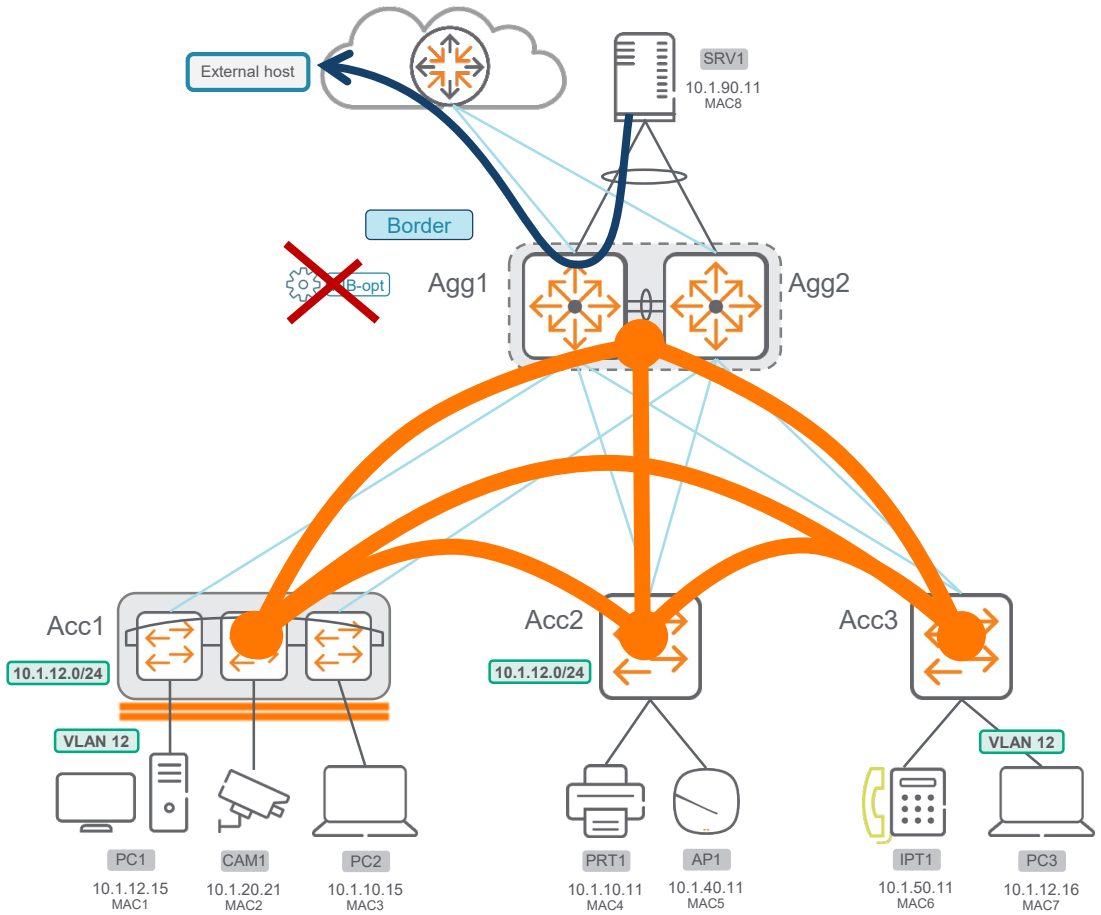
Destination VLAN presence on FIB-optimized switch



1. PC3 initiates routed traffic to PC2. The packet is routed based on RT-5 information for destination subnet 10.1.10.0/24.
   2 VTEPs own this 10.1.10.0/24 subnet, one is preferred (likely lowest router-id). Let's assume Acc1 is the preferred next-hop due to its lowest router-id.
2. Traffic is sent to Acc1 (say lowest router-id) over L3VNI.
3. A copy of the packet is send to Acc3 switch CPU. The PC2 entry is programmed in the ASIC host-table of Acc3.
4. Return packet is routed locally on Acc1 VTEP as SVI12 is local to Acc1 VTEP. The packet is not pushed to Acc3 VTEP as the 10.1.12.16/32 host route is not yet programmed in ASIC.
5. In this particular use-case, an small additional delay in establishing communication is seen and is being reduced in 10.10 maintenance release.

