



VXLAN BGP EVPN

[« VXLAN Flood and Learn | VXLAN BGP EVPN Layer 2 Fabric »](#)

Last updated: January 19, 2017

VXLAN BGP EVPN

Note:

Either configure the section [VXLAN Underlay Infrastructure Routing](#) or load its [solution configs](#) before configuring this task.

- Per the above mentioned task, the Underlay routing infrastructure should be setup as follows:
 - 5K Leafs and 7K Spines have Loopback0 interfaces with addresses 10.0.0.X/32, where X is the device number.
 - Physical links between 5Ks and 7Ks are unnumbered to Lo0.
 - Loopbacks and physical links run OSPF and PIM.
 - Lowered number Spine (e.g. 7K1) is the primary Bidir RP, and the higher numbered Spine is the secondary Bidir RP via Phantom RP.
 - BFD is enabled on physical links from Spines to Leafs, and registered to both OSPF and PIM.
 - iBGP is configured from Spines to Leafs, with Spines as iBGP Route Reflectors for IPv4 Unicast and BGP L2VPN EVPN AFIs
- Configure Classical Ethernet connectivity to the Servers as follows:
 - Configure your first 5Ks link to your first Server as an edge port and access port in VLAN 172.
 - Configure your second 5Ks link to your second Server as an edge port and access port in VLAN 172.
 - The servers should use IP addresses 172.16.0.1X/24, where X is the Server number.
- Configure VXLAN BGP EVPN on your 5Ks as follows:
 - VLAN 172 should map to VXLAN Network Identifier (VNI) 1000172.
 - Source the VXLAN Tunnel Endpoint (VTEP) from the 5Ks' Loopback0 interfaces.
 - Use the multicast group 224.1.1.1 for the replication of Broadcast, Unknown, and Multicast (BUM) traffic.
 - Use BGP for the reachability protocol under the NVE interface.
 - BGP Route Distinguishers (RDs) and Route Targets (RTs) should be auto-generated based on the Leaf's BGP AS and the VNI numbers.
- Once complete, verify that the Leaf switches are learning the remote server MAC addresses via BGP L2VPN EVPN, and that the servers have IP connectivity to each other.

Configuration [Click to collapse](#)

Note:

Nexus 5600 is required to run in store-and-forward switching mode in order to support VXLAN encapsulation. Saving the config and reloading the switch is required in order to activate this configuration change.

```
N5K1# config t
Enter configuration commands, one per line. End with CNTL/Z.
N5K1(config)# hardware ethernet store-and-fwd-switching
Enabling store-and-forward switching. Please copy the configuration and reload the switch
N5K1(config)# end
N5K1# copy run start
[#####] 100%
Copy complete, now saving to disk (please wait)...
N5K1# reload
WARNING: This command will reboot the system
Do you want to continue? (y/n) [n] y
```

```
N7K1:
!
! 7Ks run Underlay routing only and do not
! participate in VXLAN encap/decapsulation
!
nv overlay evpn
feature ospf
feature bgp
feature pim
feature bfd
ip pim rp-address 10.0.0.255 group-list 224.0.0.0/4 bidir

ip pim bfd

interface Ethernet1/1
shutdown

interface Ethernet1/2
shutdown

interface Ethernet1/3
medium p2p
bfd interval 250 min_rx 250 multiplier 4
ip unnumbered loopback0
ip router ospf 1 area 0.0.0.0
ip pim sparse-mode
no shutdown

interface Ethernet1/4
medium p2p
bfd interval 250 min_rx 250 multiplier 4
ip unnumbered loopback0
ip router ospf 1 area 0.0.0.0
ip pim sparse-mode
no shutdown

interface loopback0
ip address 10.0.0.71/32
ip router ospf 1 area 0.0.0.0
ip pim sparse-mode
```

```

interface loopback1
 ip address 10.0.0.254/31
 ip ospf network point-to-point
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode

router ospf 1
 bfd
 log-adjacency-changes

router bgp 65001
 log-neighbor-changes
 address-family 12vpn evpn
 neighbor 10.0.0.51
 remote-as 65001
 update-source loopback0
 address-family ipv4 unicast
 route-reflector-client
 address-family 12vpn evpn
 send-community extended
 route-reflector-client
 neighbor 10.0.0.52
 remote-as 65001
 update-source loopback0
 address-family ipv4 unicast
 route-reflector-client
 address-family 12vpn evpn
 send-community extended
 route-reflector-client

N7K2:
nv overlay evpn
feature ospf
feature bgp
feature pim
feature bfd

ip pim rp-address 10.0.0.255 group-list 224.0.0.0/4 bidir

ip pim bfd

interface Ethernet1/1
 shutdown

interface Ethernet1/2
 shutdown

interface Ethernet1/3
 medium p2p
 bfd interval 250 min_rx 250 multiplier 4
 ip unnumbered loopback0
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode
 no shutdown

interface Ethernet1/4
 medium p2p
 bfd interval 250 min_rx 250 multiplier 4
 ip unnumbered loopback0
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode
 no shutdown

interface loopback0
 ip address 10.0.0.72/32
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode

interface loopback1
 ip address 10.0.0.254/30
 ip ospf network point-to-point
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode

router ospf 1
 bfd
 log-adjacency-changes

router bgp 65001
 log-neighbor-changes
 address-family 12vpn evpn
 neighbor 10.0.0.51
 remote-as 65001
 update-source loopback0
 address-family ipv4 unicast
 route-reflector-client
 address-family 12vpn evpn
 send-community extended
 route-reflector-client
 neighbor 10.0.0.52
 remote-as 65001
 update-source loopback0
 address-family ipv4 unicast
 route-reflector-client
 address-family 12vpn evpn
 send-community extended
 route-reflector-client

