Introduction to VXLAN

The future path of your datacenter

Rahul Parameswaran – Technical Marketing Engineer
@rahulsp299

DGTL-BRKDCN-1645

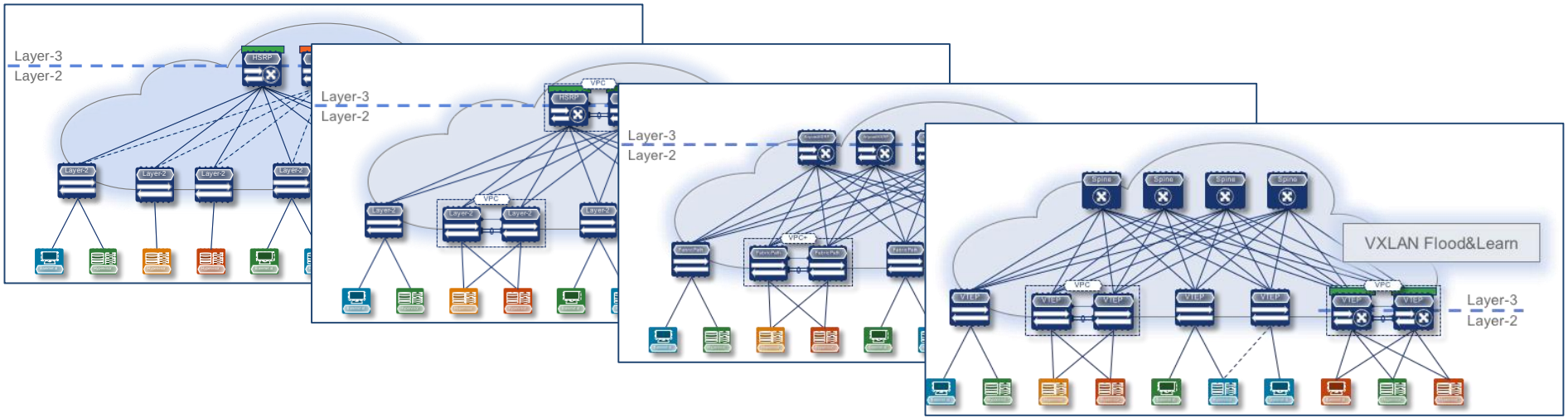
Possibilities

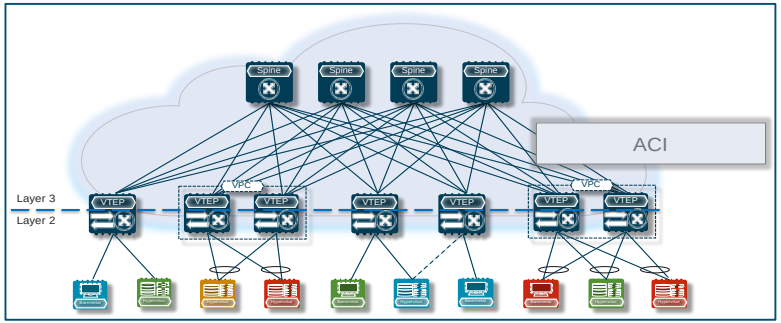
#CiscoLive

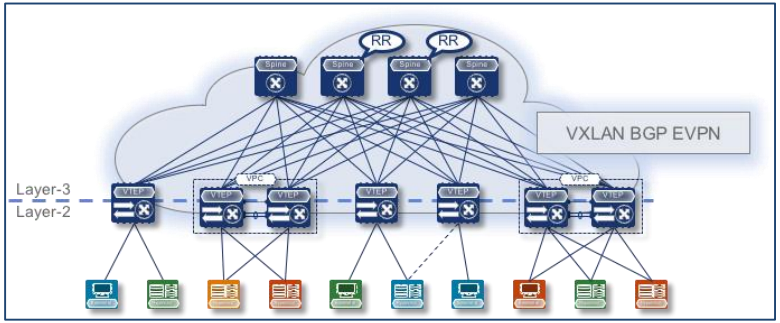
Agenda

- A short overview on Data Center Evolution
- Introduction to Overlays and VXLAN
- Understanding how MP-BGP is used as a control plane
- Packet Walk with VXLAN
- Design options and additional use cases

Data Center “Fabric” Journey







Why VXLAN Overlay

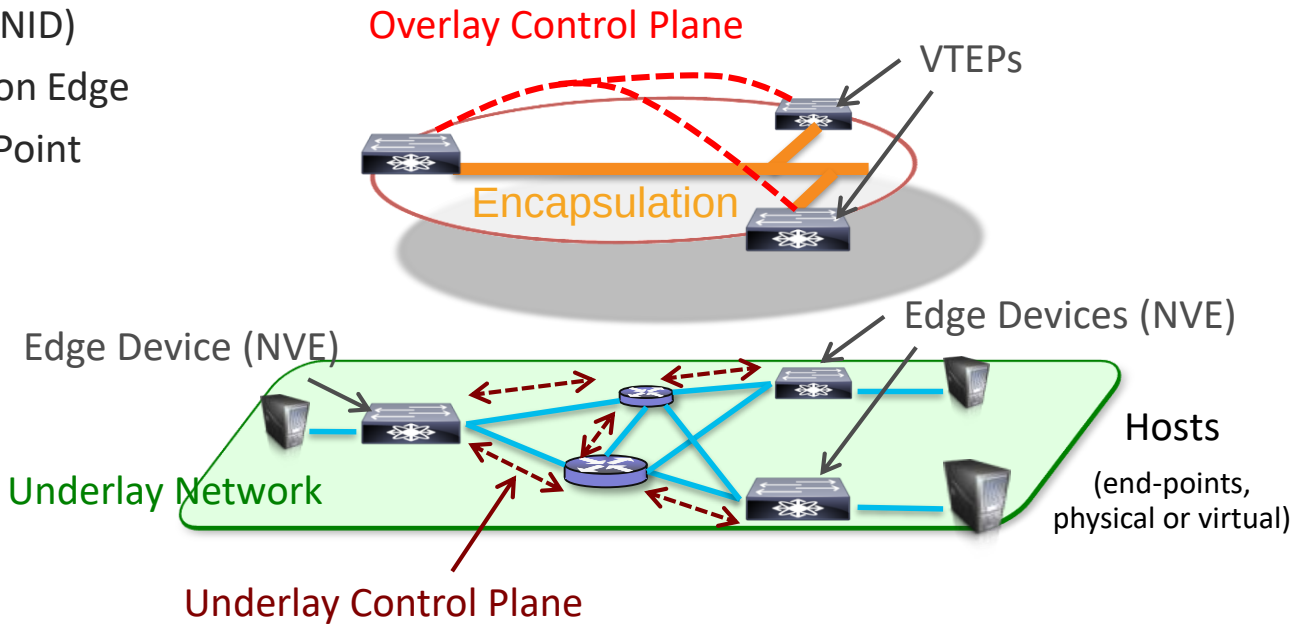
Customer Needs	VXLAN Delivered
Any workload anywhere – VLANs limited by L3 boundaries	Any Workload anywhere- across Layer 3 boundaries
VM Mobility	Seamless VM Mobility
Scale above 4k Segments (VLAN limitation)	Scale up to 16M segments
Efficient use of bandwidth	Leverages ECMP for optimal path usage over the transport network
Secure Multi-tenancy	Traffic & Address Isolation

Overlay Taxonomy

Identifier = VN Identifier (VNID)

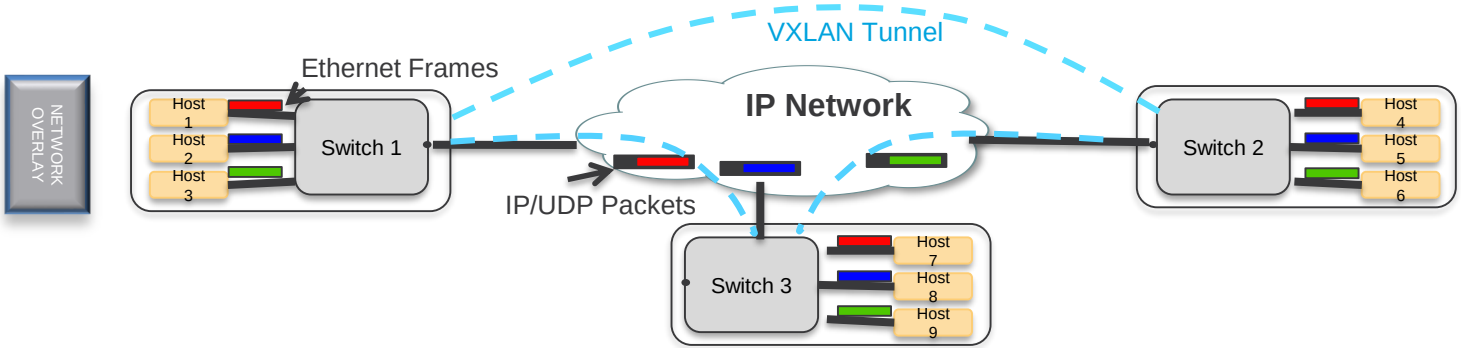
NVE = Network Virtualisation Edge

VTEP = VXLAN Tunnel End-Point

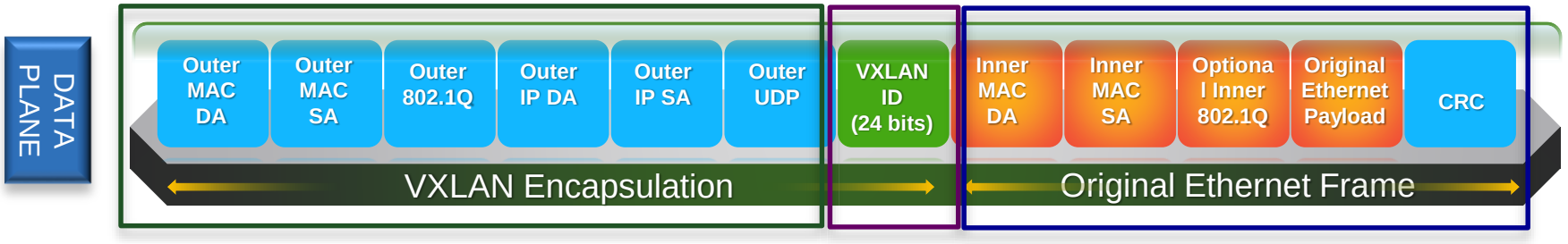


VXLAN Packet

- VXLAN is point to multi-point tunneling mechanism to extend Layer 2 networks over an IP network



