Technical Guide

Virtual Extensible LAN (VXLAN)

Feature Overview and Configuration Guide

Introduction

VXLAN tunnels carry Ethernet frames over an IP network core. You can join together two or more Layer 2 networks, to form a single Layer 2 broadcast domain, by establishing a full mesh of VXLAN tunnels between them.

VXLAN allows you to create a Layer 2 network that spans a wide area without needing to configure a VLAN for it through the network core. VXLAN offers the following advantages over a more traditional end-to-end VLAN design:

- Theoretically, up to 16 million separate VXLAN Layer 2 networks can coexist in the same network (compared to only 4094 Layer 2 networks as in the a more traditional end-to-end VLAN design).
- Network devices upstream from the VXLAN Tunnel End Points (VTEPs) do not need to maintain MAC entries for end devices. In contrast, in a traditional, end-to-end, VLAN design, core switches typically have MAC address table entries for every device in the network.

This guide describes the current AlliedWare Plus VXLAN implementation.

Contents

Products and software version that apply to this guide

This guide applies to AlliedWare Plus™ products that support VXLAN, running software version 5.5.1-2.1 or later. From this release, VXLAN is supported on the x530 Series, x950 Series, and SBx908GEN2, however the capabilities and limitations of the VXLAN feature are different between the x530 Series and the SBx908GEN2 / x950 Series. For more information, see Limitations on page 5.

Glossary

Acronym	Description
VNI	VXLAN Network Identifier. A 24-bit ID number contained in the VXLAN packet header. As the name suggests, this indicates which Layer 2 network the inner frame belongs to.
VTEP	VXLAN Tunnel End Point. A device that performs VXLAN encapsulation and decapsulation.

Supported features

AlliedWare Plus supports static VXLAN. This means that each remote VXLAN Tunnel End Point (VTEP) is specified in the device configuration. AlliedWare Plus does not support BGP-EVPN.

For x530 and SBx908GEN2 Series:

- Each VXLAN Network Identifier (VNI) can be mapped to one VLAN on each VTEP. AlliedWare Plus does not support carrying multiple Layer 2 networks over a single VNI.
- VXLAN packets are always sent from the primary IPv4 address on the global loopback interface (lo). VXLAN packets cannot be sent to IPv6 addresses or from addresses on other interfaces.
- Ingress replication is used when flooding a frame out multiple tunnels. That means a separate VXLAN packet is sent to each remote VTEP. AlliedWare Plus does not support sending VXLAN packets to multicast addresses.
- The UDP destination port in VXLAN packets is 4789.

Note: From this point on there are some differences in capabilities and limitations between product Series:

- The UDP source port in VXLAN packets varies between different overlay traffic flows.
  - For x530 Series, the range of possible port numbers is 0-63, inclusive.
  - For x950 and SBx908 GEN2 Series, the range of possible port numbers is 0-65535 inclusive.

- The x530 does not support Equal Cost MultiPath (ECMP) routing to or from remote VTEPs. Specifically, at a particular point in time:
  - the VTEP should have only one best route to each remote VTEP.
  - VXLAN packets from a particular remote VTEP should all be received through the same switchport and VLAN and from the same MAC address.
  - The VTEP does not support sending or receiving VXLAN packets through port aggregations.
- ECMP is supported for known unicast traffic on the x950 and SBx908 GEN2 Series. However, these Series do not support ECMP for Broadcast, Unknown-unicast and Multicast (BUM) traffic.

A VLAN that is mapped to a VXLAN tunnel can be in any VRF.

The following features are **not supported** on a VLAN that is mapped to a VXLAN tunnel:

- Private VLAN
- Remote mirror VLAN
- Generic VLAN Registration Protocol (GVRP)
- Multicast routing
- IPv6 routing
- Policy-based routing
- Q-in-Q
- Ethernet Protection Switched Ring (EPSR)
- DHCP, IGMP, and MLD snooping
- Intermediate System to Intermediate System routing protocol (IS-IS)
- Microsoft® Network Load Balancing (NLB)
- x950/SBx908GEN2 Series do support stacking of 2 units (but not 3).
- x950/SBx908GEN2 Series do not support VLAN classifiers on VNI-mapped VLANs.
- x950/SBx908GEN2 Series do not support local port mirroring.

The configuration instructions specify that IGMP snooping and MLD snooping should be disabled on any VLAN that is mapped to a VXLAN tunnel. Currently, disabling those features will result in multicast traffic being flooded to the CPU. Therefore, we recommend that only low volume multicast protocols be used on a VLAN that is mapped to a VXLAN tunnel.

AlliedWare Plus does not support using VXLAN at the same time as the **platform vlan-stacking-tpid** command.

Limitations

Hardware design and resource constraints apply the following limitations to the current VXLAN implementation on AlliedWare Plus switches.