```

```

N5K1:
!
! Underlay routing features

```

```

!
feature ospf
feature bgp
feature pim
feature bfd
!
ip pim rp-address 10.0.0.255 group-list 224.0.0.0/4 bidir
ip pim bfd
!
interface loopback0
 ip address 10.0.0.51/32
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode
!
interface Ethernet1/5
 no switchport
 medium p2p
 bfd interval 250 min_rx 250 multiplier 4
 ip unnumbered loopback0
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode

interface Ethernet1/6
 no switchport
 medium p2p
 bfd interval 250 min_rx 250 multiplier 4
 ip unnumbered loopback0
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode
!
router ospf 1
 bfd
 log-adjacency-changes
!
! VXLAN Overlay routing features
!
feature vn-segment-vlan-based
feature nv overlay
nv overlay evpn
hardware ethernet store-and-fwd-switching
!
vlan 172
 vn-segment 1000172
!
interface Ethernet1/1
 switchport access vlan 172
 spanning-tree port type edge
!
interface nve1
 no shutdown
 source-interface loopback0
 host-reachability protocol bgp
 member vni 1000172
 mcast-group 224.1.1.1
!
router bgp 65001
 log-neighbor-changes
 address-family 12vpn evpn
 neighbor 10.0.0.71
 remote-as 65001
 update-source loopback0
 address-family ipv4 unicast
 address-family 12vpn evpn
 send-community extended
 neighbor 10.0.0.72
 remote-as 65001
 update-source loopback0
 address-family ipv4 unicast
 address-family 12vpn evpn
 send-community extended
!
evpn
 vni 1000172 12
 rd auto
 route-target import auto
 route-target export auto

```

NSK2#

```

!
! Underlay routing features
!
feature ospf
feature bgp
feature pim
feature bfd
!
ip pim rp-address 10.0.0.255 group-list 224.0.0.0/4 bidir
ip pim bfd
!
interface loopback0
 ip address 10.0.0.52/32
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode
!
interface Ethernet1/5
 no switchport
 medium p2p
 bfd interval 250 min_rx 250 multiplier 4
 ip unnumbered loopback0
 ip router ospf 1 area 0.0.0.0
 ip pim sparse-mode

```

```

interface Ethernet1/6
  no switchport
  medium p2p
  bfd interval 250 min_rx 250 multiplier 4
  ip unnumbered loopback0
  ip router ospf 1 area 0.0.0.0
  ip pim sparse-mode
!
router ospf 1
  bfd
  log-adjacency-changes
!
! VXLAN Overlay routing features
!
feature vn-segment-vlan-based
feature nv overlay
nv overlay evpn
hardware ethernet store-and-fwd-switching
!
vlan 172
  vn-segment 1000172
!
interface Ethernet1/2
  switchport access vlan 172
  spanning-tree port type edge
!
interface nve1
  no shutdown
  source-interface loopback0
  host-reachability protocol bgp
  member vni 1000172
  mcast-group 224.1.1.1
!
router bgp 65001
  log-neighbor-changes
  address-family l2vpn evpn
  neighbor 10.0.0.71
    remote-as 65001
    update-source loopback0
  address-family ipv4 unicast
  address-family l2vpn evpn
    send-community extended
  neighbor 10.0.0.72
    remote-as 65001
    update-source loopback0
  address-family ipv4 unicast
  address-family l2vpn evpn
    send-community extended
!
evpn
  vni 1000172 12
  rd auto
  route-target import auto
  route-target export auto