Caveat #3: Aggregation Border VTEP

No external EVPN Type-2 routes for external destination

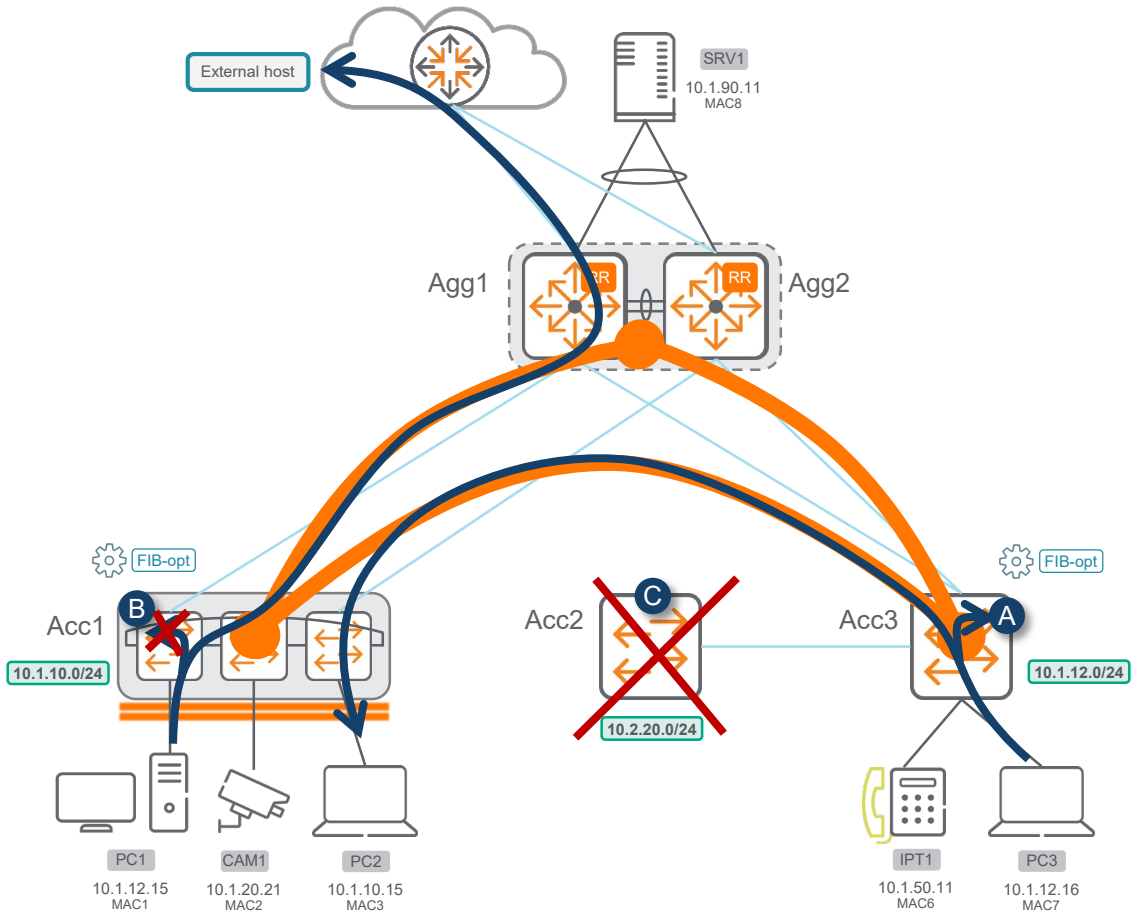


FIB-optimization is not recommended and not supported on aggregation VTEP

- If FIB-optimization is enabled on Aggregation border VTEP, any traffic to external destination is reported to CPU for programming host route.
- No Type-2 host route exist for this destination external to the fabric (only RT-5).
  Route optimization does not make sense for such external host.
- COPP fib-optimization would report lot of drops for CPU copy.

Caveat #4: COPP, border VTEP exclusion, cascaded L3 SW

COPP drops



A. For each flow initiation to an EVPN host destination, a **copy of the packet** is send to the switch CPU until the host route entry is programmed in HW.

If too many packets get copied to the CPU, the COPP policy (default 100pps rate) would drop packets before copy, and such drops before the CPU can process the information will be reported as COPP drops for AIF class.

Impact: If there are FIB-opt COPP drops, the host-route won't be programmed in the ASIC and the packet will keep be copied to CPU. This may trigger some avalanche effect…

```
6300F-4# show copp-policy default
class                drop priority rate pps burst pkts hardware rate pps
-------------------- ---- -------- -------- ---------- -----------------

fib-optimization                0        100      200        100
```

B. Each flow initiation to a destination external to the Fabric, RT-5 without any RT-2 entries, will end-up with packets always copied to CPU as not host entry is available in the EVPN routing table.

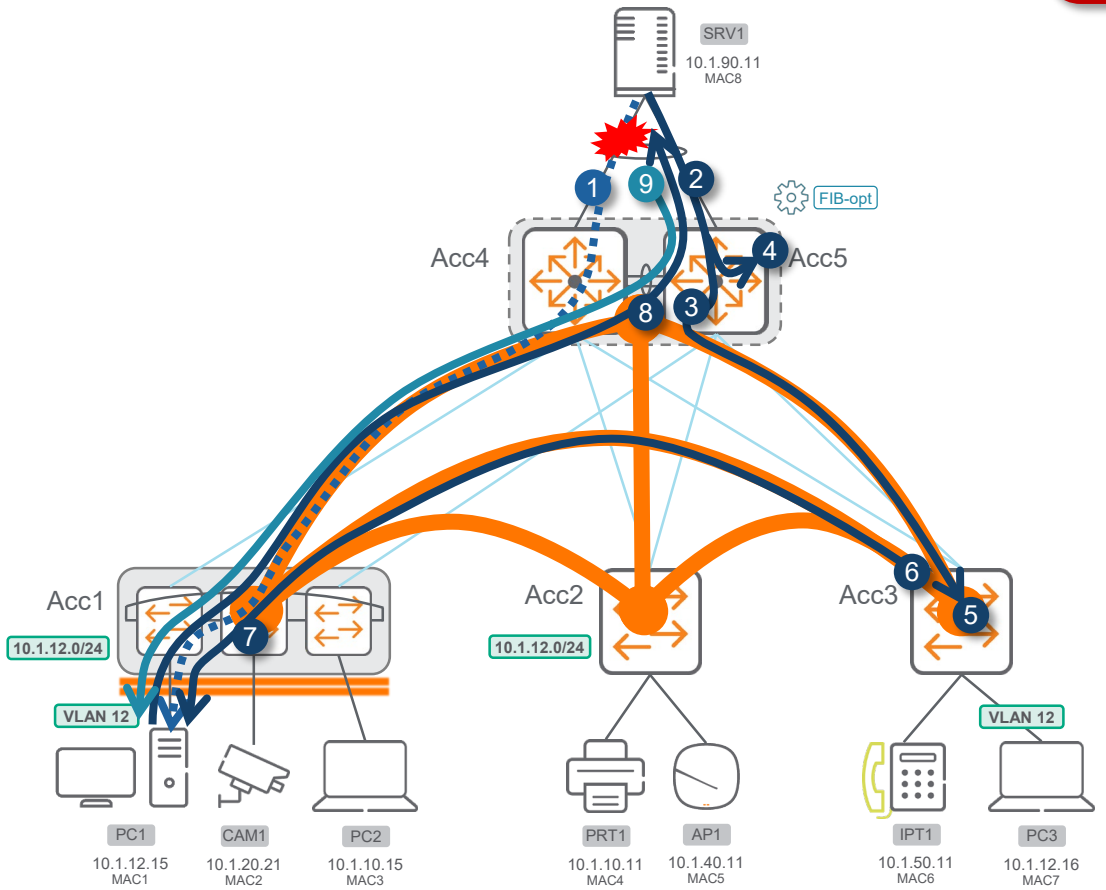
Border VTEP MUST be configured on each access VTEP in the list of next-hop exclusion, so that any destination route with border VTEP as next-hop will not be optimized and associated packet not sent to CPU.

