VXLAN Policy Based Routing (PBR)

Presenter:

Daryl Wan

Agenda

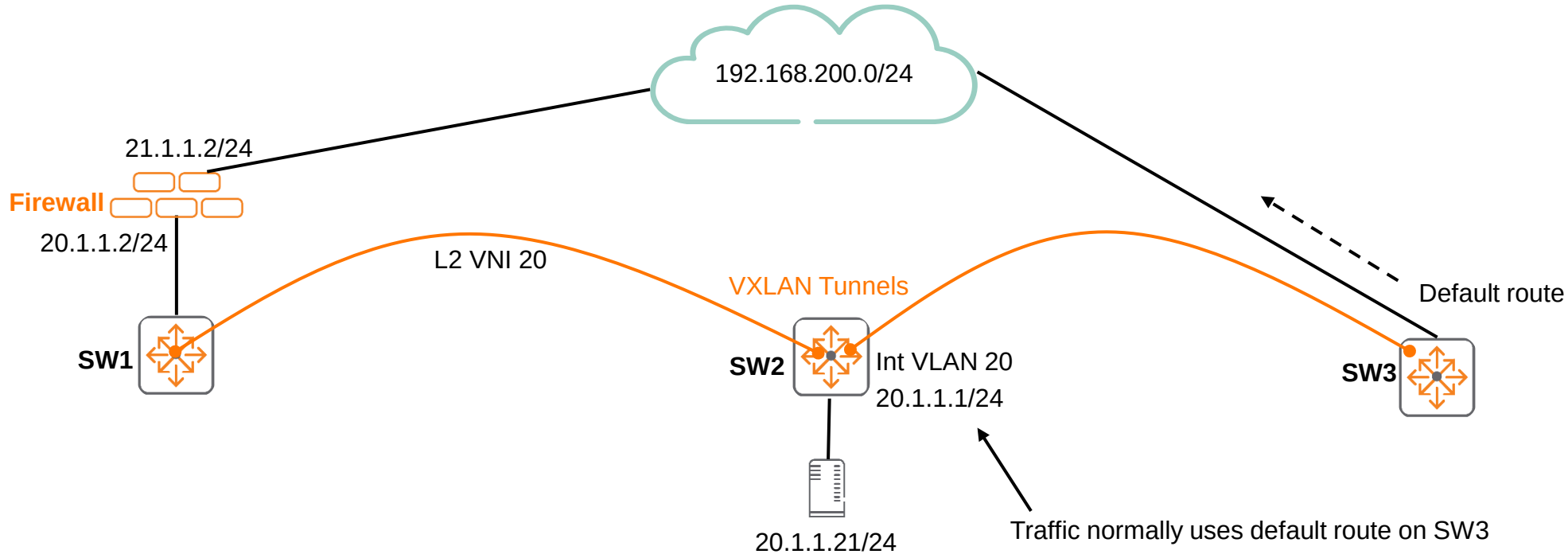
1. Overview
2. Use Cases
3. Details and Caveats
4. Configuration
5. Best Practices
6. Troubleshooting
7. Demo

Overview

VXLAN PBR (Policy Based Routing) Overview

- 10.9 adds PBR support for VXLAN deployments
- Allows L3 VTEPs to redirect traffic to desired next hop IP over an L2 VNI
  - Unidirectional PBR policy is applied inbound on an SVI
  - Another PBR policy could be used for return traffic or firewall uses NAT IP