- VXLAN uses MAC in UDP encapsulation (UDP destination port 4789)

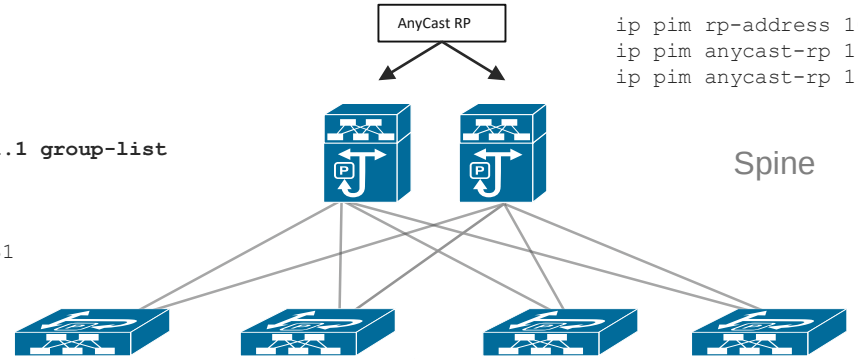


Lets Build a VXLAN Fabric

VXLAN Fabric – Creating the underlay network

IP routed Network

- Flexible topologies
- Recommend a network with redundant paths using ECMP for load sharing
- Support any routing protocols --- OSFP, IS-IS, BGP, etc.
- All proven best practices for IP routing network apply




```
ip pim rp-address 10.237.1.1 group-list 224.0.0.0/4
ip pim anycast-rp 10.237.1.1 10.255.255.101
ip pim anycast-rp 10.237.1.1 10.255.255.102
```

```
ip pim rp-address 10.237.1.1 group-list
224.0.0.0/4

interface Ethernet1/49
  ip address 192.168.1.0/31
  mtu 9216
  ip pim sparse-mode
```

```
router ospf 1
interface Ethernet1/50
  mtu 9216
  ip address 192.168.1.10/31
  ip router ospf 1 area 0.0.0.0
```

Two Modes of VXLAN

Flood-and-Learn VXLAN:

- No control plane
- Data driven flood and learning

  → Ethernet in the overlay network

- Limited scale
- Limited workload mobility
- Centralized Gateway
- Security Risk

VXLAN EVPN:

- EVPN as control plane
- VTEPs exchange L2/L3 host and subnet reachability through EVPN control plane

  → Routing protocol for both L2 and L3 forwarding

- Increased scale and stability
- Optimized workload mobility
- Distributed Anycast Gateway
- Increased Security

VXLAN BUM Traffic Handling

- BUM Traffic --- Multi-destination traffic
  - Broadcast
  - Unknown Layer-2 Unicast
  - Multicast

BUM Traffic transport mechanisms

- Multicast replication

  Requests the underlay network to run IP multicast

- Ingress unicast replication

  One unicast replica per remote VTEP

  Increase traffic load throughout the network

VXLAN with BGP EVPN Control Plane

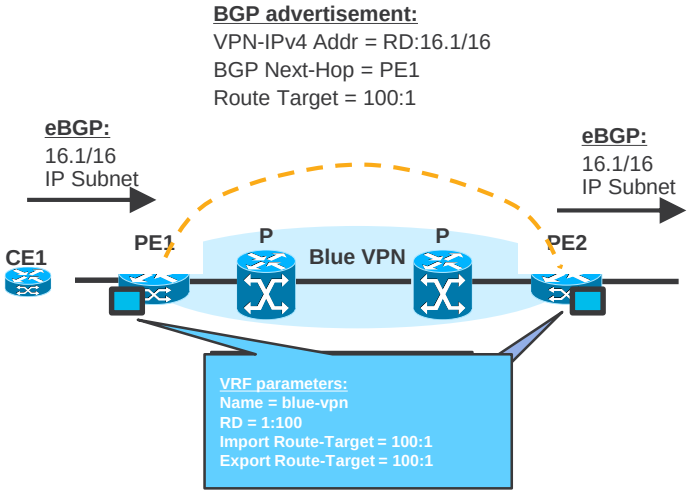
EVPN Primer --- MP-BGP Review

Virtual Routing and Forwarding (VRF)
Layer-3 segmentation for tenants' routing space

Route Distinguisher (RD):
8-byte field, VRF parameters; unique value to make VPN IP routes unique: RD + VPN IP prefix

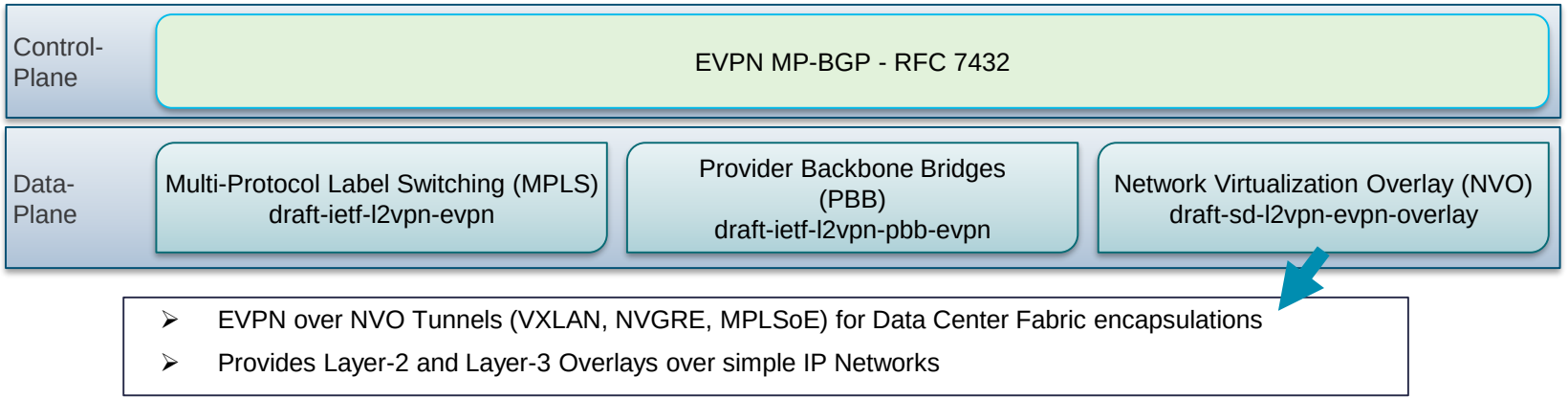
Selective distribute VPN routes:
Route Target (RT): 8-byte field, VRF parameter, unique value to define the import/export rules for VPNv4 routes

VPN Address-Family:
Distribute the MP-BGP VPN routes



What is VXLAN/EVPN?

- Standards based Overlay (VXLAN) with Standards based Control-Plane (BGP)
- Layer-2 MAC and Layer-3 IP information distribution by Control-Plane (BGP)
- Forwarding decision based on Control-Plane (minimizes flooding)
- Integrated Routing/Bridging (IRB) for Optimized Forwarding in the Overlay

