Configure VXLAN

Contents

Introduction

Prerequisites

- Requirements
- Components Used

Background Information

- Terminology
- What is VXLAN
- Why VXLANBGP EVPN
- BUM Replication Options

Configure

- Network Diagram
- Addressing
- Configurations
  - Spine-1
  - Spine-2
  - Leaf-1(vPC VTEP)
  - Leaf-2 (vPC VTEP)
  - Leaf-3 (Standalone VTEP)
  - Optional: Ingress Replication Instead of Multicast

Verify

- Underlay
- Overlay Control Plane
- NVE and VNIs
- MAC/IP Learning
- L3 VNI Routing
- vPC
- Packet Format

Troubleshoot

Related Information

Introduction

This document describes how to configure and verify VXLAN with BGP EVPN control plane on Cisco Nexus 9000 Series Switches in a spine-leaf fabric.

Prerequisites

Requirements

Cisco recommends that you have knowledge of these topics:

- Virtual Port Channel (vPC) concepts
- IP routing (OSPF or IS-IS) and BGP, including route reflectors
- Multicast routing concepts such as Protocol Independent Multicast (PIM) Sparse Mode and Rendezvous Point (RP)

Components Used

The information in this document is based on these software and hardware versions:

- Network Advantage license
- Cisco NX-OS Release 10.4(x)
- 2 x Cisco Nexus N9K-C9332D-GX2B as spines (BGP route reflectors and PIM anycast RPs)
- 3 x Cisco Nexus N9K-C93180YC-FX3 as leaves (VTEPs); Leaf-1 and Leaf-2 are a vPC pair

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, ensure that you understand the potential impact of any command.



Note: The configuration in this document applies to Nexus 9000 Cloud Scale platforms (-EX, -FX, -FX2, -FX3, -GX, -GX2 and later) in standalone NX-OS mode. It does not apply to Nexus 9000 switches in ACI mode.

Background Information

Terminology

Term	Description
VXLAN	**Virtual Extensible LAN**: which provides Layer 2 and Layer 3 overlay services over a routed IP underlay with MAC-in-UDP encapsulation (UDP port 4789).
VNI	**VXLAN Network Identifier**: A 24-bit segment

	ID. An L2 VNI maps to a VLAN (bridge domain). An L3 VNI maps to a VRF.
VTEP	**VXLAN Tunnel Endpoint**: The device that encapsulates and decapsulates VXLAN traffic.
NVE	**Network Virtualization Edge Interface** (interface nve1): The logical interface where encapsulation and decapsulation occur.
BGP EVPN	The control plane that advertises MAC, MAC/IP (Route Type 2), IP prefix (Route Type 5) and VTEP membership (Route Type 3) between VTEPs.
AnyCast Gateway	The same SVI IP and MAC address on every leaf, and the default gateway is always the local leaf.
BUM	Broadcast, Unknown unicast, and multicast traffic.

What is VXLAN

- VXLAN overlays Layer 2 and Layer 3 segments on top of any routed IP underlay.
- It uses MAC-in-UDP encapsulation, which adds 50 bytes of overhead to the original frame.
- It supports up to 16 million segments, compared to 4094 with traditional VLANs.
- The underlay uses Equal-Cost Multipath (ECMP) for all links, with no Spanning Tree Protocol between leaves and spines.

Why VXLAN BGP EVPN

The previous version of this document used *flood-and-learn*, where VTEPs learn remote MAC addresses from data-plane flooding. BGP EVPN replaces data-plane learning with a control plane and is the recommended deployment model:

