
cisco Live!
ALL IN

The bridge to possible

Introduction to VXLAN

The Future Path of Your Data Center

Tarique Shakil
Principal Technical Marketing Engineer Cloud Networking, Cisco Systems

BRKDCN-1999

Cisco Webex App

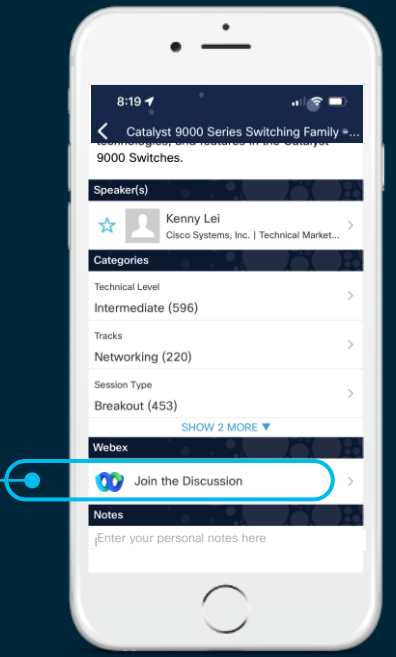
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated by the speaker until June 17, 2022.



https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-1999

Agenda

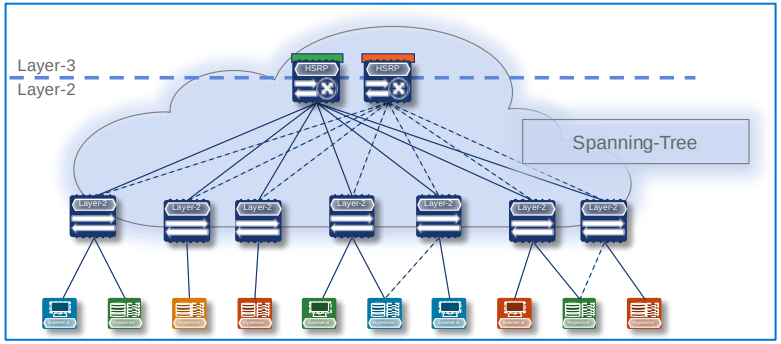
- Data Centers Challenges
- Data Center Network Overlays
- VXLAN BGP EVPN Architecture
- Configuring a VXLAN BGP EVPN Fabric
- VXLAN BGP EVPN Packet Walk
- Summary

Data Center Challenges

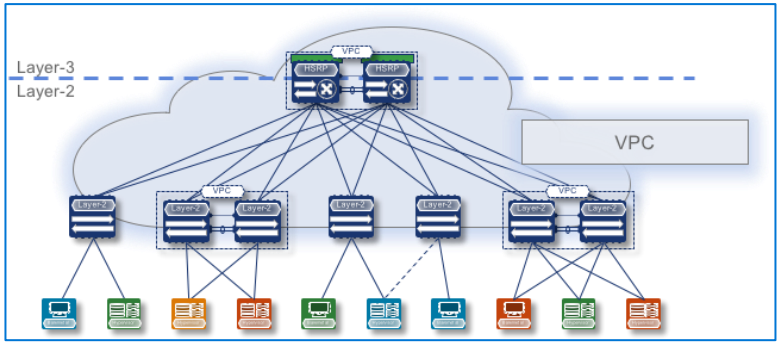
Data Center Fabric Evolution

Legacy Architectures

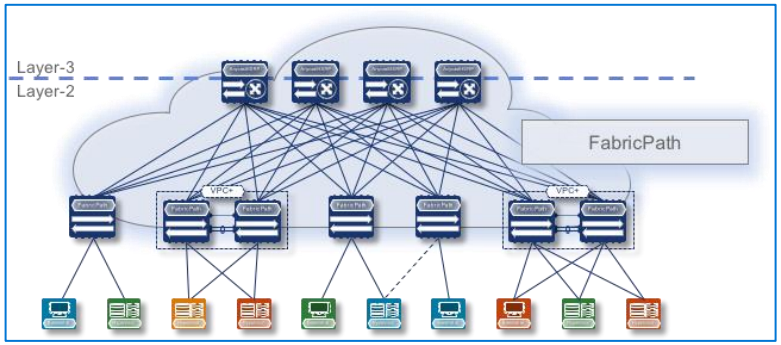
STP Based Networks



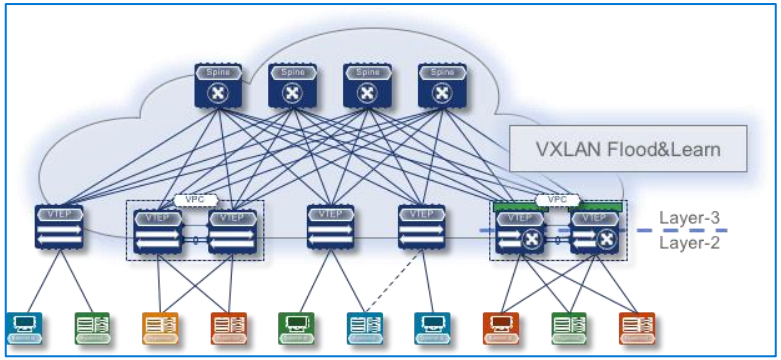
vPC Based/MCLAG Based



Fabric Path/TRILL



VXLAN Flood & Learn



Data Center Challenges

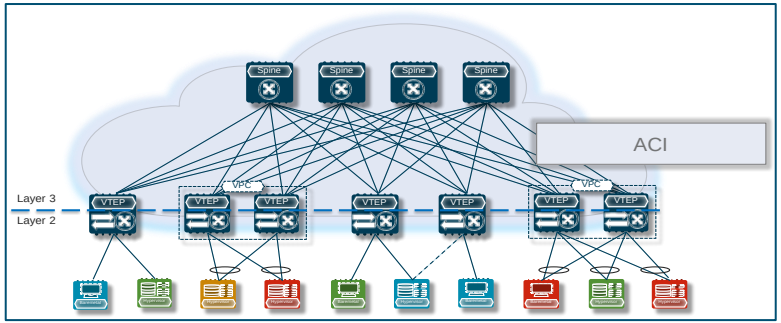
Legacy Architectures

- **Rigid Provisioning:** L4-L7 nodes always placed at distribution layer.
- **Convergence Time:** STP high convergence time. TCN MAC flush.
- **Unused Links:** STP redundant links in blocked state.
- **Suboptimal Forwarding:** path defined by root switch placement.
- **Broadcast Storms**
- **No Endpoint Mobility**
- **Scalability challenges:** Scale up vs scale out.
- **Opex/Capex:** Expensive to upgrade.
- **Lacks Elasticity:** Adapt to changing demands and requirements.

