



Configuring Layer 4 - Layer 7 Services

This chapter contains the following sections:

- [About VXLAN Layer 4 - Layer 7 Services, on page 1](#)
- [Integrating Layer 3 Firewalls in VXLAN Fabrics, on page 1](#)
- [Firewall as Default Gateway, on page 15](#)
- [Transparent Firewall Insertion, on page 16](#)
- [Firewall Clustering with VXLAN BGP EVPN, on page 22](#)
- [Service Redirection in VXLAN EVPN Fabrics, on page 25](#)

About VXLAN Layer 4 - Layer 7 Services

This chapter covers insertion of Layer 4 – Layer 7 services (firewall, load balancer, and so on) in a VXLAN fabric.

As opposed to traditional 3-tier network topologies, in which L4-L7 services are connected to the switches hosting the default gateway (aggregation/distribution), L4-L7 services in VXLAN fabrics are typically connected to the leaf or border switches, often referred to as *services leafs*.

You can attach a L4-L7 services device to a VXLAN fabric in various ways. This chapter addresses the considerations you must take depending on how the L4-L7 services device is attached and the requirements of the device and the network.

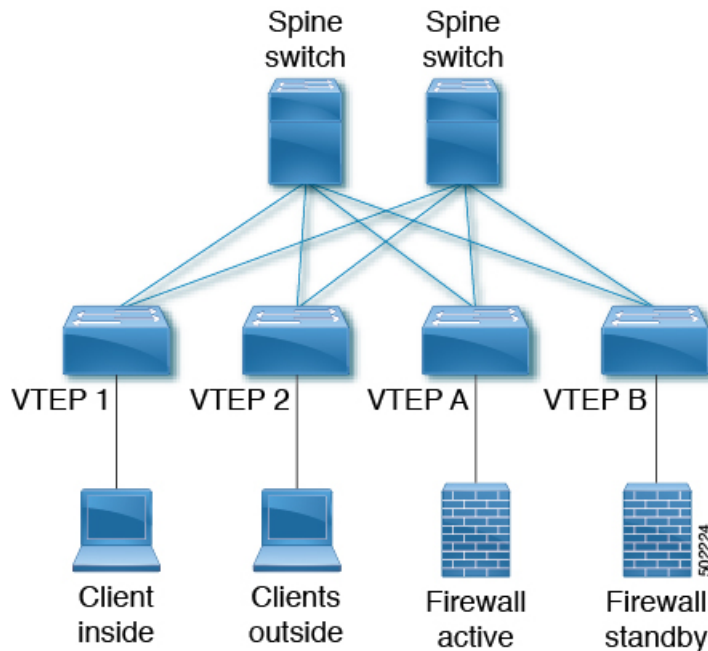
Integrating Layer 3 Firewalls in VXLAN Fabrics

This section provides details on how to integrate a firewall within a VXLAN EVPN fabric. A Layer-3 firewall involves separating different security zones.

When integrating a Layer-3 firewall in a VXLAN EVPN fabric with a distributed Anycast Gateway, each of these zones must correspond to a VRF/tenant on the fabric. The traffic within a tenant is routed by the fabric. Traffic between the tenants is routed by the firewall. This scenario often refers to an inter-tenant or tenant edge firewall.

Consider two zones: an inside zone and an outside zone. This scenario requires a VRF definition on the fabric. You can call the VRFs the inside VRF and the outside VRF. Traffic between subnets within the same VRF is routed on the VXLAN fabric using the distributed gateway. Traffic between VRFs is routed by the firewall where the rules are applied.

Figure 1: Topology Overview with Firewall Attachment



Single-Attached Firewall with Static Routing

If the firewall does not support running a routing protocol, you must have static routes on each VTEP pointing to the firewall as the next hop. The firewall also has static routes pointing to the Anycast Gateway IP as the next hop. The challenge with a static route is that the VTEP with an active firewall must be the one advertising the routes to the fabric. One way to accomplish this is to track the active firewall reachability via HMM and use this tracking to advertise routes into the fabric. When the active firewall is connected to VTEP A, VTEP A has a static route that tracks where the route is advertised if the firewall IP is learned as the HMM route. When the firewall fails and the standby firewall takes over, VTEP A learns the firewall IP using BGP, and VTEP B learns the firewall IP using HMM. VTEP A withdraws the route, and VTEP B advertises the route into the fabric. See the following example.

VTEP A and VTEP B:

```
Vlan 10
  Name inside
  Vn-segment 10010

Vlan 20
  Name outside
  Vn-segment 10020

Interface VLAN 10
  Description inside_vlan
  VRF member INSIDE
  IP address 10.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface VLAN 20
  Description outside_vlan
  VRF member OUTSIDE
```

```

IP address 20.1.1.254/24
fabric forwarding mode anycast-gateway

interface nve1
 no shutdown
 host-reachability protocol bgp
 source-interface loopback1
 member vni 10010
   mcastgroup 239.1.1.1
 member vni 10020
   mcastgroup 239.1.1.1
 member vni 1001000 associate-vrf
 member vni 1002000 associate-vrf

track 10 ip route 10.1.1.1/32 reachability hmm
 vrf member INSIDE
!
VRF context INSIDE
 Vni 1001000
 IP route 20.1.1.0/24 10.1.1.1 track 10

track 20 ip route 20.1.1.1/32 reachability hmm
 vrf member OUTSIDE
!
VRF context OUTSIDE
 Vni 1001000
 IP route 10.1.1.0/24 20.1.1.1 track 20

VTEPA# show track 10 Track 10
IP Route 20.1.1.1/32 Reachability Reachability is UP

VTEPA# show ip route 20.1.1.0/24 vrf INSIDE
IP Route Table for VRF "INSIDE"
 '*' denotes best ucast next-hop
 '**' denotes best mcast next-hop
 '[x/y]' denotes [preference/metric]
 '%<string>' in via output denotes VRF <string>

20.1.1.0/24, ubest/mbest: 1/0
 *via 10.1.1.1 [1/0], 00:00:08, static

Firewall Failure on VTEP A caused the track to go down causing VTEP A to withdraw the static
route.

VTEPA# show track 20 Track 20
IP Route 20.1.1.1/32 Reachability Reachability is DOWN

VTEPA# show ip route 20.1.1.0/24 vrf INSIDE
IP Route Table for VRF "RED"
 '*' denotes best ucast next-hop
 '**' denotes best mcast next-hop
 '[x/y]' denotes [preference/metric]
 '%<string>' in via output denotes VRF <string>

Route not found

```

Recursive Static Routes Distributed to the Rest of the Fabric

With this approach, the static routes are configured wherever the inside or outside VRF exists. As the next-hop is reachable through host routes (EVPN Route-Type2), the change of the active firewall to standby and vice versa is only seen locally and doesn't introduce any churn to the other VXLAN fabric. This approach can help to better scale and improve convergence.

Any VTEP:

```
VRF context OUTSIDE
Vni 1002000
IP route 10.1.1.0/24 20.1.1.1
! static route on VTEP pointing to Firewall next hop
! firewall VIP 20.1.1.1

VRF context INSIDE
Vni 1001000
IP route 20.1.1.0/24 10.1.1.1
! static route on VTEP pointing to Firewall next hop
! firewall VIP 10.1.1.1
```

Redistribute Static Routes into BGP and Advertise to the Rest of the Fabric

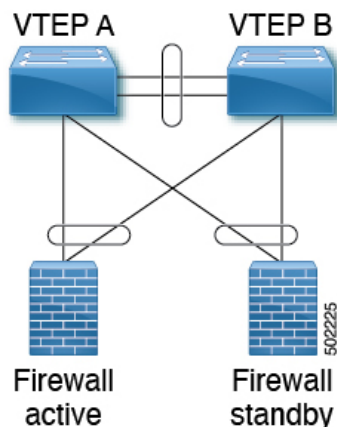
Through redistribution, we make the route toward the active firewall shown to the VTEP where it resides. The route is seen as a prefix route (EVPN Route-Type5), and as such, only the route toward the VTEP with the active firewall is seen. In the case of a firewall active/standby change, the tracking needs to detect the change and inform all of the remote VTEPs of this change. This behavior is equal to a route "delete" followed by an "add." This approach needs to notify all VTEPs with the VRF, and hence a wider churn can be seen.