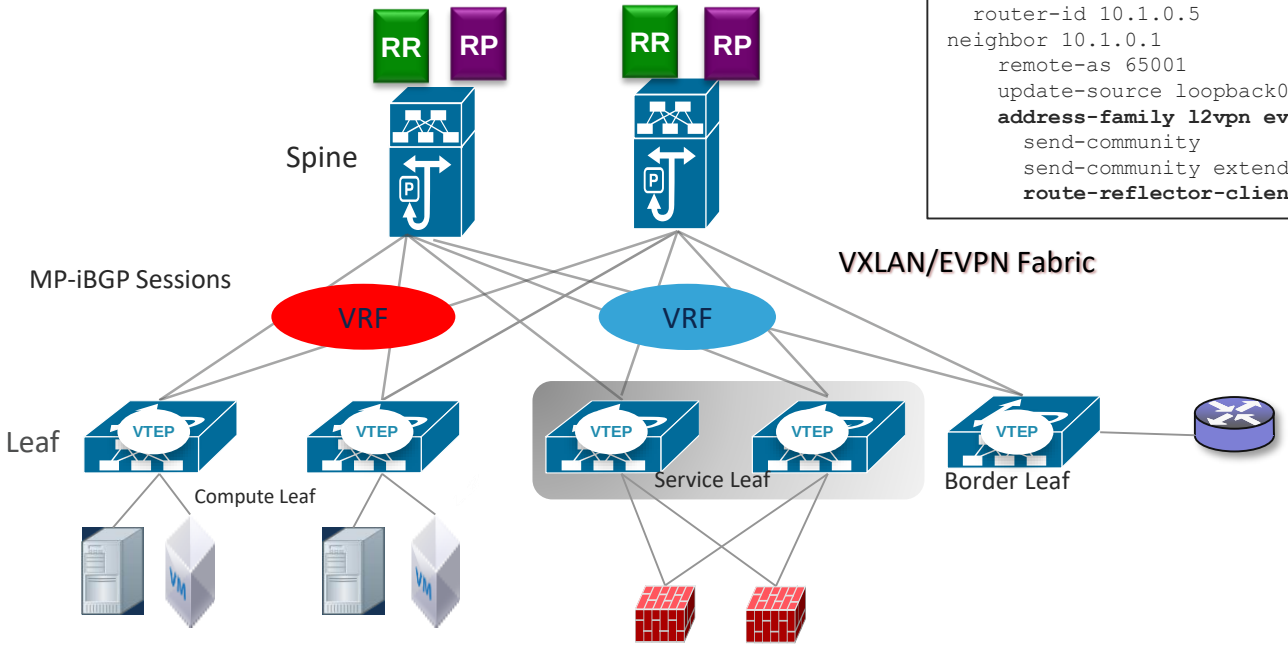


EVPN based VXLAN Fabric

RR EVPN Route Reflector

RP Rendezvous Point (Underlay)

```
! leaf bgp config
router bgp 65001
  router-id 10.1.0.4
  neighbor 10.1.0.5
    remote-as 65001
    update-source loopback0
    address-family l2vpn evpn
      send-community
      send-community extended
  vrf VRF-RED
    address-family ipv4 unicast
      advertise l2vpn evpn
    address-family ipv6 unicast
      advertise l2vpn evpn
  vrf VRF-BLUE
    address-family ipv4 unicast
      advertise l2vpn evpn
    address-family ipv6 unicast
      advertise l2vpn evpn
```




```
! spine bgp config

router bgp 65001
  router-id 10.1.0.5
neighbor 10.1.0.1
    remote-as 65001
    update-source loopback0
    address-family l2vpn evpn
      send-community
      send-community extended
      route-reflector-client
```

Configuration Snippet

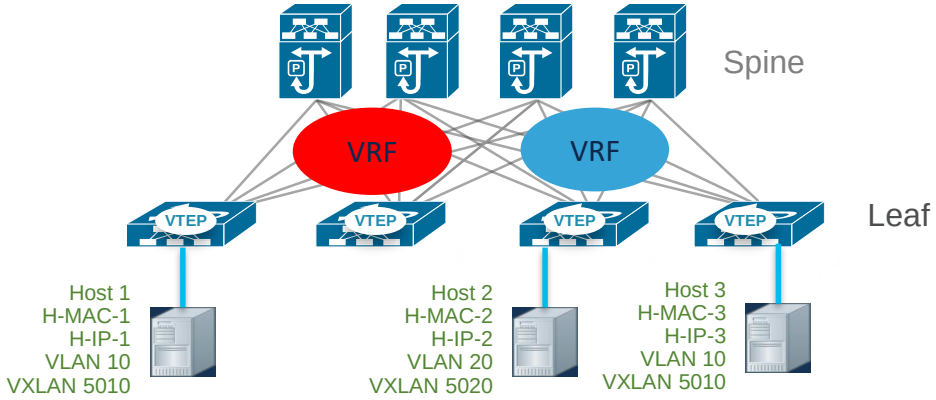
```
Vlan 10
 vn-segment 5010
Vlan 20
 vn-segment 5020

Vlan 1000
 !Layer 3 VNI
 vn-segment 9999
Vlan 2000
 !Layer 3 VNI
 vn-segment 9998

interface Vlan10
  no shutdown
  vrf member VRF-RED
  ip address 192.168.10.254/24 tag 12345
  ipv6 address 2001::1/64 tag 12345
  fabric forwarding mode anycast-gateway

interface Vlan20
  no shutdown
  vrf member VRF-BLUE
  ip address 192.168.20.254/24 tag 12345
  ipv6 address 2002::1/64 tag 12345
  fabric forwarding mode anycast-gateway

interface nve1
  source-interface loopback0
  host-reachability protocol bgp
  member vni 5010
    mcast-group 239.1.1.1
  member vni 5020
    mcast-group 239.1.1.1
  member vni 9999 associate-vrf
  member vni 9998 associate-vrf
```

Layer 2 VNI

Layer 3 VNI

Map L2VNI to NVE

Associate L3VNI to NVE



Layer 3 VNI

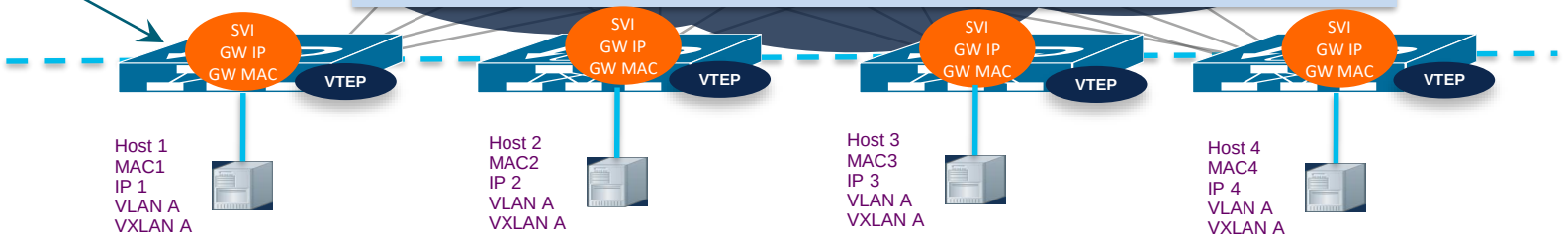
```
vrf context VRF-RED
  vni 9999
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
evpn
  vni 5010 l2
    rd auto
    route-target both auto
```

```
vrf context VRF-BLUE
  vni 9998
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
evpn
  vni 5020 l2
    rd auto
    route-target both auto
```

Distributed Anycast Gateway in MP-BGP EVPN

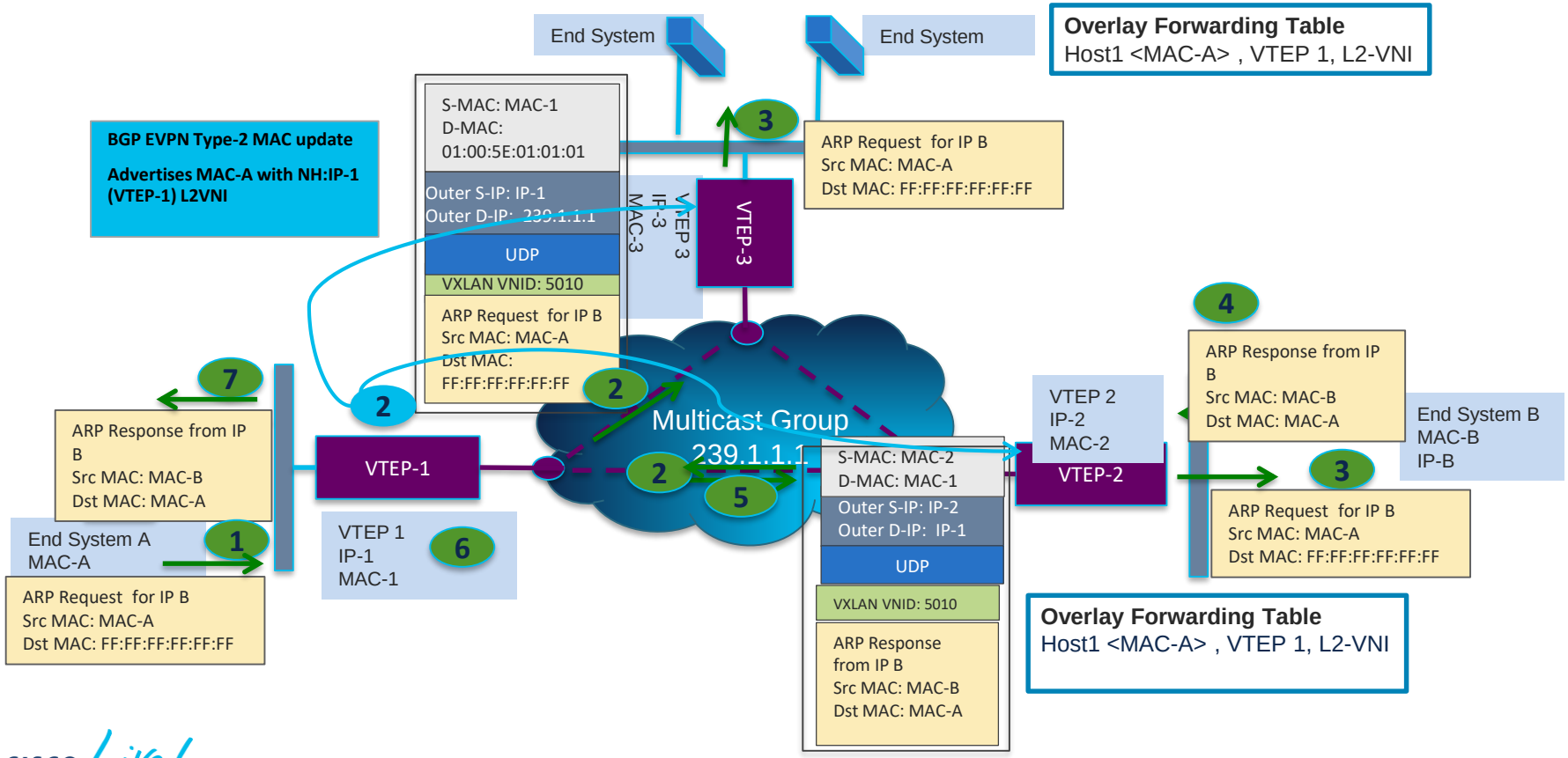
The same anycast gateway virtual IP address and MAC address are configured on all VTEPs in the VNI.

```
# VLAN to VNI mapping
vlan 20
 vn-segment 5020

# Anycast Gateway MAC, identically configured on all VTEPs
fabric forwarding anycast-gateway-mac 0002.0002.0002

# Distributed IP Anycast Gateway (SVI)
# Gateway IP address needs to be identically configured on all
VTEPs
interface vlan 20
 no shutdown
 vrf member VRF-BLUE
 ip address 192.168.20.254/24
 ipv6 address 2002::1/64
 fabric forwarding mode anycast-gateway
```