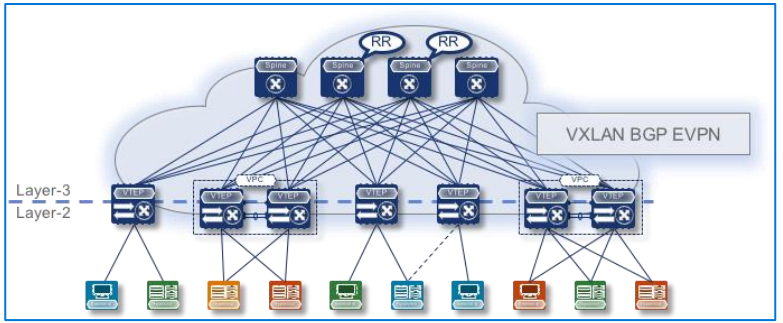
Data Center Fabric Evolution

Current Fabric Technologies

Application Centric Infrastructure – ACI



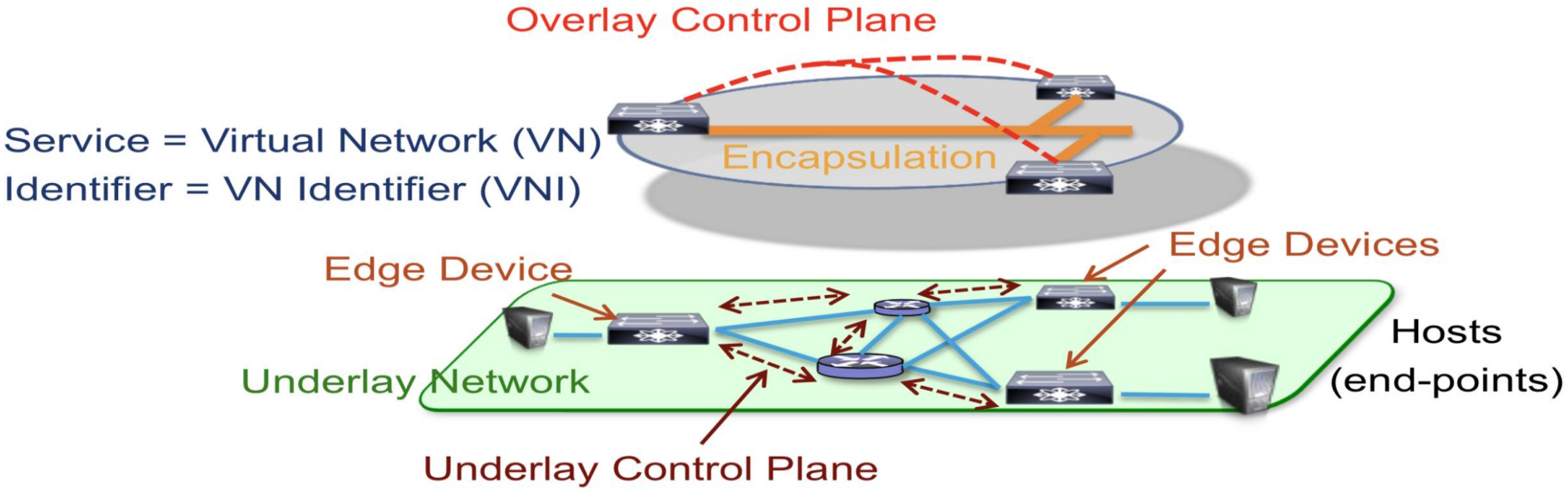
VXLAN BGP EVPN Fabrics



Focus of this session

Data Center Network Overlays

Underlay vs Overlay



Underlay is responsible for tunnel endpoint reachability while the
management of virtual tunnels is handled by the overlay.

Network Overlay Services

L2 OVERLAYS

- MPLS L2 VPNs i.e AToM, VPLS, PBB-EVPN
- Overlay Transport Virtualization OTV
- VXLAN Flood and Learn.
- VXLAN BGP EVPN (hybrid)
- L2TPv3
- Fabric Path/TRILL (MAC in MAC)
- ACI iVXLAN (hybrid)

L3 OVERLAYS

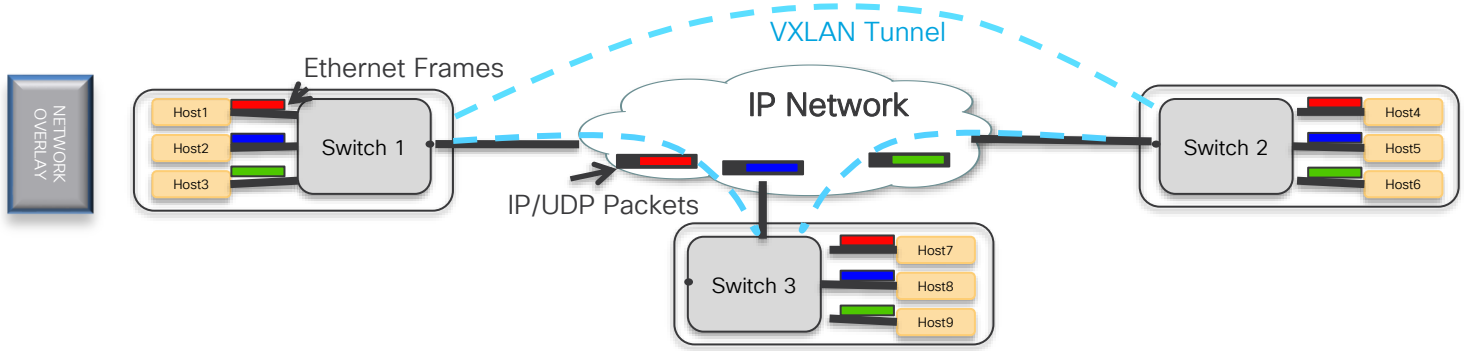
- MPLS L3 VPNs
- GRE
- LISP
- VXLAN BGP EVPN (hybrid)
- ACI iVXLAN (hybrid)

VXLAN BGP EVPN Provides Integrated Routing and Bridging (IRB) Fabric, best of L2 and L3 overlays with single overlay service.

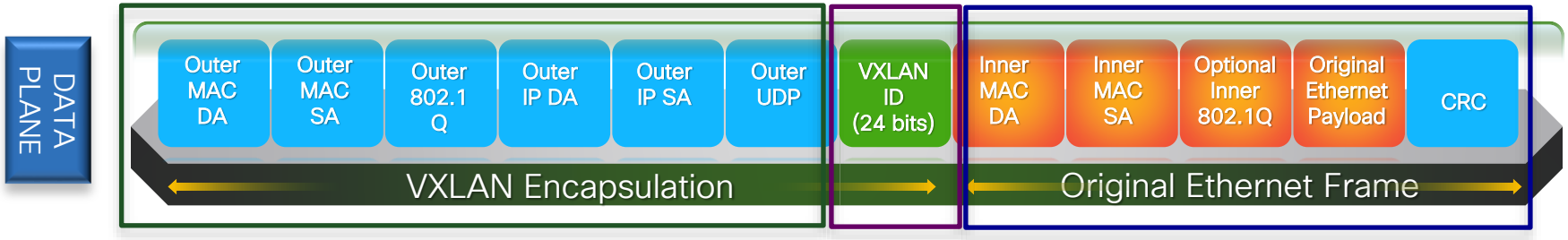
VXLAN BGP EVPN Architecture

VXLAN Data Plane Packet

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



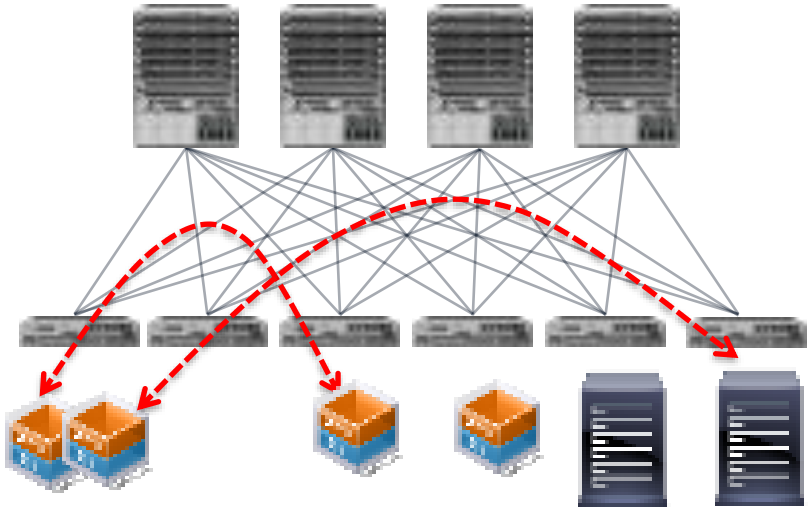
- VXLAN uses MAC in UDP encapsulation (UDP destination port 4789). VXLAN Adds 50 Bytes to the original frame.



VXLAN Overlays
Networking Components

- **VXLAN Segment**
  - VXLAN overlay network. Layer 2 Broadcast Domain.
- **VXLAN Network Identifier (VNID)**
  - Each VXLAN segment is identified by a 24-bit VNID.
- **VXLAN Tunnel Endpoint (VTEP)**
  - Tunnel Endpoint. RFC term Network Virtualization Edge.
  - Each VTEP is uniquely identified by an IP address.
  - VTEP switch when forwarding packets within the same VNID and route for inter-VNI traffic.

