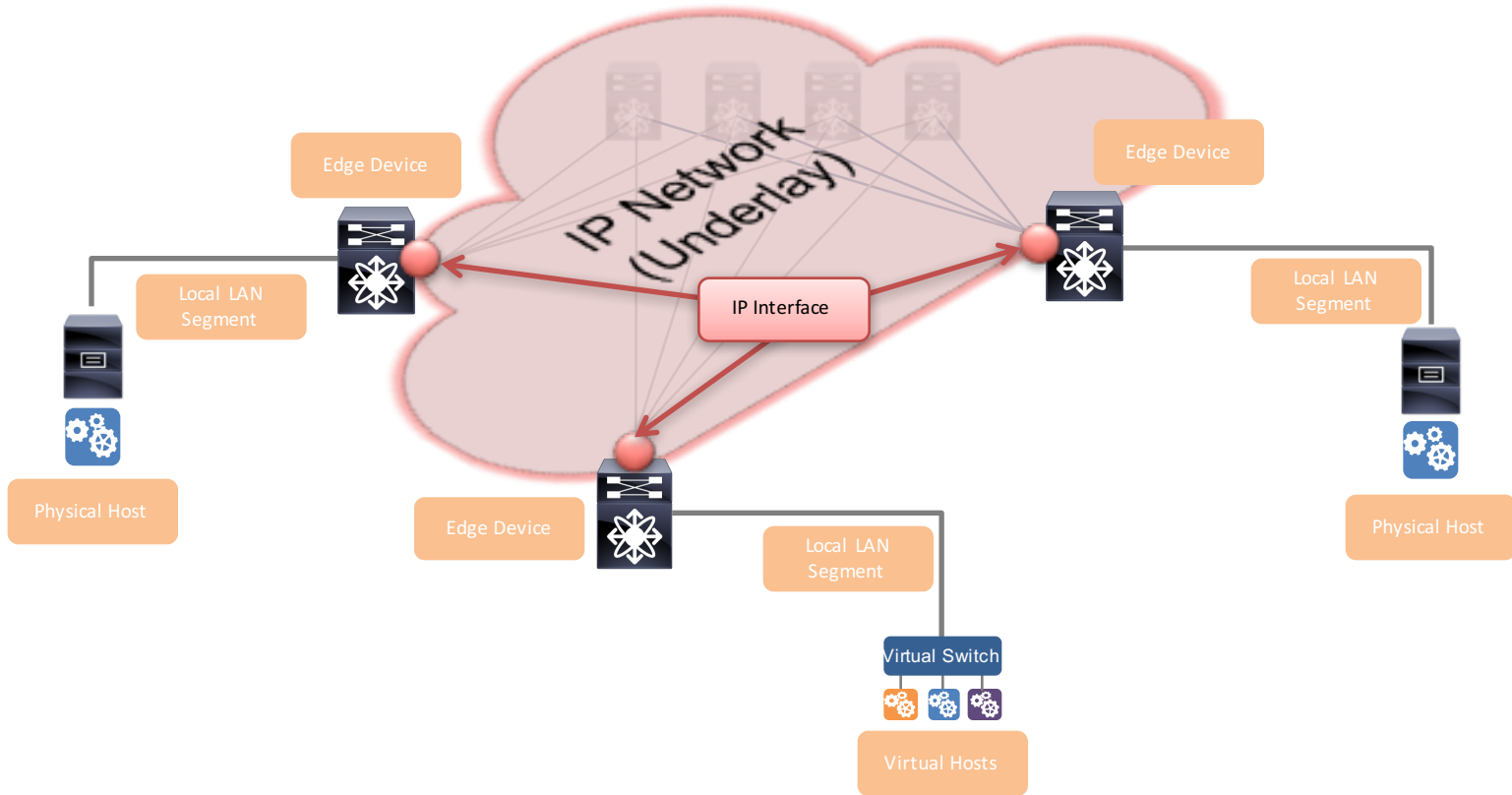




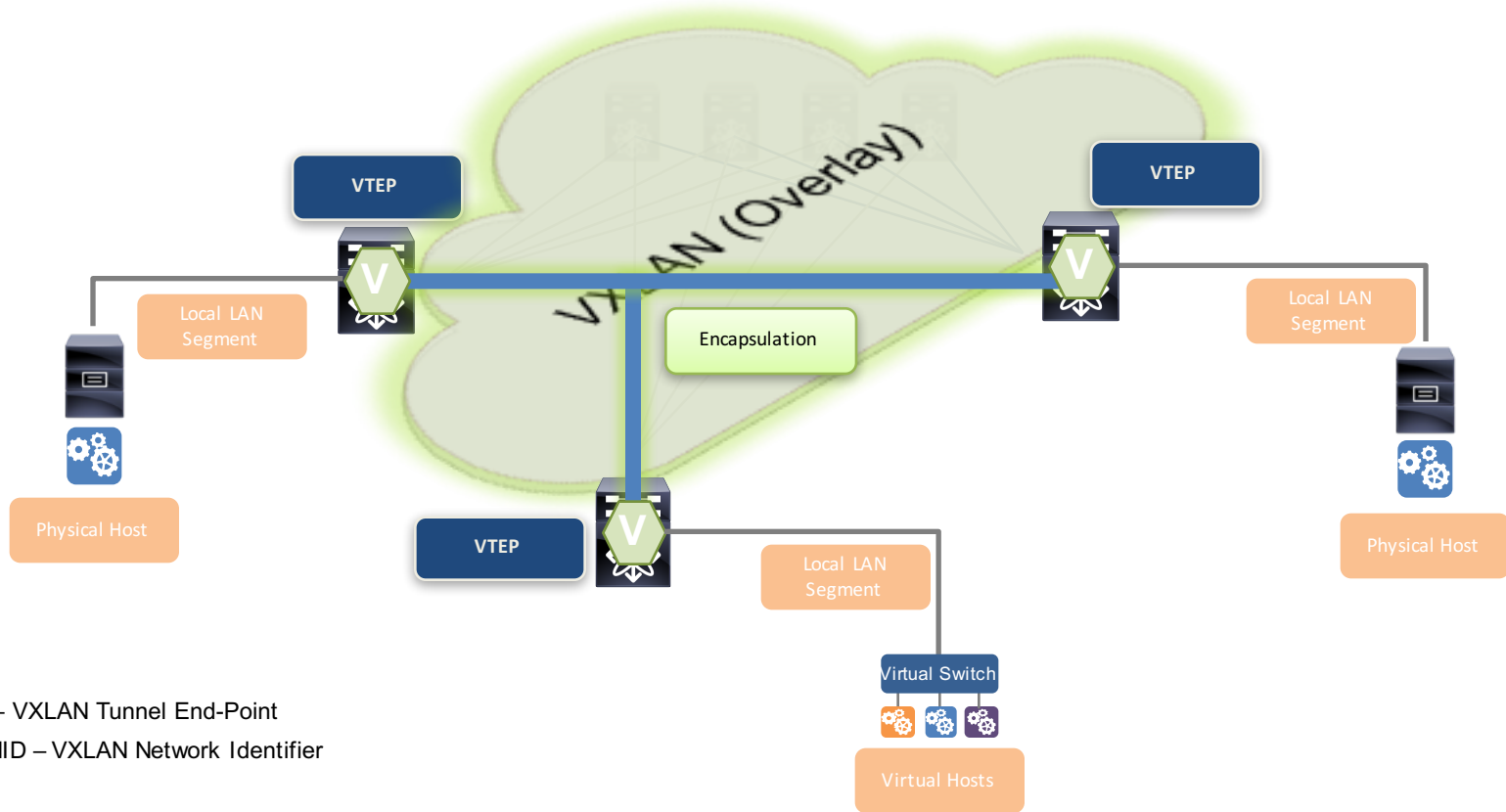
Lesson 1: VXLAN Flood and Learn

1.1 Intro to VXLAN

VXLAN Overview



VXLAN Overview



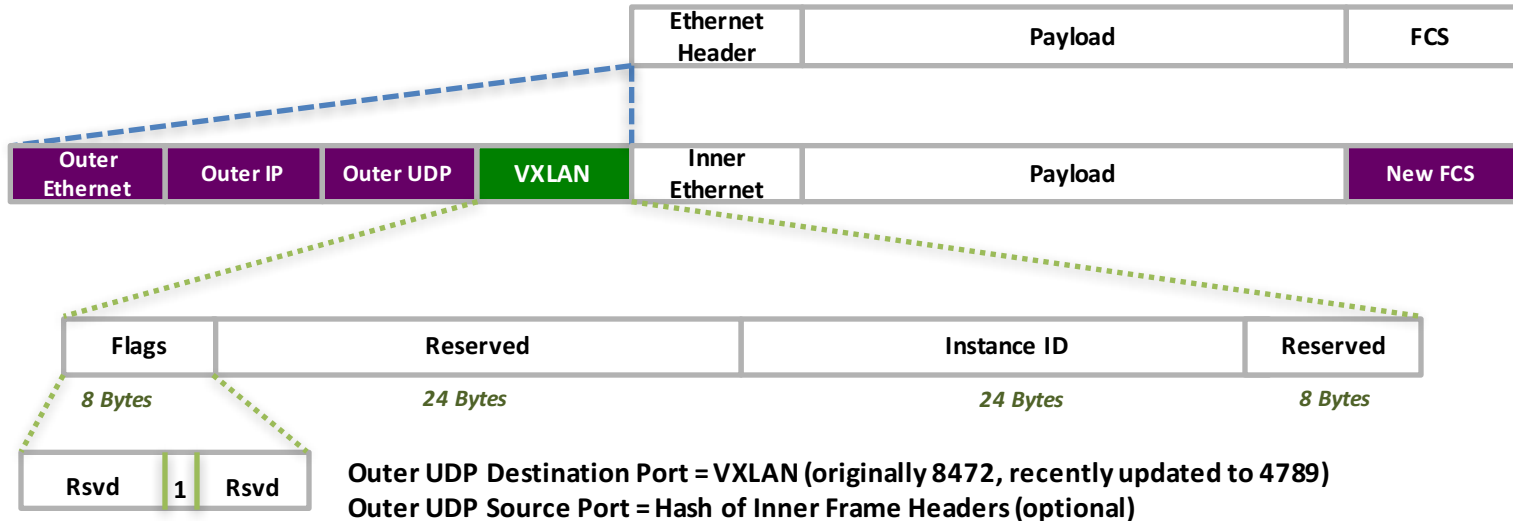
VTEP – VXLAN Tunnel End-Point

VNI/VNID – VXLAN Network Identifier

VXLAN Overview

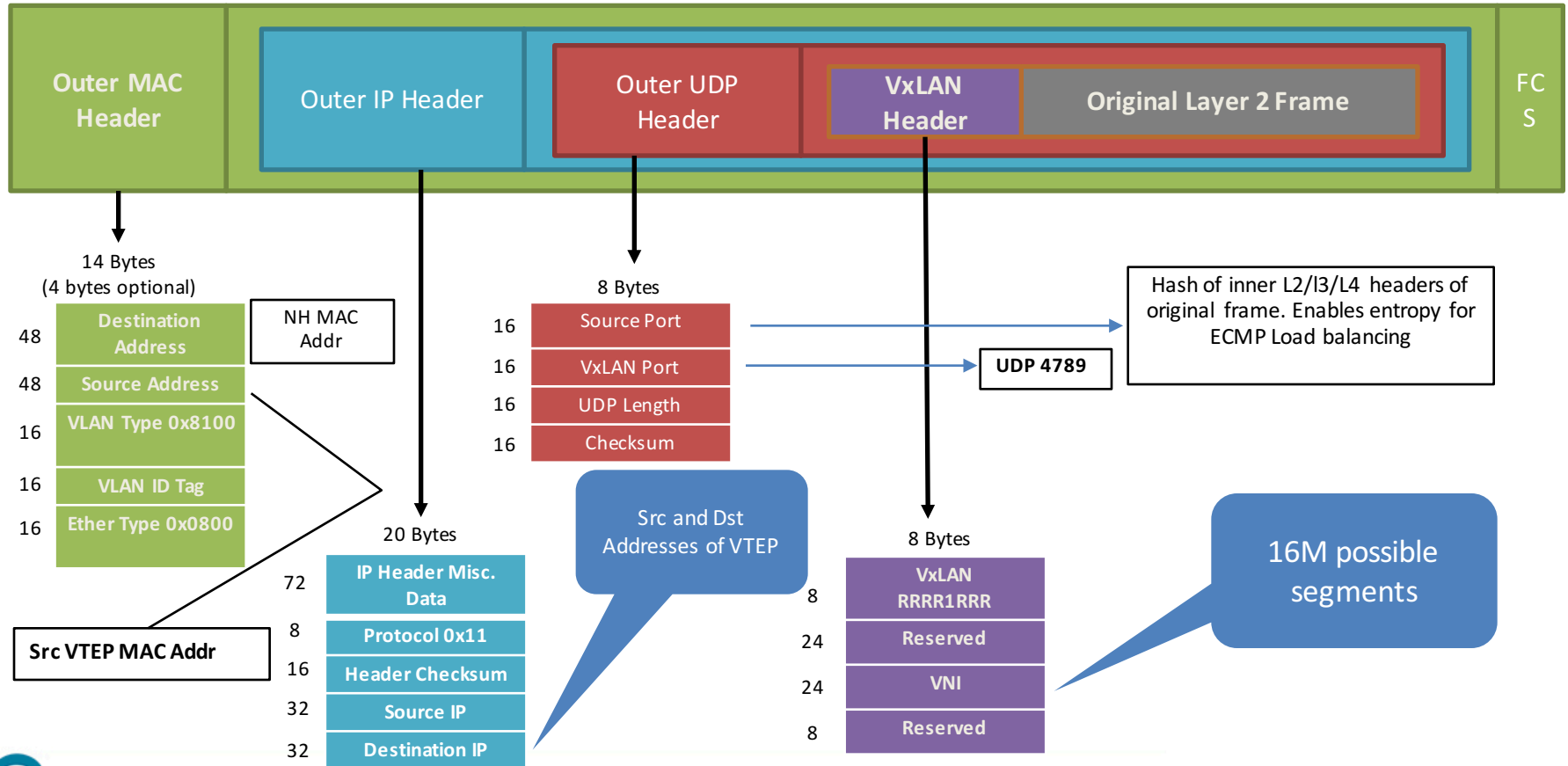
- VXLAN Overlay
 - A VXLAN Overlay or VXLAN segment is a Layer-2 broadcast domain identified by the VNID that extends or tunnels traffic from one VTEP to another.
- VXLAN Tunnel End Point (VTEP)
 - A VTEP is a device that provides both encapsulation and de-capsulation of classical Ethernet and VXLAN packets to and from a VXLAN segment
 - Each VTEP may have the following types of interfaces:
 - Switchport interfaces on the local LAN segment to support local endpoints
 - Layer-3 interfaces to the transport IP network
 - SVI interfaces
- VXLAN Gateway
 - A VTEP that bridges traffic between VXLAN segments

VXLAN Encapsulation



- The outer IP header has the source IP and destination IP of the VTEP endpoints
- The outer Ethernet header has the source MAC of the source VTEP and the destination MAC of the immediate Layer-3 next hop

VXLAN Packet Structure

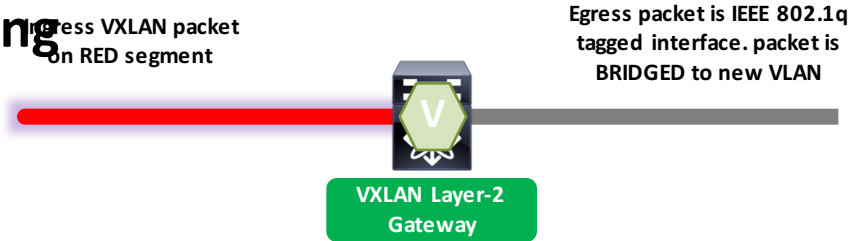


VXLAN Gateway Types

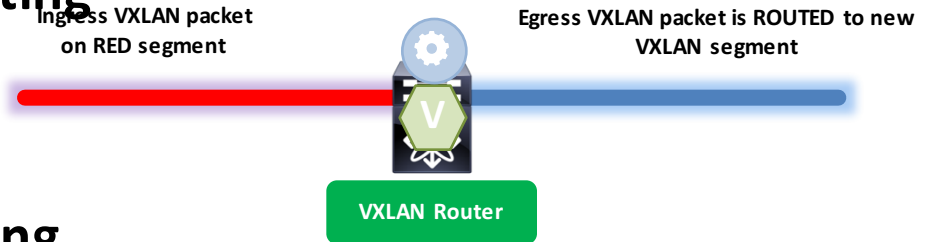