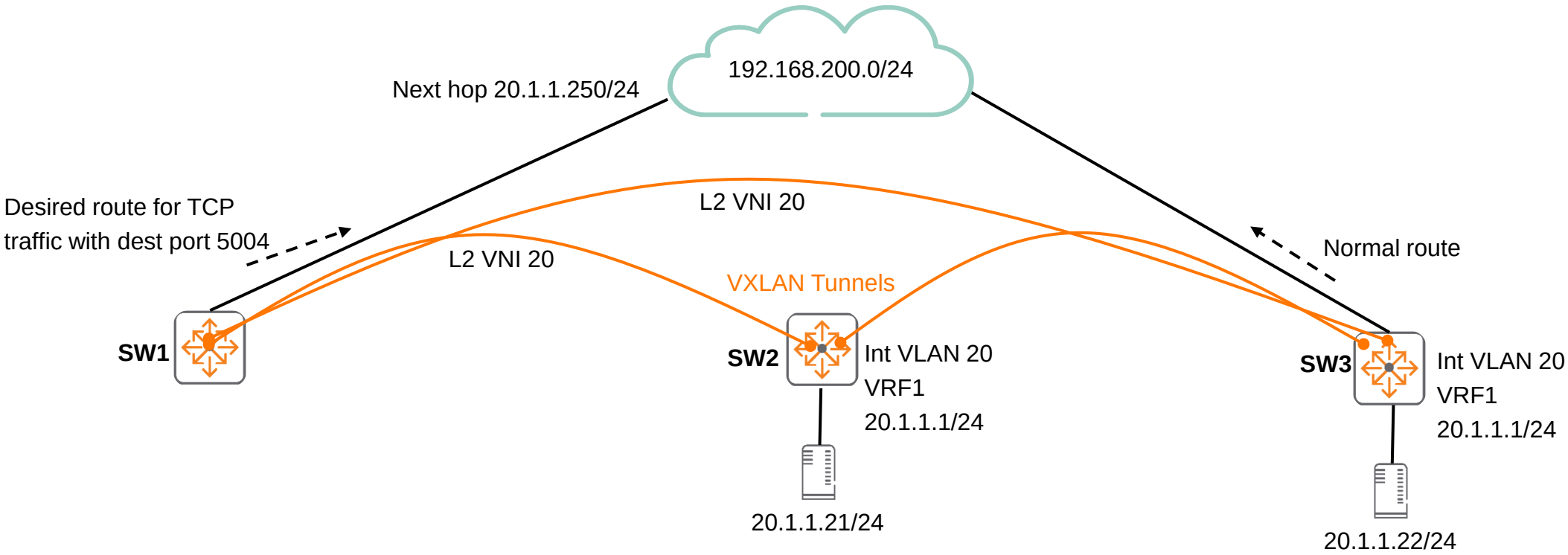
- Supported platforms:
  - 8325, 8360 and CX 10000



PBR policy applied inbound on Int VLAN 20
- Redirects traffic from host 20.1.1.21 to next hop 20.1.1.2 (via L2 VNI 20) on Firewall connected to SW1

Use Cases

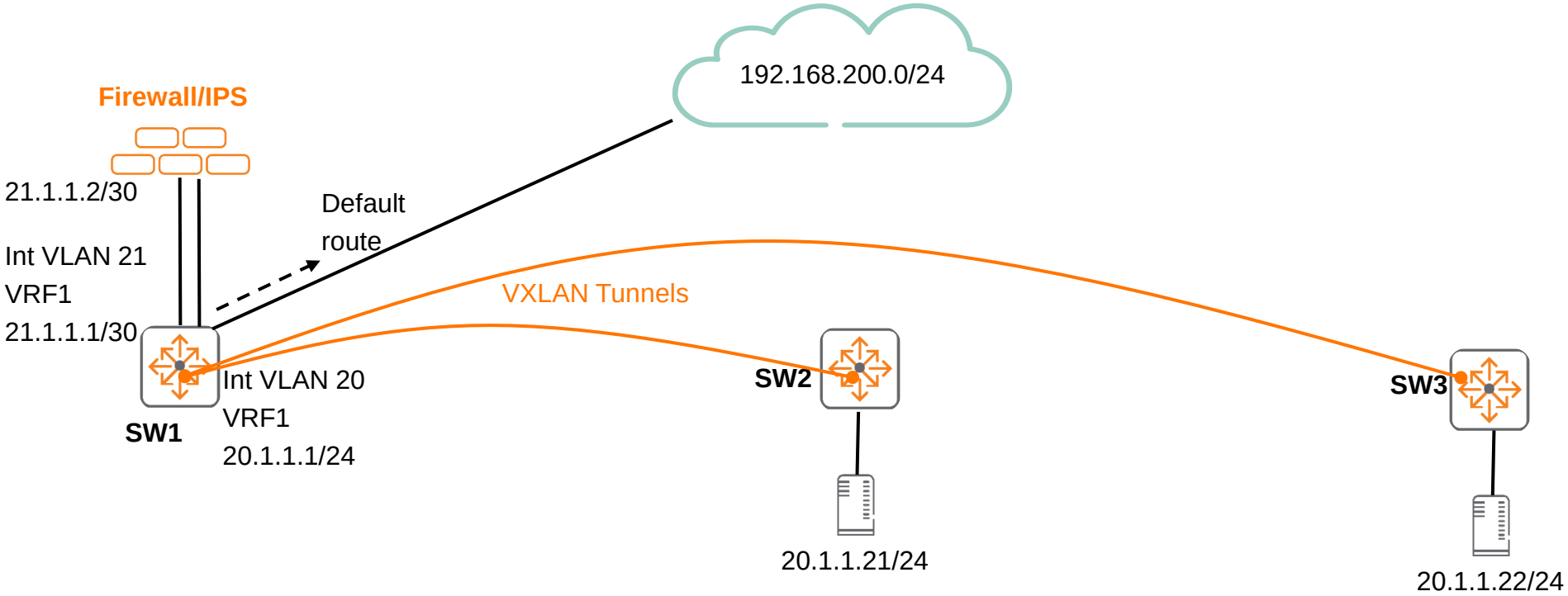
VXLAN PBR Use Case – Distributed L3 Gateways