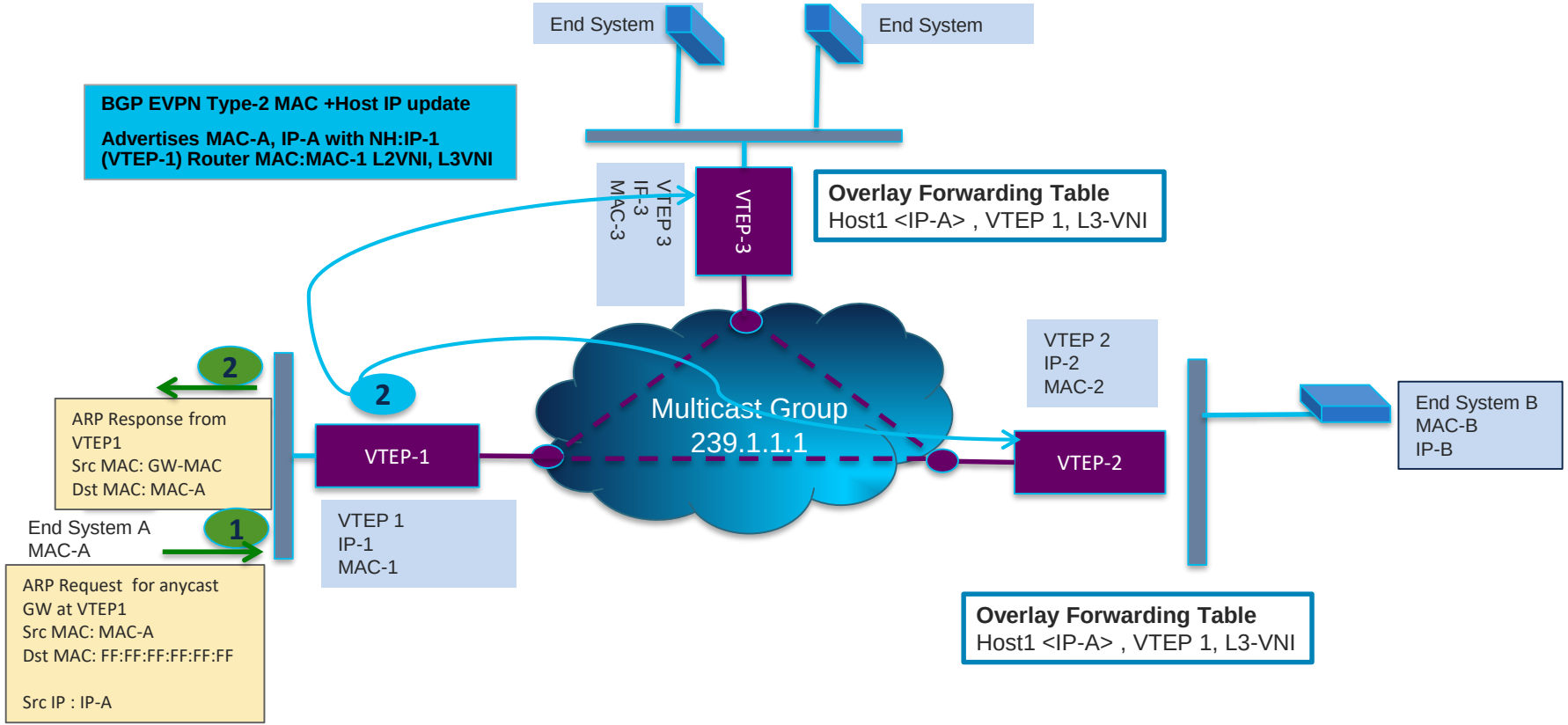
EVPN Peer and Endpoint(Host) Discovery

Triggered by Host Communication across the same VLAN/VNI (L2)



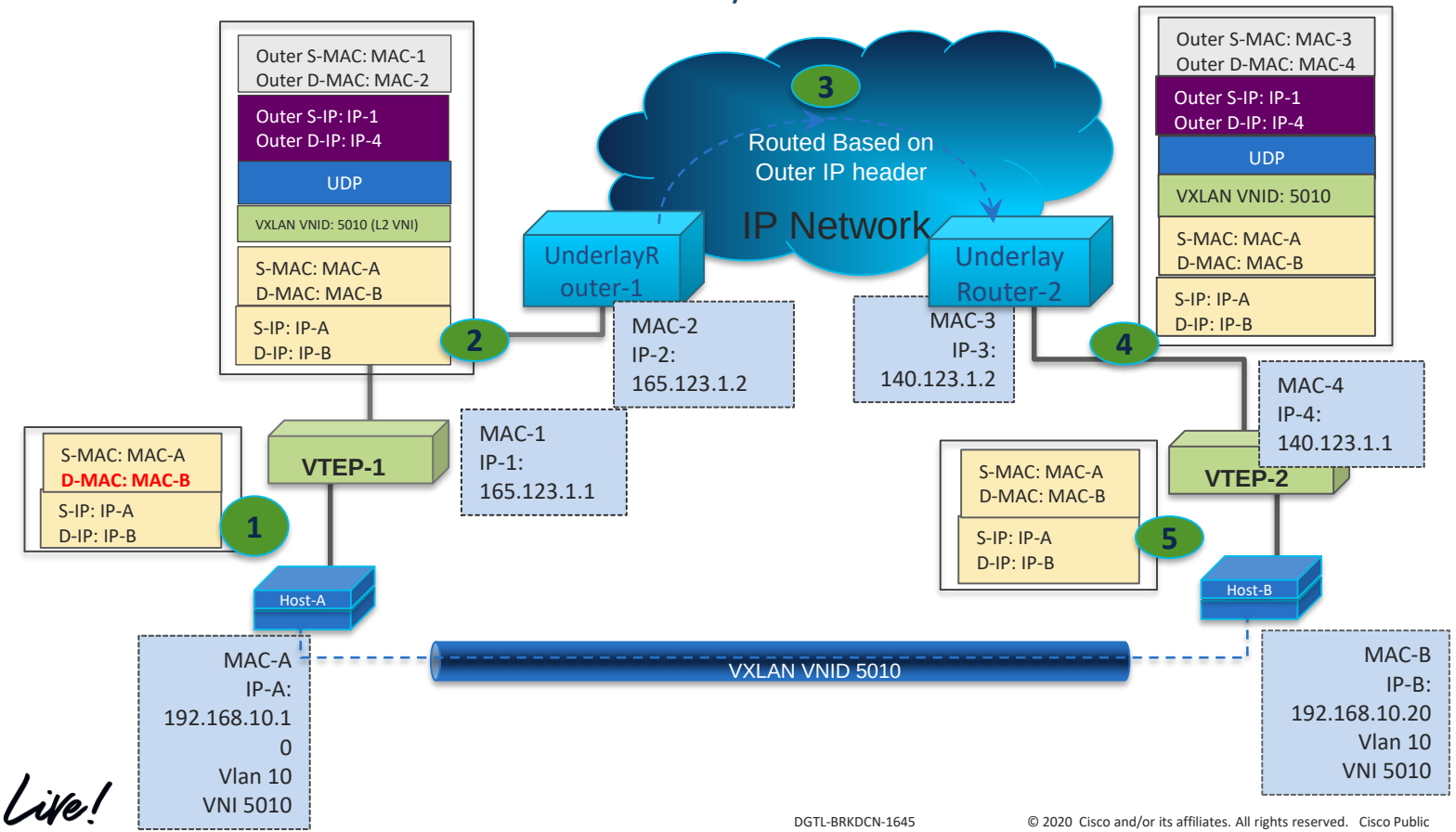
EVPN Peer and Endpoint(Host) Discovery

Triggered by Host Communication between VLAN/VNI (L3)