- Layer 2 Gateway
 - The layer 2 gateway is required when the layer 2 traffic (IEEE 802.1q tagged traffic) comes from VLAN into VxLAN segment (encapsulation) or
 - The Ingress VxLAN packet egresses out an 802.1q tagged interface (de-encapsulation), where the packet is bridged to a new VLAN.
- Layer 3 Gateway
 - A layer 3 gateway is used when there is a VxLAN to VxLAN routing
 - The ingress packet is a VxLAN packet on a routed segment but the packet egresses out on a tagged 802.1q interface and the packet is routed to a new VLAN

VXLAN Gateway Types

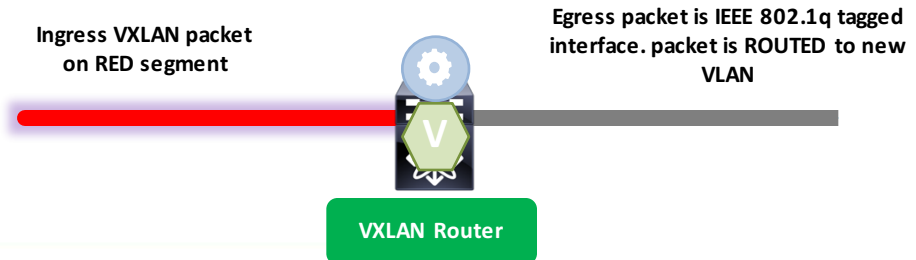
- **VXLAN to VLAN Bridging**
 - (Layer-2 Gateway)



- **VXLAN-to-VXLAN Routing**
 - (Layer-3 Gateway)

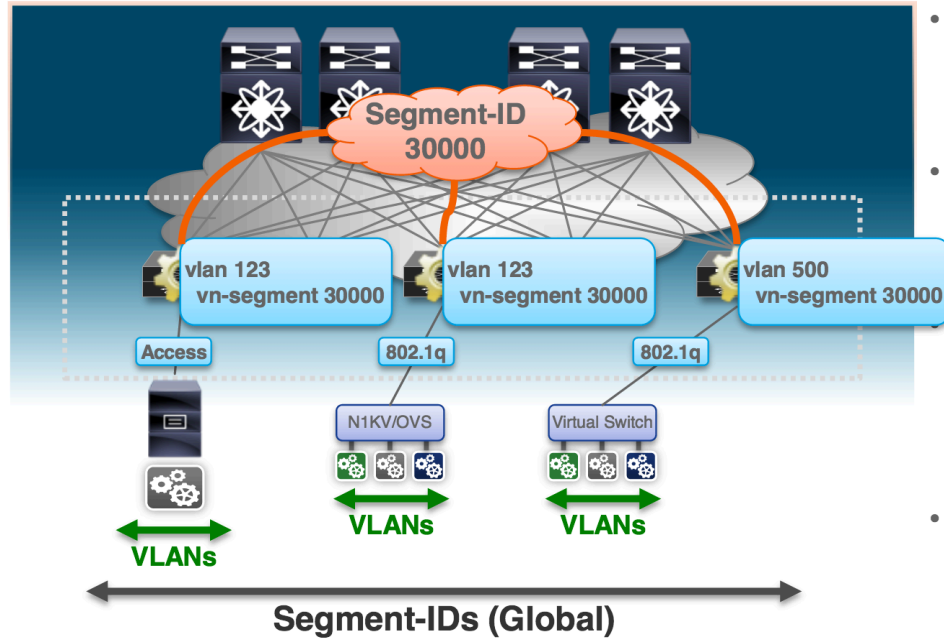


- **VXLAN-to-VLAN Routing**
 - (Layer-3 Gateway)



- VxLAN adds 50 bytes to the original Ethernet frame.
- VTEPs MUST NOT fragment VXLAN packets.
- Intermediate routers may fragment encapsulated VXLAN packets due to the larger frame size.
- The destination VTEP MAY silently discard such VXLAN fragments.
- To ensure end-to-end traffic delivery without fragmentation, it is RECOMMENDED that the MTUs (Maximum Transmission Units) across the physical network infrastructure be set to a value that accommodates the larger frame size due to the encapsulation.