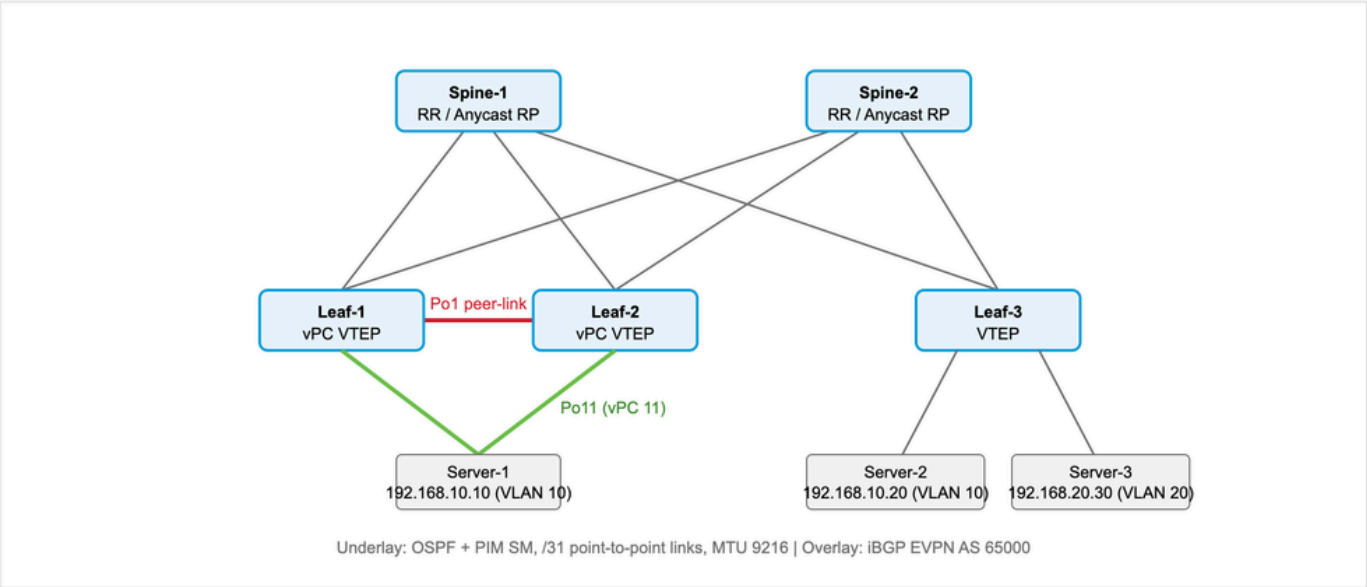
- VTEPs discover each other and learn remote MAC and IP addresses through BGP, which reduces flooding.
- ARP suppression allows local VTEP answer ARP requests for known hosts.
- Distributed anycast gateway and L3 VNI per VRF provides optimal routing between subnets at the first-hop leaf.
- Multitenancy, multihoming (vPC or EVPN ESI), and multisite extension are supported.

BUM Replication Options

Mode	Configuration	Underlay Requirement
Multicast (used in this document)	**mcast-group <group>** under the VNI	PIM Sparse Mode on all underlay links and loopbacks, and an RP (anycast RP on the spines is common)
Ingress replication	**ingress-replication protocol bgp** under the VNI	Unicast only. The ingress VTEP sends a copy to every remote VTEP that BGP Route Type 3 has learned for the VNI.

Configure

Network Diagram



Addressing

Device	Loopback0 (routing, BGP, PIM)	Loopback1 (NVE Source)	Uplinks (Eth1/49 to Spine-1, Eth1/50 to Spine-2)
Spine-1	10.0.0.1	-	Eth1/1-3 to Leaf-1/2/3: 10.1.1.0, .2, .4 /31

Spine-2	10.0.0.2	-	Eth1/1-3 to Leaf-1/2/3: 10.1.2.0, .2, .4 /31
Leaf-1	10.0.0.11	10.0.1.11, secondary 10.0.1.112	10.1.1.1/31, 10.1.2.1/31
Leaf-2	10.0.0.12	10.0.1.12, secondary 10.0.1.112	10.1.1.3/31, 10.1.2.3/31
Leaf-3	10.0.0.13	10.0.1.13	10.1.1.5/31, 10.1.2.5/31

Overlay Object	**Value**
Anycast RP	10.254.254.1 (Loopback254 on both spines), groups 239.1.1.0/24
VLAN 10 / L2 VNI	10010, mcast-group 239.1.1.10, gateway 192.168.10.1/24
VLAN 20 / L2 VNI	10020, mcast-group 239.1.1.20, gateway 192.168.20.1/24
VRF Tenant-A / L3 VNI	50000 (VLAN 500)
Anycast gateway MAC	2020.0000.00aa
vPC backup underlay SVI	VLAN 3900, 10.255.255.0/31 (Leaf-1) and 10.255.255.1/31 (Leaf-2)



Note: Use separate loopbacks for the NVE source (Loopback1), routing, and BGP peering (Loopback0). On vPC VTEPs, the NVE source loopback has a secondary IP that is the same on both peers. Remote VTEPs use this shared address (the vPC anycast VTEP) to reach the vPC pair as a single VTEP.