C. No support for cascaded non-VTEP access L3 switch.

Caveat #5: VSX use-case

Impact during link failover

Reminders for VSX use-case:
- FIB-optimization is not recommended and not supported on aggregation layer
- Port access authentication is not synchronized between VSX peers
- In campus, not so common access use-case



1. Due to hashing, SRV1 use link to Acc4 to send traffic to PC1. PC1 host route is hw-programmed in Acc4, but not yet in Acc5. Let's assume link fails.
2. Traffic moves immediately on link to Acc5.
3. Packet is sent to RT-5 destination subnet: Acc1 or Acc3 VTEPs. Let's assume Acc3 is chosen.
4. Packet is also punted to CPU for programming the hw-table.
5. Packet reaches Acc3 VTEP.
6. Packet is broadcasted to VLAN12 and associated L2VNI and sent to Acc1.
7. Packet reaches Acc1 and is forwarded to PC1.
8. Return packet.
9. Subsequent packets from SRV1 to PC1.

Although FIB-optimization is not supported on **aggregation border** VSX VTEP, FIB-optimization was tested on VSX acting as access VTEP with measured impact < 1 second without L2VNI on access L2 link failure.

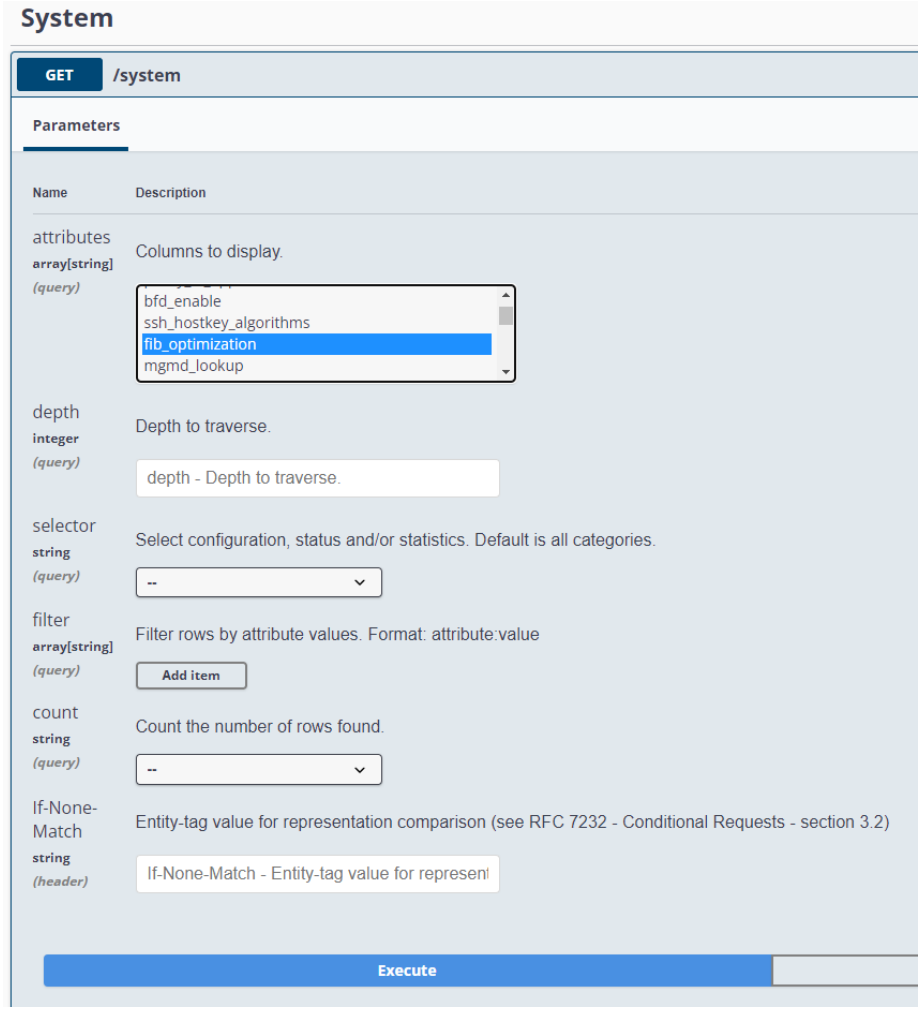
10.10 Platform Support

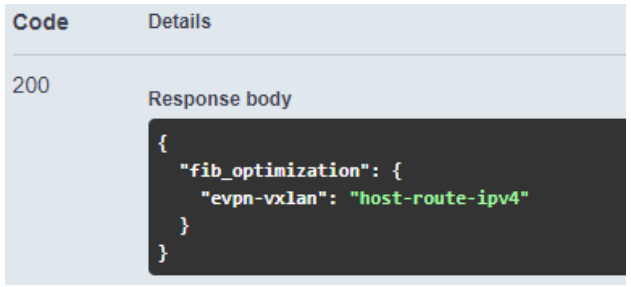
FIB optimization – IPv4 only

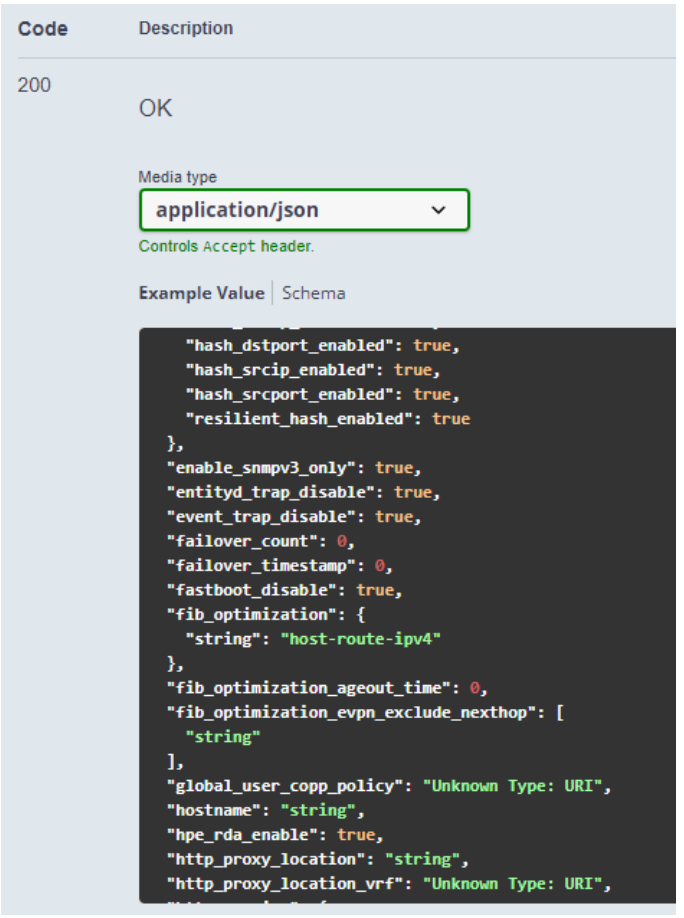
Platform	4100 6000 6100	6200	6300	6400 (v1/v2)	8320	8325	8360 (v1/v2)	8400	10000	Simulator
FIB-optim	No	No	**Yes**	**Yes**	No	No	**Yes**	No	No	No

API - system - fib_optimization

https://a.b.c.d/rest/v10.10/system?attributes=fib_optimization&depth=3







API - COPP - fib_optimization

https://a.b.c.d/rest/v10.10/system/hw_default_copp_policy/factory-default/cfg_cpes/fib_optimization?depth=3

Response body


```
{
  "burst": 200,
  "class": "fib_optimization",
  "hw_default": true,
  "priority": 0,
  "rate": 100
}
```


Code	Description
200	OK

Media type

application/json

Controls Accept header.

Example Value | Schema


```
{
  "burst": 0,
  "class": "string",
  "hw_default": true,
  "priority": 0,
  "rate": 0
}
```

API - COPP - fib_optimization - statistics

https://a.b.c.d/rest/v10.10/system?attributes=copp_statistics

Code	Details
200	Response body


```
"bfd_control_packets_passed": 0,
"bgp_packets_dropped": 0,
"bgp_packets_passed": 115341,
"captive_portal_packets_dropped": 0,