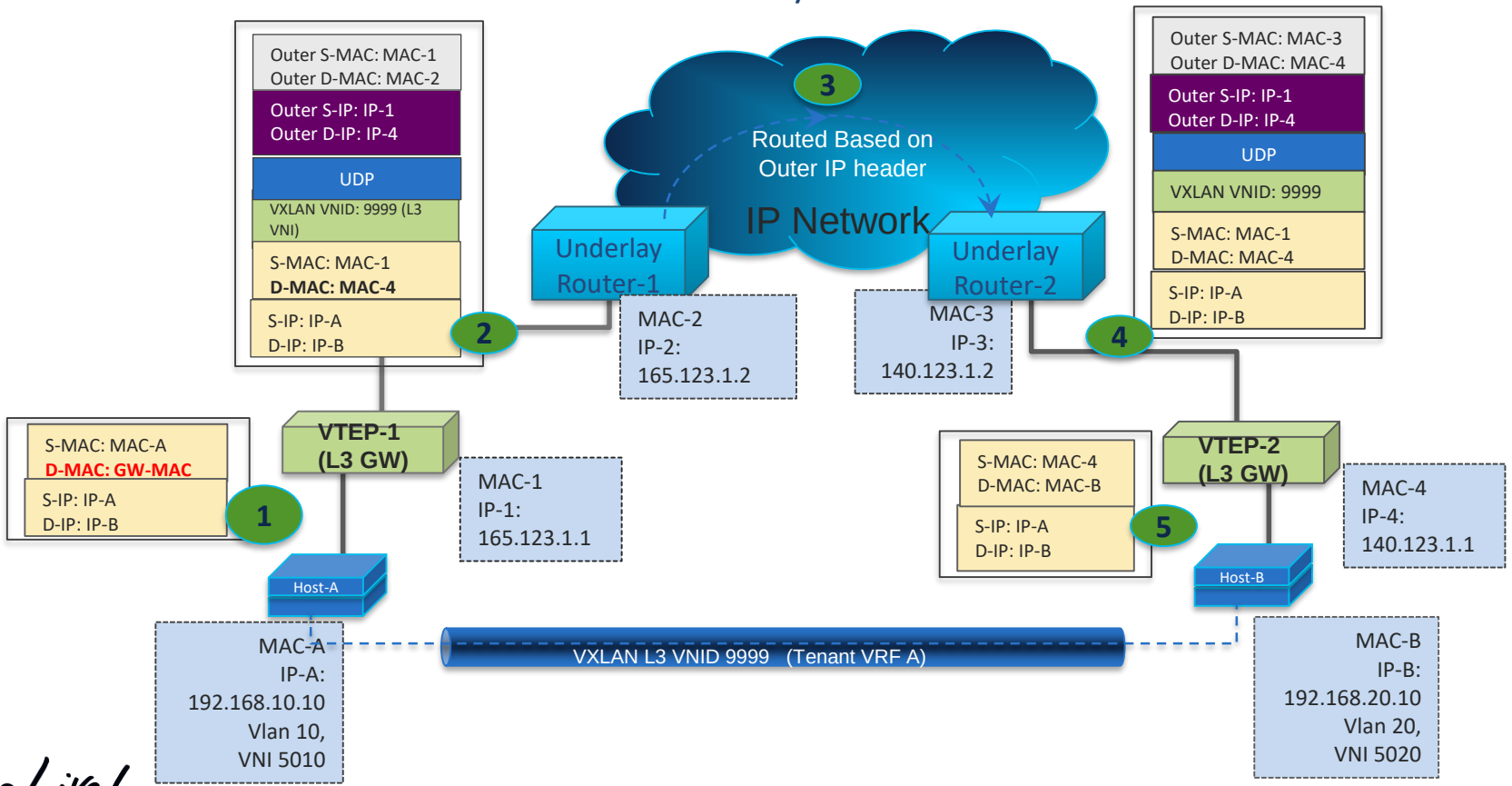
Packet Walk

Communication between hosts in same VLAN/VNI



Packet Walk

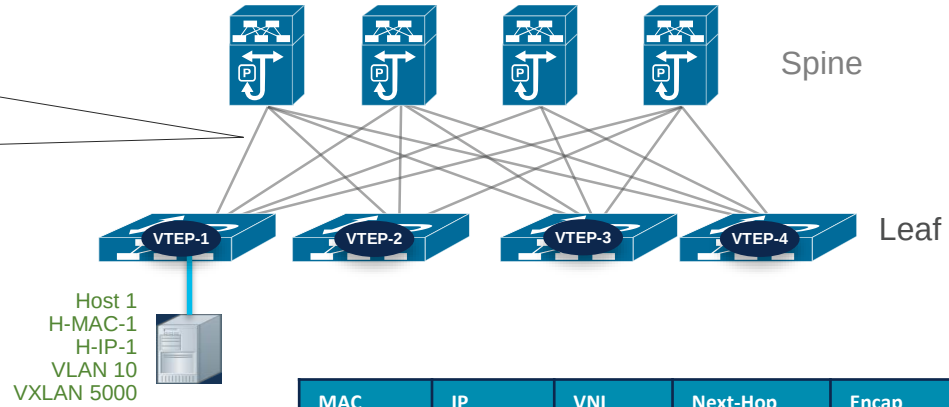
Communication between hosts in different VLAN/VNI



EVPN Control Plane --- VM Mobility

NLRI:
- Host H-MAC-1, H-IP-1
- NVE VTEP-1
- VNI 5000

Ext. Community:
- Encapsulation: VXLAN
- Cost
- Sequence number :0



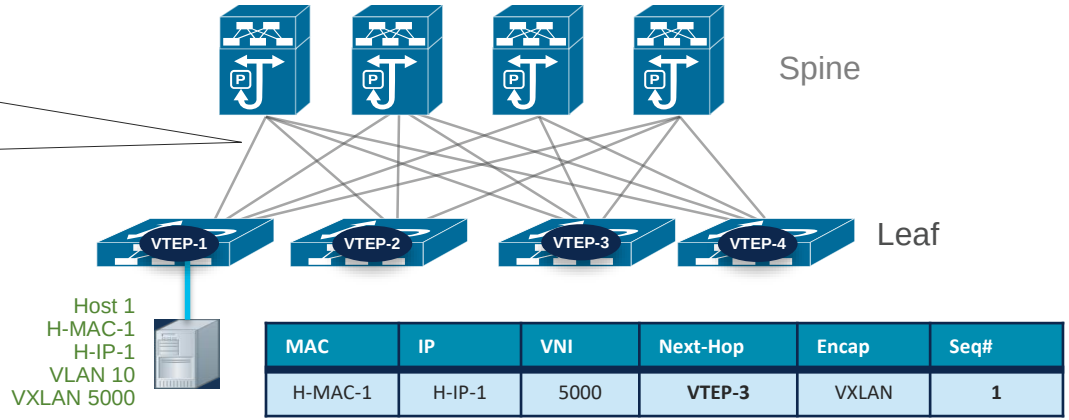
MAC	IP	VNI	Next-Hop	Encap	Seq#
H-MAC-1	H-IP-1	5000	VTEP-1	VXLAN	0

1. Host 1 attaches to VTEP-1
2. VTEP-1 detects Host1 and advertises H1 with seq #0
3. Other VTEPs learn about the host route of Host 1

EVPN Control Plane --- VM Mobility

NLRI:
- Host H-MAC-1, H-IP-1
- NVE VTEP-3
- VNI 5000

Ext. Community:
- Encapsulation: VXLAN
- Cost
- Sequence number: 1



MAC	IP	VNI	Next-Hop	Encap	Seq#
H-MAC-1	H-IP-1	5000	**VTEP-3**	VXLAN	**1**

1. Host 1 moves to VTEP-3 from VTEP-1
2. VTEP-3 detects Host 1, sends MP-BGP update for Host 1 with its own VTEP address and a new seq #1
3. Other VTEPs learn about the new route of Host 1 from VTEP 3 with a higher sequence number and prefer that update

EVPN Control Plane --- ARP Suppression

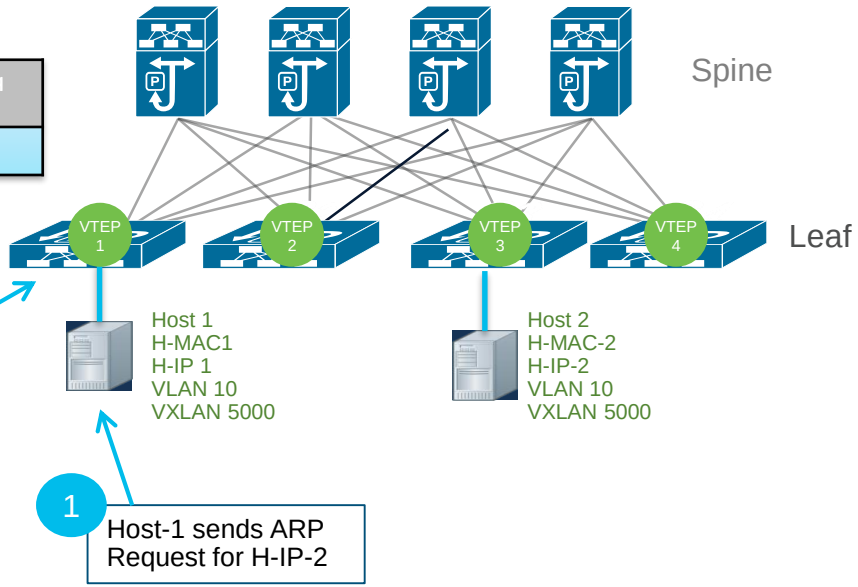
Minimize flood-&-learn behavior for host learning

MAC	IP	VNI	Next-Hop	Encap	Seq
H-MAC-2	H-IP-2	5000	VTEP-3	VXLAN	0

2

VTEP-1 receives and intercepts the ARP Request. Checks in its own host table.