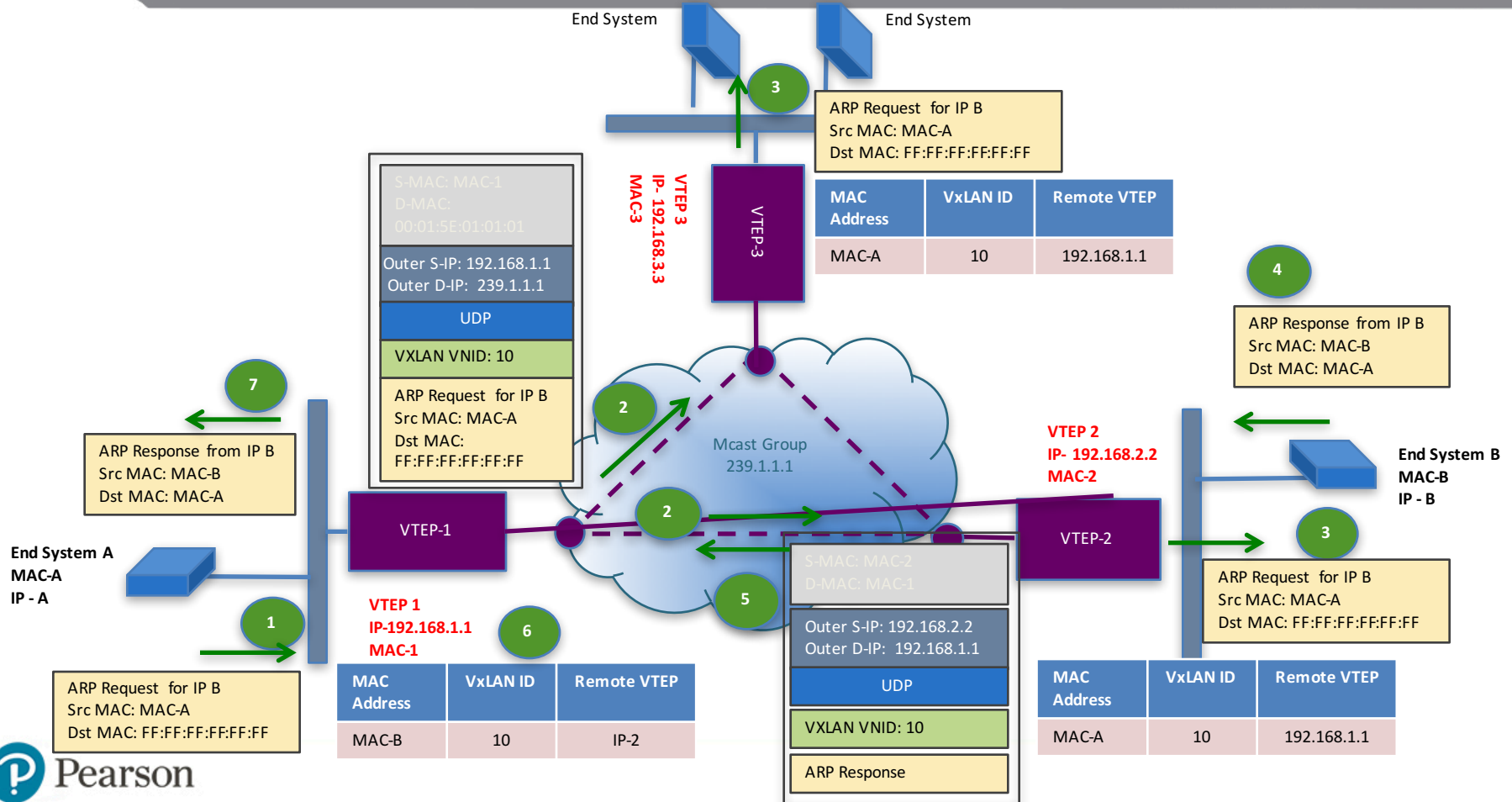
Segment ID



- Segment IDs are utilized for providing isolation at Layer 2 and Layer 3 across the Fabric.
- 802.1Q tagged frames received at the Leaf nodes from edge devices must be mapped to specific Segments.

The VLAN to Segment mapping is performed on a leaf device level VLANs become locally significant on the leaf node and 1:1 mapped to a Segment ID.

- Thus, Segment IDs are globally significant and VLAN IDs are locally significant.
- are locally significant.



VXLAN F&L Overview

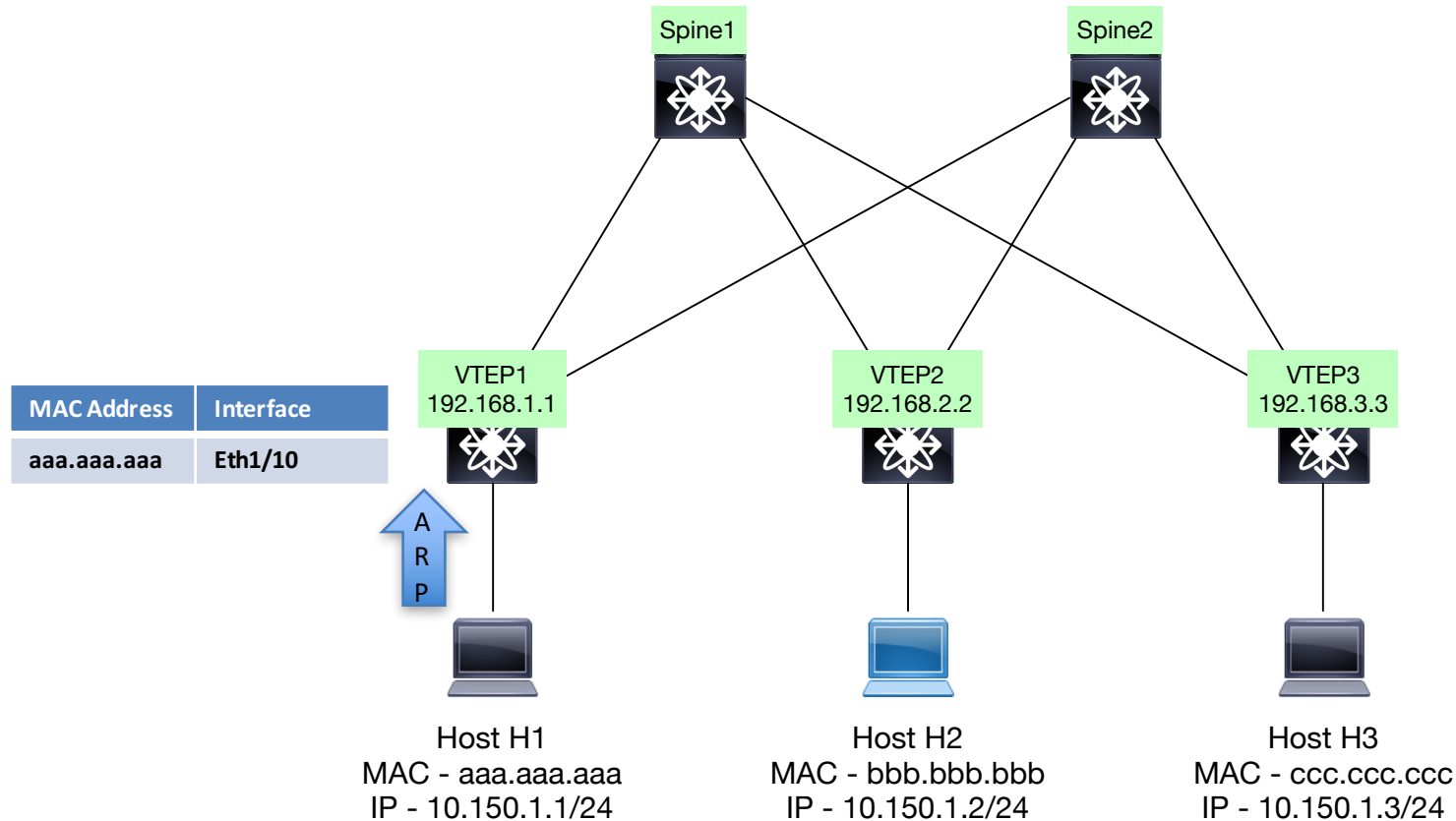
- Data Plane learning technique for VxLAN
- VNI's are mapped to a multicast group on a VTEP
- Local MACs are learnt over a VLAN (VNI) on a VTEP
- Broadcast, Unknown Unicast, Multicast (*BUM Traffic*) is flooded to the delivery multicast group for that VNI
- Remote VTEPs part of same multicast group learn host MAC, VNI and source VTEP as the next-hop for the host MAC from flooded traffic
- Unicast packets to the host MAC are sent directly to source VTEP as VxLAN encapsulated packet

Ingress Replication

- Some customers not comfortable deploying multicast in their core
- With *Ingress Replication (IR)*, BUM traffic ingress access side is replicated to remote VTEP as unicast
- Static IR VTEP tunnel is kept alive as long as the route to the VTEP is available.
- Support multiple VTEPs per VNI and a VTEP in multiple VNIs
- Up to 16 static IR VTEPs recommended
- Multicast and IR config can co-exist on the same switch nodes but on different VNI's (This is applicable for Migration Scenario. Not recommended for production.)