VTEP in a Nutshell



"VTEPs while sending traffic takes end host traffic in vlan maps to vxlan vni then encapsulates it in vxlan header. when receiving traffic from remote vxlan vtep decapsulates the vxlan packet and send it to attached host." [RFC 7348]

VXLAN Benefits

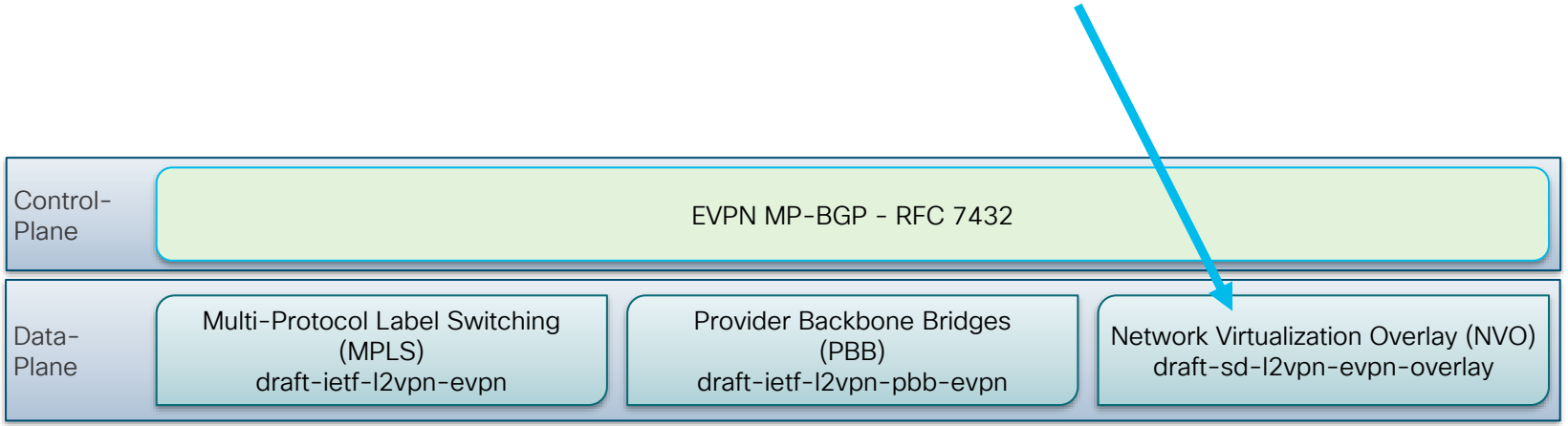
Customer Needs	VXLAN Delivered
Any workload anywhere – VLANs limited by L3 boundaries	Any Workload anywhere- across Layer 3 boundaries
VM Mobility	Seamless VM Mobility
Scale above 4k Segments (VLAN limitation)	Scale up to 16M segments
Efficient use of bandwidth	Leverages ECMP for optimal path usage over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

BGP EVPN Overview

- MP-BGP EVPN AF carries following information: MAC, IP and network prefix, VRF/VNID and VTEP IP (NLRI Next Hop).
- BGP EVPN distributes MAC,IP info avoiding flooding.
- VXLAN BGP AFI=25 (Layer 2 VPN) and SAFI = 70 (EVPN).
- VXLAN is the Tunnel Encapsulation Protocol and MP-BGP EVPN is the Control Plane for overlay distributing Layer 2 and Layer 3 routing information (MAC,IP).

BGP EVPN Standards

- EVPN over NVO Tunnels (VXLAN, NVGRE, MPLSoE) for Data Center Fabric encapsulations.
- Provides Layer-2 and Layer-3 Overlays over simple IP Networks.



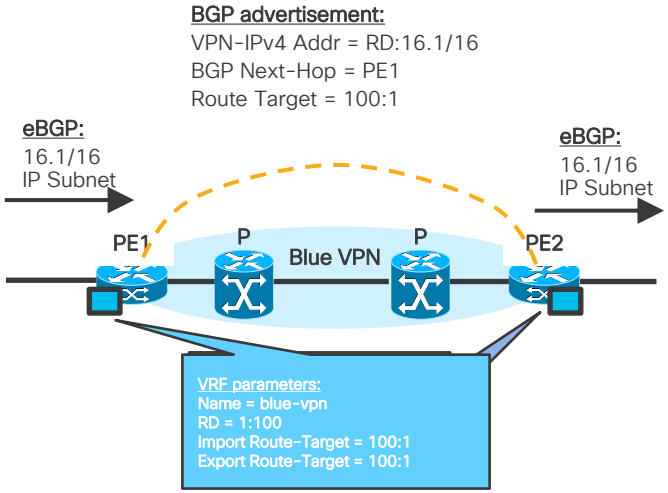
BGP EVPN Address Family

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



NX-OS VXLAN EVPN Auto RD

RD FORMAT

- The auto derived RD is constructed with the IP address of the BGP Router ID as the 4-byte administrative field (RID) and the internal VRF identifier for the 2-byte numbering field (VRF ID)
- The 2-byte numbering field for the IP-VRF uses the internal VRF ID starting at 1 and increments. VRF IDs 1 and 2 are reserved for the default VRF and the management VRF respectively. The first custom defined IP VRF uses VRF ID 3.
- The 2-byte numbering field for the MAC-VRF uses the VLAN ID + 32767, which results in 32768 for VLAN ID 1 and incrementing.

AUTO RD EXAMPLE

- Example: IP-VRF with BGP Router ID 192.0.2.1 and VRF ID 6 - RD 192.0.2.1:6.
- Example: MAC-VRF with BGP Router ID 192.0.2.1 and VLAN 20 - RD 192.0.2.1:32787

NX-OS VXLAN EVPN Auto RT

RD FORMAT

- The auto derived Route-Target is constructed with the Autonomous System Number (ASN) as the 2-byte administrative field and the Service Identifier (VNI) for the 4-byte numbering field.

AUTO RD EXAMPLE

- Example: IP-VRF within ASN 65001 and L3VNI 50001 - Route-Target 65001:50001
- Example: MAC-VRF within ASN 65001 and L2VNI 30001 - Route-Target 65001:30001

Configuring a VXLAN BGP EVPN Fabric

VXLAN BGP EVPN Configuration

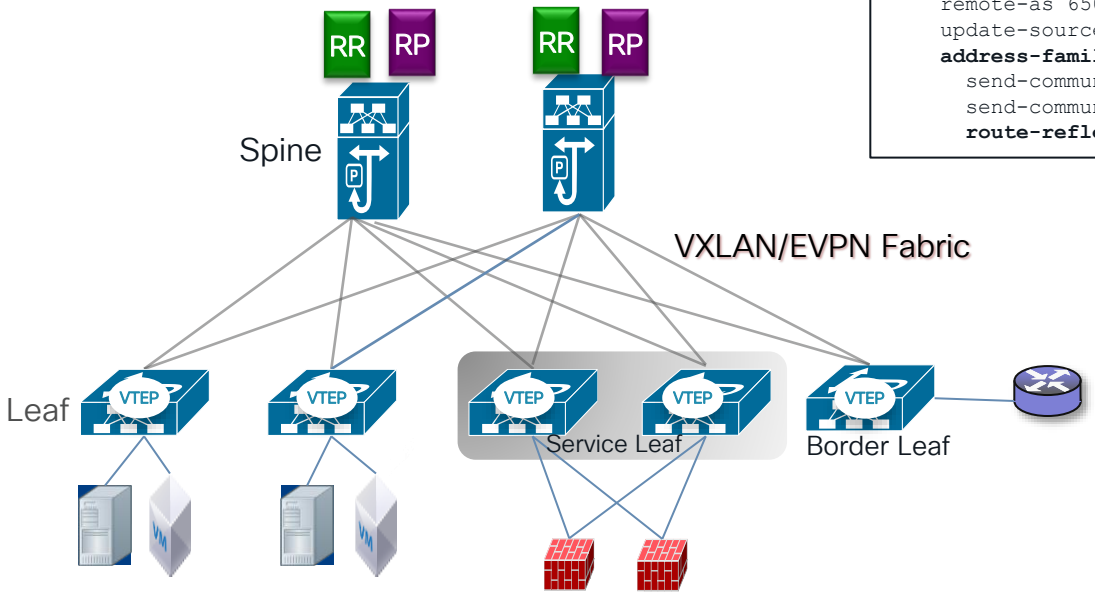
BGP on Leaf and Spine

RR EVPN Route Reflector

RP Rendezvous Point (Underlay)

```
! 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
```

```
! 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
```