Underlay VLAN not checked before performing VXLAN tunnel termination

The switch will remove the VXLAN encapsulation of ingress traffic without checking which VLAN the VXLAN packets are arriving on. Consequently, the tunnel termination process will accept VXLAN packets in any of the following cases when normally they should be dropped:

- The VLAN is not in the global VRF.
- The VLAN does not have an IP address.
- The VLAN is not configured on the port.

You can restrict which switchports VXLAN packets can arrive on using a QoS policy map as described in the section: Configuring VXLAN, Step 9 on page 9. Note that QoS policies are applied after VXLAN tunnel termination, as described in Interactions with ACLs and QoS on page 11.

However, if multiple VLANs are assigned to the same switchport, then you cannot prevent VXLAN packets arriving on an unexpected VLAN. Therefore, we recommend you do not have untrusted network traffic running over the same switchport as VXLAN packets.

VXLAN encapsulated frames may be egressed with VLAN tagged passenger frames

When normal network traffic passes through a tunnel, there isn't a VLAN tag in the inner frame. However, it is possible to get an overlay VLAN tag when a frame with multiple VLAN tags is sent through the tunnel.

- On x950 and SBx908NG Series switches, a double-tagged packet will be encapsulated and sent with only the outer overlay VLAN tag removed; the original inner VLAN tag will still be present in the passenger packet.
- On x530 Series switches, if three overlay VLAN tags are present, the passenger packet will still contain the inner-most VLAN tag.

In both cases, an AlliedWare Plus switch acting as a VTEP that subsequently receives such a tagged passenger packet, will drop the packet on ingress, but other vendors devices may behave unexpectedly.

Resource usage

Device limit and current usage can be seen at the top of **show vxlan** command output:

```
awplus#show vxlan
Resource Usage:
---------------
VTEP Usage                          3/127
VNI Usage                           5/1000
VTEP multiplied by VNI Usage        15/8192
Hardware Virtual Port Usage         128/16384
```

VTEP usage

The VXLAN Tunnel End Point

- On x950 and SBx908 GEN2 Series switches -**127** VTEPs.

  This limit is reduced for every ECMP uplink and aggregator member port used. For example, you can only have 63 VTEPs if there are 2 ECMP uplinks which are using switchports. This will be reduced to 31 VTEPs if those ECMP uplinks are using aggregators with 2 member switchports.

- On x530 Series switches - none enforced

  We recommend 20 VTEPs and a maximum of 40 VTEPs for standard Ethernet frames. We also recommend 4 VTEPs and a maximum of 8 VTEPs when using Jumbo Ethernet frames. Exceeding these values can cause BUM traffic to not be replicated to all VTEPs because the port buffers are exceeded. It is also limited by Hardware Virtual Ports, as explained below.

VNI usage

The VXLAN network identifier (VNI)—used to uniquely identify the VXLAN

- On x950 and SBx908 GEN2 Series switches -**1000** VNIs
- On x530 Series switches, there is no strict enforcement. However, this will be limited by Hardware Virtual Ports (as described below).

VTEP multiplied by VNI usage

- On x950 and SBx908 GEN2 Series switches - **8192**
- On x530 Series switches - none enforced

Hardware Virtual Port usage

The way this resource is used is different between the platforms.

- On x950 and SBx908 GEN2 Series switches - **16384** entries

  To support VXLAN, some hardware tables use up an entry for every single VLAN that is assigned to a port.

  For example, if 20 VLANs are assigned to one port then 20 entries will be used. If those same VLANs are assigned to a different port then another 20 entries will be used. This means that not as many VLANs can be configured and assigned to ports compared to what might be considered normal without **platform l3-vxlan** enabled. We recommend not exceeding 16000 e.g. 4000 VLANs assigned to 4 ports or 400 VLANs assigned to 10 ports.

- On x530 Series switches - **896** entries

  To support VXLAN, hardware tables use an entry per VNI per VTEP per ECMP uplink nexthop.

Configuring VXLAN

This section provides some basic configuration steps.

1. **Configure the hardware for VXLAN support**

Enter the following commands on the VTEPs (the switches performing VXLAN encapsulation and decapsulation).

For SBx908GEN2/x950 devices, enable VXLAN with the following command:

```
awplus(config)# platform l3-vxlan enable
```

For x530 devices use the following commands:

```
awplus(config)# platform acls-to-vlanclassifiers {more-vlan-
classifiers|half-and-half|more-acls} tunnel-support
```

Note: The parameter **tunnel-support** is important, it must be included to use VXLAN. You can choose a balance of VLAN classifiers and ACLs that suits you.

```
awplus(config)# platform inter-chip-header 4-word
awplus(config)# platform jumboframe
```

Save the configuration and reboot the switch

2. **Configure VLANs, IP addresses, and routing for the core network**

Configure all switches in the core network to provide IPv4 connectivity between the VTEPs.