VXLAN Packet Flow

Multicast Enabled Core
VNI 10000 - 239.1.1.1



VXLAN Packet Flow – con't

Multicast Enabled Core
VNI 10000 - 239.1.1.1

Ethernet Header

S-MAC: MAC-1 (VTEP-1 MAC)
DMAC: 00:01:5E:01:01:01

IP Header

Outer S-IP: 192.168.1.1
Outer D-IP: 239.1.1.1

UDP Header

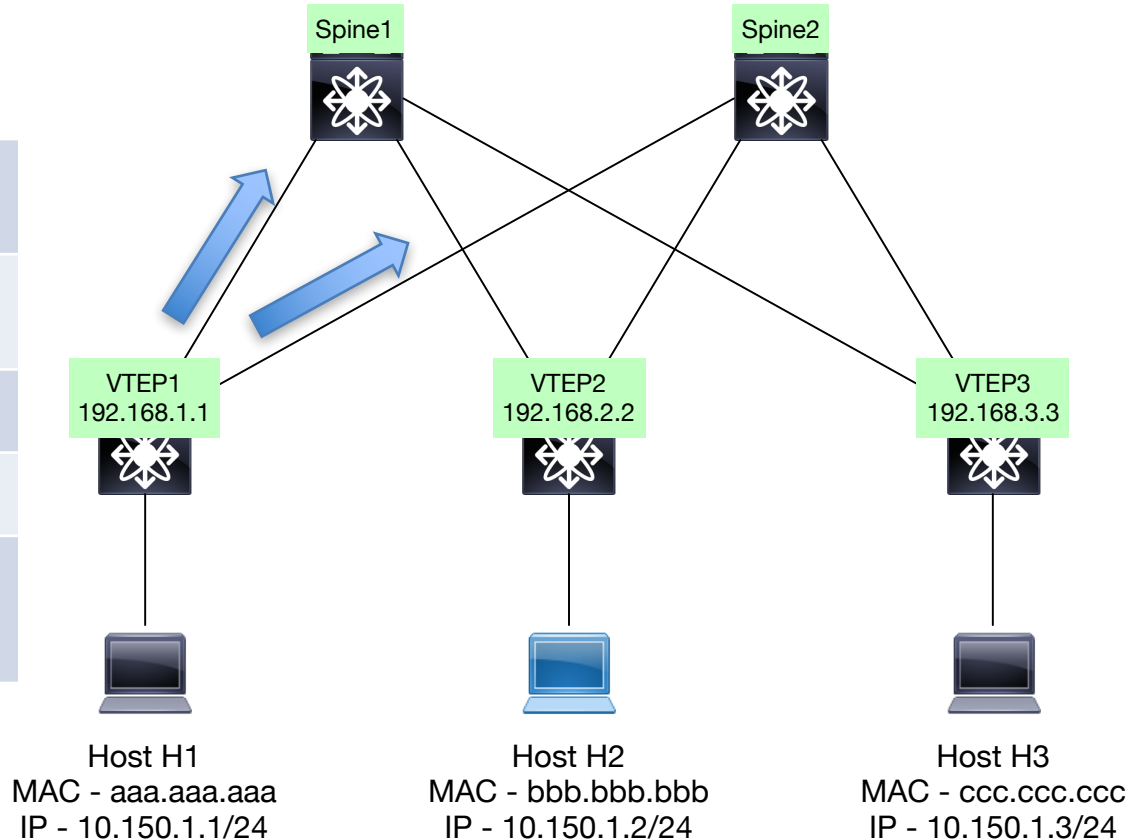
DST Port 4789

VXLAN Header

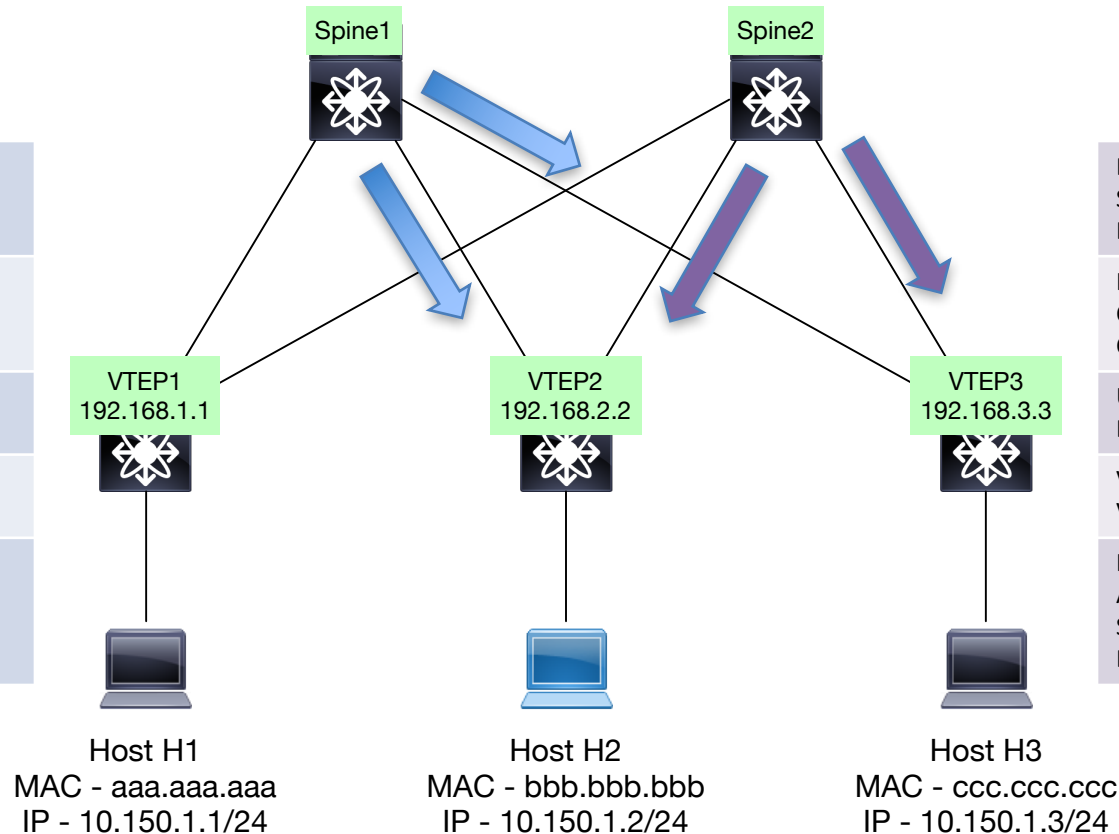
VXLAN VNID: 10000

Inner Frame

ARP Request for IP 10.150.1.2
SRC MAC: aaa.aaa.aaa
Dst MAC: FF:FF:FF:FF:FF:FF



VXLAN Packet Flow – con't



Ethernet Header

S-MAC: Spine1-intf-MAC
DMAC: 00:01:5E:01:01:01

IP Header

Outer S-IP: 192.168.1.1
Outer D-IP: 239.1.1.1

UDP Header

DST Port 4789

VXLAN Header

VXLAN VNID: 10000

Inner Frame

ARP Request for IP 10.150.1.2
SRC MAC: aaa.aaa.aaa
Dst MAC: FF:FF:FF:FF:FF:FF

Multicast Enabled Core
VNI 10000 - 239.1.1.1

Ethernet Header

S-MAC: Spine2-intf-MAC
DMAC: 00:01:5E:01:01:01

IP Header

Outer S-IP: 192.168.1.1
Outer D-IP: 239.1.1.1

UDP Header

DST Port 4789

VXLAN Header

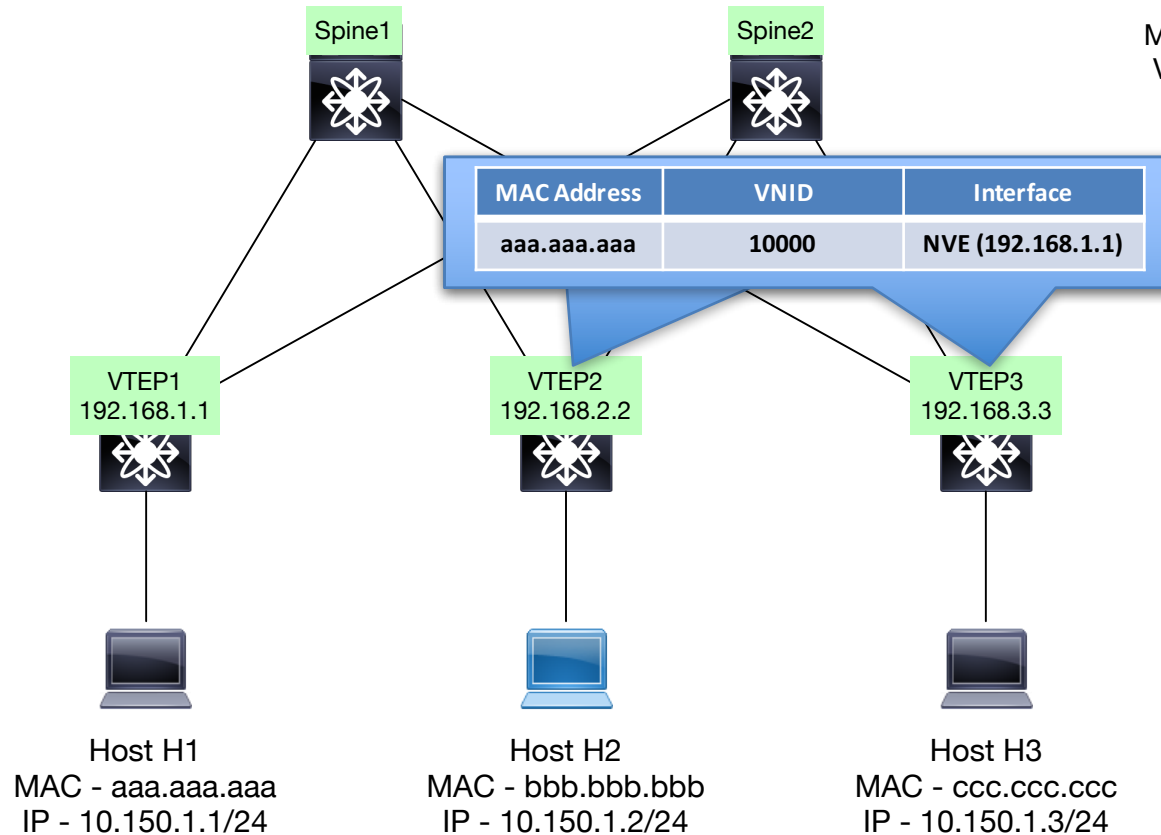
VXLAN VNID: 10000