VTEP A and VTEP B:

```
router bgp 65000
vrf OUTSIDE
address-family ipv4 unicast
redistribute static route-map Static-to-BGP
```

Dual-Attached Firewall with Static Routing

Figure 2: Dual-Attached Firewall with Static Routing



VTEP A and VTEP B:

```

Vlan 10
  Name inside
  Vn-segment 10010

Vlan 20
  Name outside
  Vn-segment 10020

interface nve1
  no shutdown
  host-reachability protocol bgp
  source-interface loopback1
  member vni 10010
    mcastgroup 239.1.1.1
  member vni 10020
    mcastgroup 239.1.1.1
  member vni 1001000 associate-vrf
  member vni 1002000 associate-vrf

Interface VLAN 10
  Description inside_vlan
  VRF member INSIDE
  IP address 10.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface VLAN 20
  Description outside_vlan
  VRF member OUTSIDE
  IP address 20.1.1.254/24
  fabric forwarding mode anycast-gateway

VRF context INSIDE
  Vni 1001000
  IP route 20.1.1.0/24 10.1.1.1
  ! static route on VTEP pointing to Firewall next hop
  ! firewall VIP 10.1.1.1
VRF context OUTSIDE
  Vni 1002000
  IP route 10.1.1.0/24 20.1.1.1
  ! static route on VTEP pointing to Firewall next hop
  ! firewall VIP 20.1.1.1

router bgp 65000
  vrf INSIDE
    address-family ipv4 unicast
      redistribute static route-map INSIDE-to-BGP
  vrf OUTSIDE
    address-family ipv4 unicast
      redistribute static route-map OUTSIDE-to-BGP

```

Single-Attached Firewall with eBGP Routing

If the firewall supports BGP, one option is to use BGP as a protocol between the firewall and the service VTEP. Peering using the anycast IP is not supported. The recommended design is to use dedicated loopback IPs on each VTEP and peer using the loopback. As long as the loopback interfaces are not advertised via

EVPN, the same IP address could be used on all of the belonging VTEPs. We recommend using individual IP addresses on a per-VTEP basis.

Reachability to the loopback from the firewall can be configured using a static route on the firewall, pointing to the Anycast Gateway IP on the VTEPs.

In the following example, an eBGP peering is established from the VTEPs, which are in AS 65000, and the firewall in AS 65002. The BGP peering with iBGP is not supported.



Note When having eBGP peering to active/standby firewalls connected to different VTEPs, **export-gateway-ip** must be enabled.

Do not use Anycast Gateway for BGP peerings.

VTEP A:

```
Vlan 10
  Name inside
  Vn-segment 10010

Vlan 20
  Name outside
  Vn-segment 10020

Interface VLAN 10
  Description inside_vlan
  VRF member INSIDE
  IP address 10.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface loopback100
  Vrf member INSIDE
  Ip address 172.16.1.253/32

Interface VLAN 20
  Description outside_vlan
  VRF member OUTSIDE
  IP address 20.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface loopback101
  Vrf member OUTSIDE
  Ip address 172.18.1.253/32

router bgp 65000
  vrf INSIDE
  ! peer with Firewall Inside
  neighbor 10.1.1.0/24 remote-as 65123
  update-source loopback100
  ebgp-multihop 5
  address-family ipv4 unicast
    local-as 65051 no-prepend replace-as

  vrf OUTSIDE
  ! peer with Firewall Outside
  neighbor 20.1.1.0/24 remote-as 65123
  update-source loopback101
  ebgp-multihop 5
```

```

address-family ipv4 unicast
  local-as 65052 no-prepend replace-as

```

VTEP B:

```

Vlan 10
  Name inside
  Vn-segment 10010

Vlan 20
  Name outside
  Vn-segment 10020
Interface VLAN 10
  Description inside_vlan
  VRF member INSIDE
IP address 10.1.1.254/24
fabric forwarding mode anycast-gateway

Interface loopback100
  Vrf member INSIDE
  Ip address 172.16.1.254/32

Interface VLAN 20
  Description outside_vlan
  VRF member OUTSIDE
IP address 20.1.1.254/24
fabric forwarding mode anycast-gateway

Interface loopback101
  Vrf member OUTSIDE
  Ip address 172.18.1.254/32

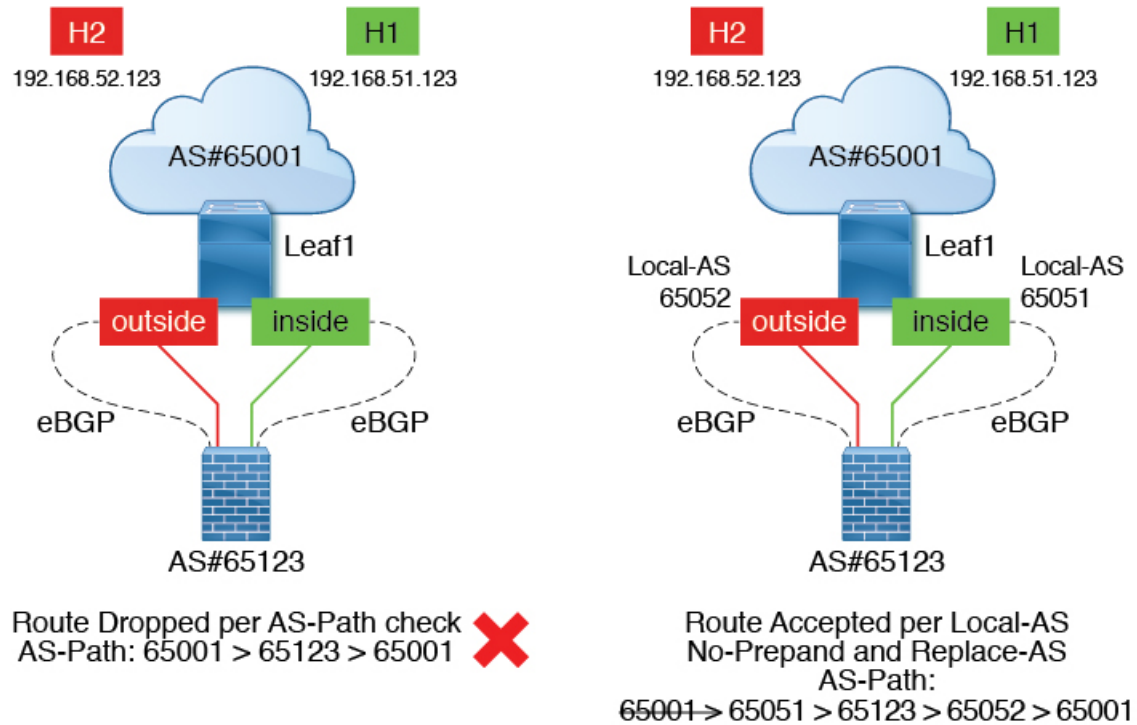
router bgp 65000
  vrf INSIDE
    ! peer with Firewall Inside
    neighbor 10.1.1.0/24 remote-as 65123
    update-source loopback100
    ebgp-multihop 5
    address-family ipv4 unicast
      local-as 65051 no-prepend replace-as

  vrf OUTSIDE
    ! peer with Firewall Outside
    neighbor 20.1.1.0/24 remote-as 65123
    update-source loopback101
    ebgp-multihop 5
    address-family ipv4 unicast
      local-as 65052 no-prepend replace-as

```

With the VXLAN fabric generally being in a single BGP Autonomous System (AS), the AS of the inside VRF and the outside VRF is the same. BGP does not install routes that are received from its own AS. Therefore, we need to adjust the AS-path to override this rule. Various approaches exist, including disabling the rule that BGP drops routes from its own AS, which has further implications to the network. To keep all of the BGP protection mechanics in place, the “local-as” approach allows you to mimic routes being originated from a different AS. We recommend inserting the “local-as #ASN# no-prepend replace-as” on each firewall peering with different “local-as” per VRF.