"captive_portal_packets_passed": 0,
"client_onboard_packets_dropped": 0,
"client_onboard_packets_passed": 0,
"default_packets_dropped": 0,
"default_packets_passed": 1433,
"dfp_collector_packets_dropped": 0,
"dfp_collector_packets_passed": 0,
"dhcp_packets_dropped": 0,
"dhcp_packets_passed": 0,
"erps_packets_dropped": 0,
"erps_packets_passed": 0,
"fib_optimization_packets_dropped": 0,
"fib_optimization_packets_passed": 1047297,
"icmp_broadcast_ipv4_packets_dropped": 0,
"icmp_broadcast_ipv4_packets_passed": 523647,
"icmp_multicast_ipv6_packets_dropped": 0,
"icmp_multicast_ipv6_packets_passed": 13220,
"icmp_security_ipv6_packets_dropped": 0,
"icmp_security_ipv6_packets_passed": 0,
"icmp_unicast_ipv4_packets_dropped": 0,
```

Configuration

Configuration

fib-optimization commands

- Command options

```
6300(config)# fib-optimization
  ageout-time  Configure Fib Optimization route age-out time
  evpn-vxlan   EVPN Configuration

6300(config)# fib-optimization evpn-vxlan
  exclude-nexthop  Configure next-hops to be excluded from FIB optimization
  host-route       Enable FIB optimization for host Routes

6300(config)# fib-optimization evpn-vxlan host-route
  ipv4  Enable FIB optimization for IPV4 Routes

6300(config)# fib-optimization evpn-vxlan host-route ipv4
  <cr>
```

Configuration

Enabling fib-optimization

- Enabling

```
6300(config)# fib-optimization evpn-vxlan host-route ipv4
```

- Disabling

```
6300(config)# no fib-optimization evpn-vxlan host-route ipv4
```

- After fib-optimization is enabled, 2 cases:
  - **without clear bgp**: the current host routes are optimized after the ageout timer (removed from ASIC after ageout time)
  - **with clear bgp**: the host routes are immediately optimized and not programmed in ASIC

Configuration

Age-out timer

- Default ageout-time is 90 seconds

```
6300(config)# fib-optimization ageout-time
  <60-3600>  Route age-out time in seconds (Default: 90 seconds)
```

- What about reducing the age-out timer ?
  - **Benefit**: free-up very frequently the hw-table to keep accommodating high number of new destination hosts
  - **Drawback**: if traffic to many hosts stops and resumes regularly under a time-period higher and close to ageout time, there will be more CPU demand to process aging of host entries.