```

Verification

The final verification in VXLAN with BGP L2VPN EVPN signaling, like VXLAN Flood and Learn, is the MAC address table of the Leaf switches. If the network is working properly, MAC addresses connected to remote Leafs that are in the same VXLAN segment (i.e. the same VNI) will be listed as reachable through the VXLAN tunnel (i.e. the NVE interface) via the next-hop of the remote Leaf's Loopback interface (i.e. the VTEP).

In the below output we can see that Leafs N5K1 and N5K2 learn the remote MAC addresses of Server 2 and Server 1 respectively.

```

N5K1# show mac address-table dynamic vlan 172
Legend:
  * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
  age - seconds since last seen,+ - primary entry using vPC Peer-Link
  VLAN    MAC Address    Type    age    Secure NTFY  Ports/SWID.SSID.LID
-----+-----+-----+-----+-----+-----
* 172    001b.2188.8075  dynamic  0      F F    nve1/10.0.0.52
* 172    001b.218d.3d98  dynamic  100    F F    Eth1/1

N5K2# show mac address-table dynamic vlan 172
Legend:
  * - primary entry, G - Gateway MAC, (R) - Routed MAC, O - Overlay MAC
  age - seconds since last seen,+ - primary entry using vPC Peer-Link
  VLAN    MAC Address    Type    age    Secure NTFY  Ports/SWID.SSID.LID
-----+-----+-----+-----+-----+-----
* 172    001b.2188.8075  dynamic  320    F F    Eth1/2
* 172    001b.218d.3d98  dynamic  0      F F    nve1/10.0.0.51

```

One of the main differences between this example and the previous Flood and Learn approach is that the VTEPs are explicitly signaled in the Control Plane with BGP, as opposed to discovered in the Data Plane. This can be seen in the below output.

```

N5K1# show nve peers
Interface Peer-IP      State LearnType Uptime  Router-Mac
-----+-----+-----+-----+-----+-----
nve1     10.0.0.52           Up    CP        02:27:33 n/a

N5K1# show nve peers detail
Details of nve Peers:
-----+-----+-----+-----+-----+-----
Peer-IP: 10.0.0.52
NVE Interface      : nve1
Peer State         : Up
Learn Type         : CP
Peer Uptime        : 02:27:38

```

```

Router-Mac      n/a
Peer First VNI   : 1000172
Time since Create : 02:27:38
Configured VNIs : 1000172
Provision State  : add-complete
Route-Update     : Yes
Peer Flags       : None
Learn Src        : RNH
Learnt CP VNIs  : 1000172
Peer-ifindex-resp : Yes
-----

```

Each MAC address within the VLAN to VNI mapping space on the Leaf switch generates a unique BGP route. This is basically analogous to doing a **redistribute connected** into BGP in the IP routing world. These routes can be verified as follows.

```

N5K1# show bgp l2vpn evpn
BGP routing table information for VRF default, address family L2VPN EVPN
BGP table version is 10, local router ID is 10.0.0.51
Status: s-suppressed, x-deleted, S-stale, d-dampened, h-history, *-valid, >-best
Path type: i-internal, e-external, c-confed, l-local, a-aggregate, r-redist, I-injected
Origin codes: i - IGP, e - EGP, ? - incomplete, | - multipath, & - backup

   Network          Next Hop          Metric      LocPrf      Weight Path
Route Distinguisher: 10.0.0.51:32939 (L2VNI 1000172)
*>i[2]:[0]:[0]:[48]:[001b.2188.8075]:[0]:[0.0.0.0]/216
      10.0.0.52                100          0 i
*>l[2]:[0]:[0]:[48]:[001b.218d.3d98]:[0]:[0.0.0.0]/216
      10.0.0.51                100         32768 i

Route Distinguisher: 10.0.0.52:32939
*>i[2]:[0]:[0]:[48]:[001b.2188.8075]:[0]:[0.0.0.0]/216
      10.0.0.52                100          0 i
* i      10.0.0.52                100          0 i