Configurations

These configurations show only the parts relevant to the underlay and VXLAN. Configure management, AAA, and other features based on your requirements.

Spine-1

```
nv overlay evpn
feature ospf
feature bgp
feature pim
feature nv overlay

ip pim rp-address 10.254.254.1 group-list 239.1.1.0/24
ip pim anycast-rp 10.254.254.1 10.0.0.1
ip pim anycast-rp 10.254.254.1 10.0.0.2

interface loopback0
  description Routing / BGP / Anycast-RP member
  ip address 10.0.0.1/32
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

interface loopback254
  description Anycast-RP
  ip address 10.254.254.1/32
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

interface Ethernet1/1
  description to Leaf-1 Eth1/49
  no switchport
  mtu 9216
  medium p2p
  ip address 10.1.1.0/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown

interface Ethernet1/2
  description to Leaf-2 Eth1/49
  no switchport
  mtu 9216
  medium p2p
  ip address 10.1.1.2/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown

interface Ethernet1/3
  description to Leaf-3 Eth1/49
  no switchport
  mtu 9216
  medium p2p
  ip address 10.1.1.4/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown

router ospf UNDERLAY
  router-id 10.0.0.1
```

```
router bgp 65000
  router-id 10.0.0.1
  address-family l2vpn evpn
  template peer VTEP
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community
      send-community extended
      route-reflector-client
  neighbor 10.0.0.11
    inherit peer VTEP
  neighbor 10.0.0.12
    inherit peer VTEP
  neighbor 10.0.0.13
    inherit peer VTEP
```

Spine-2

```
interface loopback0
  ip address 10.0.0.2/32
interface Ethernet1/1
  ip address 10.1.2.0/31
interface Ethernet1/2
  ip address 10.1.2.2/31
interface Ethernet1/3
  ip address 10.1.2.4/31
router ospf UNDERLAY
  router-id 10.0.0.2
router bgp 65000
  router-id 10.0.0.2
```

Leaf-1 (vPC VTEP)

```
nv overlay evpn
feature ospf
feature bgp
feature pim
feature fabric forwarding
feature interface-vlan
feature vn-segment-vlan-based
feature lacp
feature vpc
feature nv overlay

fabric forwarding anycast-gateway-mac 2020.0000.00aa
ip pim rp-address 10.254.254.1 group-list 239.1.1.0/24

vlan 10
  vn-segment 10010
```

```
vlan 20
  vn-segment 10020
vlan 500
  name L3VNI-Tenant-A
  vn-segment 50000
vlan 3900
  name vPC-backup-underlay

system nve infra-vlans 3900

vrf context Tenant-A
  vni 50000
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn

vpc domain 10
  peer-switch
  role priority 10
  peer-keepalive destination 172.16.0.2 source 172.16.0.1 vrf management
  delay restore 150
  peer-gateway
  ip arp synchronize
  ipv6 nd synchronize

interface port-channel1
  description vPC peer-link
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 10,20,500,3900
  spanning-tree port type network
  vpc peer-link

interface port-channel11
  description Server-1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 10,20
  spanning-tree port type edge trunk
  mtu 9216
  vpc 11

interface Vlan10
  no shutdown
  vrf member Tenant-A
  no ip redirects
  ip address 192.168.10.1/24
  fabric forwarding mode anycast-gateway

interface Vlan20
  no shutdown
  vrf member Tenant-A
  no ip redirects
  ip address 192.168.20.1/24
  fabric forwarding mode anycast-gateway

interface Vlan500
  description L3VNI Tenant-A
  no shutdown
  vrf member Tenant-A
  no ip redirects
```

```
  ip forward

interface Vlan3900
  description vPC backup underlay
  no shutdown
  mtu 9216
  no ip redirects
  ip address 10.255.255.0/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

interface nve1
  no shutdown
  host-reachability protocol bgp
  advertise virtual-rmac
  source-interface loopback1
  member vni 10010
    suppress-arp
    mcast-group 239.1.1.10
  member vni 10020
    suppress-arp
    mcast-group 239.1.1.20
  member vni 50000 associate-vrf

interface Ethernet1/1
  description Server-1
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 10,20
  mtu 9216
  channel-group 11 mode active
  no shutdown

interface Ethernet1/53-54
  description vPC peer-link to Leaf-2
  switchport
  switchport mode trunk
  switchport trunk allowed vlan 10,20,500,3900
  channel-group 1 mode active
  no shutdown

interface Ethernet1/49
  description to Spine-1 Eth1/1
  no switchport
  mtu 9216
  medium p2p
  ip address 10.1.1.1/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown

interface Ethernet1/50
  description to Spine-2 Eth1/1
  no switchport
  mtu 9216
  medium p2p
  ip address 10.1.2.1/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
```

```
  no shutdown

interface loopback0
  description Routing / BGP
  ip address 10.0.0.11/32
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

interface loopback1
  description NVE source
  ip address 10.0.1.11/32
  ip address 10.0.1.112/32 secondary
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

router ospf UNDERLAY
  router-id 10.0.0.11

router bgp 65000
  router-id 10.0.0.11
  address-family l2vpn evpn
    advertise-pip
  template peer SPINE-RR
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community
      send-community extended
  neighbor 10.0.0.1
    inherit peer SPINE-RR
  neighbor 10.0.0.2
    inherit peer SPINE-RR
  vrf Tenant-A
    address-family ipv4 unicast
      advertise l2vpn evpn

evpn
  vni 10010 l2
    rd auto
    route-target import auto
    route-target export auto
  vni 10020 l2
    rd auto
    route-target import auto
    route-target export auto
```