    There is a internal mechanism to age-out host entries in batch in order to protect the CPU while removing lot of host entries (ex: 10K+) at the same time. This will induce some additional delay in aging an entry (i.e. ageout timer + batching_delay).

Configuration

Excluding host routes behind identified VTEP next-hop from fib-optimization

- Exclusion

```
6300(config)# fib-optimization evpn-vxlan exclude-nexthop
  A.B.C.D  IP address of the next-hop to be excluded
```

- Exclusion removal

```
6300(config)# no fib-optimization evpn-vxlan exclude-nexthop
  A.B.C.D  IP address of the next-hop to be excluded
```

- Host routes of which next-hop is configured in fib-optimization exclusion are programmed in ASIC as soon as EVPN RT-2 is learnt.
- It is not possible, in this release, to exclude a list of hosts.

Configuration

vsx-sync fib-optimization

```
vsx
   system-mac 00:00:00:01:02:01
   inter-switch-link lag 256
   role primary
   vsx-sync fib-optimization
```

```
8360(config-vsx)# vsx-sync
  aaa                           Sync all AAA instances
  acl-log-timer                 Sync access-list log timer instance
  arp-security                  Sync all ARP security configurations
  bfd-global                    Sync all BFD global configuration
  bgp                           Sync all BGP, ip aspath list, community list,
                                prefix list, route map configurations
  control-plane-acls            Sync all Control-plane Access-list instances
  copp-policy                   Sync all CoPP instances
  dcb-global                    Sync global configurations for DCB features
                                (DCBx, PFC and ETS)
  dhcp-relay                    Sync all DHCP RELAY instances
  dhcp-server                   Sync all DHCPv4-Server and DHCPv6-Server
                                instances
  dhcp-snooping                 Sync all DHCPv4-Snooping and DHCPv6-Snooping
                                instances.
  dns                           Sync all DNS instances
  evpn                          Sync all evpn configurations
  fib-optimization              Sync all FIB optimization configurations
  gbp                           Sync for all GBP
  hardware-high-capacity-tcam   Sync High capacity TCAM/LPM configuration
  icmp-tcp                      Sync all icmp and tcp instances
  keychain                      Sync all keychain configurations
  lldp                          Sync all LLDP instances
  loop-protect-global           Sync all Loop-protect global configuration
  mac-lockout                   Sync all mac lockout configurations
  macsec                        Sync all MACsec and MKA policies
  mclag-interfaces              Sync QoS, LACP, Loop-Protect, LAG description,
                                sFlow, STP, Rate-Limits, Vlans, ACLs, MACsec,
                                private-vlan-port-type and Portfilters for MCLAG
                                interface instances
  mdns-sd-global                Sync all mDNS configurations
  mgmd-global                   Sync all MGMD global instances
  msdp-global                   Sync all MSDP global instances
  nd-snooping                   Sync all ND-Snooping instances.
  neighbor                      Sync all IPv4 and IPv6 neighbor configurations
  ospf                          Sync all OSPF instances
  pim                           Sync router PIM context configuration
  policy-global                 Sync all policy global instances
  qos-global                    Sync all QoS global instances
  rip                           Sync all RIP configurations
  route-map                     Sync all ip aspath list, community list, prefix
                                list, route map configurations
  sflow-global                  Sync all sFlow global instances
  snmp                          Sync all SNMP instances
  ssh                           Sync all SSH instances
  static-routes                 Sync all Static Routes instances
  stp-global                    Sync all STP Global Configuration
  time                          Sync all time instances
  udp-forwarder                 Sync all UDP FORWARDER instances
  vrrp                          Sync all VRRP instances
  vsx-global                    Sync all VSX global configuration
  <cr>
```

Best Practices

FIB optimization

Best practices

Campus

- It is recommended to **enable fib-optimization** when there is no or little east-west overlay traffic (<50% of traffic)
- If there is lot of clients participating to east-west overlay traffic (>50% of traffic), **disable fib-optimization**.
  It may be counter-productive and drops in COPP policy should be closely monitored if enabled under such condition.
- Use default age-out timer (90s).
- Use next-hop exclusion for identified next-hop hosting destination of frequent regular traffic.
- On each access VTEP switch where fib-optimization is enabled, configure “exclude-nexthop” with the border VTEP IP address.

DataCenter

- It is not recommended to enable fib-optimization as there is typically lot of workloads participating in east-west traffic in a given VRF.
- If enabled, special care is required on:
  - COPP policy drops.
  - Sub-optimum traffic path to optimized host-route (used RT-5, might reach wrong VTEP, then L2VNI)
  - VSX MCLAG link failover