```

In the above output we see that there are two routes, one to MAC address 001b.2188.8075 which is advertised by VTEP 10.0.0.52, and the second to MAC address 001b.218d.3d98 which is advertised by VTEP 10.0.0.51. Note that the route to the MAC advertised by 10.0.0.52 has multiple paths. This is due to the fact that both Spines (N7K1 and N7K2 in this specific topology) are learning the MAC address route from Leaf N5K2, and reflecting the route to Leaf N5K1.

Just like regular IPv4 Unicast BGP routing, only one of these paths is used by BGP - the Best Path - due to the BGP Best Path Selection Algorithm. However, since both paths have the same next hop (the remote VTEP's Loopback0), it doesn't really matter which specific path is installed, just that a path *is* installed.

The specifics of the path selection can be seen below.

```

N5K1# show bgp l2vpn evpn 001b.2188.8075
Route Distinguisher: 10.0.0.51:32939 (L2VNI 1000172)
BGP routing table entry for [2]:[0]:[0]:[48]:[001b.2188.8075]:[0]:[0.0.0.0]/216, version 10
Paths: (1 available, best #1)
Flags: (0x00021a) on xmit-list, is in l2rib/evpn, is not in HW,

  Advertised path-id 1
  Path type: internal, path is valid, is best path
    Imported from 10.0.0.52:32939:[2]:[0]:[0]:[48]:[001b.2188.8075]:[0]:[0.0.0.0]/216
  AS-Path: NONE, path sourced internal to AS
    10.0.0.52 (metric 9) from 10.0.0.71 (10.0.0.71)
    Origin IGP, MED not set, localpref 100, weight 0
    Received label 1000172
  Extcommunity:
    RT:65001:1000172
    ENCAP:8
  Originator: 10.0.0.52 Cluster list: 10.0.0.71

  Path-id 1 not advertised to any peer

Route Distinguisher: 10.0.0.52:32939
BGP routing table entry for [2]:[0]:[0]:[48]:[001b.2188.8075]:[0]:[0.0.0.0]/216, version 9
Paths: (2 available, best #1)
Flags: (0x000202) on xmit-list, is not in l2rib/evpn, is not in HW, , is locked

  Advertised path-id 1
  Path type: internal, path is valid, is best path
  AS-Path: NONE, path sourced internal to AS
    10.0.0.52 (metric 9) from 10.0.0.71 (10.0.0.71)
    Origin IGP, MED not set, localpref 100, weight 0
    Received label 1000172
  Extcommunity:
    RT:65001:1000172
    ENCAP:8
  Originator: 10.0.0.52 Cluster list: 10.0.0.71

  Path type: internal, path is valid, not best reason: Neighbor Address
  AS-Path: NONE, path sourced internal to AS
    10.0.0.52 (metric 9) from 10.0.0.72 (10.0.0.72)
    Origin IGP, MED not set, localpref 100, weight 0
    Received label 1000172
  Extcommunity:
    RT:65001:1000172
    ENCAP:8
  Originator: 10.0.0.52 Cluster list: 10.0.0.72

  Path-id 1 not advertised to any peer

```

In the above output we see the detailed attributes of the two paths, and that path #1 learned from 10.0.0.71 (N7K1) is chosen as best. The three most important fields in the above output are the MAC address of the destination (001b.2188.8075), the next-hop to get to the VTEP (10.0.0.52), and the Route Target (65001:1000172).

The Route Target is a BGP Extended Community whose value is auto-generated (per our configuration) to encode the BGP ASN and the VNI numbers. The use case of this field in VXLAN BGP is the exact same logic as MPLS L2VPN or L3VPN MP-BGP, which is to decide whether or not we need to install the route in the forwarding table (the MAC address table in our case) when we receive it from our peers. The idea is that if we have VNIs A & B configured, and another Leaf has VNIs B & C configured, we ideally should only install MAC addresses for VNI B, as this is the only MAC address table that we both have in common. The result is an optimization of the amount of MAC address entries that are actually written to TCAM on the Leaf switch.