Figure 3: eBGP AS-Path Check



Dual-Attached Firewall with eBGP Routing

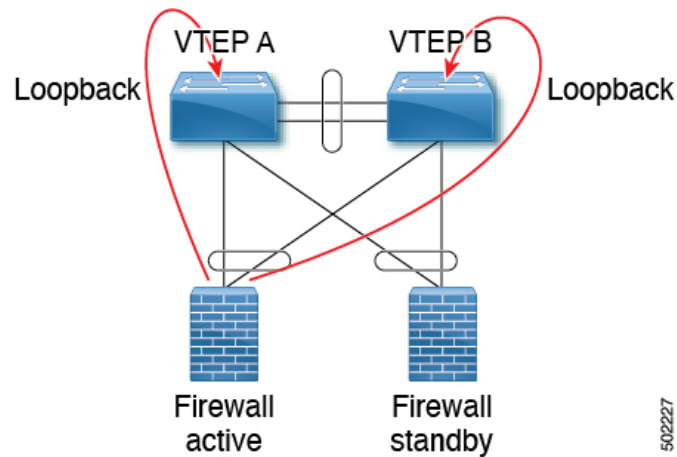
If the firewall supports BGP, one option is to use BGP as a protocol between the firewall and the service VTEP. Peering using the anycast IP is not supported. The recommended design is to use dedicated loopback IPs on each VTEP and peer using the loopback. As long as the loopback interfaces are not advertised via EVPN, the same IP address could be used on all of the belonging VTEPs. We recommend using individual IP addresses on a per-VTEP basis. For vPC environments, it is required.

Reachability to the loopback from the firewall can be configured using a static route on the firewall, pointing to the Anycast Gateway IP on the VTEPs.

In vPC deployments, you must have a per-VRF peering via a vPC peer-link. In addition to the per-VRF peering, you can enable the advertisement of prefix routes (EVPN Route-Type 5) using the **advertise-pip** command. For vPC with fabric peering, the per-VRF peering is not necessary, and the advertisement of prefix routes (EVPN Route-Type5) is required.

In the following example, an eBGP peering is established from the VTEPs, which are in AS 65000, and the firewall in AS 65002. The BGP peering with iBGP is not supported.

Figure 4: Dual-Attached Firewall with eBGP



Note When having eBGP peering to active/standby firewalls connected to different VTEPs, **export-gateway-ip** must be enabled.

Do not use Anycast Gateway for BGP peerings.

VTEP A:

```
Vlan 10
  Name inside
  Vn-segment 10010

Vlan 20
  Name outside
  Vn-segment 10020

Interface VLAN 10
  Description inside_vlan
  VRF member INSIDE
  IP address 10.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface loopback100
  Vrf member INSIDE
  Ip address 172.16.1.253/32

Interface VLAN 20
  Description outside_vlan
  VRF member OUTSIDE
  IP address 20.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface loopback101
  Vrf member OUTSIDE
  Ip address 172.18.1.253/32

router bgp 65000
vrf INSIDE
  ! peer with Firewall Inside
```

```

neighbor 10.1.1.0/24 remote-as 65123
update-source loopback100
ebgp-multihop 5
address-family ipv4 unicast
  local-as 65051 no-prepend replace-as

vrf OUTSIDE
  ! peer with Firewall Outside
neighbor 20.1.1.0/24 remote-as 65123
update-source loopback101
ebgp-multihop 5
address-family ipv4 unicast
  local-as 65052 no-prepend replace-as

```

VTEP B:

```

Vlan 10
  Name inside
  Vn-segment 10010

Vlan 20
  Name outside
  Vn-segment 10020

Interface VLAN 10
  Description inside_vlan
  VRF member INSIDE
  IP address 10.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface loopback100
  Vrf member INSIDE
  Ip address 172.16.1.254/32

Interface VLAN 20
  Description outside_vlan
  VRF member OUTSIDE
  IP address 20.1.1.254/24
  fabric forwarding mode anycast-gateway

Interface loopback101
  Vrf member OUTSIDE
  Ip address 172.18.1.254/32

router bgp 65000
  vrf INSIDE
  ! peer with Firewall Inside
  neighbor 10.1.1.0/24 remote-as 65123
  update-source loopback100
  ebgp-multihop 5
  address-family ipv4 unicast
    local-as 65051 no-prepend replace-as

  vrf OUTSIDE
  ! peer with Firewall Outside
  neighbor 20.1.1.0/24 remote-as 65123
  update-source loopback101
  ebgp-multihop 5
  address-family ipv4 unicast
    local-as 65052 no-prepend replace-as

```

Per-VRF Peering via vPC Peer-Link

VTEP A and VTEP B:

```

vlan 3966
! vlan use for peering between the vPC VTEPS

vlan 3967
! vlan use for peering between the vPC VTEPS

system nve infra-vlans 3966,3967

interface vlan 3966
vrf member INSIDE
ip address 100.1.1.1/31

interface vlan 3967
vrf member OUTSIDE
ip address 100.1.2.1/31

router bgp 65000
vrf INSIDE
neighbor 100.1.1.0 remote-as 65000
update-source vlan 3966
next-hop self
address-family ipv4 unicast

vrf OUTSIDE
neighbor 100.1.2.0 remote-as 65000
update-source vlan 3967
next-hop self
address-family ipv4 unicast

```

The routes learned in each VRF are advertised to the rest of the fabric via BGP EVPN updates.

Single-Attached Firewall with OSPF

The following example shows a configuration snippet from VTEP A running OSPF peering with the firewall. SVIs are defined on the VTEP for both inside and outside VRFs. The VTEP peers with the firewall on each of these VRFs dynamically learn routing information to go from one VRF to the other.

VTEP A and VTEP B:

```

vlan 10
name inside
vn-segment 10010

vlan 20
name outside
vn-segment 10020

interface VLAN 10
Description inside_vlan
VRF member INSIDE
IP address 10.1.1.254/24
IP router ospf 1 area 0
fabric forwarding mode anycast-gateway

Interface VLAN 20
Description outside_vlan
VRF member OUTSIDE

```

```

IP address 20.1.1.254/24
IP router ospf 1 area 0
fabric forwarding mode anycast-gateway

interface nve1
no shutdown
host-reachability protocol bgp
source-interface loopback1
member vni 10010
  mcastgroup 239.1.1.1
member vni 10020
  mcastgroup 239.1.1.1
member vni 1001000 associate-vrf
member vni 1002000 associate-vrf

router ospf 1
router-id 192.168.1.1
vrf INSIDE
VRF OUTSIDE

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

10.1.1.0/24, ubest/mbest: 1/0
  *via 20.1.1.1 Vlan20, [110/41], 1w5d, ospf-1, intra

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

20.1.1.0/24, ubest/mbest: 1/0
  *via 10.1.1.1 Vlan10, [110/41], 1w5d, ospf-1, intra

```

This route is then redistributed into BGP and advertised through the EVPN fabric so that all other VTEPs have all routes in each VRF pointing to VTEP A as the next hop.

Redistribute OSPF Routes into BGP and Advertise to the Rest of the Fabric

VTEP A and VTEP B:

```

router bgp 65000
vrf OUTSIDE
  address-family ipv4 unicast
    redistribute ospf 1 route-map OUTSIDEOSPF-to-BGP
vrf INSIDE
  address-family ipv4 unicast
    redistribute ospf 1 route-map INSIDEOSPF-to-BGP

VTEPA# show ip route 10.1.1.0/24 vrf OUTSIDE

10.1.1.0/24 ubest/mbest: 1/0
  *via 10.1.1.18%default, [200/41], 1w1d, bgp-65000, internal, tag 65000 (evpn) segid:
200100 tunnelid: 0xa010112 encap: VXLAN

```

Traffic is VXLAN encapsulated from VTEP to services VTEP and decapsulated and sent to the firewall. The firewall enforces the rules and sends the traffic to the services VTEP on the inside VRF. This traffic is then VXLAN encapsulated and sent to the destination VTEP where traffic is decapsulated and sent to the end client.

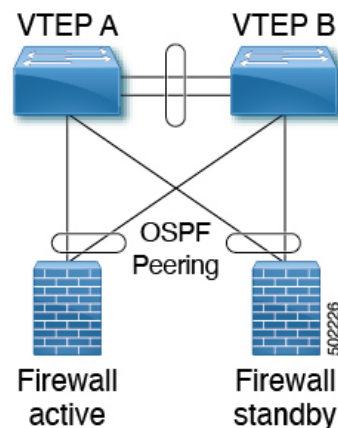
Firewall Failover

When the active firewall fails and the standby firewall takes over, routes are withdrawn from service VTEP A and advertised to the fabric by service VTEP B.

Dual-Attached Firewall with OSPF

Cisco NX-OS supports dynamic OSPF peering over vPC using Layer 3, which enables firewall connectivity using vPC and establishes OSPF peering over this link. The VLAN used to establish peering between the Cisco Nexus 9000 switches and the firewall must be a non-VXLAN-enabled VLAN.

Figure 5: Dual-Attached Firewall with OSPF



Note Do not use Anycast Gateway for OSPF adjacencies.