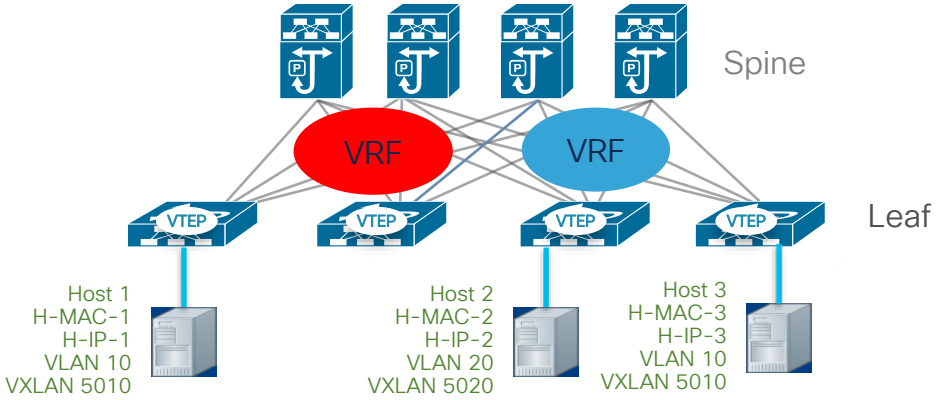
VXLAN BGP EVPN Configuration

VTEP Virtual Networks and Overlay Interface

```
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

IP VRF

MAC VRF

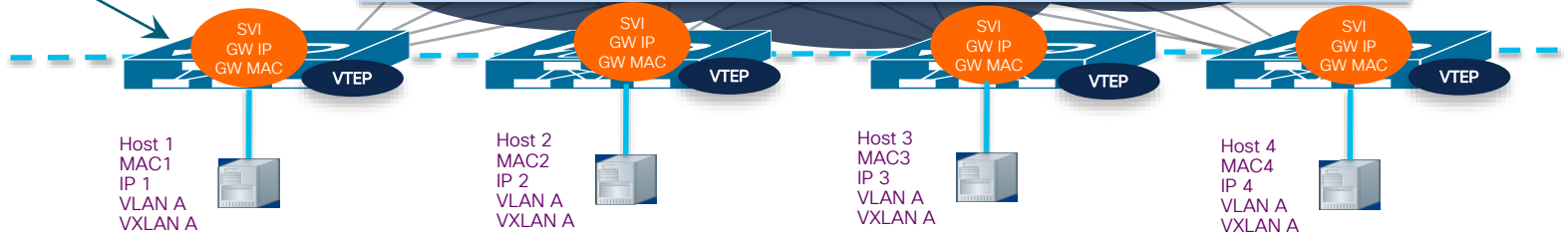
```
vrf context VRF-RED
  vni 9999
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn
evpn
  vni 5010 l2
    rd auto
    route-target both auto
```

```
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

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

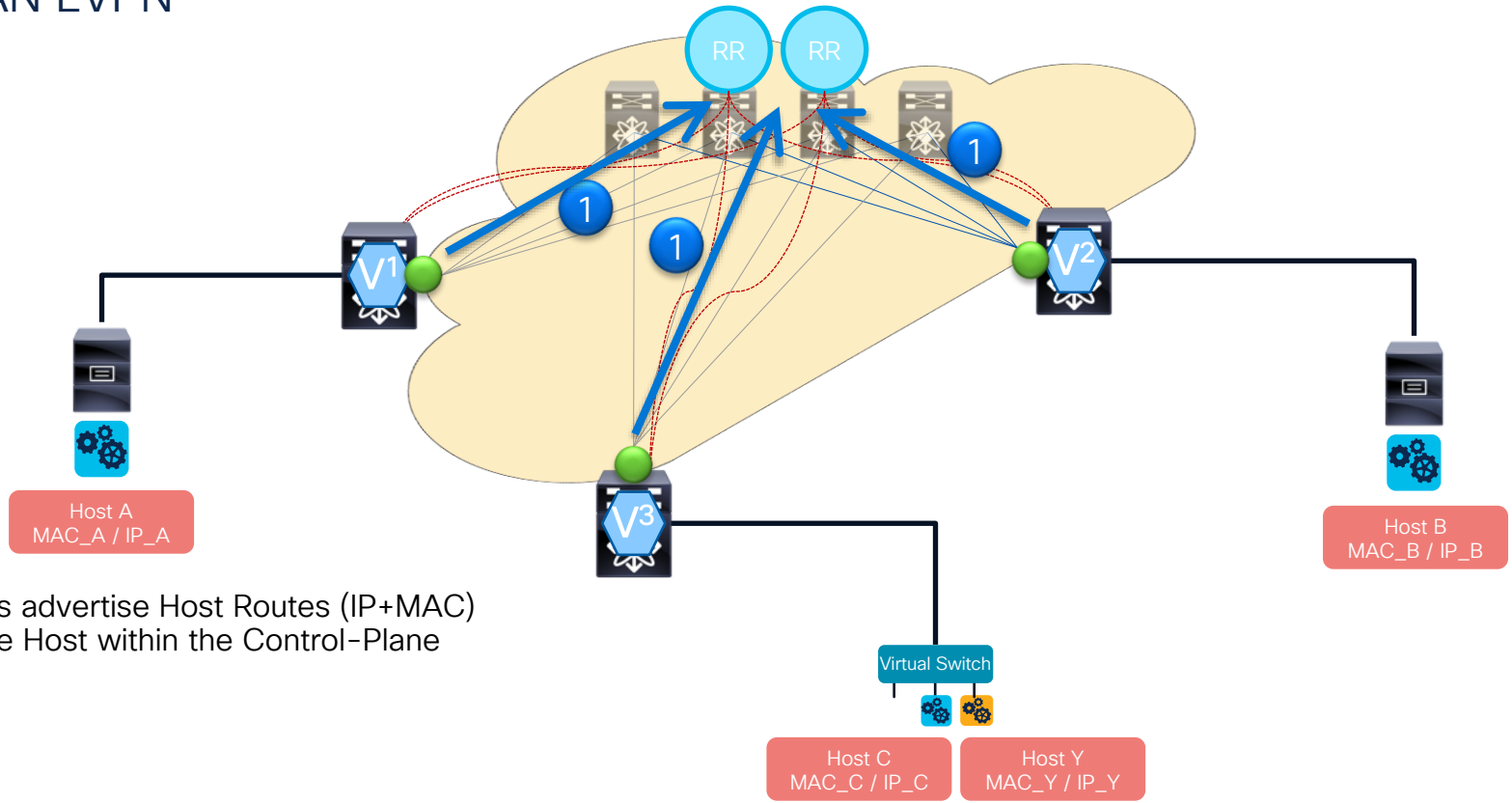
```
# VLAN to VNI mapping
vlan 20
 vn-segment 5020

# 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
```



VXLAN BGP EVPN Packet Walk

Protocol Learning & Distribution (1)