Inner Frame

ARP Request for IP 10.150.1.2
SRC MAC: aaa.aaa.aaa
Dst MAC: FF:FF:FF:FF:FF:FF

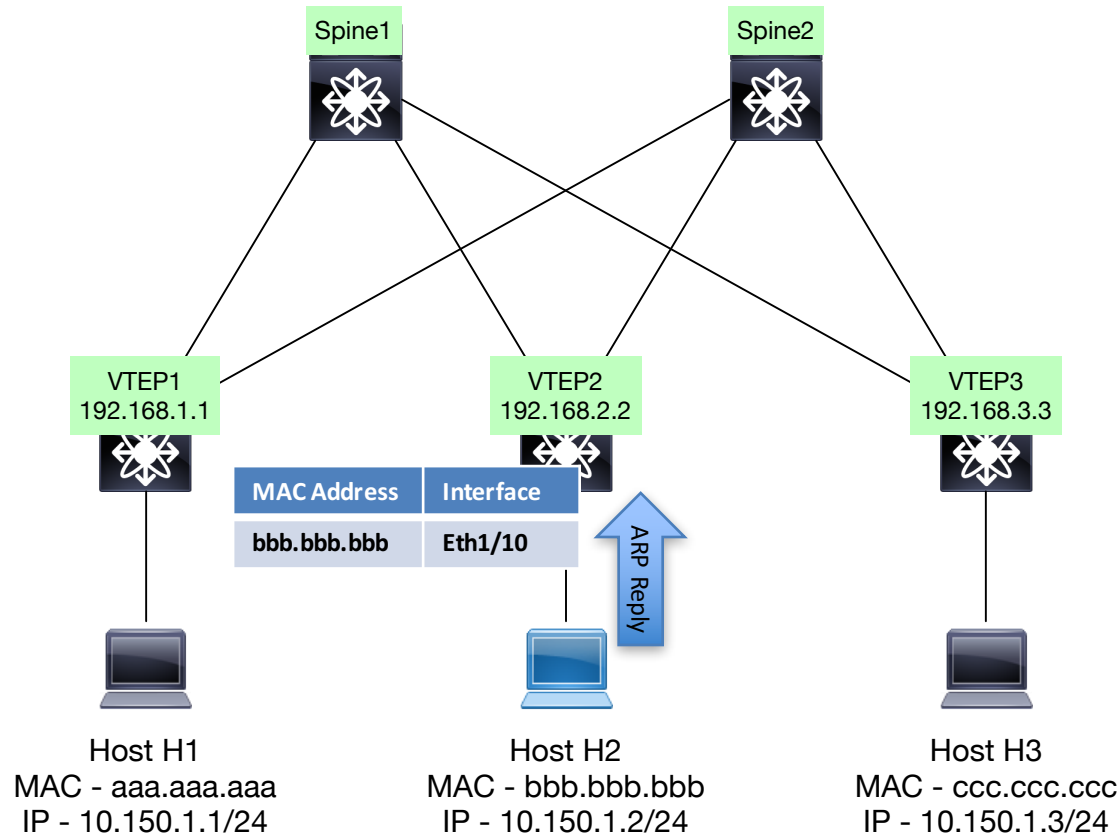
VXLAN Packet Flow – con't

Multicast Enabled Core
VNI 10000 - 239.1.1.1

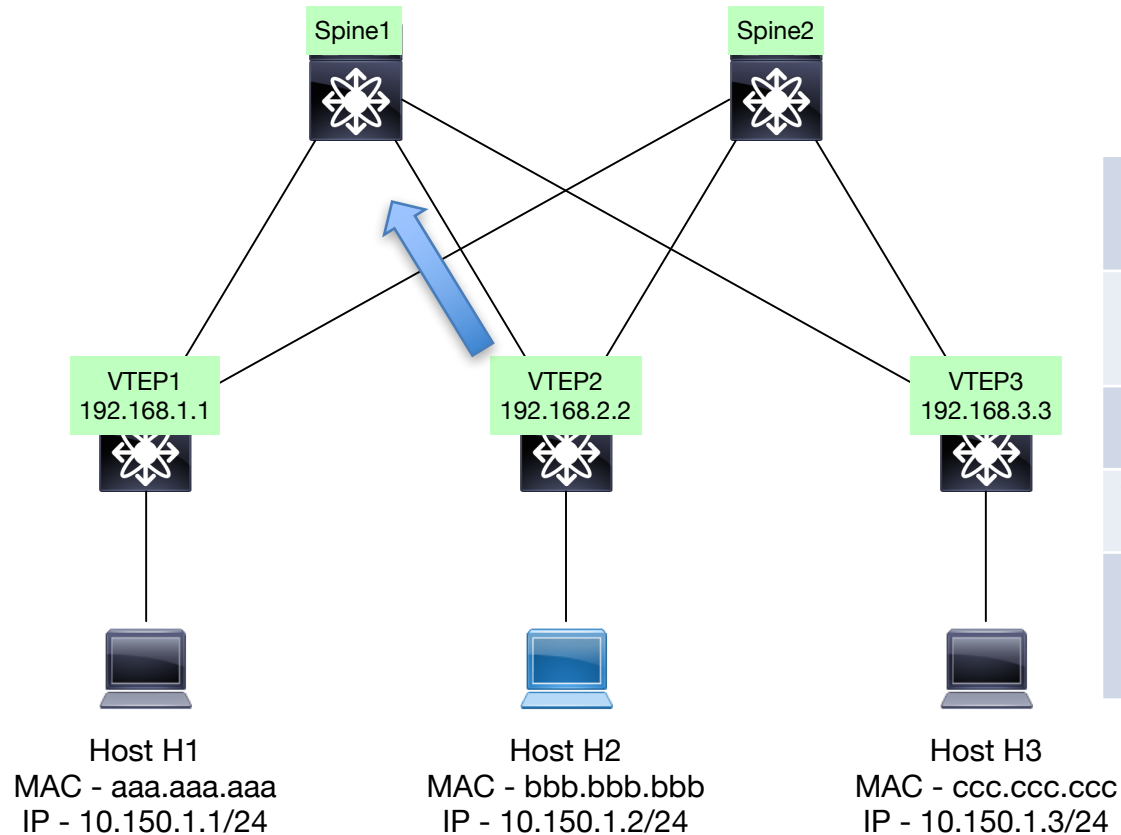


VXLAN Packet Flow – con't

Multicast Enabled Core
VNI 10000 - 239.1.1.1



VXLAN Packet Flow – con't



Multicast Enabled Core
VNI 10000 - 239.1.1.1

Ethernet Header

S-MAC: MAC-2 (VTEP-2 MAC)
DMAC: Spine1-intf-MAC

IP Header

Outer S-IP: 192.168.2.2
Outer D-IP: 192.168.1.1

UDP Header

DST Port 4789

VXLAN Header

VXLAN VNID: 10000

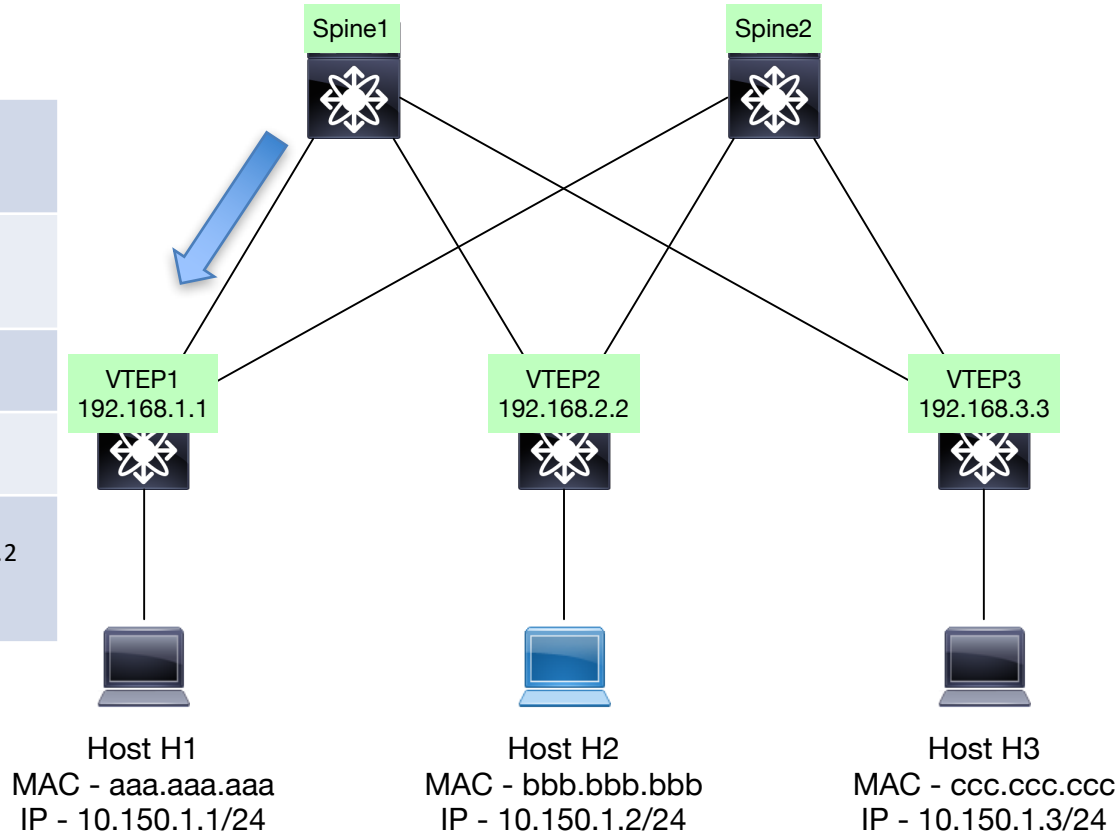
Inner Frame

ARP Request for IP 10.150.1.2
SRC MAC: bbb.bbb.bbb
Dst MAC: aaa.aaa.aaa

VXLAN Packet Flow – con't

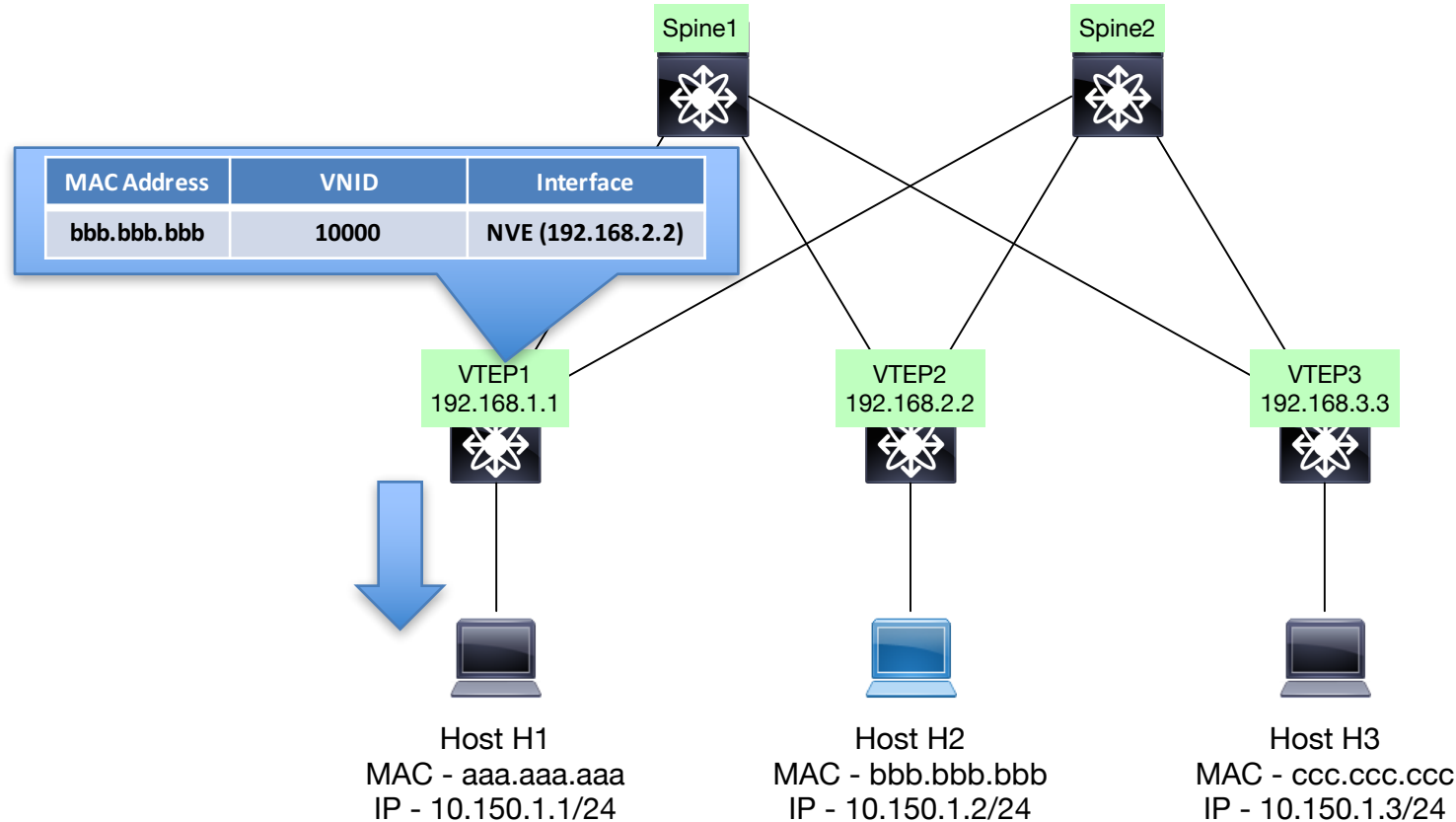
Multicast Enabled Core
VNI 10000 - 239.1.1.1

Ethernet Header S-MAC: Spine1-intf-MAC DMAC: MAC-1
IP Header Outer S-IP: 192.168.2.2 Outer D-IP: 192.168.1.1
UDP Header DST Port 4789
VXLAN Header VXLAN VNID: 10000
Inner Frame ARP Request for IP 10.150.1.2 SRC MAC: bbb.bbb.bbb Dst MAC: aaa.aaa.aaa