Note: On Cloud Scale vPC VTEPs, the backup underlay SVI (VLAN 3900) must be listed in **system nve infra-vlans**. This VLAN must be in the default VRF, must not have a vn-segment, and must be allowed on the peer-link. It carries VXLAN traffic over the peer-link for orphan ports and when one vPC peer loses all its uplinks. Do not configure infra VLANs whose IDs are 512 apart, for example 10 and 522.

advertise-pip and **advertise virtual-rmac** form the vPC peers, which advertise type-5 (prefix) routes and orphan-host routes with their primary IP (PIP) and their own router MAC. Without these, traffic towards a host behind only one peer can be sent to the peer that cannot reach it.

ARP suppression (**suppress-arp**) is supported only for VNIs where the VTEP is the distributed anycast gateway for that VNI.

Leaf-2 (vPC VTEP)

```
vpc domain 10
  role priority 20
  peer-keepalive destination 172.16.0.1 source 172.16.0.2 vrf management

interface Vlan3900
  ip address 10.255.255.1/31

interface Ethernet1/49
  ip address 10.1.1.3/31
interface Ethernet1/50
  ip address 10.1.2.3/31

interface loopback0
  ip address 10.0.0.12/32
interface loopback1
  ip address 10.0.1.12/32
  ip address 10.0.1.112/32 secondary

router ospf UNDERLAY
  router-id 10.0.0.12
router bgp 65000
  router-id 10.0.0.12
```

⚠ **Caution**: The NVE configuration (VNIs, mcast-groups, source interface, secondary IP) must be identical on both vPC peers. A mismatch is a vPC Type-1 consistency failure and the NVE interface or VNIs are suspended. Check by running the **show vpc consistency-parameters global** command.