- Traffic to 192.168.200.0/24 normally uses route on SW3
- Desire to use a different link to destination, only for certain traffic flows

- PBR policy applied inbound on **Int VLAN 20 of SW2 and SW3**
- Redirect **TCP traffic with destination port 5004** to next hop 20.1.1.250 connected to SW1
- All other traffic towards 192.168.200.0/24 continues to use SW3

VXLAN PBR Use Case – Centralized L3 Gateways



- Traffic to 192.168.200.0/24 normally uses default route on SW1
- Desire to inspect traffic from certain IPs

- PBR policy applied inbound on **Int VLAN 20 of SW1**
- Redirect **UDP traffic from both source IPs 20.1.1.21 and 20.1.1.22** to next hop 21.1.1.2 on Firewall/IPS connected to SW1
- All other source IPs towards 192.168.200.0/24 continues to use default route on SW1

Details

VXLAN PBR Details

- Requirements:
  - Applied on L3 VTEP
    - Either centralized L3 gateway or distributed L3 gateways
  - PBR policy is applied inbound on an SVI
  - PBR next hop IP is over L2 VNI or directly connected interface
  - Directly connected ARP entry required for next hop IP on L3 VTEP

- Supports
  - Class matches based on available parameters in AOS-CX,
    - e.g. UDP or TCP ports, source and destination IP ranges and DCSP values
  - Default and non-Default VRF
  - IPv4 and IPv6
  - VSX

- Specific to CX 10000
  - A security policy has to be created in PSM for traffic to be redirected via PBR
    - PBR action is done after Elba/DPU traffic inspection
  - If traffic is not allowed via security policy in PSM, it is dropped, therefore cannot be redirected to PBR

PBR Caveats

```
class ip pbr-class
    90 match any 17.181.0.0/255.255.0.0 any count
!
pbr-action-list pbr-al
    10 nexthop 172.16.200.200
    20 default-nexthop 172.16.250.250
```

- PBR does not support remote routers as next-hop or default-next-hop routers (a.k.a recursive)
- PBR only supports routers that are on a directly connected network

- nexthop
  - Sets the next hop for routing the packet
- default-nexthop
  - Sets the next hop for routing the packet when there is no explicit route for its destination
  - Overrides a system default route if already configured and also applies if there is no system default route

VXLAN PBR Caveats

```
class ip pbr-class
    90 match any 17.181.0.0/255.255.0.0 any count
!
pbr-action-list pbr-al
    10 nexthop 172.16.200.200
    20 default-nexthop 172.16.250.250
```

- Next-hop and default-nexthop entries in the action list are continuously monitored for reachability
  - Probing occurs every 5 seconds
  - If none of the next hops are reachable, the routing table is utilized
- If the desire is to drop traffic when next hops are unreachable, add “interface null”