Troubleshooting

show fib-optimization

status / age-out timer / excluded nexthops

```
6300# show fib-optimization configuration
Address family       : EVPN IPv4
Operation status     : Enabled
Route age-out time   : 90
Excluded nexthops    :
```

show ip route fib-optimization

Optimized host-routes

```
6300# show ip route fib-optimization vrf VRF1

EVPN ipv4 host routes optimized by Aruba Intelligent Forwarding

Origin Codes: C - connected, S - static, L - local
              R - RIP, B - BGP, O - OSPF
Type Codes:   E - External BGP, I - Internal BGP, V - VPN, EV - EVPN
              IA - OSPF internal area, E1 - OSPF external type 1
              E2 - OSPF external type 2

VRF: VRF1

Prefix              Nexthop          Interface     VRF(egress)       Origin/    Distance/    Age
                                                                     Type       Metric
-------------------------------------------------------------------------------------------------
-
10.1.10.1/32        192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.10.10/32       192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.10.13/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:34m:35s
10.1.10.15/32       192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.10.18/32       192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.11.1/32        192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.11.11/32       192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.12.14/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:34m:37s
10.1.12.15/32       192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.3/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.4/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.5/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.6/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.103/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.105/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.21.5/32     192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.21.8/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.21.105/32   192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.1/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.2/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.101/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.103/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s

Total Route Count : 22
```

```
6300# show ip route fib-optimization summary all-vrfs

 IPv4 Routes Optimized Summary

 VRF name :  VRF1
 Number of evpn routes optimized :   22
```

All host routes that are listed here are not in the ASIC !

show ip route fib-optimization

Traffic aging-out

```
6300# show ip route fib-optimization 10.1.10.15 vrf VRF1

VRF: VRF1

   Prefix           : 10.1.10.15/32                          VRF(egress)        : -
   Nexthop          : 192.168.2.5                            Interface          : -
   Origin           : bgp                                    Type               : bgp_evpn
   Distance         : 200                                    Metric             : 0
   Age              : 02h:47m:35s                            Tag                : 0
   Encap Type       : vxlan                                  Encap Details      : l3vni 100001
```


```
6300# show ip route fib-optimization 10.1.10.15 vrf VRF1

No EVPN ipv4 host routes optimized
```


```
6300# show ip route fib-optimization 10.1.10.15 vrf VRF1

No EVPN ipv4 host routes optimized
```


```
6300# show ip route fib-optimization 10.1.10.15 vrf VRF1

VRF: VRF1

   Prefix           : 10.1.10.15/32                          VRF(egress)        : -
   Nexthop          : 192.168.2.5                            Interface          : -
   Origin           : bgp                                    Type               : bgp_evpn
   Distance         : 200                                    Metric             : 0
   Age              : 02h:47m:35s                            Tag                : 0
   Encap Type       : vxlan                                  Encap Details      : l3vni 100001
```

show ip route

still includes host routes

routing table still include host routes

```
6300# show ip route vrf VRF1

Displaying ipv4 routes selected for forwarding

Origin Codes: C - connected, S - static, L - local
              R - RIP, B - BGP, O - OSPF
Type Codes:   E - External BGP, I - Internal BGP, V - VPN, EV - EVPN
              IA - OSPF internal area, E1 - OSPF external type 1
              E2 - OSPF external type 2

VRF: VRF1

Prefix              Nexthop          Interface     VRF(egress)       Origin/    Distance/    Age
                                                                     Type       Metric
-------------------------------------------------------------------------------------------------
10.1.10.0/24        192.168.2.5      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.10.1/32        192.168.2.5      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.10.10/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.10.13/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:04m:31s
10.1.10.15/32       192.168.2.5      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.10.18/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.11.0/24        192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.11.1/32        192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.11.11/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
10.1.12.0/24        -                vlan12        -                 C          [0/0]        -
10.1.12.1/32        -                vlan12        -                 L          [0/0]        -
10.1.12.14/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:04m:32s
10.1.12.15/32       192.168.2.5      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.11.3/32     192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.11.4/32     192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.11.5/32     192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.11.6/32     192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.11.103/32   192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.11.105/32   192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.21.5/32     192.168.2.5      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.21.7/32     -                loopback12    -                 L          [0/0]        -
192.168.21.8/32     192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.21.105/32   192.168.2.5      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.41.1/32     192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.41.2/32     192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.41.101/32   192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.41.103/32   192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.101.0/30    192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s
192.168.110.0/31    192.168.2.8      -             -                 B/EV       [200/0]      00h:41m:56s

Total Route Count : 29
```

l3-resmgrd aif/show is reflecting the routing table: 29 routes

```
6300:/home/admin# ovs-appctl -t l3-resmgrd aif/show
VRF          PREFIX                              DP-STATE     FROM       CPU-COPY   ENABLE-AGEOUT   AIF-OPTIMISATION   NEXTHOP-ID   FORWARDING-ID
-------------------------------------------------------------------------------------------------------------------------------------------------------
VRF1         192.168.41.2/32                     disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.12.15/32                       disabled     bgp        false      true            inactive           0xc537fd39   N/A
VRF1         192.168.110.0/31                    single       bgp        true       false           not_applicable     0xa37fb3ee   0x000000A2
VRF1         10.1.12.0/24                        connected    connected  false      false           not_applicable     0x712e6591   N/A
VRF1         192.168.41.1/32                     disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.11.1/32                        disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         192.168.11.105/32                   disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.10.1/32                        disabled     bgp        false      true            inactive           0xc537fd39   N/A
VRF1         10.1.11.0/24                        single       bgp        true       false           not_applicable     0xa37fb3ee   0x000000A2
VRF1         192.168.41.101/32                   disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.12.1/32                        disabled     local      false      false           not_applicable     0x712e6591   N/A
VRF1         192.168.11.3/32                     disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.12.14/32                       disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.11.11/32                       disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.10.0/24                        single       bgp        true       false           not_applicable     0xc537fd39   0x000000A1
VRF1         192.168.21.105/32                   disabled     bgp        false      true            inactive           0xc537fd39   N/A
VRF1         10.1.10.18/32                       disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         192.168.11.4/32                     disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         192.168.21.8/32                     disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         192.168.21.7/32                     disabled     local      false      false           not_applicable     0xde469609   N/A
VRF1         192.168.21.5/32                     disabled     bgp        false      true            inactive           0xc537fd39   N/A
VRF1         192.168.11.5/32                     disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.10.15/32                       disabled     bgp        false      true            inactive           0xc537fd39   N/A
VRF1         192.168.11.6/32                     disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         192.168.101.0/30                    single       bgp        true       false           not_applicable     0xa37fb3ee   0x000000A2
VRF1         10.1.10.13/32                       disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         10.1.10.10/32                       disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         192.168.11.103/32                   disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
VRF1         192.168.41.103/32                   disabled     bgp        false      true            inactive           0xa37fb3ee   N/A
```