VTEP A:

```
Vlan 10
  Name inside

Vlan 20
  Name outside

Interface VLAN 10
  Description inside_vlan
  VRF member INSIDE
  IP address 10.1.1.253/24
  Ip router ospf 1 area 0

Interface VLAN 20
  Description outside_vlan
  VRF member OUTSIDE
  IP address 20.1.1.253/24
  Ip router ospf 1 area 0
```

```

vpc domain 100
 layer3 peer-router
 peer-gateway
 peer-switch
 peer-keepalive destination x.x.x.x source x.x.x.x peer-gateway
 ipv6 nd synchronize
 ip arp synchronize

router ospf 1
 vrf INSIDE VRF OUTSIDE

```

VTEP B:

```

Vlan 10
 Name inside

```

```

Vlan 20
 Name outside

```

```

Interface VLAN 10
 Description inside_vlan
 VRF member INSIDE
 IP address 10.1.1.254/24
 Ip router ospf 1 area 0

```

```

Interface VLAN 20
 Description outside_vlan
 VRF member OUTSIDE
 IP address 20.1.1.254/24
 Ip router ospf 1 area 0

```

```

vpc domain 100
 layer3 peer-router
 peer-gateway
 peer-switch
 peer-keepalive destination x.x.x.x source x.x.x.x peer-gateway
 ipv6 nd synchronize
 ip arp synchronize

router ospf 1
 vrf INSIDE VRF OUTSIDE

```

```

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

10.1.1.0/24, ubest/mbest: 1/0
 *via 20.1.1.1 Vlan20, [110/41], 1w5d, ospf-1, intra

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

20.1.1.0/24, ubest/mbest: 1/0
 *via 10.1.1.1 Vlan10, [110/41], 1w5d, ospf-1, intra

```

Redistribute OSPF Routes into BGP and Advertise to the Rest of the Fabric

VTEP A and VTEP B:

```
router bgp 65000
vrf OUTSIDE
  address-family ipv4 unicast
    redistribute ospf 1 route-map OUTSIDEOSPF-to-BGP
vrf INSIDE
  address-family ipv4 unicast
    redistribute ospf 1 route-map INSIDEOSPF-to-BGP
```

Firewall as Default Gateway

In this deployment model, the VXLAN fabric is a Layer 2 fabric, and the default gateway resides on the firewall.

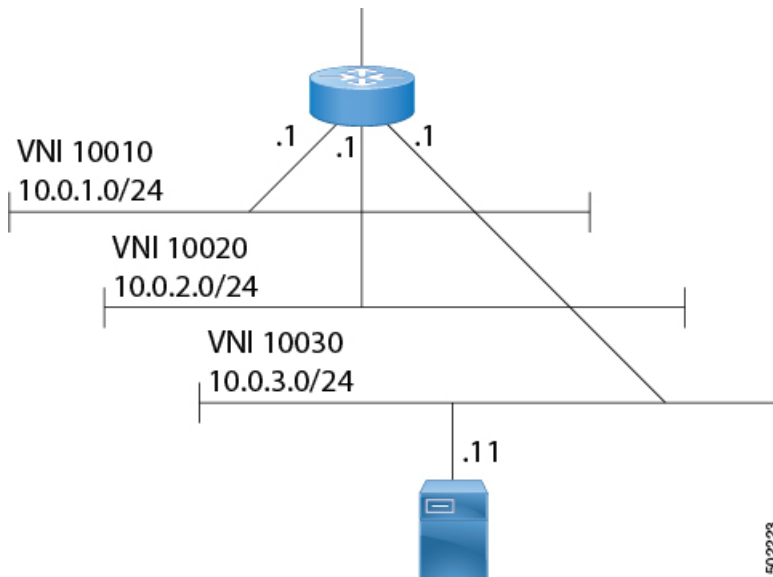
For example:

```
vlan 10
  name WEB
  vn-segment 10010
vlan 20
  name APPLICATION
  vn-segment 10020
vlan 30
  name DATABASE
  vn-segment 10030

interface nve1
  no shutdown
  host-reachability protocol bgp
  source-interface loopback1
  member vni 10010
    mcastgroup 239.1.1.1
  member vni 10020
    mcastgroup 239.1.1.1
  member vni 10030
    mcastgroup 239.1.1.1
```

The firewall has a logical interface in each VNI and is the default gateway for all endpoints. Every inter-VNI communication flows through the firewall. Take special care with the sizing of the firewall so that it does not become a bottleneck. Therefore, use this design in environments with low-bandwidth requirements.

Figure 6: Firewall as Default Gateway with a Layer-2 VXLAN Fabric



Transparent Firewall Insertion

Transparent firewalls or Layer 2 firewalls (including IPS/IDS) typically bridge between an inside VLAN and outside VLAN and inspect traffic as it traverses through them. VLAN stitching is done by placing the default gateway for the service on the inside VLAN. The Layer 2 reachability to this gateway is done on the outside VLAN.

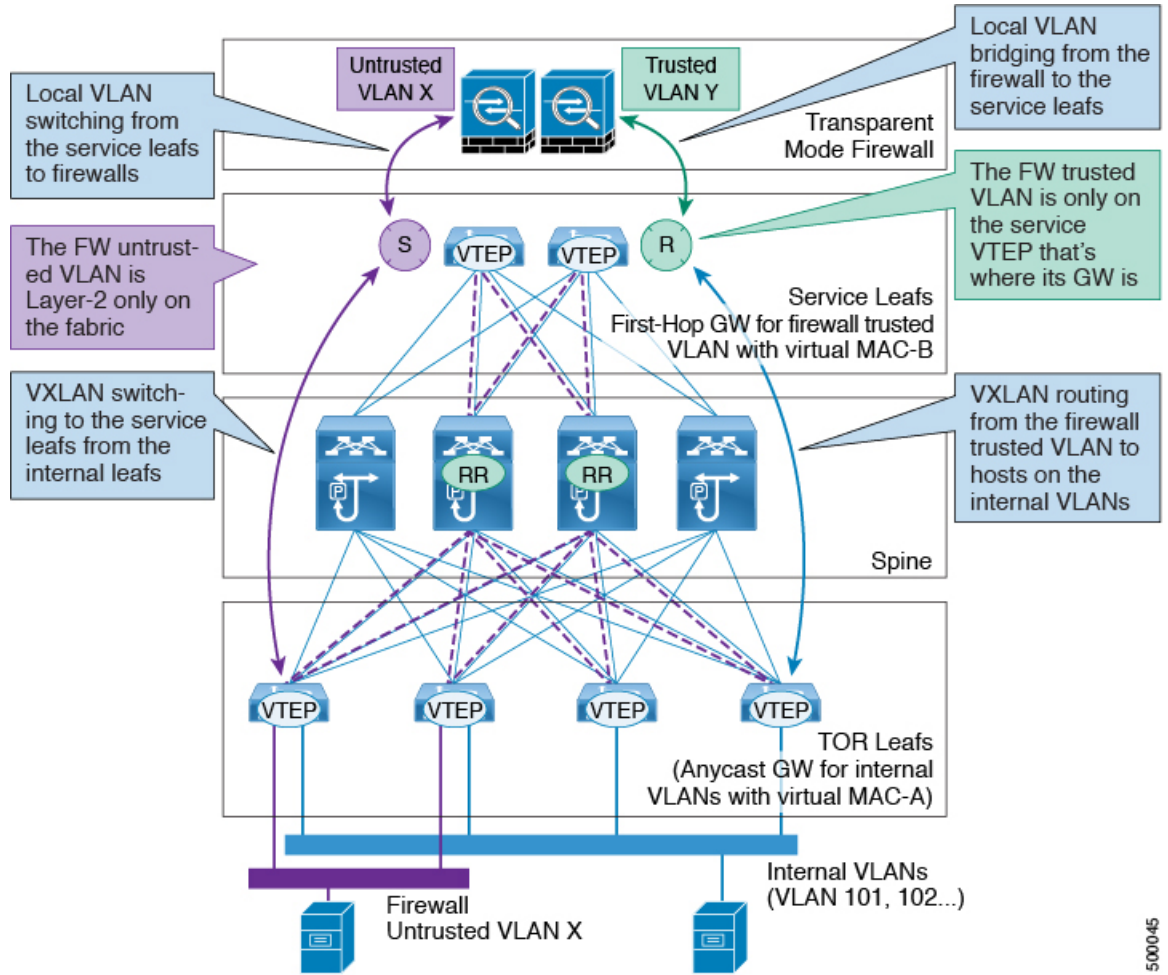
Overview of EVPN with Transparent Firewall Insertion

The topology contains the following types of VLANs:

- Internal VLAN (a regular VXLAN on ToR leafs with Anycast Gateway)
- Firewall untrusted VLAN X
- Firewall trusted VLAN Y

In this topology, the traffic that goes from VLAN X to other VLANs must go through a transparent Layer 2 firewall that is attached to the service leafs. This topology utilizes an approach of an untrusted VLAN X and a trusted VLAN Y. All ToR leafs have a Layer 2 VNI VLAN X. There is no SVI for VLAN X. The service leafs that are connected to the firewall have Layer 2 VNI VLAN X, non-VXLAN VLAN Y, and SVI Y with an HSRP gateway.