- If it has an match for H-IP-2, it'll send ARP response on behave of Host-2
- If it doesn't have a match for H-IP-2, it'll forward the ARP request to remote VTEPs via multicast encap or head-end replication

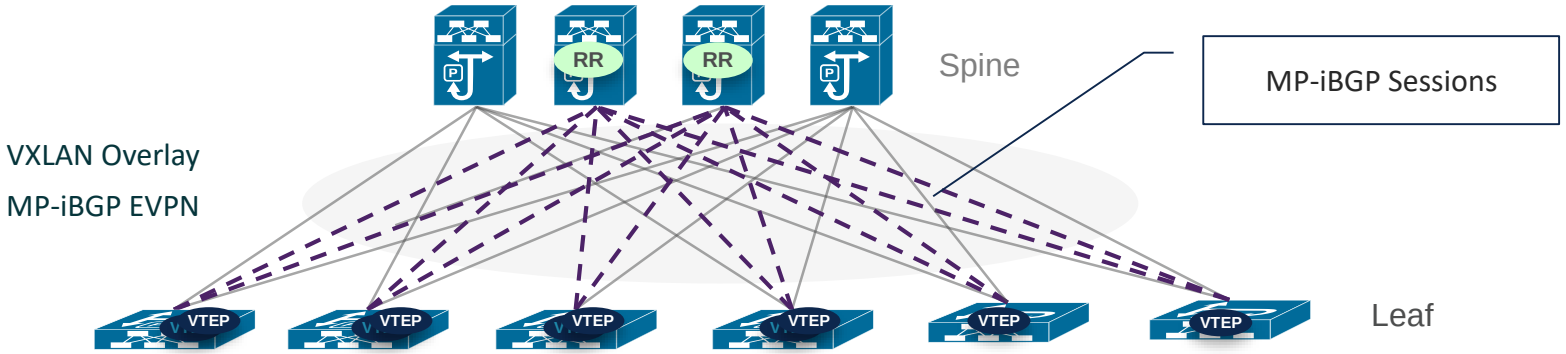


Functions of VXLAN/EVPN

Host/Network Reachability Advertisement	Advertise host/network reachability information through control protocol (MP-BGP)
VTEP Security & Authentication	Authenticate VTEPs through BGP peer authentication
Distributed Anycast Gateway	Seamless and Optimal vm-mobility
ARP Suppression	Early ARP termination Localize ARP learning process Minimize network flooding

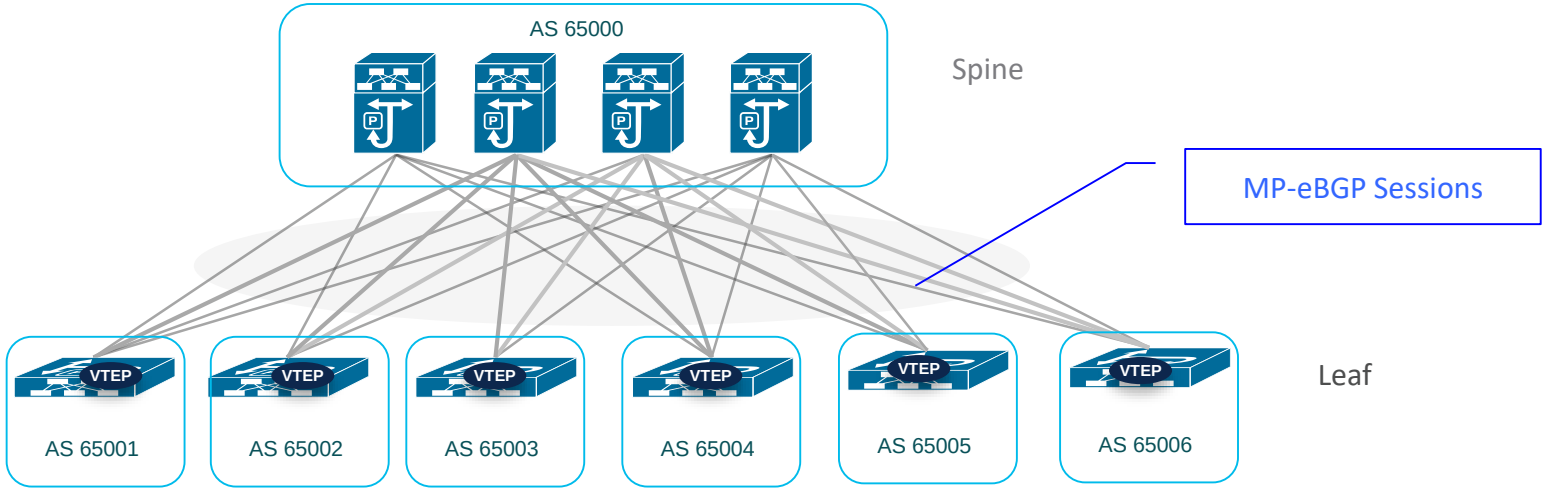
Design Options and Use case

VXLAN Fabric Design with MP-iBGP EVPN



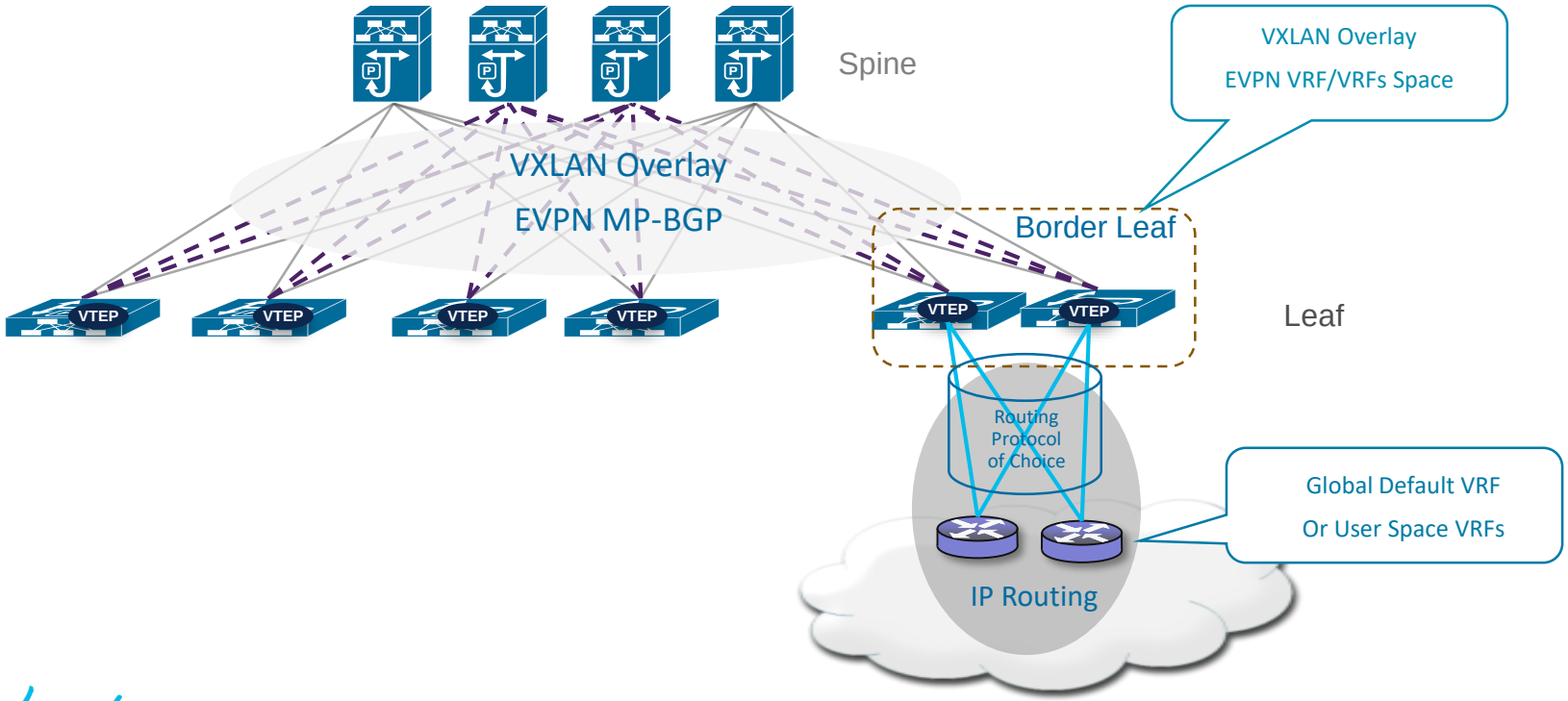
- VTEP Functions are on leaf layer
- Spine nodes are iBGP route reflector
- Spine nodes don’t need to be VTEP