Leaf-3 (Standalone VTEP)

```
nv overlay evpn
feature ospf
feature bgp
feature pim
feature fabric forwarding
feature interface-vlan
feature vn-segment-vlan-based
feature nv overlay

fabric forwarding anycast-gateway-mac 2020.0000.00aa
ip pim rp-address 10.254.254.1 group-list 239.1.1.0/24

vlan 10
```

```
  vn-segment 10010
vlan 20
  vn-segment 10020
vlan 500
  name L3VNI-Tenant-A
  vn-segment 50000

vrf context Tenant-A
  vni 50000
  rd auto
  address-family ipv4 unicast
    route-target both auto
    route-target both auto evpn

interface Vlan10
  no shutdown
  vrf member Tenant-A
  no ip redirects
  ip address 192.168.10.1/24
  fabric forwarding mode anycast-gateway

interface Vlan20
  no shutdown
  vrf member Tenant-A
  no ip redirects
  ip address 192.168.20.1/24
  fabric forwarding mode anycast-gateway

interface Vlan500
  no shutdown
  vrf member Tenant-A
  no ip redirects
  ip forward

interface nve1
  no shutdown
  host-reachability protocol bgp
  source-interface loopback1
  member vni 10010
    suppress-arp
    mcast-group 239.1.1.10
  member vni 10020
    suppress-arp
    mcast-group 239.1.1.20
  member vni 50000 associate-vrf

interface Ethernet1/1
  description Server-2
  switchport
  switchport access vlan 10
  spanning-tree port type edge
  no shutdown

interface Ethernet1/2
  description Server-3
  switchport
  switchport access vlan 20
  spanning-tree port type edge
  no shutdown

interface Ethernet1/49
  description to Spine-1 Eth1/3
```

```
  no switchport
  mtu 9216
  medium p2p
  ip address 10.1.1.5/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown

interface Ethernet1/50
  description to Spine-2 Eth1/3
  no switchport
  mtu 9216
  medium p2p
  ip address 10.1.2.5/31
  ip ospf network point-to-point
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode
  no shutdown

interface loopback0
  ip address 10.0.0.13/32
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

interface loopback1
  description NVE source
  ip address 10.0.1.13/32
  ip router ospf UNDERLAY area 0.0.0.0
  ip pim sparse-mode

router ospf UNDERLAY
  router-id 10.0.0.13

router bgp 65000
  router-id 10.0.0.13
  template peer SPINE-RR
    remote-as 65000
    update-source loopback0
    address-family l2vpn evpn
      send-community
      send-community extended
  neighbor 10.0.0.1
    inherit peer SPINE-RR
  neighbor 10.0.0.2
    inherit peer SPINE-RR
  vrf Tenant-A
    address-family ipv4 unicast
      advertise l2vpn evpn

evpn
  vni 10010 l2
    rd auto
    route-target import auto
    route-target export auto
  vni 10020 l2
    rd auto
    route-target import auto
    route-target export auto
```

Optional: Ingress Replication Instead of Multicast

If the underlay does not run multicast, replace **mcast-group** with BGP-based ingress replication on every VTEP. The PIM and the RP configuration are not required for the L2 VNIs.

```
interface nve1
  member vni 10010
    suppress-arp
    ingress-replication protocol bgp
  member vni 10020
    suppress-arp
    ingress-replication protocol bgp
```

Verify

Use this section to confirm that your configuration works properly.

The Cisco CLI Analyzer (registered customers only) supports certain **show** commands. Use the Cisco CLI Analyzer to view an analysis of **show** command outputs.

Area	**Commands**
Underlay	show ip ospf neighbors, show ip route 10.0.1.13, show ip pim neighbor, show ip pim rp
Overlay Control Plane	show bgp l2vpn evpn summary, show bgp l2vpn evpn
NVE	show nve interface nve1 detail, show nve peers, show nve vni
MAC/IP Learning	show mac address-table vlan 10, show l2route evpn mac all, show l2route evpn mac-ip all, show ip arp suppression-cache detail
L3 VNI	show ip route vrf Tenant-A, show bgp l2vpn evpn route-type 2, show bgp l2vpn evpn route-type 5
Multicast Replication	show ip mroute 239.1.1.10

vPC	show vpc, show vpc consistency-parameters global, show system nve infra-vlans

These outputs are from a steady state after traffic was sent between Server-1, Server-2 and Server-3.