Overview of EVPN with Transparent Firewall Insertion



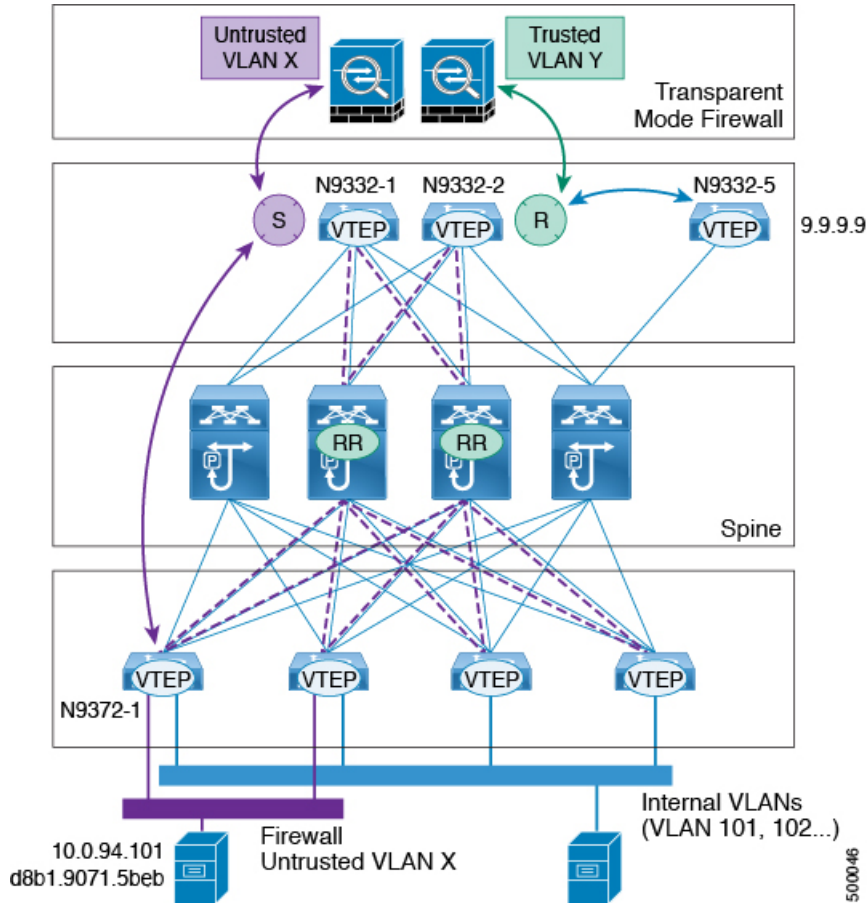
500045



Note For VXLAN EVPN, we recommend using the distributed Anycast Gateway with transparent firewall insertion. Doing so allows all VLANs to be VXLAN enabled. When using an HSRP/VRRP-based First-Hop Gateway, the VLAN for the SVI can't be VXLAN enabled and should reside on a vPC pair for redundancy.

EVPN with Transparent Firewall Insertion Example

Example of EVPN with Transparent Firewall Insertion



- Host in VLAN X: 10.1.94.101
- ToR leaf: N9372-1
- Service leaf in vPC: N9332-1 and N9332-2
- Border leaf: N9332-5

ToR Leaf Configuration

```

vlan 94
vn-segment 100094

interface nve1
 member vni 100094
 mcastgroup 239.1.1.1

router bgp 64500
 routerid 1.1.2.1
 neighbor 1.1.1.1 remote-as 64500
 address-family l2vpn evpn
 send-community extended
  
```

```

neighbor 1.1.1.2 remote-as 64500
address-family l2vpn evpn
  send-community extended
vrf Ten1
  address-family ipv4 unicast
  advertise l2vpn evpn

evpn
vni 100094 l2
  rd auto
  route-target import auto
  route-target export auto

```

Service Leaf 1 Configuration Using HSRP

```

vlan 94
description untrusted_vlan
  vn-segment 100094

vlan 95
  description trusted_vlan

vpc domain 10
  peer-switch
  peer-keepalive destination 10.1.59.160
  peer-gateway
  auto-recovery
  ip arp synchronize

interface Vlan2
description vpc_backup_svi_for_overlay
  no shutdown
  no ip redirects
  ip address 10.10.60.17/30
  no ipv6 redirects
  ip router ospf 100 area 0.0.0.0
  ip ospf bfd
  ip pim sparsemode

interface Vlan95
description SVI_for_trusted_vlan
  no shutdown
  mtu 9216
  vrf member Ten-1
  no ip redirects
  ip address 10.0.94.2/24
  hsrp 0
  preempt priority 255
  ip 10.0.94.1

interface nve1
  member vni 100094
  mcast-group 239.1.1.1

router bgp 64500
  routerid 1.1.2.1
  neighbor 1.1.1.1 remote-as 64500
  address-family l2vpn evpn
  send-community extended
  neighbor 1.1.1.2 remote-as 64500
  address-family l2vpn evpn
  send-community extended
vrf Ten-1
  address-family ipv4 unicast

```

```

network 10.0.94.0/24 /*advertise /24 for SVI 95 subnet; it is not VXLAN anymore*/
advertise l2vpn evpn

```

```

evpn
vni 100094 l2
rd auto
route-target import auto
route-target export auto

```

Service Leaf 2 Configuration Using HSRP

```

vlan 94
description untrusted_vlan
vnsegment 100094

vlan 95
description trusted_vlan

vpc domain 10
peer-switch
peer-keepalive destination 10.1.59.159
peer-gateway
auto-recovery
ip arp synchronize

interface Vlan2
description vpc_backup_svi_for_overlay
no shutdown
no ip redirects
ip address 10.10.60.18/30
no ipv6 redirects
ip router ospf 100 area 0.0.0.0
ip pim sparsemode

interface Vlan95
description SVI_for_trusted_vlan
no shutdown
mtu 9216
vrf member Ten-1
no ip redirects
ip address 10.0.94.3/24
hsrp 0
preempt priority 255
ip 10.0.94.1

interface nve1
member vni 100094
mcastgroup 239.1.1.1

router bgp 64500
router-id 1.1.2.1
neighbor 1.1.1.1 remote-as 64500
address-family l2vpn evpn
send-community extended
neighbor 1.1.1.2 remote-as 64500
address-family l2vpn evpn
send-community extended
vrf Ten-1
address-family ipv4 unicast
network 10.0.94.0/24 /*advertise /24 for SVI 95 subnet; it is not VXLAN anymore*/
advertise l2vpn evpn

evpn
vni 100094 l2

```

```
rd auto
route-target import auto
route-target export auto
```

Show Command Examples

Display information about the ingress leaf learned local MAC from host:

```
switch# sh mac add vl 94 | i 5b|MAC
* primary entry, G - Gateway MAC, (R) Routed - MAC, O - Overlay MAC
VLAN MAC Address Type age Secure NTFY Ports
* 94 d8b1.9071.5beb dynamic 0 F F Eth1/1
```

Display information about the service leaf found MAC of host:



Note In VLAN 94, the service leaf learned the host MAC from the remote peer by BGP.

```
switch# sh mac add vl 94 | i VLAN|eb

VLAN MAC Address Type age Secure NTFY Ports
* 94 d8b1.9071.5beb dynamic 0 F F nvel(1.1.2.1)

switch# sh mac add vl 94 | i VLAN|eb

VLAN MAC Address Type age Secure NTFY Ports
* 94 d8b1.9071.5beb dynamic 0 F F nvel(1.1.2.1)

switch# sh mac add vl 95 | i VLAN|eb

VLAN MAC Address Type age Secure NTFY Ports
+ 95 d8b1.9071.5beb dynamic 0 F F Po300

switch# sh mac add vl 95 | i VLAN|eb

VLAN MAC Address Type age Secure NTFY Ports
+ 95 d8b1.9071.5beb dynamic 0 F F Po300
```