VXLAN Packet Flow – con't

Multicast Enabled Core
VNI 10000 - 239.1.1.1

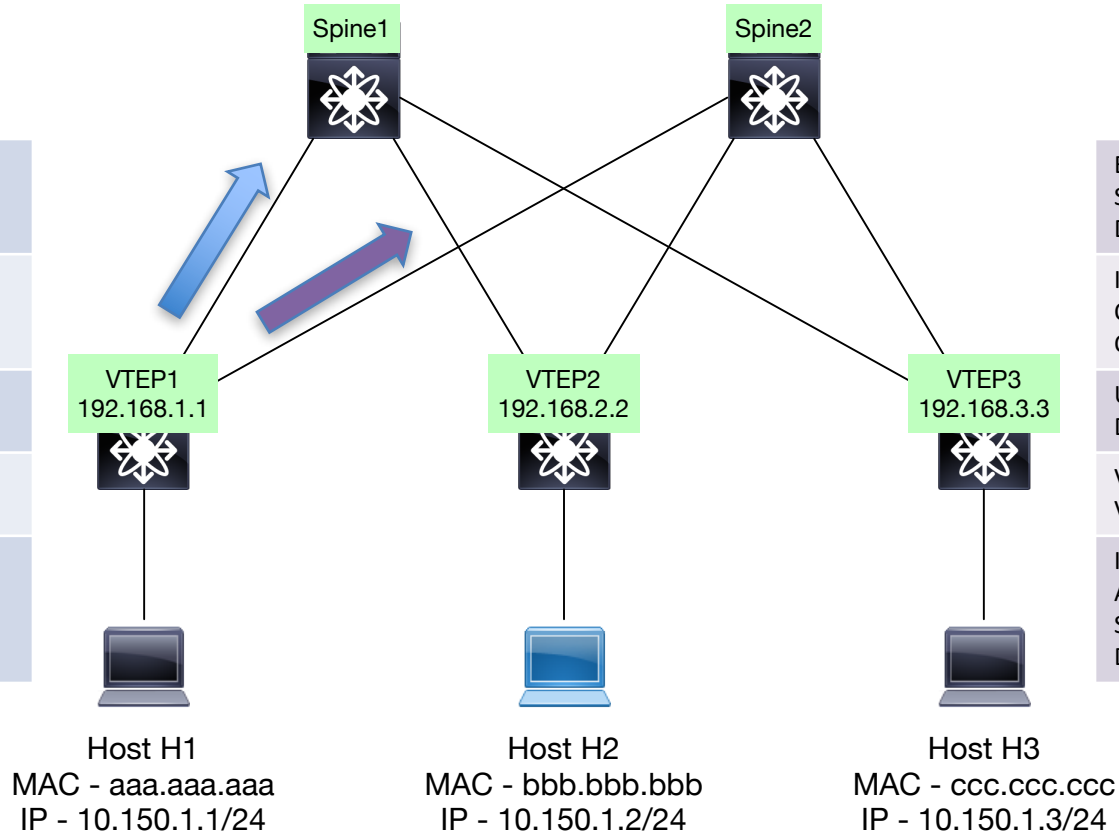


VXLAN – Flood and Learn

- Data Plane learning technique for VXLAN
- VXLAN Network Identifier (VNI's) are mapped to a multicast group on a VTEP
- Local MACs are learnt over a VLAN (VNI) on a VTEP
- Broadcast, Unknown Unicast, Multicast (*BUM Traffic*) is flooded to the delivery multicast group for that VNI
- Remote VTEPs part of same multicast group learn host MAC, VNI and source VTEP as the next-hop for the host MAC from flooded traffic
- Unicast packets from remote VTEP to the source host are sent directly to source VTEP as VxLAN encapsulated packet

VXLAN Packet Flow – Ingress Replication

Ethernet Header
S-MAC: MAC-1
DMAC: Spine1-intf-MAC
IP Header
Outer S-IP: 192.168.1.1
Outer D-IP: 192.168.2.2
UDP Header
DST Port 4789
VXLAN Header
VXLAN VNID: 10000
Inner Frame
ARP Request for IP 10.150.1.2
SRC MAC: aaa.aaa.aaa
Dst MAC: FF:FF:FF:FF:FF:FF



Ethernet Header
S-MAC: MAC-1
DMAC: Spine2-intf-MAC
IP Header
Outer S-IP: 192.168.1.1
Outer D-IP: 192.168.3.3
UDP Header
DST Port 4789
VXLAN Header
VXLAN VNID: 10000
Inner Frame
ARP Request for IP 10.150.1.2
SRC MAC: aaa.aaa.aaa
Dst MAC: FF:FF:FF:FF:FF:FF

Ingress Replication

- Not all devices might support multicast in their core
 - Or network operators / administrators might not be comfortable deploying multicast
- With *Ingress Replication (IR)*, Broadcast, Unknown unicast, Multicast (BUM) traffic ingress access side is replicated to remote VTEP as unicast
- Static IR VTEP tunnel is kept alive as long as the route to the VTEP is available.
- Support multiple VTEPs per VNI and a VTEP in multiple VNIs



Lesson 1: VXLAN Flood and Learn

1.2 Configuring and Verifying VXLAN Flood & Learn

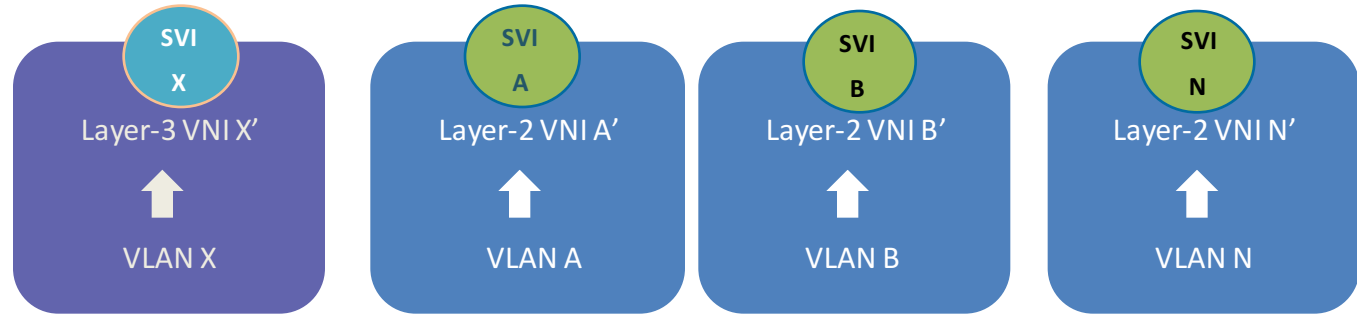


Lesson 2: VXLAN-EVPN

2.1 Intro to VXLAN-EVPN

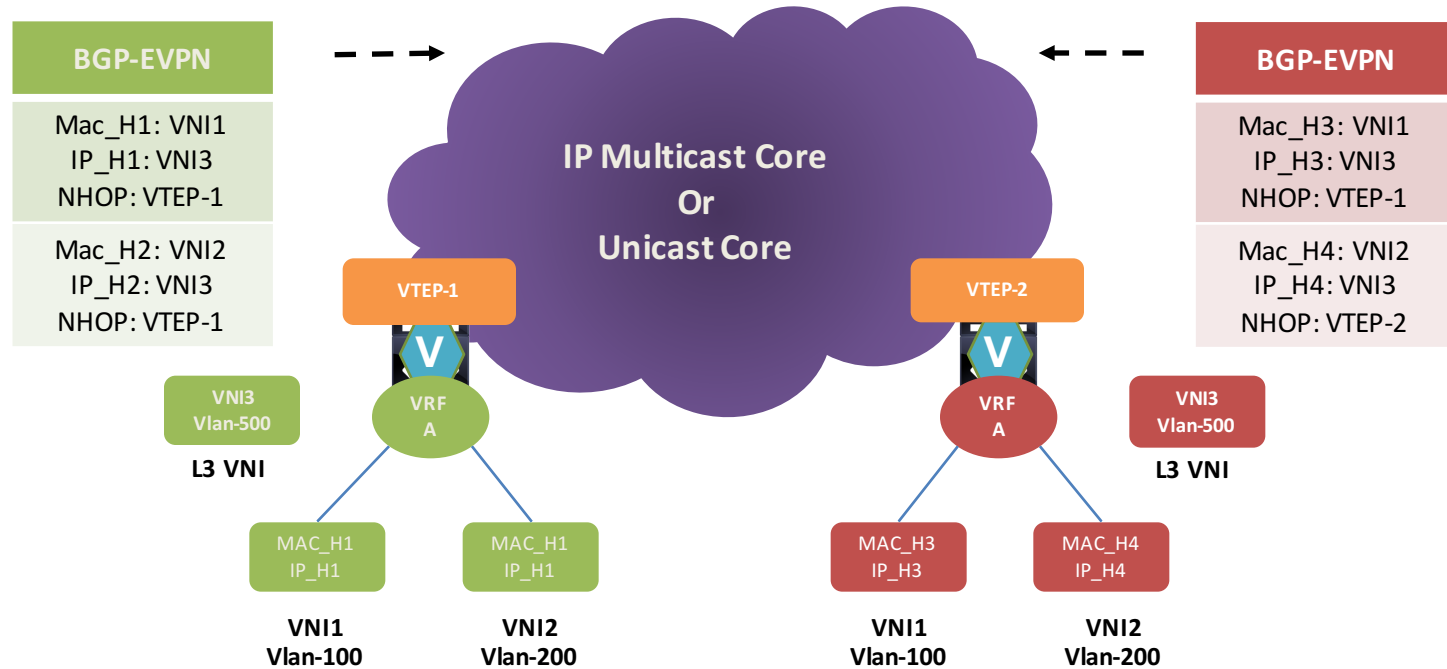
Tenant

Tenant 1 (VRF 1)