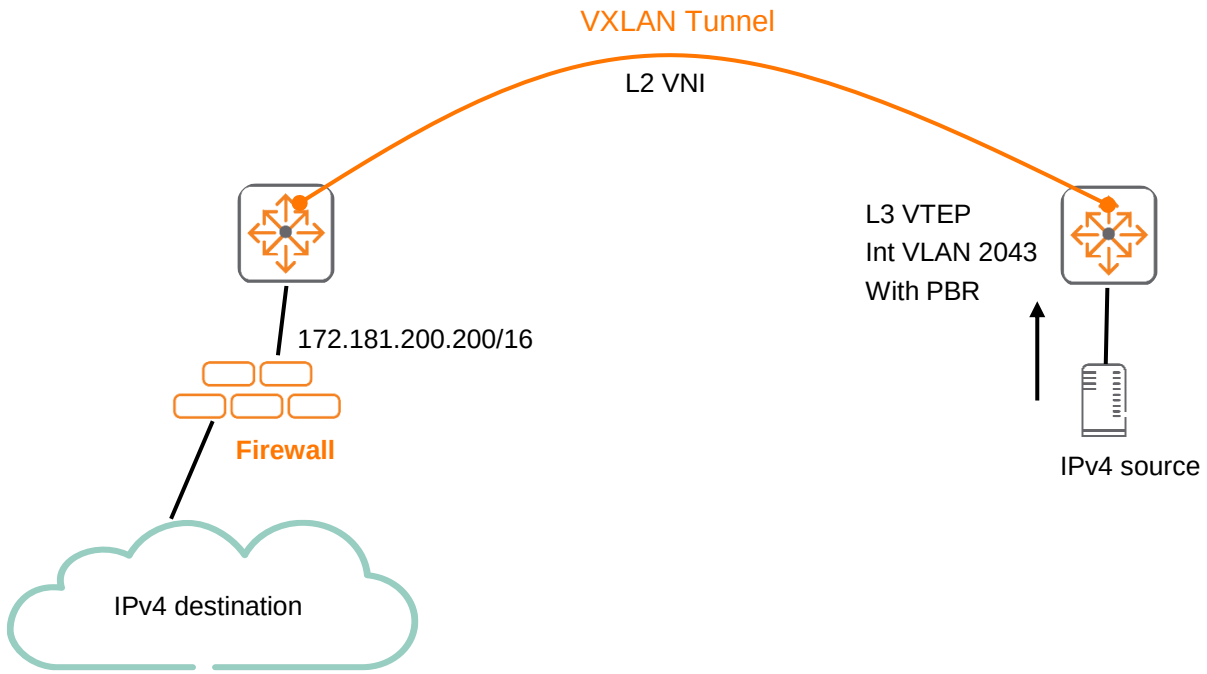
```
class ip pbr-class
    90 match any 17.181.0.0/255.255.0.0 any count
!
pbr-action-list pbr-al
    10 nexthop 172.16.200.200
    20 default-nexthop 172.16.250.250
    80 interface null
```

- Nexthop has to be learnt via ARP on directly connected L2 VNI, it cannot be learnt via remote EVPN ARP