Underlay

```
Leaf-1# show ip pim rp
PIM RP Status Information for VRF "default"
BSR disabled
Auto-RP disabled
BSR RP Candidate policy: None
BSR RP policy: None
Auto-RP Announce policy: None
Auto-RP Discovery policy: None

Anycast-RP 10.254.254.1 members:
  10.0.0.1  10.0.0.2

RP: 10.254.254.1, (0),
 uptime: 1d01h   priority: 255,
 RP-source: (local),
 group ranges:
 239.1.1.0/24
```

Overlay Control Plane

```
Leaf-1# show bgp l2vpn evpn summary
BGP summary information for VRF default, address family L2VPN EVPN
BGP router identifier 10.0.0.11, local AS number 65000
BGP table version is 58, L2VPN EVPN config peers 2, capable peers 2

Neighbor        V    AS    MsgRcvd    MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
10.0.0.1        4 65000       1520       1488       58    0    0 1d01h    9
10.0.0.2        4 65000       1519       1490       58    0    0 1d01h    9
```

NVE and VNIs

```
Leaf-1# show nve interface nve1 detail
Interface: nve1, State: Up, encapsulation: VXLAN
 VPC Capability: VPC-VIP-Only [notified]
 Local Router MAC: 5254.0011.1b08
 Host Learning Mode: Control-Plane
 Source-Interface: loopback1 (primary: 10.0.1.11, secondary: 10.0.1.112)
```

```
 Source Interface State: Up
 Virtual RMAC Advertisement: Yes
 NVE Flags:
 Interface Handle: 0x49000001
 Source Interface hold-down-time: 180
 Source Interface hold-up-time: 30
 Remaining hold-down time: 0 seconds
 Virtual Router MAC: 0200.0a00.0170

Leaf-1# show nve peers
Interface Peer-IP                                 State LearnType Uptime   Router-Mac
--------- --------------------------------------  ----- --------- -------- -----------------
nve1      10.0.1.13                               Up    CP        1d01h    5254.0013.1b08

Leaf-1# show nve vni
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       S-ND - Suppress ND
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication
       HYB - Hybrid IRB mode

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ -----
nve1      10010    239.1.1.10        Up    CP   L2 [10]            SA
nve1      10020    239.1.1.20        Up    CP   L2 [20]            SA
nve1      50000    n/a               Up    CP   L3 [Tenant-A]
```



Note: With BGP EVPN, NVE peers appear when a remote VTEP advertises routes for a VNI that is configured locally (LearnType CP). Leaf-3 sees the vPC pair as one peer, the AnyCast VTEP 10.0.1.112. With advertise-pip, Leaf-3 also sees the vPC primary IPs 10.0.1.11 and 10.0.1.12 for type-5 and orphan routes.

MAC/IP Learning

```
Leaf-1# show l2route evpn mac all
Flags -(Rmac):Router MAC (Stt):Static (L):Local (R):Remote (V):vPC link
(Dup):Duplicate (Spl):Split (Rcv):Recv (AD):Auto-Delete (D):Del Pending
(S):Stale (C):Clear, (Ps):Peer Sync (O):Re-Originated (Nho):NH-Override
(Pf):Permanently-Frozen, (Orp): Orphan

Topology    Mac Address    Prod   Flags         Seq No     Next-Hops
----------- -------------- ------ ------------- ---------- ----------------
10          0050.5601.0a0a Local  L,            0          Po11
10          0050.5603.0a14 BGP    SplRcv        0          10.0.1.13 (Label: 10010)
20          0050.5603.1e1e BGP    SplRcv        0          10.0.1.13 (Label: 10020)

Leaf-1# show ip arp suppression-cache detail
Flags: + - Adjacencies synced via CFSoE
       L - Local Adjacency
       R - Remote Adjacency
       L2 - Learnt over L2 interface
       PS - Added via L2RIB, Peer Sync
```

```
       RO - Dervied from L2RIB Peer Sync Entry

Ip Address      Age      Mac Address    Vlan Physical-ifindex    Flags    Remote Vtep Addrs

192.168.10.10   00:02:11 0050.5601.0a0a   10 port-channel11       L
192.168.10.20   00:05:42 0050.5603.0a14   10 (null)               R        10.0.1.13
```

L3 VNI Routing

```
Leaf-1# show ip route vrf Tenant-A 192.168.20.30
IP Route Table for VRF "Tenant-A"
'*' denotes best ucast next-hop
'**' denotes best mcast next-hop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

192.168.20.30/32, ubest/mbest: 1/0
    *via 10.0.1.13%default, [200/0], 00:10:21, bgp-65000, internal, tag 65000, segid: 50000 tunnelid: 0
```