Display information about service leaf learned ARP for host on VLAN 95:

```
switch# sh ip arp vrf ten-1
Address      Age      MAC Address      Interface
10.0.94.101 00:00:26 d8b1.9071.5beb  Vlan95
```

Service Leaf learns 9.9.9.9 from EVPN.

```
switch# sh ip route vrf ten-1 9.9.9.9
IP Route Table for VRF "Ten-1"
'*' denotes best ucast nexthop
'**' denotes best mcast nexthop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

9.9.9.9/32, ubest/mbest: 1/0
  *via 1.1.2.7%default, [200/0], 02:57:27, bgp64500,internal, tag 65000 (evpn) segid: 10011
  tunnelid: 0x1
  010207 encap: VXLAN
```

Display information about the border leaf learned host routes by BGP:

```
switch# sh ip route 10.0.94.101

IP Route Table for VRF "default"
'*' denotes best ucast nexthop
'***' denotes best mcast nexthop
'[x/y]' denotes [preference/metric]
'%<string>' in via output denotes VRF <string>

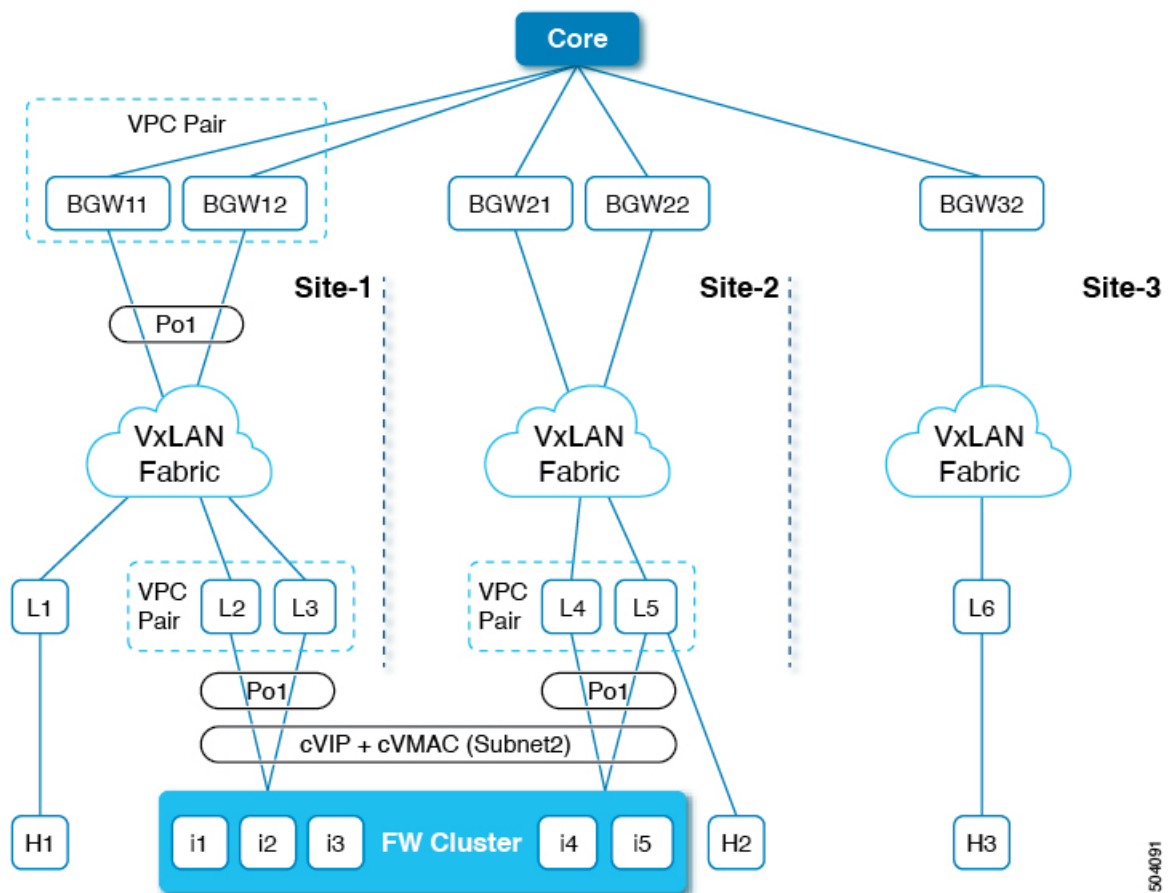
10.0.94.0/24, ubest/mbest: 1/0
   *via 10.100.5.0, [20/0], 03:14:27, bgp65000,external, tag 6450
```

Firewall Clustering with VXLAN BGP EVPN

This section provides details on how to configure a firewall cluster that spans across multiple sites running a VXLAN fabric with a BGP EVPN control plane.

The following topology illustrates the firewall clustering with VXLAN EVPN.

Figure 7: Firewall clustering with VXLAN EVPN



This topology covers the following:

- Firewall cluster consists of multiple instances that act as a single device.
- Routed Access to firewall can be through a different or same subnet.
- Firewall employs a L2 port-channel spanned across all instances.
- A common ESI represents all vPC port-channels that connect to the firewall cluster.
- Single VIP/VMAC is present across all instances.
- BGP-EVPN VXLAN overlay per site is stitched at Border Gateways.
- Anycast forwarding of Active-to-Active instances within the same site and Active-to-Backup access to firewall across sites for traffic flows is supported.
- Each site has a single vPC pair connected to the cluster with a port-channel interface assigned to it.
- The cluster VIP and cluster VMAC are advertised into the VXLAN EVPN fabric as BGP EVPN Route Target-2s (with the ESI set to the configured value on each vPC's port-channel interface). The next hop of the Route Target-2 is the vPC pair's VTEP VIP address.
- Each site may have multiple clusters. The clusters are attached to the vPC pair with their individual port-channels with unique ESIs.
- Each cluster has its own cVIP and cVMAC that are advertised into the VXLAN EVPN fabric as BGP EVPN Route Target -2s (with the ESI set to the configured value on its vPC's Port-channel interface).
- A cluster may have multiple VLANs on the port-channel connected to the vPC pair. Each cVIP/cVMAC learnt on a VLAN is advertised with its corresponding L2VNI as a Route T-2 EVPN route.
- VIP and VMAC (Firewall Hosts) are attached to a single spanned Ether-channel.
- Spanned Ether-channel extends across sites.
- Anycast forwarding to VIP is determined by leverage of existing BGP path attributes and best-path selection.

On the VTEP leafs attached to the firewall cluster, BGP uses a route-map to attach a community to firewall cluster-related EVPN EAD/ES (Type-1) and MAC/IP (Type-2) routes.

```
router bgp 12000
 address-family l2vpn evpn
 originate-map set_esi
 template peer SITE-BGW
  remote-as 12000
  update-source loopback1
  address-family l2vpn evpn
  send-community
  send-community extended
 template peer VTEP-PEERS
  remote-as 12000
  update-source loopback1
  address-family l2vpn evpn
  send-community
  send-community extended
```

On the border gateways, BGP uses a route-map to match the firewall clustering community attached to EVPN EAD/ES (Type-1) and MAC/IP (Type-2) routes.

```
router bgp 11000
```

```

bestpath as-path multipath-relax
neighbor 111.111.10.1 remote-as 12000
peer-type fabric-external
address-family l2vpn evpn
  send-community
  send-community extended
  route-map preserve_esi out
  rewrite-evpn-rt-asn

```

On the VTEP leafs attached to the firewall cluster, you need to configure a route-map to attach a community to firewall cluster-related EVPN EAD/ES (Type-1) and MAC/IP (Type-2) routes.

```

route-map set_esi permit 10
  match tag 100000
  match evpn route-type 1 2
  set community 23456:12345
route-map set_esi permit 15

```



Caution The **match tag** command in a route-map associated with route-map *<name>* out BGP command under neighbor address-family mode is only supported if configured under address-family l2vpn evpn.

On the border gateways, you need to configure separate route-maps for fabric-internal and fabric-external peers to match the firewall clustering community attached to EVPN EAD/ES (Type-1) and MAC/IP (Type-2) routes.