VXLAN Fabric Design with MP-eBGP EVPN

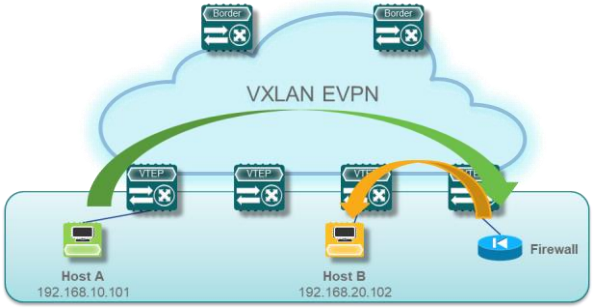


- VTEP Functions are on leaf layer
- Spine nodes are MP-eBGP Peers to VTEP leafs
- Spine nodes don’t need to be VTEP
- VTEP leafs can be in the same or different BGP AS’s

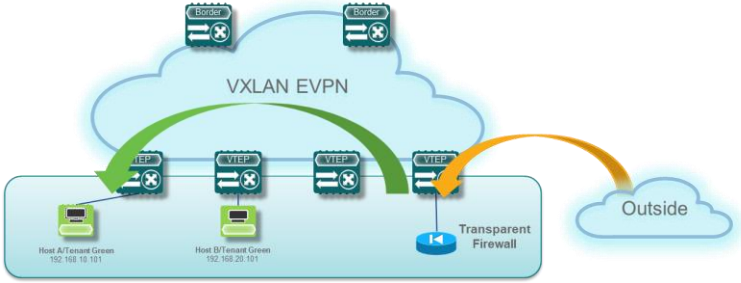
VXLAN Fabric - External Routing



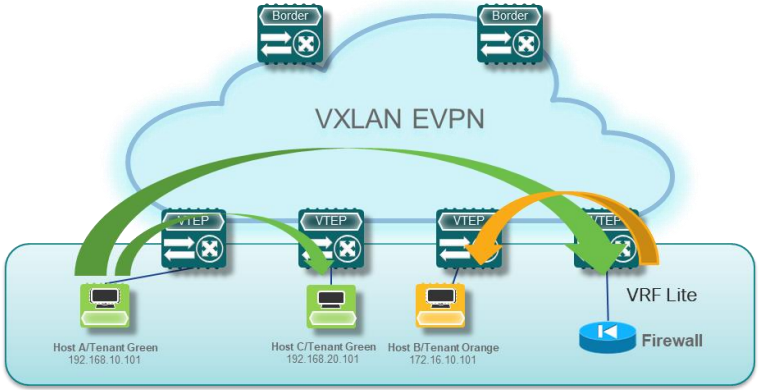
VXLAN Fabric – Service Insertion



Firewall as a default gateway : Centralized Gateway- Firewall bottleneck



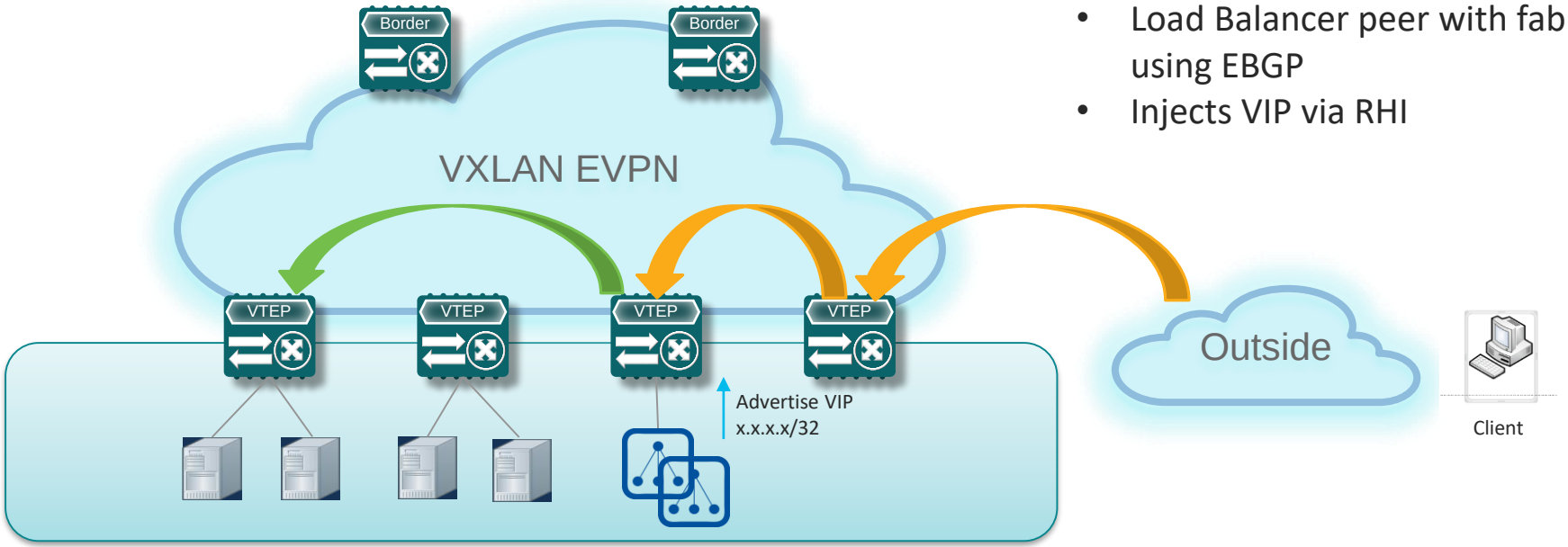
Transparent Firewall : Inspect and then bridge Traffic between “dirty” VLAN and “clean” VLAN



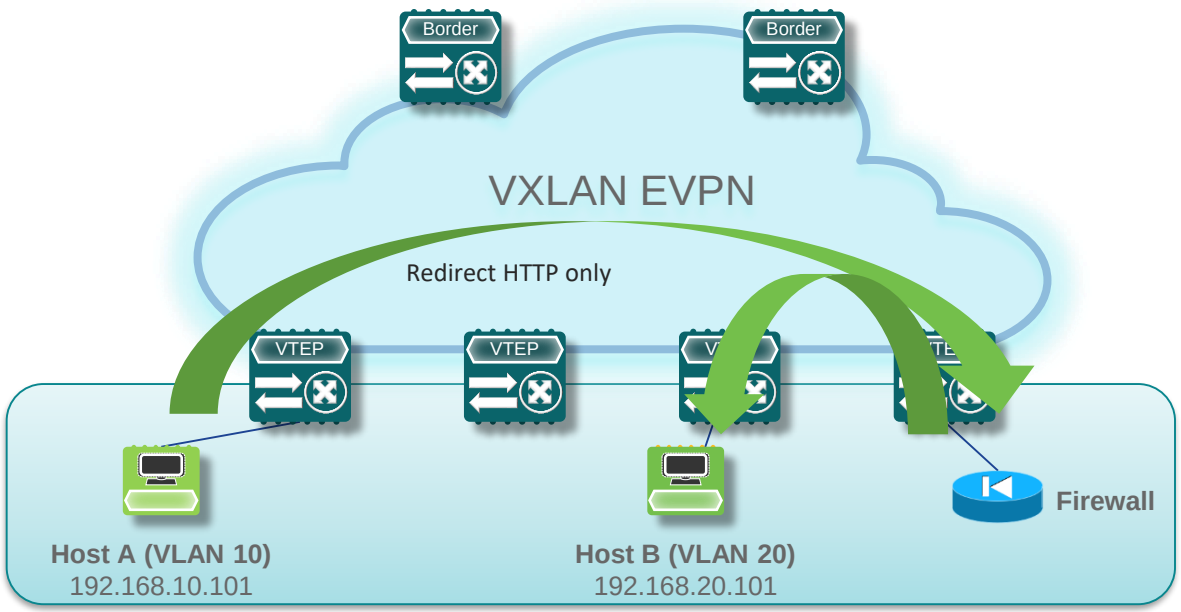
Tenant Edge Firewall: Traffic between Tenants/VRFs routed via the firewall

VXLAN Fabric – Service Insertion

- Load Balancer Integration
- Load Balancer peer with fabric using EBGP
- Injects VIP via RHI



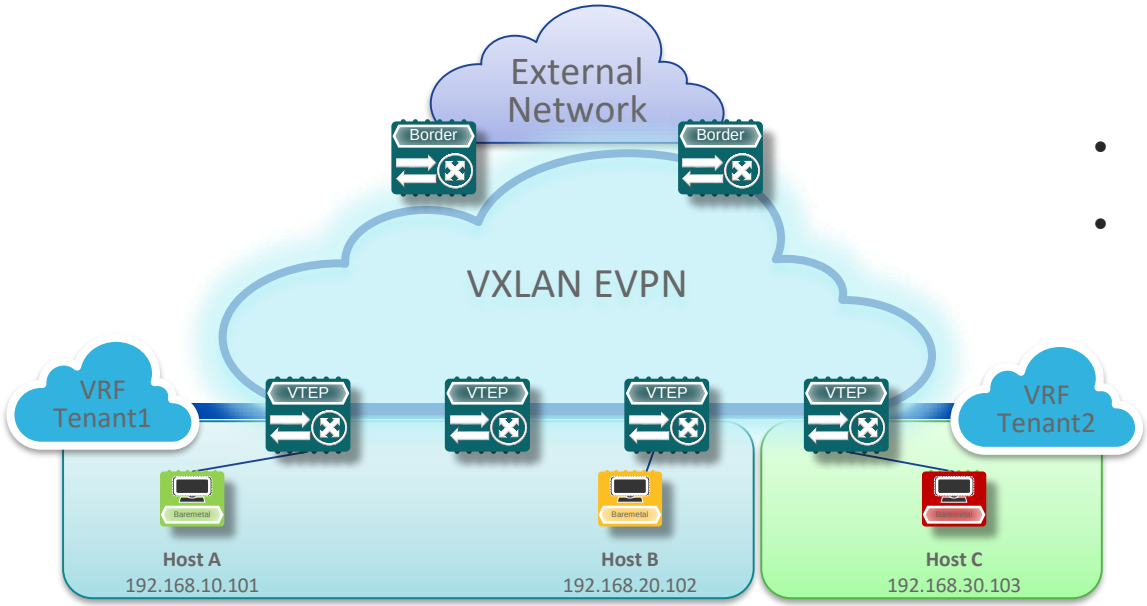
VXLAN Fabric – Selective Traffic Redirection



- Leverages Policy Based Redirect
- Inter VLAN traffic bypass default routing lookup and redirected
- Service Redirection to Load Balancers, Firewalls etc.