---

vPC

```
Leaf-1# show vpc consistency-parameters global | include Nve|Src|Sec|Host
Nve1 Adm St, Src Adm St,    1     Up, Up, 10.0.1.112, CP,    Up, Up, 10.0.1.112, CP,
Sec IP, Host Reach, VMAC          TRUE, Disabled,               TRUE, Disabled,
Adv, SA,mcast l2, mcast           0.0.0.0, 0.0.0.0, Disabled,   0.0.0.0, 0.0.0.0, Disabled,
Nve Vni Configuration       1     10010-10020                   10010-10020

Leaf-1# show system nve infra-vlans
Currently active infra-vlans: 3900
```

Packet Format

All VXLAN packets are sourced from the NVE source loopback of the ingress VTEP. For hosts on vPC ports, this is the shared secondary IP (10.0.1.112). Known unicast traffic goes to the NVE IP of the remote VTEP. BUM traffic goes to the multi-cast group of the VNI, or with the ingress replication, and a unicast copy goes to each remote VTEP.

Header	**Example: Server-1 to Server-2 (VLAN 10, L2 VNI)**	**Example: Server-1 to Server-3 (Routed, L3 VNI)**
Outer Ethernet	Leaf-1 uplink MAC to spine MAC	Leaf-1 uplink MAC to spine MAC

Outer IP	10.0.1.112 to 10.0.1.13	10.0.1.112 to 10.0.1.13 (10.0.1.11 to 10.0.1.13 if the traffic is from an orphan port and advertise-pip is used)
Outer UDP	Source port is a hash of the inner flow (entropy for ECMP); destination port 4789	Same
VXLAN	Flags 0x08 (I flag), VNI 10010	Flags 0x08, VNI 50000
Inner Ethernet	Server-1 MAC to Server-2 MAC	Leaf-1 router MAC to Leaf-3 router MAC
Inner IP	192.168.10.10 to 192.168.10.20	192.168.10.10 to 192.168.20.30 (TTL decremented)



Note: VXLAN adds 50 bytes to the original frame. Set the underlay MTU (all spine and leaf uplinks, and the vPC backup SVI) to at least 50 bytes more than the largest host MTU. 9216 is recommended.

Troubleshoot

Symptom	**Check**	**Common Cause**
NVE peer not in show nve peers	show bgp l2vpn evpn summary, show ip route <remote NVE IP>, show nve vni	BGP EVPN session down; remote NVE loopback not advertised in the underlay; VNI not configured on both VTEPs; missing send-community extended
NVE interface down or VNIs suspended on a vPC VTEP	show nve interface nve1 detail, show vpc consistency-parameters global	NVE configuration mismatch between peers; the source loopback is still in the hold-down timer after a reload; secondary IP missing or different
Hosts in the same VNI cannot communicate, or ARP fails	show l2route evpn mac all, show ip arp suppression-cache detail, show ip mroute <group>	RT-2 not received (route-target or VNI mismatch); BUM not replicated (no PIM neighbor, no

		RP, PIM missing on the NVE loopback); suppress-arp on a VNI without an anycast gateway SVI
Routing between VNIs fails	show ip route vrf Tenant-A, show nve vni (L3 VNI state), show bgp l2vpn evpn route-type 5	L3 VNI VLAN or SVI missing or shut; vni missing under the VRF context; advertise l2vpn evpn missing under the BGP VRF; route-target both auto evpn missing
Traffic fails only for orphan ports, or after one vPC peer loses its uplinks	show system nve infra-vlans, show ip ospf neighbors vrf default	Backup underlay SVI missing, not in system nve infra-vlans, or not allowed on the peer-link
Large packets dropped, small packets work	show interface <uplink> (TX errors), ping with DF bit set between NVE loopbacks	Underlay MTU too small for VXLAN overhead



Note: These items are not supported as VXLAN uplinks:

- Layer 3 subinterfaces
- SVIs that are not configured in system nve infra-vlans

The NVE source loopback must be in the default VRF.

Related Information

- Cisco Nexus 9000 Series NX-OS VXLAN Configuration Guide, Release 10.5(x)
- Understand and Configure Nexus 9000 vPC with Best Practices
- Technical Support & Documentation - Cisco Systems