- The core network must be in the global VRF.
- Each VTEP must have an IPv4 address on the global loopback interface (lo).

3. **Adjust Maximum Transmission/Receive Unit settings throughout the core network**

The core network needs to allow larger-than-normal packet sizes to account for VXLAN encapsulation overhead. In a typical network, you should set the maximum packet size in the core to at least 1550 bytes (50 bytes more than normal).

Set the Maximum Transmission Unit (MTU) for VLAN interfaces with the **mtu** command in interface configuration mode. That should be done for VLAN interfaces that may handle VXLAN packets (both on the VTEPs and on other core network devices). However, VLAN interfaces that are not part of the core network do not need to be adjusted.

```
awplus(config)# interface vlan100
awplus(config-if)# mtu 1550
```

You also need to increase the Maximum Receive Unit (MRU) for all switchports that are part of the core network. You have already done this for the VTEPs in step 1 when you entered the **platform jumboframe** command. However, you still need to configure any other core network devices.

The command to increase the MRU of switchports varies between products. Refer to the Switching Feature Overview and Configuration Guide, Support for Jumbo Frames section for an overview. Refer to the product's command reference to see what commands are available.

4. **Create edge VLANs**

- On each VTEP, create one or more VLANs that you want to join to VXLAN tunnels. You can place those VLANs in any VRF of your choosing.
- Assign those VLANs to switchports if required. It is also OK to have an edge VLAN that's not assigned to any switchports. (You might do that if you only want the VTEP to route packets in or out of the tunnel.)

5. **Disable IGMP/MLD Snooping for edge VLANs on the VTEPs**

AlliedWare Plus does not support IGMP snooping or MLD snooping on VLANs that are joined to VXLAN tunnels. This limitation only applies to the VTEP; it is OK to use IGMP/MLD snooping on other switches connected to the same VLAN. This limitation only applies to VLANs that will be joined to VXLAN tunnels (in a later step with the **map-access** command); it is OK to use IGMP/MLD snooping on other VLANs.

IGMP snooping is enabled by default so you should explicitly disable it where it is not supported. If the VLAN is in the global VRF then MLD snooping is also enabled by default and, again, you should explicitly disable it.

Those features can be disabled on a per-VLAN basis with these commands:

```
awplus(config)# interface vlan100
awplus(config-if)# no ip igmp snooping
awplus(config-if)# no ip mld snooping
```

Refer to the command reference manual for more information and alternative ways of enabling/disabling those features.

6. **Configure VXLAN source interface on VTEPs**

Enter the following commands on the VTEPs:

```
awplus(config)# nvo vxlan
awplus(config-vxlan)# source-interface lo
```

These commands configure the global loopback interface (lo) as the source interface for VXLAN tunnels. This means the following things:

- When the switch sends a VXLAN packet, the source IP address will be the primary IPv4 address on the loopback interface.
- The switch expects incoming VXLAN packets to be addressed to the primary IPv4 address on the loopback interface.

Currently, only the global loopback interface can be the source interface.

7. **Configure the remote VTEPs**

Each VTEP needs to be given information about the other VTEPs it will be communicating with. Use the following command to specify each remote VTEP:

```
awplus(config-vxlan)# remote-vtep <vtep-name> <vtep-address>
```

Give each remote VTEP a unique name. The **flood-list** command (which is part of the next configuration step) will refer back to this name. VXLAN packets will be sent to (and received from) the IP address specified in this command. This must be the primary IPv4 address on the remote VTEP's source interface.

8. **Join VLANs to VXLAN Network Identifiers (VNIs)**

Use the **map-access** command to link a particular VLAN to a particular VNI. You can enter the command multiple times to specify multiple mappings.

If you want all of the remote VTEPs to be part of a Layer 2 network then enter the command as below. If, in the future, you create more remote VTEPs (with the **remote-vtep** command) then they will automatically become part of the Layer 2 network.

```
awplus(config-vxlan)# map-access vlan <1-4094> vni <1-16777215>
```

Alternatively, you can select which remote VTEPs will participate in the Layer 2 network with a flood list. Each flood list command can contain the names of any number of remote VTEPs.

```
awplus(config-vxlan)# flood-list <list-name> [vteps <vtep-name>...]
awplus(config-vxlan)# map-access vlan <1-4094> vni <1-16777215> flood-
list <list-name>
```

9. **Add QoS policies to mitigate VXLAN limitations**

We recommend you use QoS policies to mitigate the limitation described in the following section:

- Underlay VLAN not checked before performing VXLAN tunnel termination on page 5.