Configuration

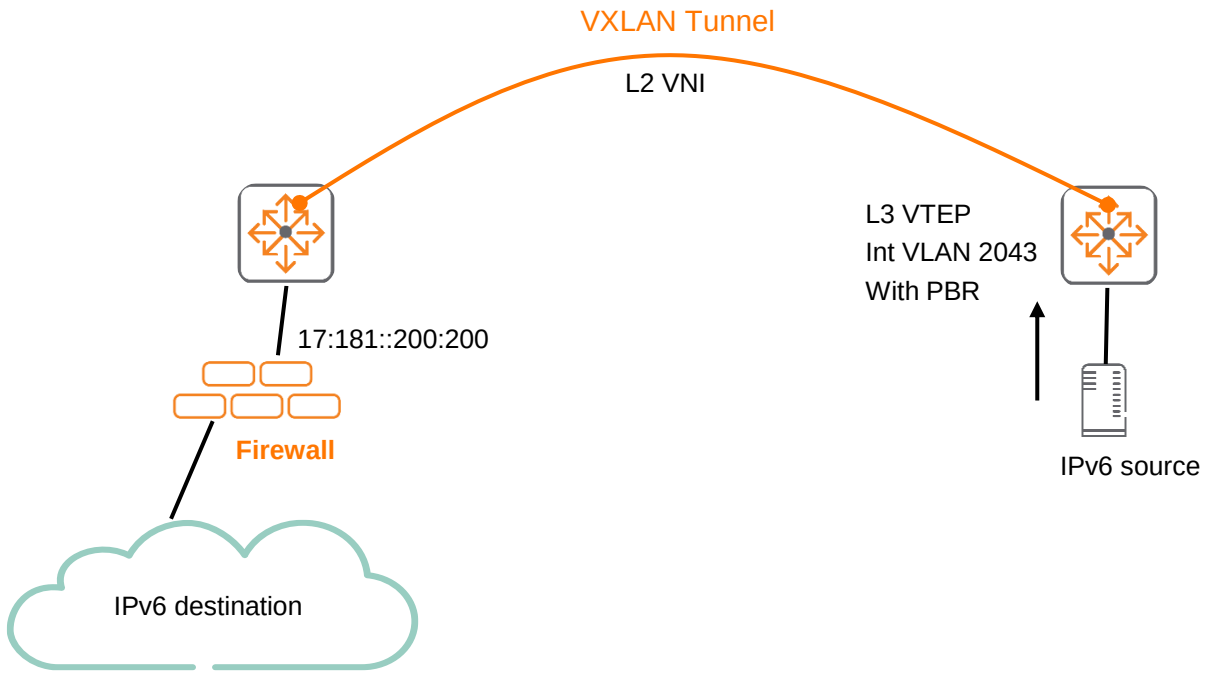
VXLAN PBR IPv4 Configuration Example

```
class ip pbr-class
    10 ignore any 17.181.1.101 any count
    20 ignore any 17.182.1.101 any count
    30 ignore any 17.185.1.101 any count
    40 ignore any 18.5.1.101 any count
    90 match any 17.181.0.0/255.255.0.0 any count
    100 match any 17.182.0.0/255.255.0.0 any count
    110 match any 17.185.0.0/255.255.0.0 any count
    120 match any 18.5.0.0/255.255.0.0 any count
!
pbr-action-list pbr-al
    10 nexthop 172.181.200.200
    20 default-nexthop 172.181.250.250
!
policy pbr-policy
    10 class ip pbr-class action pbr pbr-al
!
interface vlan 2043
    apply policy pbr-policy routed-in
    vrf attach vrf1
    ip address 17.181.0.1/16
    active-gateway ip mac 00:00:01:00:01:17
    active-gateway ip 17.181.0.254
    ip ospf 1 area 0.0.0.0
```



VXLAN PBR IPv6 Configuration Example

```
class ipv6 pbr-class-v6
    10 ignore any 17:181:1::2 any count
    20 ignore any 17:182:1::2 any count
    30 ignore any 17:185:1::2 any count
    40 ignore any 18:5:1::2 any count
    50 match any 17:181::0:0/32 any count
    60 match any 17:182::0:0/32 any count
    70 match any 17:185::0:0/32 any count
    80 match any 18:5::0:0/32 any count
!
pbr-action-list pbr-al-v6
    10 nexthop 17:181::200:200
    20 default-nexthop 17:181::250:250
!
policy pbr-policy
    20 class ipv6 pbr-class-v6 action pbr pbr-al-v6
!
interface vlan 2043
    apply policy pbr-policy routed-in
    vrf attach vrf1
    ipv6 address 17:181::1/32
    active-gateway ipv6 mac 00:00:01:00:01:17
    active-gateway ipv6 17:181:0::254
    active-gateway ipv6 fe80::2043
    ipv6 ospfv3 1 area 0.0.0.0
```