Matching outbound L2VPN/EVPN route-map to fabric-internal peers:

```

route-map preserve_esi permit 10
  match community preserve_esi
  match evpn route-type 2
  set esi unchanged
route-map preserve_esi permit 15
route-map preserve_esi permit 30

```

Matching outbound L2VPN/EVPN route-map to fabric-external peers:

```

route-map preserve_esi_external permit 10
  match community preserve_esi
  match evpn route-type 2
  set esi unchanged
route-map preserve_esi_external permit 15
  match community preserve_esi
  match evpn route-type 1
route-map preserve_esi_external permit 20
  match evpn route-type 1
  match route-type local
route-map preserve_esi_external deny 25
  match evpn route-type 1
route-map preserve_esi_external permit 30

```

The ethernet-segment can be configured only under vPC port-channel.

```

interface port-channel 100
  ethernet-segment vpc
  esi <esi> [ tag <uint >]
interface port-channel 200
  ethernet-segment vpc
  esi system-mac <system-mac> <local-identifier> [tag <uint>]

```


A common ESI represents all vPC port-channels that connect to the firewall cluster. You can configure ESI under a vPC port-channel.

```
evpn esi multihoming
port-channel 100
  ethernet-segment 1
    system-mac aa.bb.cc <anycast-host>
```

Keep the same system-mac for all vPC port-channels that host the same firewall cluster.

For more firewall information, see [Integrating Layer 3 Firewalls in VXLAN Fabrics](#).

Service Redirection in VXLAN EVPN Fabrics

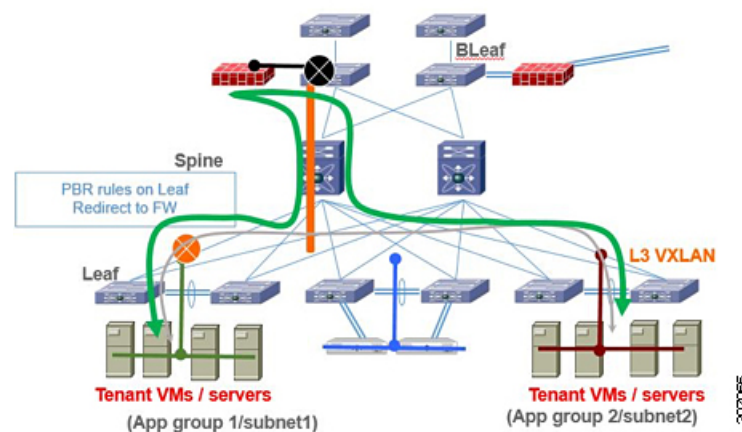
Today, insertion of service appliances (also referred to as service nodes or service endpoints) such as firewalls, load-balancers, etc are needed to secure and optimize applications within a data center. This section describes the Layer 4-Layer 7 service insertion and redirection features offered on VXLAN EVPN fabrics that provides sophisticated mechanisms to onboard and selectively redirect traffic to these services.

Use of Policy-Based Redirect for Services Insertion

Policy-based redirect (PBR) provides a mechanism to bypass a routing table lookup and redirect traffic to a next-hop IP reachable over VXLAN. The feature enables service redirection to Layer 4-Layer 7 devices such as firewalls and load balancers.

PBR involves configuring a route-map with rules that dictate where traffic must be forwarded. The route map is applied on the tenant SVI to influence traffic coming from the host-facing interfaces to a next hop reachable via the fabric.

In scenarios where traffic is coming to a VTEP from the overlay and needs to be redirected to another next hop, the PBR policy must be applied on the fabric facing Layer-3 VNI Interface.



In the previous figure, communication between App group 1 and App group 2 takes place via inter-VLAN/VNI routing in the tenant VRF by default. If there is a requirement where traffic from App group 1 to App group 2 must go through a firewall, a PBR policy can be used to redirect traffic. The example in section “Configuration Example for Policy-Based Redirect” provides the necessary configuration that redirects the traffic flow.

This VXLAN PBR functionality is very basic and lacks many of the required functionality for proper insertion of services in VXLAN fabric. Hence the recommendation is to instead look at ePBR for all the reasons explained in [Enhanced-Policy Based Redirect \(ePBR\)](#), on page 30 section.

Guidelines and Limitations for Policy-Based Redirect

The following guidelines and limitations apply to PBR over VXLAN.

- The following platforms support PBR over VXLAN:
 - Cisco Nexus 9332C and 9364C switches
 - Cisco Nexus 9300-EX switches
 - Cisco Nexus 9300-FX/FX2/FX3 switches
 - Cisco Nexus 9300-GX switches
 - Cisco Nexus 9300-GX2 switches
 - Cisco Nexus 9332D-H2R switches
 - Cisco Nexus 93400LD-H1 switches
 - Cisco Nexus 9364C-H1 switches
 - Cisco Nexus 9504 and 9508 switches with -EX/FX line cards
- Beginning with Cisco NX-OS Release 10.2(3)F, the VXLAN PBR feature is supported with VXLANv6 on all TOR switches.
- PBR over VXLAN doesn't support the following features: VTEP ECMP, and the **load-share** keyword in the **set {ip | ipv6} next-hop ip-address** command.

Enabling the Policy-Based Redirect Feature

To configure basic PBR, in cases where the advanced (and recommended) ePBR functions are not deployed, see the following sections:

- [Enabling the Policy-Based Redirect Feature](#), on page 26
- [Configuring a Route Policy](#), on page 27
- [Verifying the Policy-Based Redirect Configuration](#), on page 28
- [Configuration Example for Policy-Based Redirect](#), on page 29

Before you begin

Enable the policy-based redirect feature before you can configure a route policy.

SUMMARY STEPS

1. **configure terminal**
2. **[no] feature pbr**
3. (Optional) **show feature**
4. (Optional) **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: switch# configure terminal	Enters global configuration mode.
Step 2	[no] feature pbr Example: switch(config)# feature pbr	Enables the policy-based routing feature.
Step 3	(Optional) show feature Example: switch(config)# show feature	Displays enabled and disabled features.
Step 4	(Optional) copy running-config startup-config Example: switch(config)# copy running-config startup-config	Saves this configuration change.

Configuring a Route Policy

You can use route maps in policy-based routing to assign routing policies to the inbound interface. Cisco NX-OS routes the packets when it finds a next hop and an interface.



Note The switch has a RAACL TCAM region by default for IPv4 traffic.

Before you begin

Configure the RAACL TCAM region (using TCAM carving) before you apply the policy-based routing policy. For instructions, see the “Configuring ACL TCAM Region Sizes” section in the [Cisco Nexus 9000 Series NX-OS Security Configuration Guide, Release 9.2\(x\)](#).

SUMMARY STEPS

1. **configure terminal**
2. **interface** *type slot/port*
3. **{ip | ipv6} policy route-map** *map-name*
4. **route-map** *map-name* [**permit** | **deny**] [*seq*]
5. **match {ip | ipv6} address access-list-name** *name* [*name...*]
6. **set ip next-hop** *address1*
7. **set ipv6 next-hop** *address1*
8. (Optional) **set interface null0**
9. (Optional) **copy running-config startup-config**

DETAILED STEPS

	Command or Action	Purpose
Step 1	configure terminal Example: switch# configure terminal	Enters global configuration mode.
Step 2	interface <i>type slot/port</i> Example: switch(config)# interface ethernet 1/2	Enters interface configuration mode.
Step 3	{ip ipv6} policy route-map <i>map-name</i> Example: switch(config-inf)# ip policy route-map Testmap	Assigns a route map for IPv4 or IPv6 policy-based routing to the interface.
Step 4	route-map <i>map-name</i> [permit deny] [<i>seq</i>] Example: switch(config-inf)# route-map Testmap	Creates a route map or enters route-map configuration mode for an existing route map. Use <i>seq</i> to order the entries in a route map.
Step 5	match {ip ipv6} address <i>access-list-name name</i> [<i>name...</i>] Example: switch(config-route-map)# match ip address access-list-name ACL1	Matches an IPv4 or IPv6 address against one or more IPv4 or IPv6 access control lists (ACLs). This command is used for policy-based routing and is ignored by route filtering or redistribution.
Step 6	set ip next-hop <i>address1</i> Example: switch(config-route-map)# set ip next-hop 192.0.2.1	Sets the IPv4 next-hop address for policy-based routing.
Step 7	set ipv6 next-hop <i>address1</i> Example: switch(config-route-map)# set ipv6 next-hop 2001:0DB8::1	Sets the IPv6 next-hop address for policy-based routing.
Step 8	(Optional) set interface null0 Example: switch(config-route-map)# set interface null0	Sets the interface that is used for routing. Use the null0 interface to drop packets.
Step 9	(Optional) copy running-config startup-config Example: switch(config-route-map)# copy running-config startup-config	Saves this configuration change.