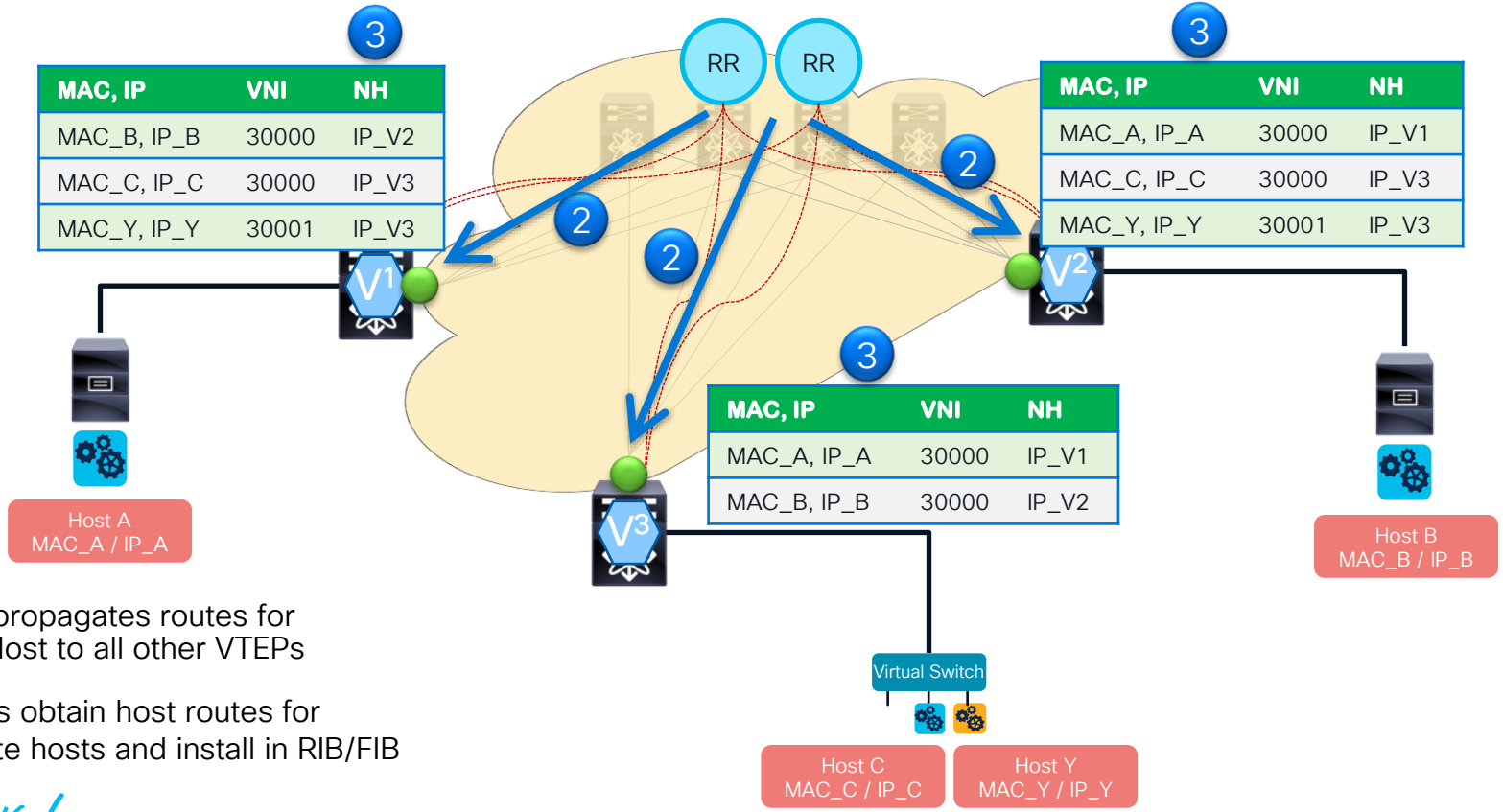
VXLAN EVPN



1 VTEPs advertise Host Routes (IP+MAC) for the Host within the Control-Plane

Protocol Learning & Distribution (2)