Best Practices

Best Practices

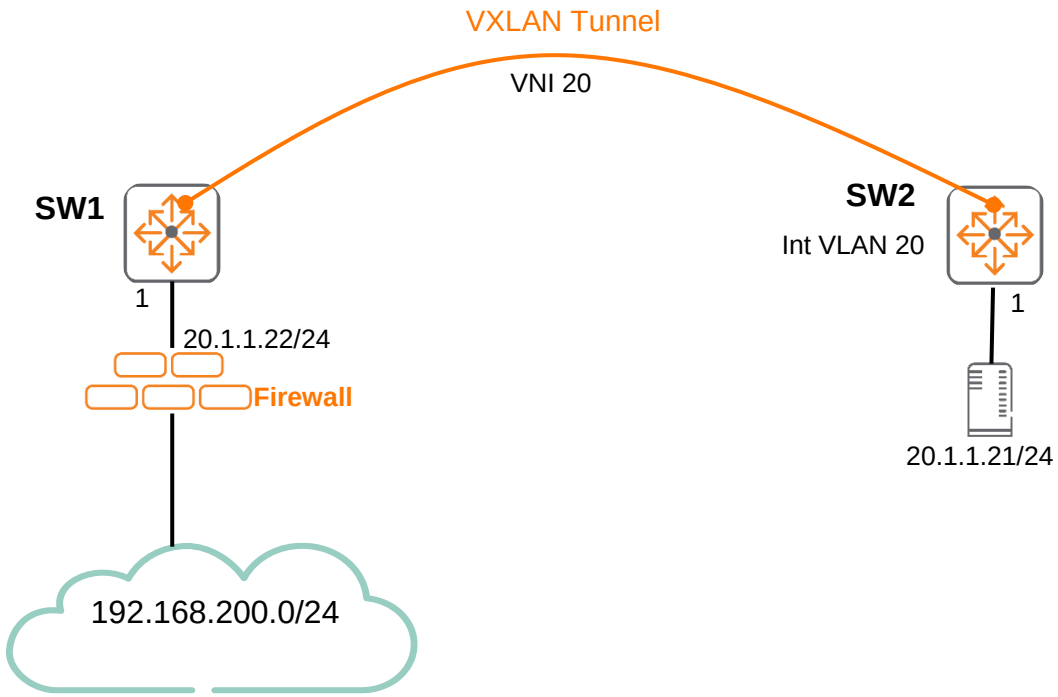
```
class ip pbr-class
    10 ignore any 17.181.1.101 any count
    20 ignore any 17.182.1.101 any count
    30 ignore any 17.185.1.101 any count
    40 ignore any 18.5.1.101 any count
    90 match any 17.181.0.0/255.255.0.0 any count
    100 match any 17.182.0.0/255.255.0.0 any count
    110 match any 17.185.0.0/255.255.0.0 any count
    120 match any 18.5.0.0/255.255.0.0 any count
```

- Add “count” to verify your class matches experience any hits

Troubleshooting

VXLAN PBR Troubleshooting

- Have a topology diagram ready
- Ensure IPs, interface details are included
- Check physical cabling and generate “show tech” when opening a TAC case
- Check network: show LLDP neighbor, ensure underlay network works using ping and traceroute between loopbacks and interfaces, fix any issues found



- Recommended troubleshooting flow

1. Check VXLAN PBR configs are correctly configured
2. Verify VXLAN PBR configs are correctly applied
3. Check next hop is reachable (ARP entry exists) via L2 VNI
4. Check hitcounts on class matches
5. Verify traffic is sent to next hop IP

1. Check VXLAN PBR configs are correctly configured

– Refer to config section for IPv4 and IPv6 sample configs

2. Verify VXLAN PBR configs are correctly applied

– Check that your expected VRF, SVI, policy, class, action-list, nexthop (active) appears as expected

– "show pbr summary" shows only active nexthops

```
SW2# sh pbr sum
VRF
      Port
                Policy
                          Class
                                    PBR
                                           Sequence  Type              Nexthop
-------------------------------------------------------------------------------
VRF1
      vlan20
                pbr-policy
                          pbr-class
                                    pbr-al
                                                10  nexthop           20.1.1.22 (active)
```

– "show pbr interface" shows both active and inactive nexthops

```
SW2# sh pbr int vlan20
VRF
      Port
                Policy
                          Class
                                    PBR
                                           Sequence  Type              Nexthop
-------------------------------------------------------------------------------
VRF1
      vlan20
                pbr-policy
                          pbr-class
                                    pbr-al
                                                 5  nexthop           192.168.3.10
                                                10  nexthop           20.1.1.22 (active)
```

3. Check next hop is reachable (ARP entry exists) via L2 VNI

– Ensure desired nexthop ARP entry is not known via EVPN

```
SW2# sh arp evpn vrf VRF1

IPv4 Address     MAC                Port         Physical Port               State      VRF
-------------------------------------------------------------------------------------------
Total Number Of EVPN ARP Entries Listed: 0.
```

– ARP entry for nexthop has to be learnt via L2 VNI (directly connected) and not via EVPN

```
SW2# sh arp vrf VRF1

IPv4 Address     MAC                Port         Physical Port               State      VRF
-------------------------------------------------------------------------------------
20.1.1.2         00:50:56:8e:29:03  vlan20       vxlan1(192.168.0.6)         permanent  VRF1
20.1.1.22        00:50:56:8e:3f:cc  vlan20       vxlan1(192.168.0.6)         permanent  VRF1
20.1.1.21        00:50:56:8e:a6:95  vlan20       1/1/1                       reachable  VRF1

Total Number Of ARP Entries Listed: 3.
```