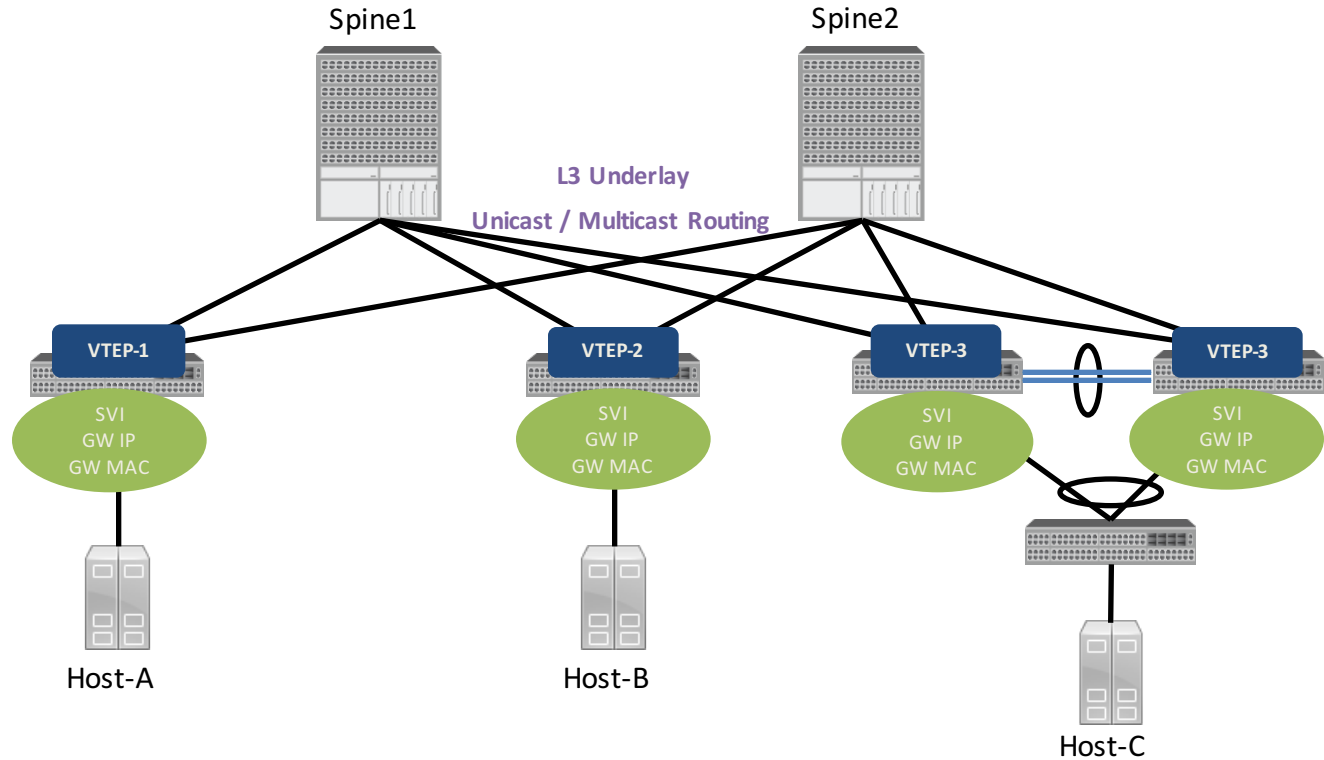


- 1 Layer-3 VNI per Tenant (VRF) for routing
- VNI X' is used for routed packets
- 1 Layer-2 VNI per Layer-2 segment
- Multiple Layer-2 VNIs per tenant
- VNI A' and B' are used for bridged packets

MP-BGP - EVPN



Distributed Anycast Gateway



Distributed Anycast Gateway - Config

- All VTEPs has same IP address for an L2 VNI
- Anycast Gateway MAC is global to each VTEP for all VNI's for all Tenants
- One virtual MAC / VTEP
- All VTEPs should have same virtual MAC address

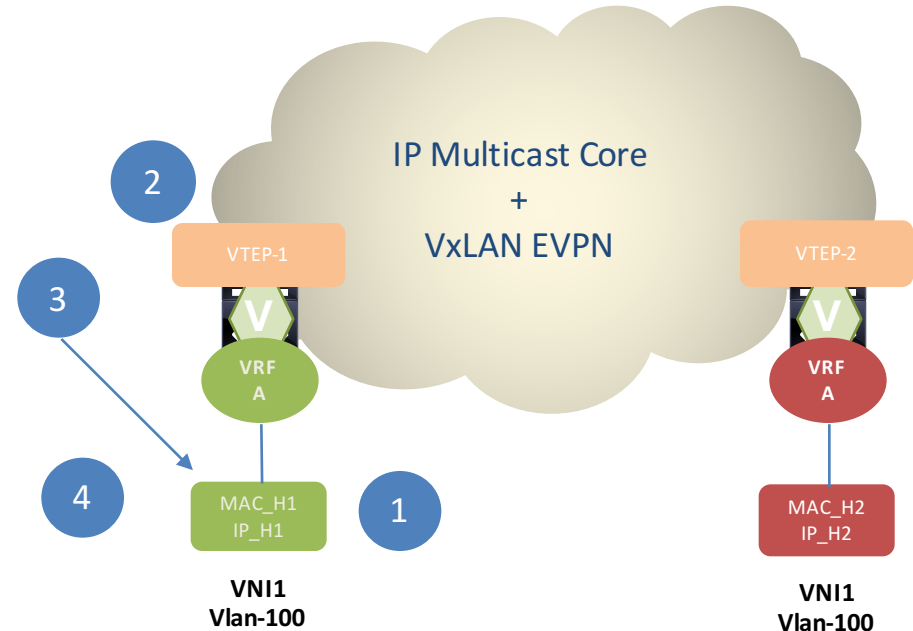
```
fabric forwarding anycast-gateway-mac 0001.0001.0001
!  
interface Vlan100  
no shutdown  
vrf context test-evpn-tenant  
ip address 172.16.1.254/24  
fabric forwarding mode anycast-gateway
```

VXLAN Gateway Types

- Layer 2 Gateway
 - The layer 2 gateway is required when the layer 2 traffic (IEEE 802.1q tagged traffic) comes from VLAN into VXLAN segment (encapsulation) or
 - The Ingress VXLAN packet egresses out an 802.1q tagged interface (de-encapsulation), where the packet is bridged to a new VLAN
- Layer 3 Gateway
 - A layer 3 gateway is used when there is a VXLAN to VXLAN routing
 - The ingress packet is a VXLAN packet on a routed segment but the packet egresses out on a tagged 802.1q interface and the packet is routed to a new VLAN

ARP Suppression

- Hosts send out Gratuitous ARP (GARP) when they come online
- Local leaf node receives GARP, creates local ARP cache and advertises to other leaf by BGP as route type 2
- Remote leaf node puts IP-MAC info into remote ARP cache and suppresses incoming ARP request for this IP
- If IP info not found in ARP suppression cache table, VTEP floods the ARP request to other VTEPs



EVPN Route-Types

- BGP Ethernet Virtual Private Network (EVPN) uses 5 different route types for IP prefixes and advertisement
 - Type 1 - Ethernet Auto-Discovery (A-D) route
 - **Type 2 - MAC advertisement route → L2 VNI MAC/MAC-IP**
 - Type 3 - Inclusive Multicast Route → **EVPN IR, Peer Discovery**
 - Type 4 - Ethernet Segment Route
 - **Type 5 - IP Prefix Route → L3 VNI Route**
- Route type 2 or MAC Advertisement route is for MAC and ARP resolution advertisement, MAC or MAC-IP
- Route type 5 or IP Prefix route will be used for the advertisement of prefixes, IP only

Route TYPE - 8

Length - 10

Route Type Specific

EVPN BGP Route Type 2 Fields

- Ethernet Tag ID, MAC Address Length, MAC Address, IP Address Length, and IP Address fields are considered to be part of the prefix in the NLRI.
- Ethernet Segment Identifier, MPLS Label1, and MPLS Label2 are treated as route attributes, not part of the "route". Both the IP and MAC address lengths are in bits.

```
Leaf1#show bgp l2vpn evpn 523d.e706.ae1b
BGP routing table information for VRF default, address family L2VPN
EVPN
Route Distinguisher: 10000:1      (L2VNI 10000)
BGP routing table entry for
[2]:[0]:[0]:[48]:[8c60.4f93.5ffc]:[0]:[0.0.0.0]/216, version 8
Paths: (1 available, best #1)
Flags: (0x00010a) on xmit-list, is not in l2rib/evpn

Advertised path-id 1
Path type: local, path is valid, is best path, no labeled nexthop
AS-Path: NONE, path locally originated
192.168.1.1 (metric 0) from 0.0.0.0 (192.168.1.1)
Origin IGP, MED not set, localpref 100, weight 32768
Received label 10000
Extcommunity: RT:10000:1
```

Route Distinguisher – 8 byte

Ethernet Segment ID – 10 byte

Ethernet Tag ID – 4 byte

MAC Address Length – 1 byte

MAC Address – 6 byte

IP Address Length – 1 byte

IP Address – 0, 4, 16 byte

MPLS Label 1 – 3 byte, L2VNI

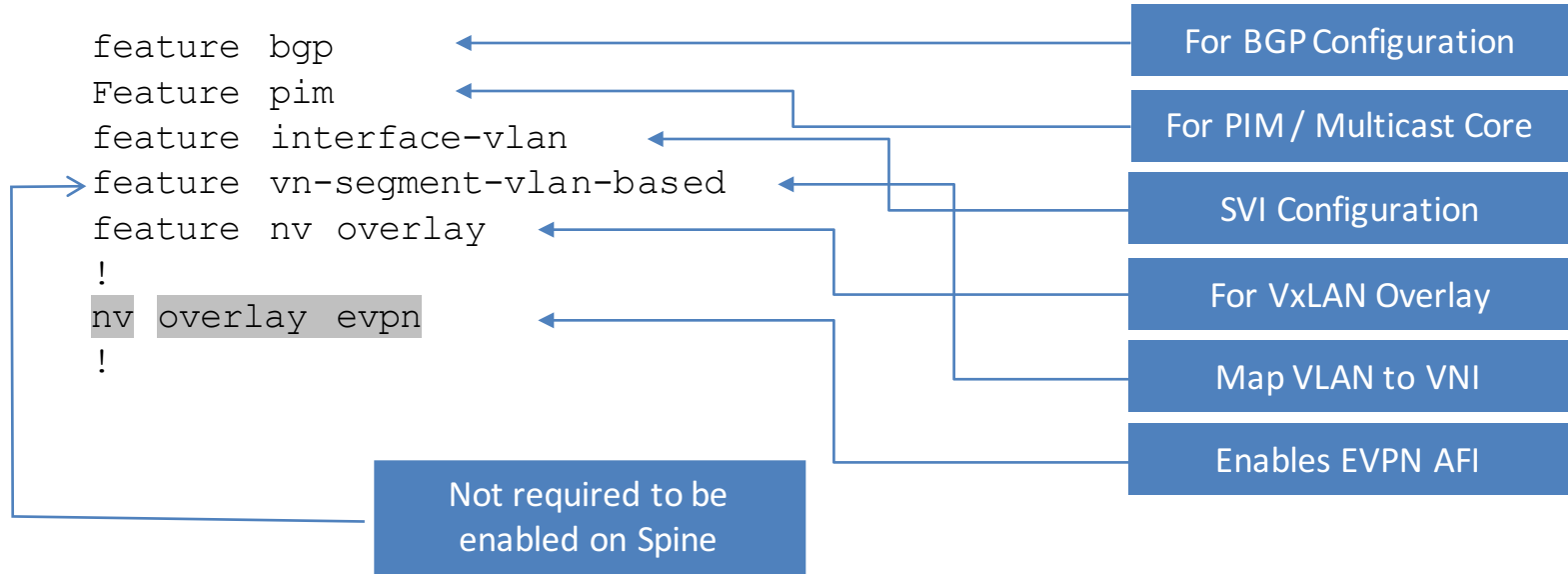
MPLS Label 2 – 3 byte L3VNI



Lesson 2: VXLAN-EVPN

2.2 Configuring VXLAN-EVPN

Feature Enablement



Underlay Configuration (2-AS Design)