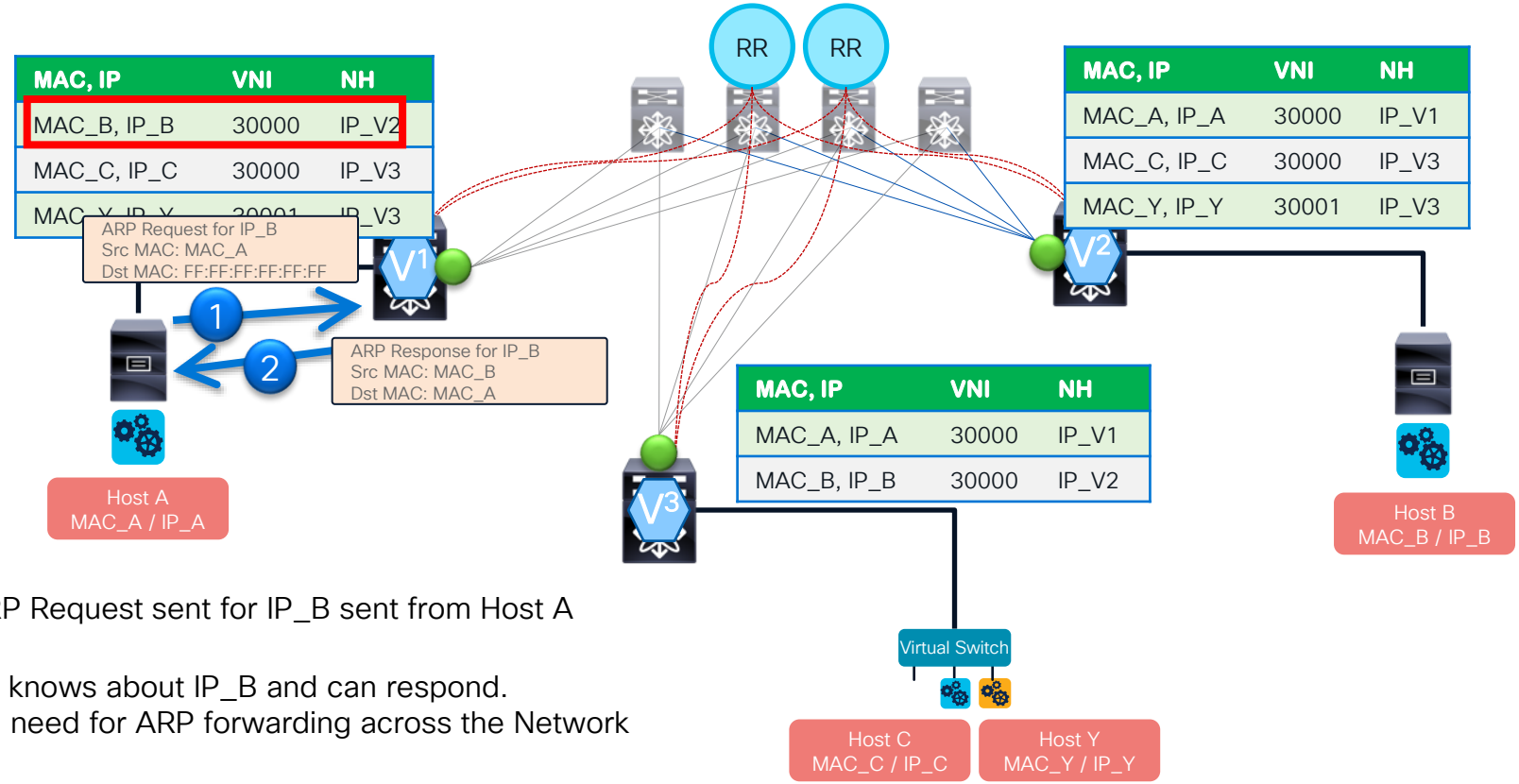
VXLAN EVPN



2 BGP propagates routes for The Host to all other VTEPs

3 VTEPs obtain host routes for remote hosts and install in RIB/FIB

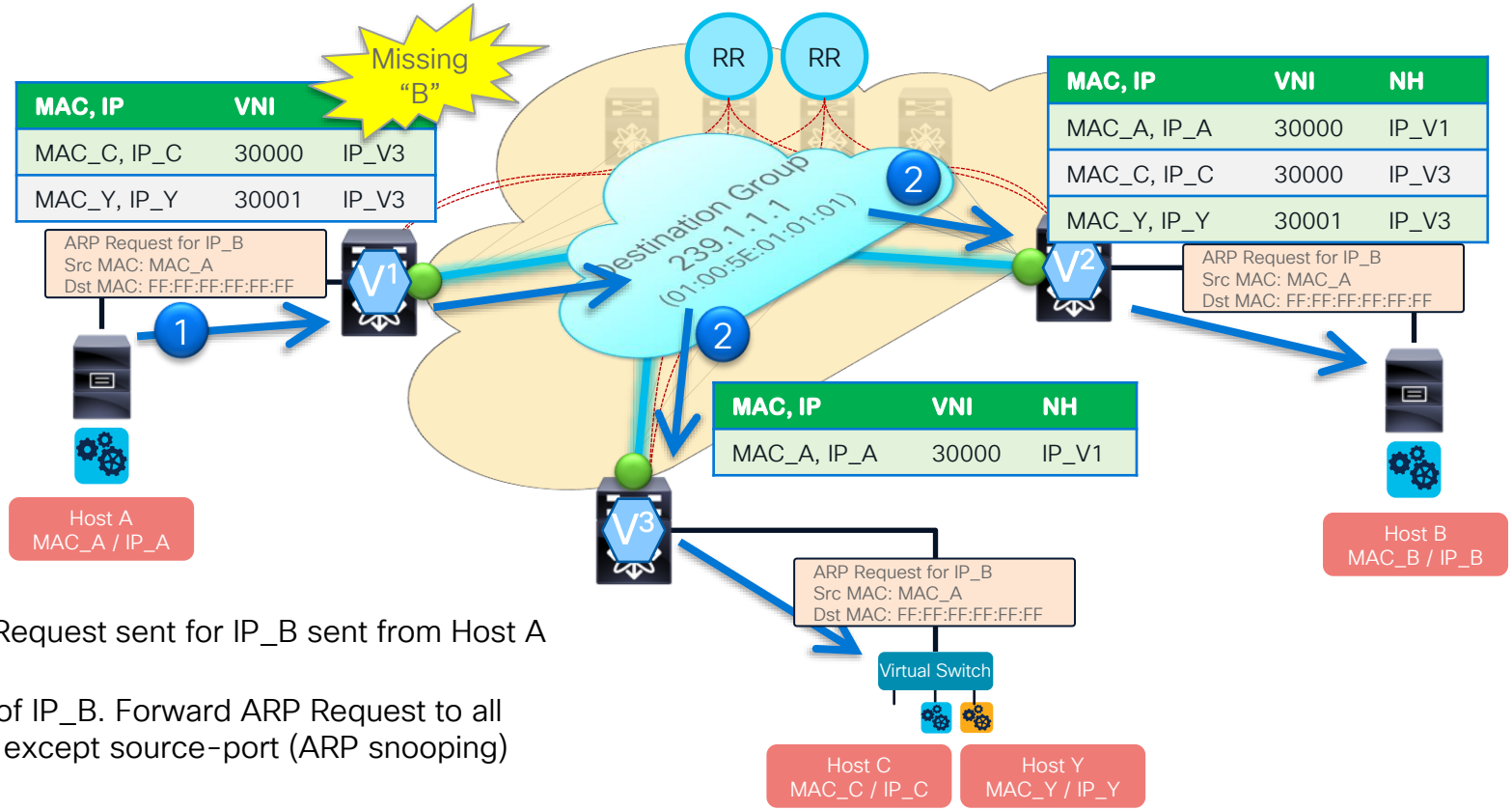
ARP Suppression



1. ARP Request sent for IP_B sent from Host A

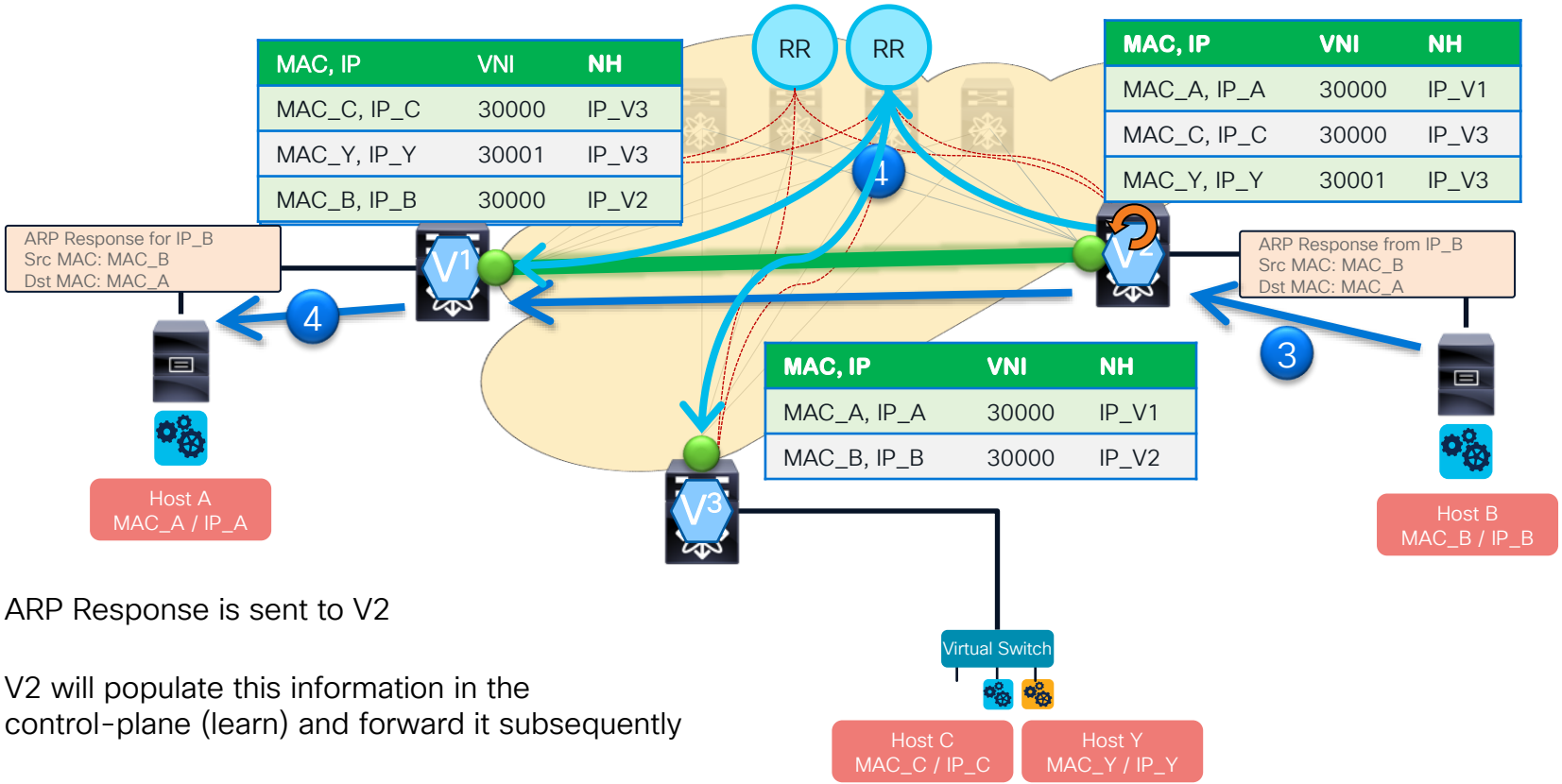
2. V1 knows about IP_B and can respond.
No need for ARP forwarding across the Network

ARP Handling on Silent Host (1)



1. ARP Request sent for IP_B sent from Host A
2. Miss of IP_B. Forward ARP Request to all Ports except source-port (ARP snooping)

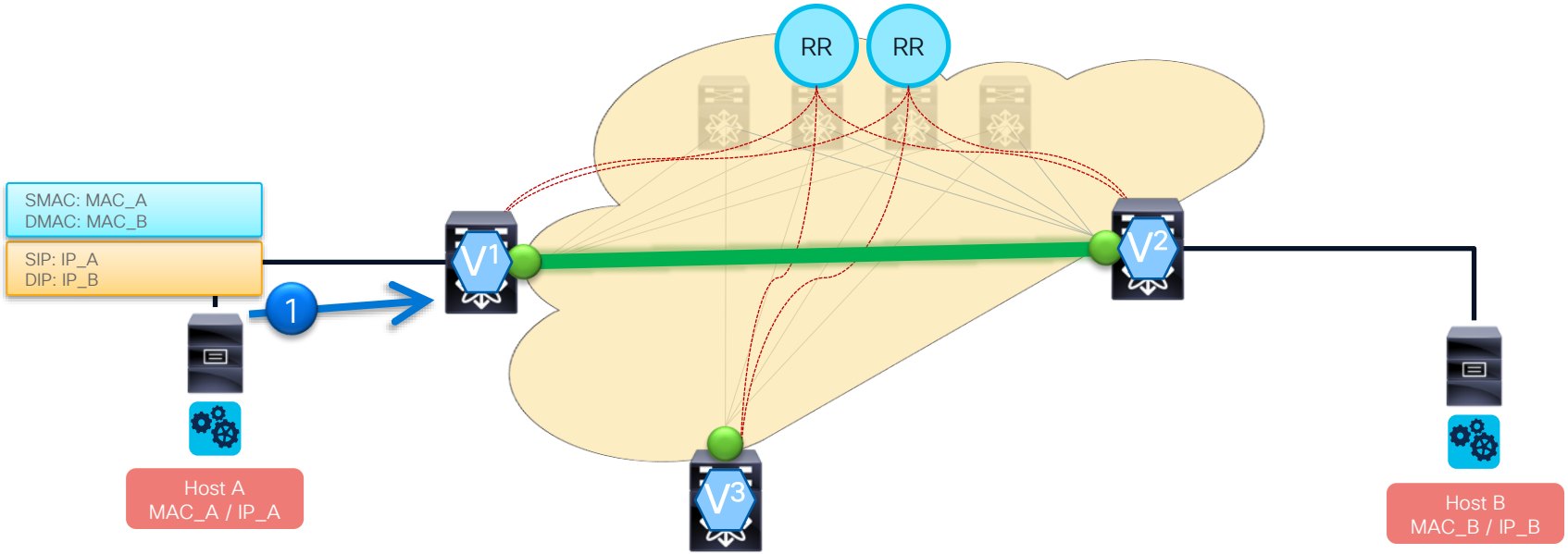
ARP Handling on Silent Host (2)