22 optimized host routes

```
6300# show ip route fib-optimization vrf VRF1

EVPN ipv4 host routes optimized by Aruba Intelligent Forwarding
Origin Codes: C - connected, S - static, L - local
              R - RIP, B - BGP, O - OSPF
Type Codes:   E - External BGP, I - Internal BGP, V - VPN, EV - EVPN
              IA - OSPF internal area, E1 - OSPF external type 1
              E2 - OSPF external type 2
VRF: VRF1

Prefix              Nexthop          Interface     VRF(egress)       Origin/    Distance/    Age
                                                                     Type       Metric
-------------------------------------------------------------------------------------------------
10.1.10.1/32        192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.10.10/32       192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.10.13/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:34m:35s
10.1.10.15/32       192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.10.18/32       192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.11.1/32        192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.11.11/32       192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
10.1.12.14/32       192.168.2.8      -             -                 B/EV       [200/0]      00h:34m:37s
10.1.12.15/32       192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.3/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.4/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.5/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.6/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.103/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.11.105/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.21.5/32     192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.21.8/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.21.105/32   192.168.2.5      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.1/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.2/32     192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.101/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s
192.168.41.103/32   192.168.2.8      -             -                 B/EV       [200/0]      02h:44m:09s

Total Route Count : 22
```

COPP

Before active traffic

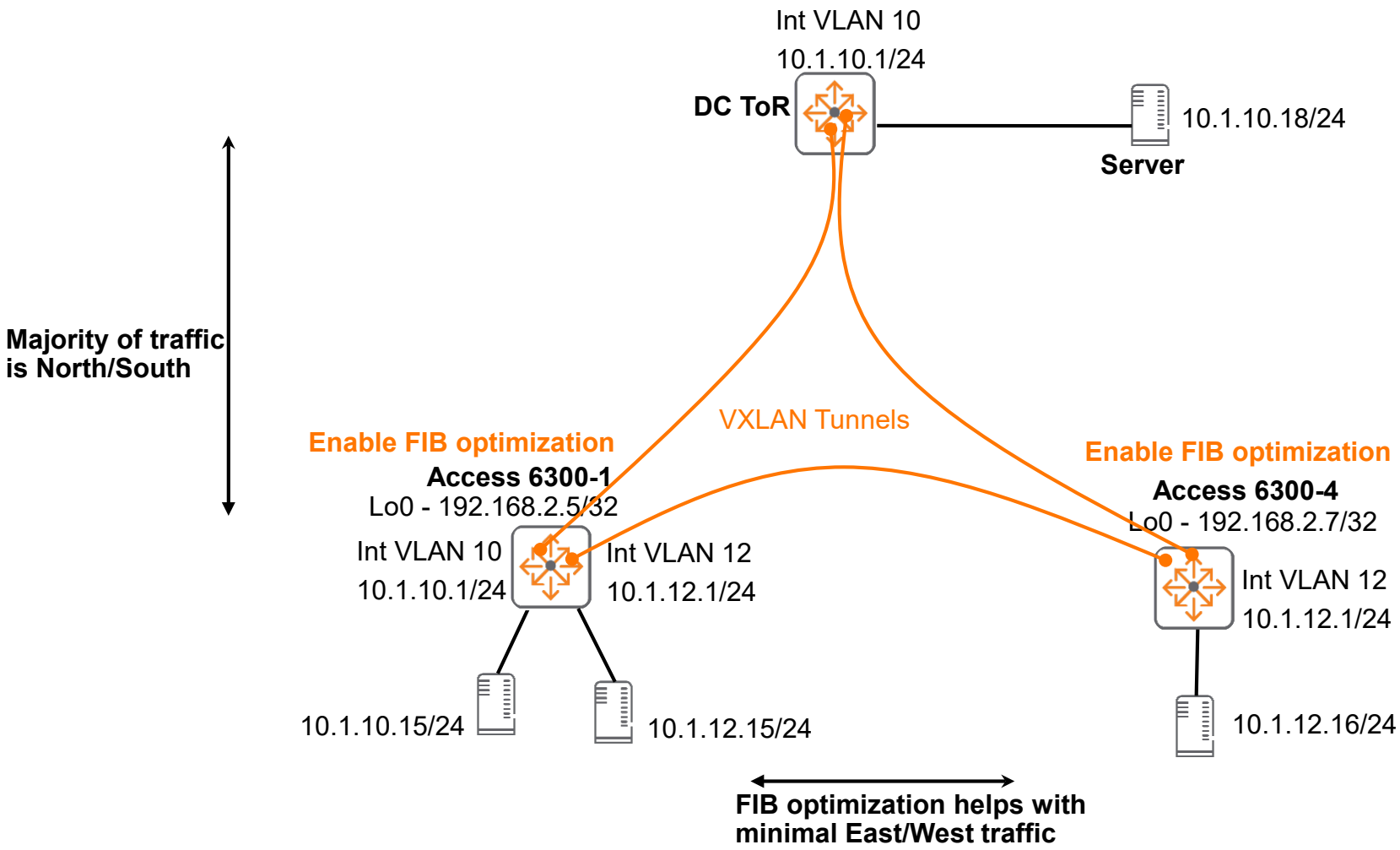
```
6300# show copp-policy statistics class fib-optimization
Statistics for CoPP policy 'default':
Class: fib-optimization
Description: Forwarding Information Base (FIB) Optimization.
    priority             : 0
    rate (pps)           : 100
    burst size (pkts)    : 200

    packets passed   : 663307            packets dropped  : 0
```