Leaf

```
router bgp 65000
  router-id 192.168.1.1
  address-family ipv4 unicast
    network 1.1.1.1/32
    network 192.168.1.1/32
  address-family l2vpn evpn
  neighbor 10.1.101.10
    remote-as 65001
    address-family ipv4 unicast
      allowas-in 3
      disable-peer-as-check
  neighbor 10.1.201.20
    remote-as 65001
    address-family ipv4 unicast
      allowas-in 3
      disable-peer-as-check
```

Spine

```
router bgp 65001
  router-id 192.168.1.1
  address-family ipv4 unicast
    network 10.10.10.10/32
    network 192.168.10.10/32
  address-family l2vpn evpn
  neighbor 10.1.201.1
    remote-as 65000
    address-family ipv4 unicast
      allowas-in 3
      disable-peer-as-check
  neighbor 10.1.202.2
    remote-as 65000
    address-family ipv4 unicast
      allowas-in 3
      disable-peer-as-check
```

Leaf Node Configuration – L2 VNI

```
vlan 100
  vn-segment 10000

! Create L2 VNI

evpn
  vni 10000 12
    rd 10000:1
    route-target import 10000:1
    route-target export 10000:1
!
interface nve1
  no shutdown
  source-interface loopback0
  host-reachability protocol bgp
  member vni 10000
    mcast-group 239.1.1.1
    suppress-arp
```

```
fabric forwarding anycast-gateway-mac 0001.0001.0001
interface Vlan100
  no shutdown
  vrf member EVPN-TENANT
  ip address 100.1.1.254/24
  fabric forwarding mode anycast-gateway
!
router bgp 65000
neighbor 10.10.10.10
  remote-as 65001
  update-source loopback0
  ebgp-multihop 3
  address-family l2vpn evpn
    allowas-in 3
    disable-peer-as-check
    send-community extended
vrf EVPN-TENANT
  address-family ipv4 unicast
    advertise l2vpn evpn
```

Leaf Node Configuration – L3 VNI

```
vlan 500
  vn-segment 50000
!
vrf context EVPN-TENANT
  vni 50000
  rd 50000:1
  address-family ipv4 unicast
    route-target import 20000:1
    route-target import 20000:1 evpn
    route-target export 20000:1
    route-target export 20000:1 evpn
!
interface Vlan500
  no shutdown
  vrf member EVPN-TENANT
  ip forward
!
```

```
interface nve1
  no shutdown
  source-interface loopback0
  host-reachability protocol bgp
  member vni 50000 associate-vrf
!
interface loopback200
  vrf member EVPN-TENANT
  ip address 200.1.1.1/32
!
router bgp 65000
  vrf EVPN-TENANT
    address-family ipv4 unicast
      network 200.1.1.1/32
      advertise l2vpn evpn
```

Spine Node Configuration

```
interface loopback0
  ip address 192.168.10.10/32
  ip pim sparse-mode
!
interface loopback1
  ip address 192.168.50.50/32
  ip pim sparse-mode
!
ip pim rp-address 192.168.50.50 group-list
                                     239.1.1.0/24
ip pim anycast-rp 192.168.50.50 192.168.10.10
ip pim anycast-rp 192.168.50.50 192.168.20.20
```

```
router bgp 65001
  address-family l2vpn evpn
    nexthop route-map permitall
    retain route-target all
  neighbor 1.1.1.1
    remote-as 65000
    update-source loopback0
    ebgp-multihop 3
  address-family l2vpn evpn
    disable-peer-as-check
    send-community extended
    route-map permitall out
!
route-map permitall permit 10
  set ip next-hop unchanged
```


L2 & L3 VNI for IPv6

```
vlan 100
  vn-segment 10000
vlan 200
  vn-segment 20000
evpn
  vni 10000 12
    rd 10000:1
    route-target import 10000:1
    route-target export 10000:1
!
vrf context EVPN-TENANT
  vni 20000
  rd 20000:1
address-family ipv4 unicast
  route-target both 20000:1
  route-target both 20000:1 evpn
address-family ipv6 unicast
  route-target both 20000:1
  route-target both 20000:1 evpn
!
interface Vlan200
  no shutdown
  vrf member EVPN-TENANT
  ip forward
  ipv6 address use-link-local-only
```

```
fabric forwarding anycast-gateway-mac
0001.0001.0001
interface Vlan100
  no shutdown
  vrf member EVPN-TENANT
  ip address 100.1.1.254/24
  ipv6 address 2001::1/64
  fabric forwarding mode anycast-gateway
!
interface nve1
  no shutdown
  source-interface loopback0
  host-reachability protocol bgp
  member vni 10000
    mcast-group 239.1.1.1
  member vni 20000 associate-vrf
!
router bgp 100
vrf EVPN-TENANT
  address-family ipv4 unicast
    advertise l2vpn evpn
  address-family ipv6 unicast
    advertise l2vpn evpn
!
vpc domain 10
  ipv6 nd synchronize
```



Lesson 2: VXLAN-EVPN

2.2 Verifying VXLAN-EVPN



2. VXLAN-EVPN

2.3 Troubleshooting VXLAN-EVPN

BUM Traffic over Multicast Core

- BUM Traffic – Broadcast traffic (ARP request and other broadcast packets), Multicast traffic from hosts, etc.
- Check the multicast group associated with L2VNI
- Get the source VTEP IP and Router MAC
- Check if ARP Suppression is enabled or not
- Ask the right Questions:
 - Check if the VTEP is a VPC VTEP or Standalone VTEP?
 - Is the issue seen for IPv4 hosts or IPv6 hosts?
 - Know the trigger and understand if the issue is continuously reproducible or not?

Troubleshooting BUM Traffic

```
LEAF45# ethanalyzer local interface inband capture-fil "arp" limit-captured-frames 0
```

Capturing on inband

```
2018-05-21 14:52:36.289960 20:20:00:00:00:aa -> 54:7f:ee:07:e1:41 ARP 10.150.1.254 is  
at 20:20:00:00:00:aa
```

```
2018-05-21 14:52:36.295037 54:7f:ee:07:e1:41 -> ff:ff:ff:ff:ff:ff ARP Who has  
10.150.1.35? Tell 10.150.1.36
```

```
2018-05-21 14:52:36.295425 2c:54:2d:f6:0f:bc -> 54:7f:ee:07:e1:41 ARP 10.150.1.35 is at  
2c:54:2d:f6:0f:bc
```

```
2018-05-21 14:52:38.127594 54:7f:ee:07:e1:41 -> ff:ff:ff:ff:ff:ff ARP Who has  
10.150.1.38? Tell 10.150.1.36
```

```
LEAF45# debug logfile arp
```

```
LEAF45# debug ip arp packet
```

```
14:52:38.127774 arp: (context 3) Receiving packet from Vlan1501, logical interface  
Vlan1501 physical interface port-channel36, (prty 6) Hrd type 1 Prot type 800 Hrd len 6  
Prot len 4 OP 1, Pkt size 46
```

```
14:52:38.127811 arp: Src 547f.ee07.e141/10.150.1.36 Dst ffff.ffff.ffff/10.150.1.38
```

Troubleshooting BUM Traffic

```
LEAF45# show ip mroute 231.1.150.1 10.0.0.204 detail
```

```
IP Multicast Routing Table for VRF "default"
```

```
Total number of routes: 8
```

```
Total number of (*,G) routes: 2
```

```
Total number of (S,G) routes: 5
```

```
Total number of (*,G-prefix) routes: 1
```

```
(10.0.0.204/32, 231.1.150.1/32), uptime: 00:14:01, nve(0) mrib(0) ip(0) pim(1)
```

```
RPF-Source: 10.0.0.204 [0/0]
```

```
Data Created: No
```

```
Received Register stop
```

```
VXLAN Flags
```

```
VXLAN Encap
```

```
VPC Flags
```

```
RPF-Source Forwarder
```

```
Stats: 51/2601 [Packets/Bytes], 27.200 bps
```

```
Stats: Active Flow
```

```
Incoming interface: loopback1, RPF nbr: 10.0.0.204
```

```
Outgoing interface list: (count: 1) (bridge-only: 0)
```

```
Ethernet1/50, uptime: 00:09:52, pim
```

Troubleshooting BUM Traffic

```
LEAF43# show ip mroute 231.1.150.1 10.0.0.204 detail
```

```
IP Multicast Routing Table for VRF "default"
```

```
Total number of routes: 8
```

```
Total number of (*,G) routes: 2
```

```
Total number of (S,G) routes: 5
```

```
Total number of (*,G-prefix) routes: 1
```

```
(10.0.0.204/32, 231.1.150.1/32), uptime: 00:03:19, ip(0) mrib(1) pim(0)
```

```
RPF-Source: 10.0.0.204 [3/110]
```

```
Data Created: Yes
```

```
VXLAN Flags
```

```
VXLAN Decap
```

```
Stats: 1/51 [Packets/Bytes], 0.000 bps
```

```
Stats: Inactive Flow
```

```
Incoming interface: Ethernet1/50, RPF nbr: 10.0.0.42
```

```
Outgoing interface list: (count: 1) (bridge-only: 0)
```

```
nve1, uptime: 00:03:19, mrib
```

Troubleshooting BUM Traffic

```
LEAF44# show ip mroute 231.1.150.1 10.0.0.204 detail
```

```
IP Multicast Routing Table for VRF "default"
```

```
Total number of routes: 8
```

```
Total number of (*,G) routes: 2
```

```
Total number of (S,G) routes: 5
```

```
Total number of (*,G-prefix) routes: 1
```

```
(10.0.0.204/32, 231.1.150.1/32), uptime: 00:03:57, ip(0) mrib(1) pim(0)
```

```
RPF-Source: 10.0.0.204 [3/110]
```

```
Data Created: Yes
```

```
VXLAN Flags
```

```
VXLAN Decap
```

```
VPC Flags
```

```
RPF-Source Forwarder
```

```
Stats: 1/51 [Packets/Bytes], 0.000 bps
```

```
Stats: Inactive Flow
```

```
Incoming interface: Ethernet1/50, RPF nbr: 10.0.0.42
```

```
Outgoing interface list: (count: 1) (bridge-only: 0)
```

```
nve1, uptime: 00:03:57, mrib
```


Troubleshooting BUM Traffic

```
LEAF45(config)# monitor session 1
LEAF45(config-monitor)# source interface ethernet 1/50
LEAF45(config-monitor)# destination interface sup-eth 0
LEAF45(config-monitor)# no shut
LEAF45(config-monitor)# end
```

```
LEAF45# ethanalyzer local interface inband capture-filter "host 231.1.150.1" limit-
captured-frames 0
```

Capturing on inband

```
2018-05-21 16:21:01.985236    10.0.0.204 -> 231.1.150.1  UDP Source port: 41316
Destination port: 4789
```

Use the **detail** option with **ethanalyzer** to see the whole packet