3 ARP Response is sent to V2

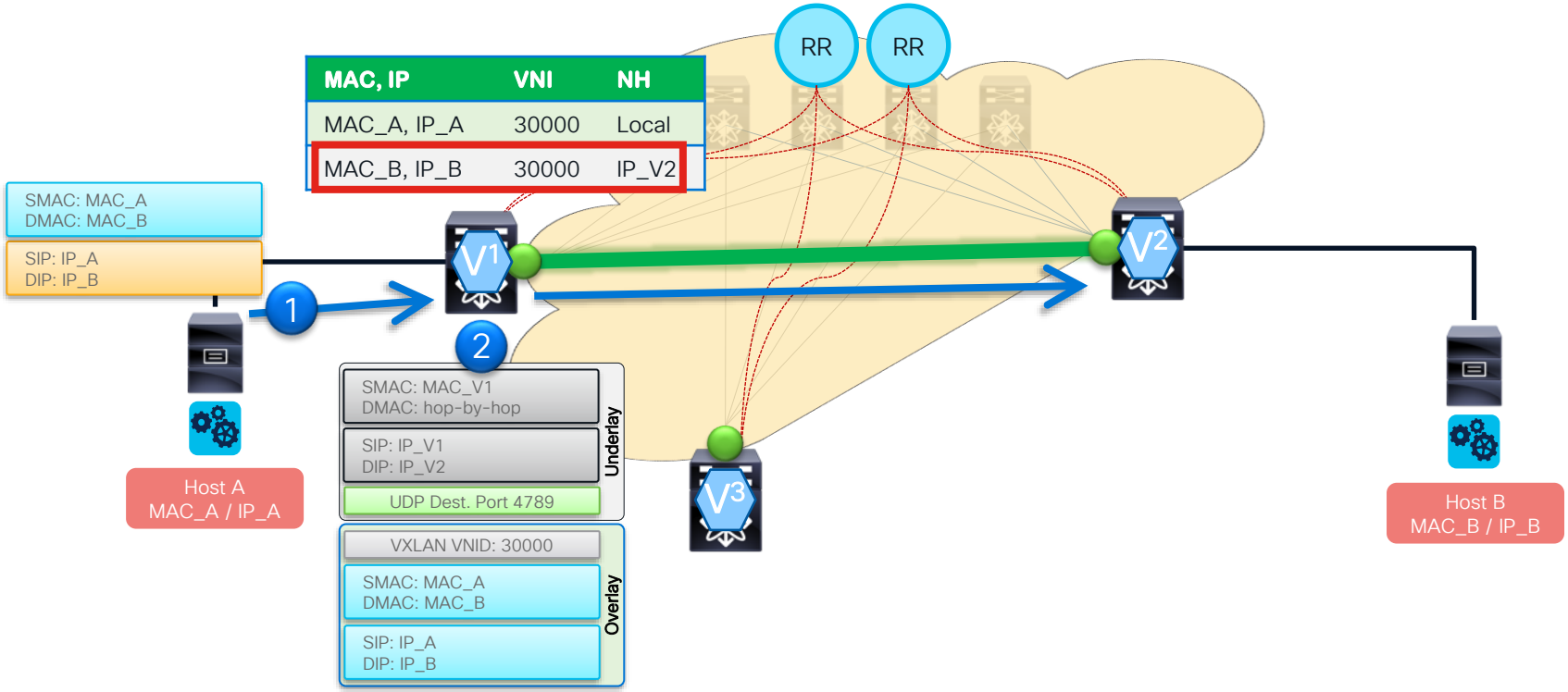
4 V2 will populate this information in the control-plane (learn) and forward it subsequently

Packet Forwarding Bridging



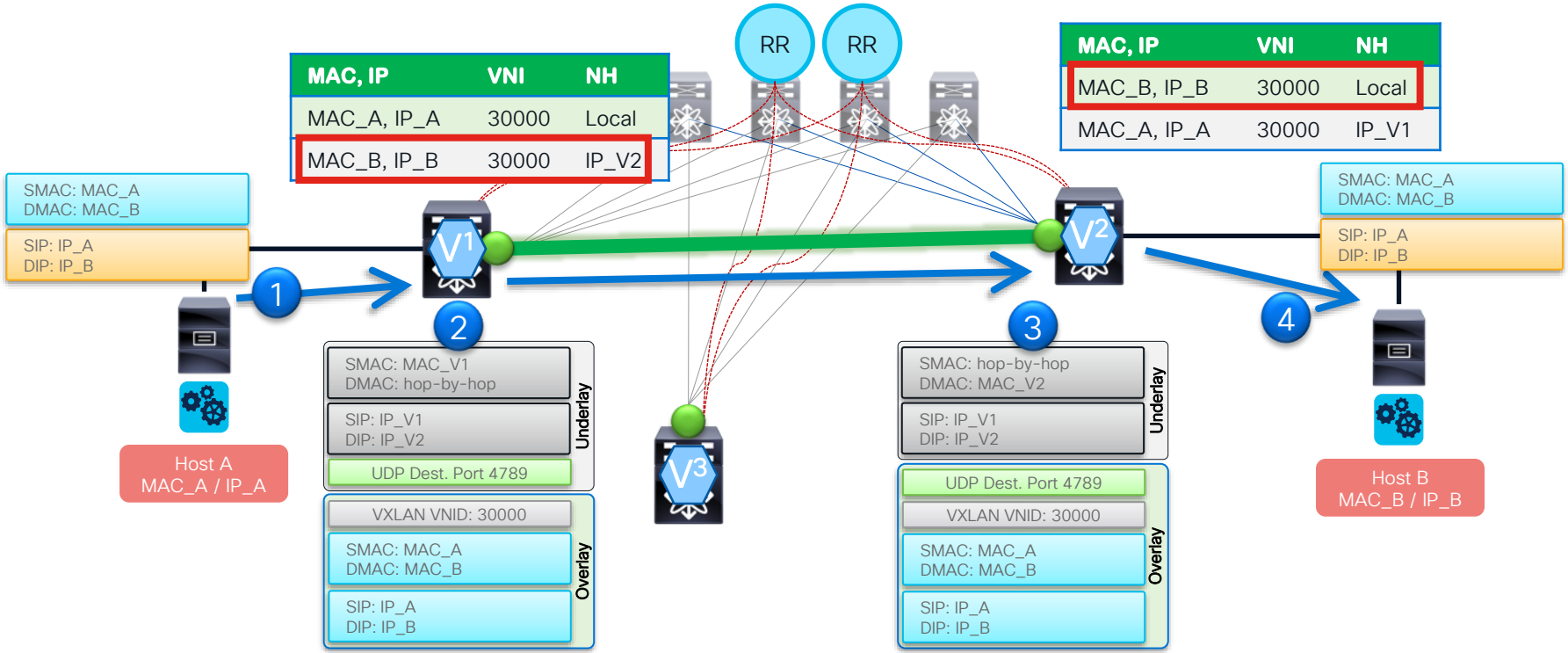
1 Host A needs to send packets to Host B sends it to VTEP V1