After traffic starts

```
6300# show copp-policy statistics class fib-optimization
Statistics for CoPP policy 'default':
Class: fib-optimization
Description: Forwarding Information Base (FIB) Optimization.
    priority             : 0
    rate (pps)           : 100
    burst size (pkts)    : 200

    packets passed   : 663312            packets dropped  : 0
```

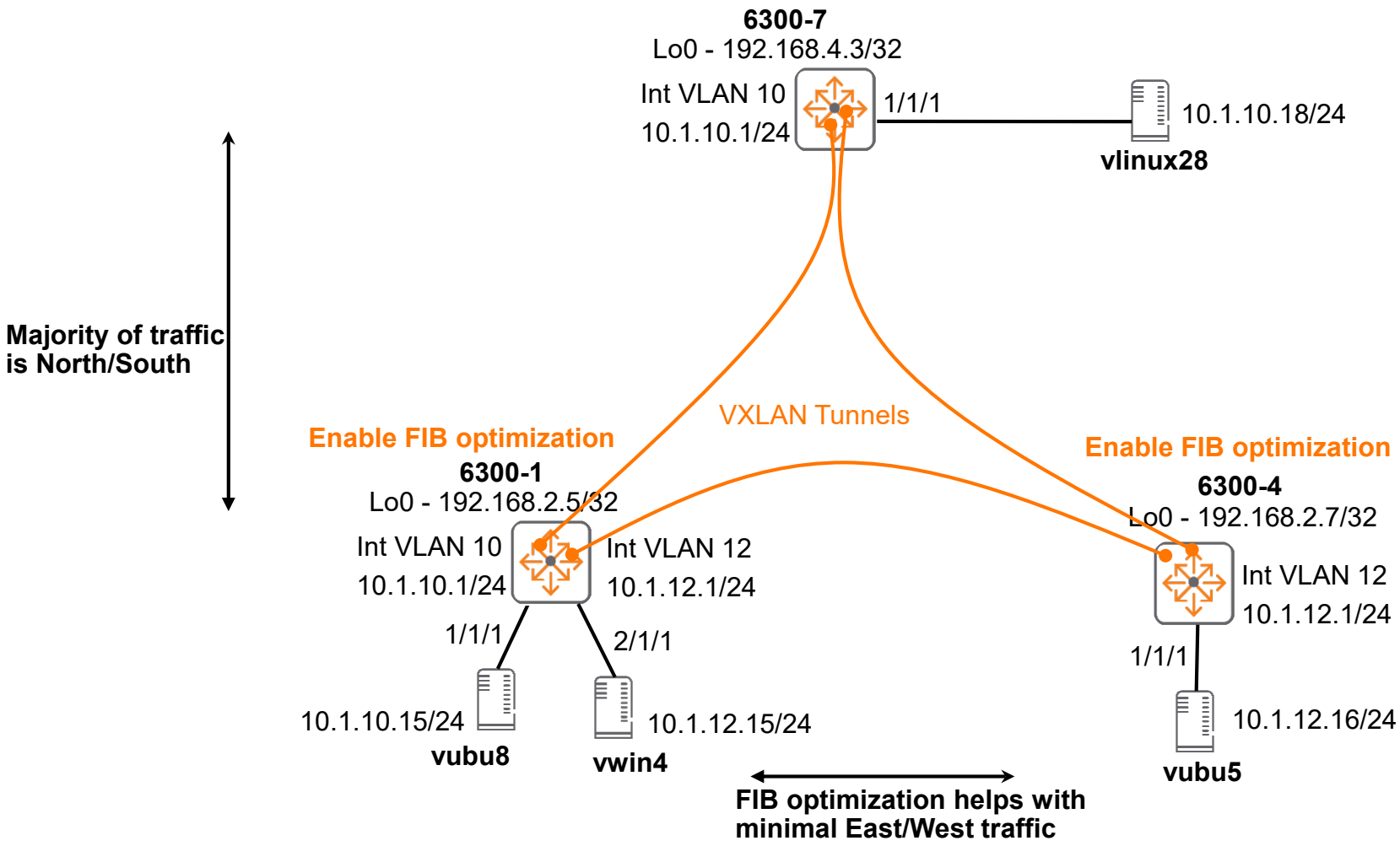
Demonstration

FIB-optimization Demo



- Show FIB optimization is not active for NS traffic on 6300-4
- Show FIB optimization for east/west traffic between clients on the same subnet or different subnets on different VTEPs on 6300-4
- Show CLI to prevent optimization of host routes on 6300-4 towards 6300-1

FIB-optimization Demo



- Show FIB optimization is not active for NS traffic on 6300-4
- Show FIB optimization for east/west traffic between clients on the same subnet or different subnets on different VTEPs on 6300-4
- Show CLI to prevent optimization of host routes on 6300-4 towards 6300-1

Resources

Feature/Solution References

- User Guides update:
  - VXLAN (10.10: https://www.arubanetworks.com/techdocs/AOS-CX/10.10/PDF/vxlan.pdf)

aruba
a Hewlett Packard Enterprise company

Thank you

vincent.giles@hpe.com
daryl.wan@hpe.com