Verifying the Policy-Based Redirect Configuration

To display the policy-based redirect configuration information, perform one of the following tasks:

Command	Purpose
show [ip ipv6] policy [name]	Displays information about an IPv4 or IPv6 policy.
show route-map [name] pbr-statistics	Displays policy statistics.

Use the **route-map map-name pbr-statistics** command to enable policy statistics. Use the **clear route-map map-name pbr-statistics** command to clear these policy statistics.

Configuration Example for Policy-Based Redirect

Perform the following configuration on all tenant VTEPs, excluding the service VTEP.

```
feature pbr

ipv6 access-list IPV6_App_group_1
10 permit ipv6 any 2001:10:1:1::0/64

ip access-list IPV4_App_group_1
10 permit ip any 10.1.1.0/24

ipv6 access-list IPV6_App_group_2
10 permit ipv6 any 2001:20:1:1::0/64

ip access-list IPV4_App_group_2
10 permit ip any 20.1.1.0/24

route-map IPV6_PBR_Appgroup1 permit 10
  match ipv6 address IPV6_App_group_2
  set ipv6 next-hop 2001:100:1:1::20 (next hop is that of the firewall)

route-map IPV4_PBR_Appgroup1 permit 10
  match ip address IPV4_App_group_2
  set ip next-hop 10.100.1.20 (next hop is that of the firewall)

route-map IPV6_PBR_Appgroup2 permit 10
  match ipv6 address IPV6_App_group1
  set ipv6 next-hop 2001:100:1:1::20 (next hop is that of the firewall)

route-map IPV4_PBR_Appgroup2 permit 10
  match ip address IPV4_App_group_1
  set ip next-hop 10.100.1.20 (next hop is that of the firewall)

interface Vlan10
! tenant SVI appgroup 1
vrf member appgroup
ip address 10.1.1.1/24
no ip redirect
ipv6 address 2001:10:1:1::1/64
no ipv6 redirects
fabric forwarding mode anycast-gateway
ip policy route-map IPV4_PBR_Appgroup1
ipv6 policy route-map IPV6_PBR_Appgroup1
interface Vlan20
! tenant SVI appgroup 2
vrf member appgroup
ip address 20.1.1.1/24
no ip redirect
ipv6 address 2001:20:1:1::1/64
no ipv6 redirects
fabric forwarding mode anycast-gateway
```

```
ip policy route-map IPV4_PBR_Appgroup2
ipv6 policy route-map IPV6_PBR_Appgroup2
```

On the service VTEP, the PBR policy is applied on the tenant VRF SVI. This ensures the traffic post decapsulation will be redirected to firewall.

```
feature pbr
```

```
ipv6 access-list IPV6_App_group_1
10 permit ipv6 any 2001:10:1:1::0/64
```

```
ip access-list IPV4_App_group_1
10 permit ip any 10.1.1.0/24
```

```
ipv6 access-list IPV6_App_group_2
10 permit ipv6 any 2001:20:1:1::0/64
```

```
ip access-list IPV4_App_group_2
10 permit ip any 20.1.1.0/24
```

```
route-map IPV6_PBR_Appgroup1 permit 10
  match ipv6 address IPV6_App_group_2
  set ipv6 next-hop 2001:100:1:1::20 (next hop is that of the firewall)
```

```
route-map IPV6_PBR_Appgroup permit 20
  match ipv6 address IPV6_App_group1
  set ipv6 next-hop 2001:100:1:1::20 (next hop is that of the firewall)
```

```
route-map IPV4_PBR_Appgroup permit 10
  match ip address IPV4_App_group_2
  set ip next-hop 10.100.1.20 (next hop is that of the firewall)
```

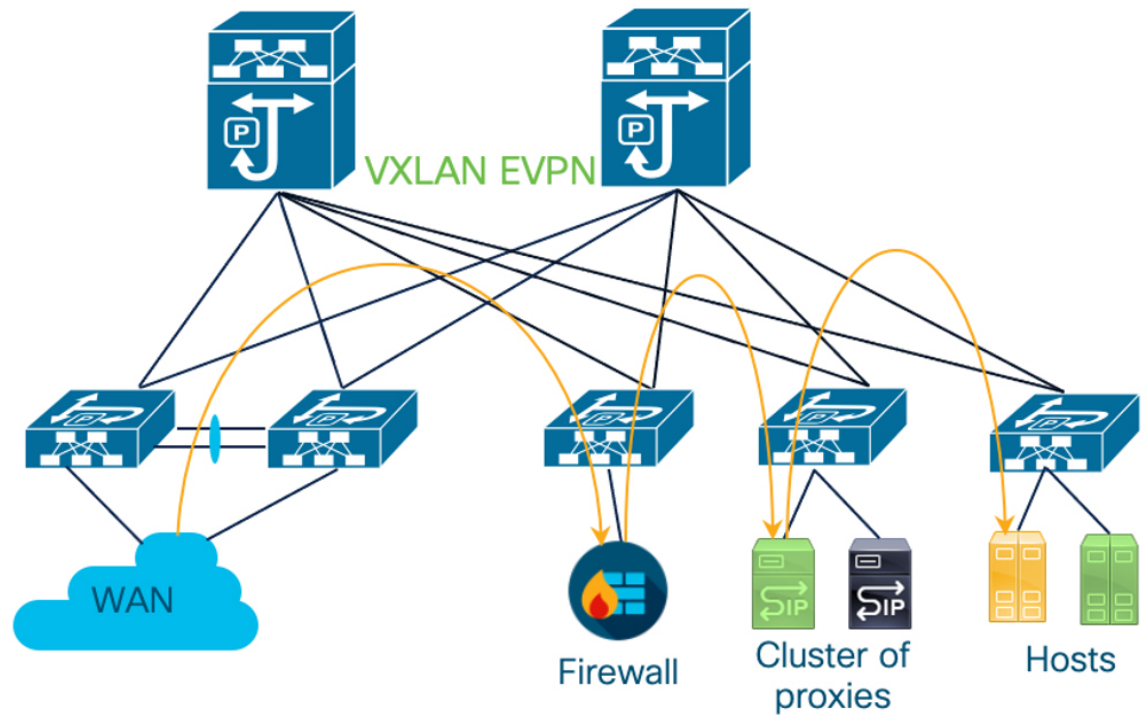
```
route-map IPV4_PBR_Appgroup permit 20
  match ip address IPV4_App_group_1
  set ip next-hop 10.100.1.20 (next hop is that of the firewall)
```

```
interface vlan1000
!L3VNI SVI for Tenant VRF
vrf member appgroup
ip forward
ipv6 forward
ipv6 ipv6 address use-link-local-only
ip policy route-map IPV4_PBR_Appgroup
ipv6 policy route-map IPV6_PBR_Appgroup
```

Enhanced-Policy Based Redirect (ePBR)

VXLAN PBR as a solution to selectively redirect traffic can only cater to simple traffic redirection requirements. For more complex use cases like service chaining, symmetric load-balancing, or tracking health of service appliances, usage of PBR becomes difficult. The challenge with service chaining using PBR is that it requires the user to create unique policies per node and manage the redirection rules manually across all the nodes in the chain. Also, given the stateful nature of the service nodes, the PBR rules must ensure symmetry for the reverse traffic, and this adds additional complexity to the configuration and management of the PBR policies.

Enhanced Policy-Based Redirect (ePBR) provides a comprehensive solution to insert service nodes, selectively redirect and load-balance traffic. ePBR provides a simplified workflow to create traffic chains and load-balancing rules along with providing options for probing/monitoring the health of service appliances and taking corrective action in the event of failure. ePBR is supported in both single and multi-site VXLAN EVPN deployments.



In this Figure, selective traffic originating from WAN is chained to a firewall and then the traffic is load-balanced across a cluster of proxies before forwarding toward the destination hosts. ePBR ensures symmetry is maintained for a given flow by making sure that traffic in both forward and reverse direction is redirected to the same service endpoint in the cluster of TCP proxies.

For more detailed information, guidelines and configuration examples on ePBR, see [Cisco Nexus 9000 Series NX-OS ePBR Configuration Guide](#) and [Layer 4 to Layer 7 Service Redirection with Enhanced Policy-Based Redirect White Paper](#).