– If above conditions are not seen, resolve them before proceeding further

4. Check hitcounts on class matches

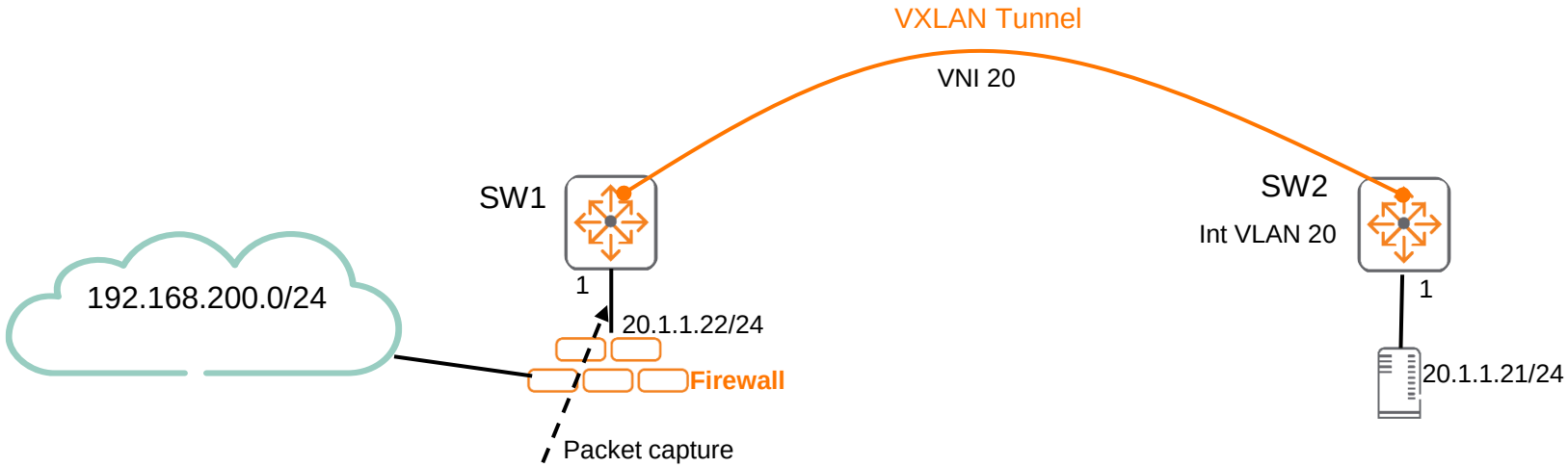
- Verify packet hitcounts on class matches
- If it hits, traffic will utilize PBR policy

```
SW2(config)# sh policy hitcounts pbr-policy
Statistics for Policy pbr-policy:

VRF default
interface vlan 2043-2044,2047,2121 (routed-in):
     Matched Packets  Configuration
10 class ip pbr-class action pbr pbr-al
                   0  10 ignore any 17.181.1.101 any count
             1139144  20 ignore any 17.182.1.101 any count
             2278288  30 ignore any 17.185.1.101 any count
                   0  40 ignore any 18.5.1.101 any count
                   0  90 match any 17.181.0.0/255.255.0.0 any count
             2754487  100 match any 17.182.0.0/255.255.0.0 any count
             3213497  110 match any 17.185.0.0/255.255.0.0 any count
             2754402  120 match any 18.5.0.0/255.255.0.0 any count
20 class ipv6 pbr-class-v6 action pbr pbr-al-v6
                   0  10 ignore any 17:181:1::2 any count
             1163876  20 ignore any 17:182:1::2 any count
             1163876  30 ignore any 17:185:1::2 any count
                   0  40 ignore any 18:5:1::2 any count
                   0  50 match any 17:181::0:0/32 any count
             2321754  60 match any 17:182::0:0/32 any count
             4179051  70 match any 17:185::0:0/32 any count
             1393001  80 match any 18:5::0:0/32 any count
```

5. Verify traffic is sent to next hop IP

– Packet captures (port mirror) might be required to check if traffic is sent to next hop IP