Packet Forwarding Bridging



2 VTEP V1 knows IP_B it encapsulates the frame with VNID 30000 and send it to V2

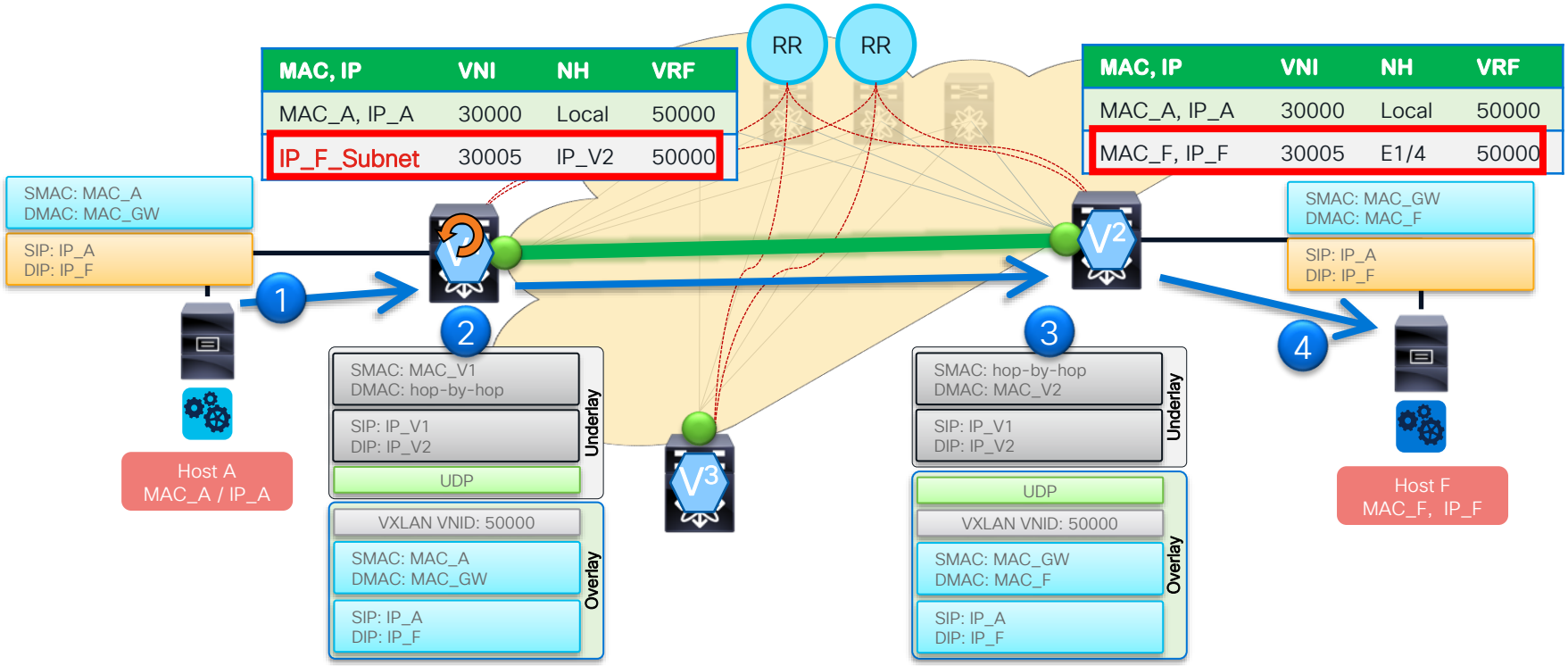
Packet Forwarding Bridging



3 VTEP V2 receives the frame looks up IP_B and MAC_B

4 VTEP V2 then forwards the frame to the Host B

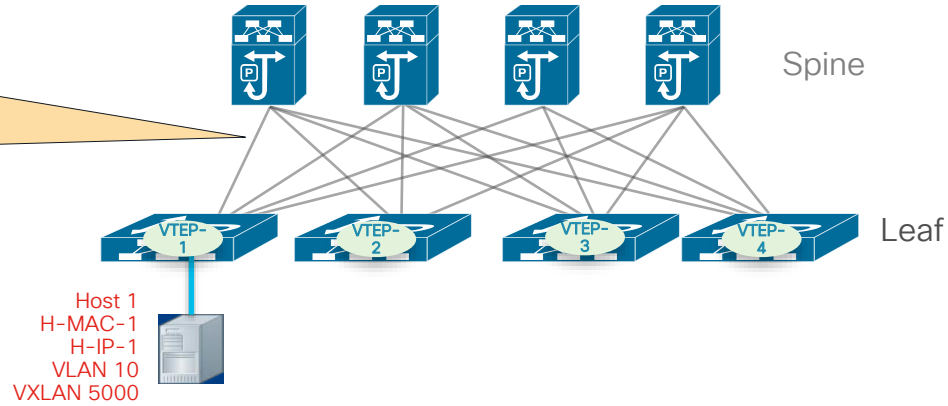
Packet Forwarding Routing



BGP 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



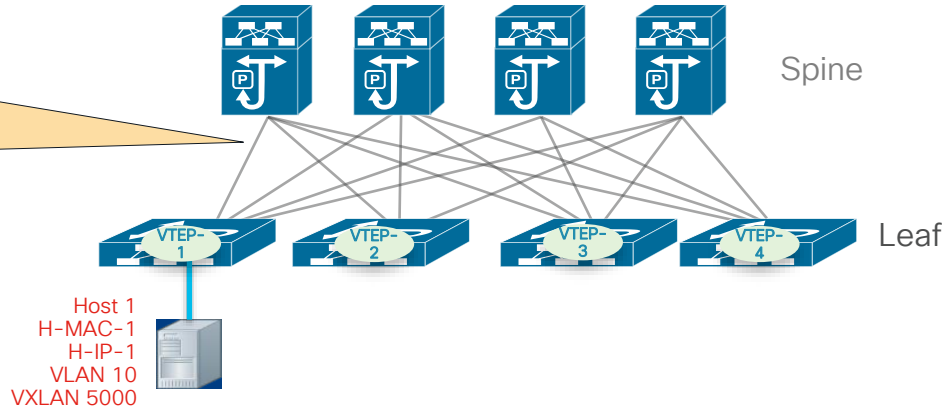
1. Host 1 attaches to VTEP-1
2. VTEP-1 detects Host 1 and advertises Host 1 with seq #0
3. Other VTEPs learn about Host 1 thru BGP updates

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

BGP 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



1. Host 1 moves to VTEP-3 from VTEP-1
2. VTEP-3 detects Host , sends MP-BGP updates for Host 1 with its own VTEP address and a new Seq #1
3. Other VTEPs learn about the new route to Host 1 thru BGP updates

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

Summary

Key Takeaways

- VXLAN BGP EVPN in the Cisco Nexus 9000 is a coupling of industry standard VXLAN Network Virtualization Overlay and BGP EVPN.
- VXLAN BGP EVPN provides the following benefits
  - Mobility, Networking Segmentation, Full Cross Sectional BW, IRB, ECMP and integration of physical and virtual networking.
- The VXLAN BGP EVPN is a CLOS Fabric with switches with VTEPs at the edge that connect to the services decoupling the core routing from the edge.
- BGP EVPN control plane provides capabilities such as ARP Suppression which reduced ARP broadcast across fabric and VM Mobility.

Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!
- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.
- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.

Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

Learn

Cisco U.
IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning
Subscription-based product, technology, and certification training

Cisco Modeling Labs
Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network
Resource community portal for certifications and learning

Train

Cisco Training Bootcamps
Intensive team & individual automation and technology training programs

Cisco Learning Partner Program
Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training
Accelerated curriculum of product, technology, and certification courses

Certify

Cisco Certifications and Specialist Certifications
Award-winning certification program empowers students and IT Professionals to advance their technical careers

Cisco Guided Study Groups
180-day certification prep program with learning and support

Cisco Continuing Education Program
Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

cisco Live!

Thank you


cisco Live!
ALL IN