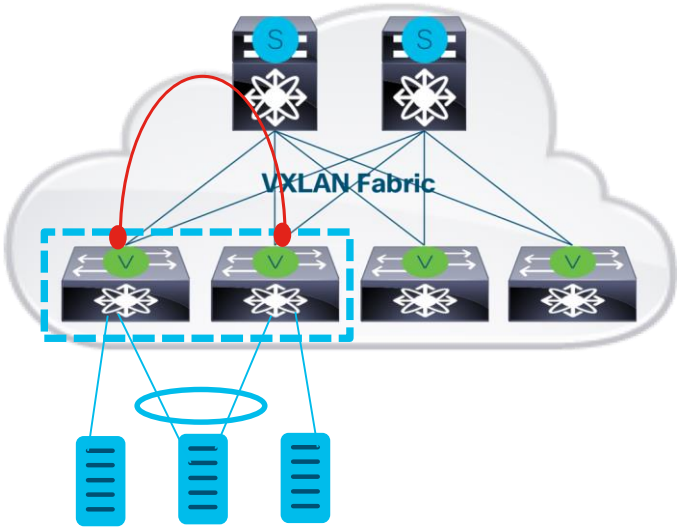
VXLAN Fabric – Centralized Route Leaking

- Extranet Support



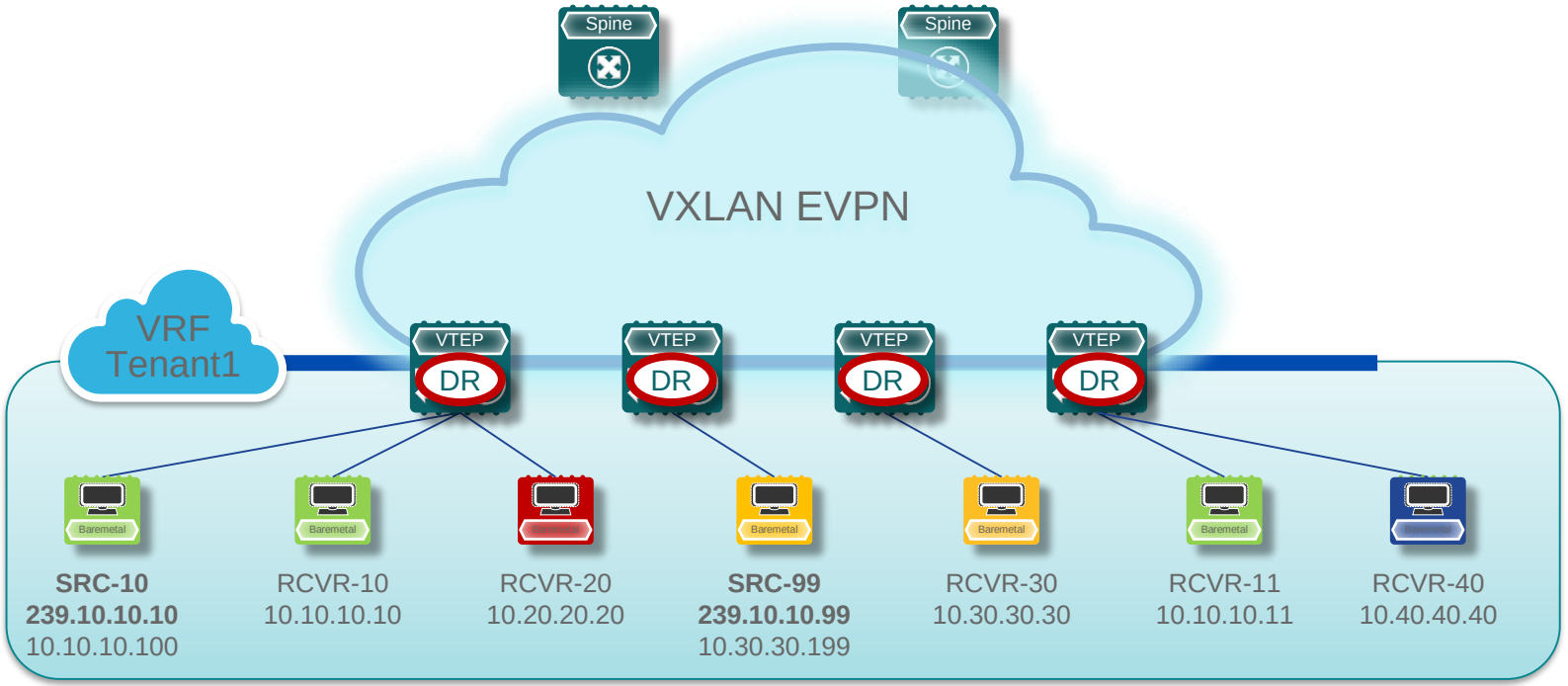
- Use Cases – Shared Services, External Connectivity
- VRF to VRF or VRF to Default
- Centralize Location for leaking routes

Peerlink-Less VPC

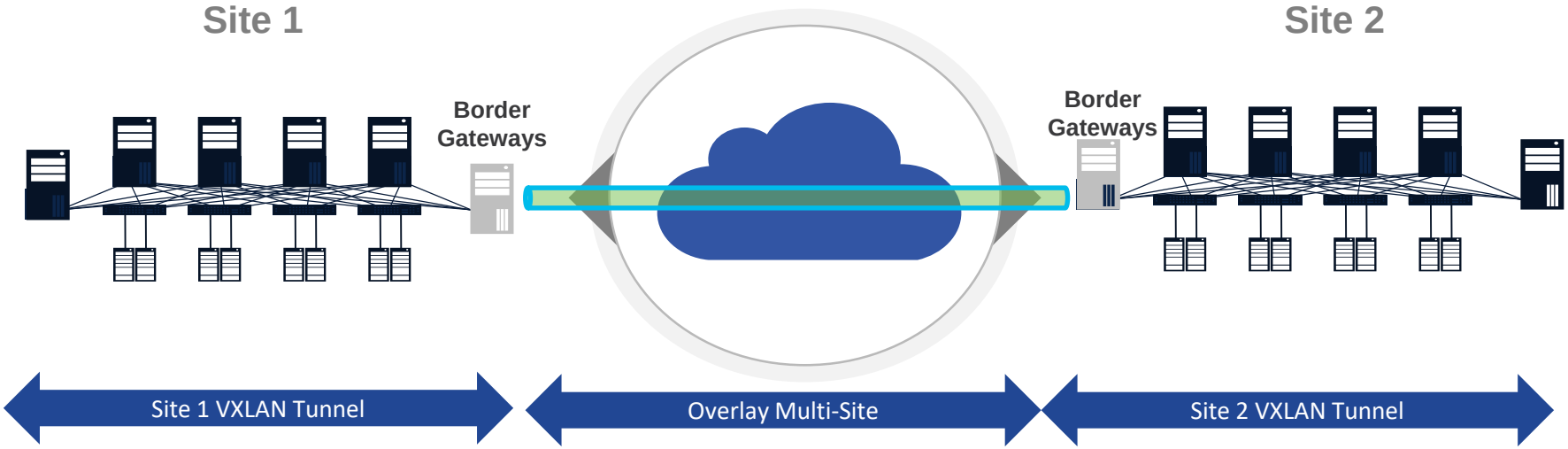


- Enhanced dual-homing solution without wasting physical ports
- Preserve traditional vPC characteristics

VXLAN Fabric – Tenant Routed Multicast



VXLAN EVPN Multi-Site



Scale through Hierarchical Forwarding

Fault Containment	Convergence independent of Network Size	Separate Admin Domains	Single Box

Summary

Summary

- VXLAN enables scalable Data Center fabrics
- BGP EVPN with VXLAN provides a robust control plane enabling multi-tenancy, VM mobility , optimizing traffic forwarding
- Seamless integration with service nodes such as Firewalls and Load balancers and ability to provide shared services
- Fabric can cater to multicast traffic in the overlay
- VXLAN as a DCI with Multi-Site

Thank you

Possibilities

#CiscoLive