– Config to mirror traffic

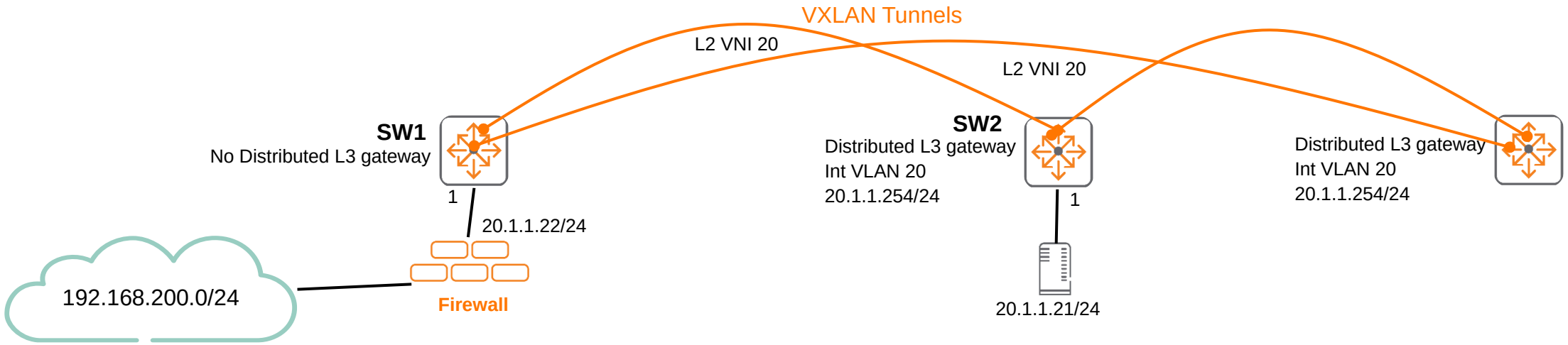
```
mirror session 1
    enable
    destination interface 1/1/40
    source interface 1/1/51 both
```

2	0.000116	20.1.1.220	20.1.1.2	ICMP	111 Destination unreachable (Network unreachable)
19	10.979978	20.1.1.21	192.168.200.10	ICMP	74 Echo (ping) request id=0x0001, seq=840/18435, t
27	15.562575	20.1.1.21	192.168.200.10	ICMP	74 Echo (ping) request id=0x0001, seq=841/18691, t
38	20.564799	20.1.1.21	192.168.200.10	ICMP	74 Echo (ping) request id=0x0001, seq=842/18947, t
42	23.808591	20.1.1.220	20.1.1.2	ICMP	111 Destination unreachable (Network unreachable)
45	24.814537	20.1.1.220	20.1.1.2	ICMP	111 Destination unreachable (Network unreachable)
6	1.913537	fd00:192:168:20::23	ff02::1:ff00:1	ICMPv6	86 Neighbor Solicitation for fd00:192:168:20::1 fro
7	2.474589	fd00:192:168:20::23	ff02::1:ff00:1	ICMPv6	86 Neighbor Solicitation for fd00:192:168:20::1 fro
8	3.478320	fd00:192:168:20::23	ff02::1:ff00:1	ICMPv6	86 Neighbor Solicitation for fd00:192:168:20::1 fro

```
> Frame 27: 74 bytes on wire (592 bits), 74 bytes captured (592 bits) on interface \Device\NPF_{E417155C-6750-4480-B8C6-9C1CFC
> Ethernet II, Src: ArubaaHe_bb:41:00 (90:20:c2:bb:41:00), Dst: VMware_8e:3f:cc (00:50:56:8e:3f:cc)
> Internet Protocol Version 4, Src: 20.1.1.21, Dst: 192.168.200.10
> Internet Control Message Protocol
```

Demo

VXLAN PBR Demo



- Demo flow
  - Generate traffic from 20.1.1.21 to 192.168.200.20 (routed out SW1 towards non firewall interface)
  - Show wireshark on firewall before VXLAN PBR
  - Enable PBR on SW2 (traffic towards 192.168.200.20 should now have firewall 20.1.1.22 as next hop)
  - Show wireshark on firewall after VXLAN PBR

- Note:
  - If SW1 is a distributed L3 gateway with SVI20, ARP entry for 20.1.1.22 on SW2 will be learnt using EVPN ARP
  - SW1 should not have SVI20 so that 20.1.1.22 is learnt via directly connected L2 VNI on SW2

aruba
a Hewlett Packard Enterprise company

Thank you

daryl.wan@hpe.com