In the example configuration below, VXLAN packets are only expected to arrive on port1.0.1. If a VXLAN packet arrives on another switchport and it matches a configured tunnel, then the inner frame will be dropped by the ACL. However, these rules to not prevent VXLAN packets arriving on an unexpected VLAN through port1.0.1. Therefore, we recommend you do not have untrusted non-VXLAN traffic arriving on the same port as VXLAN packets.

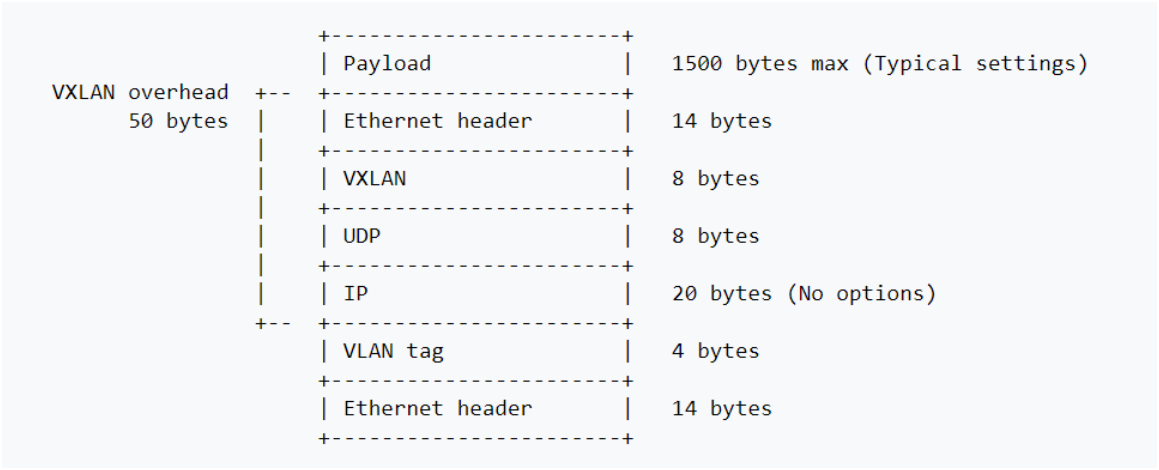
Table 1: Example QoS policy configuration for VXLAN

```
awplus(config)#access-list hardware ACL_DROP_VXLAN
awplus(config-ip-hw-acl)#deny mac any any tunnel-terminated
awplus(config-ip-hw-acl)#exit

awplus(config)#! Ports that VXLAN packets are not expected to arrive on.
awplus(config)#! This port range should not include stacking links.
awplus(config)#interface port1.0.2-port1.0.26
awplus(config-if)#access-group ACL_DROP_VXLAN
awplus(config-if)#exit
```

Maximum Transmission/Receive Unit

The core network needs to allow larger-than-normal packet sizes to account for the extra bytes added by VXLAN encapsulation. Refer to the section: Configuring VXLAN, Adjust Maximum Transmission/Receive Unit settings throughout the core network on page 7 for the commands to do this.

The diagram below shows that VXLAN adds 50 bytes of overhead when there is no overlay VLAN tag. (AlliedWare Plus never adds an overlay VLAN tag.) Therefore, the maximum VXLAN packet size in a typical network is 1550 bytes.



Interactions with ACLs and QoS

For the x530

Hardware ACLs and QoS policies are applied to ingress traffic after VXLAN tunnel termination. Therefore, when frames arrive through a VXLAN tunnel, the ACLs/QoS see only the inner frame without any VXLAN encapsulation.

Also, the tunnel termination process assigns the inner frame to the VLAN that the tunnel is mapped to. So, for example, if you configure **map-access vlan 20 vni 5000** then the class-map command **match vlan 20** will match frames arriving through VXLAN tunnels with VNI 5000.

If an incoming VXLAN packet does not match any tunnel, then the tunnel termination process will ignore it. In that case, ACLs/QoS see the entire frame including the VXLAN encapsulation.

For the SBx908GEN2

If you create an ACL which matches on both MAC address and VLAN tag, and apply it on an VXLAN uplink, then any received VXLAN packets will be processed in the following way:

- The portion of the ACL which matches against MAC address matches against the VXLAN passenger frame.
- The portion of the ACL which matches against VLAN matches against the underlay VLAN tag.

Monitoring VXLAN

The **show mac address-table** command displays MAC address tables. The highlighted entry below shows that frames addressed to e01a.ea3d.5188 will be forwarded into a VXLAN tunnel.

```
awplus#show mac address-table
VLAN port            mac            fwd
1    CPU          e01a.ea5b.dc5b   forward   static
102  port1.0.1    0013.1929.dd02   forward   dynamic
102  port1.0.1    0013.1929.dd43   forward   dynamic
102  CPU          e01a.ea5b.dc5b   forward   static
112  CPU          e01a.ea5b.dc5b   forward   static
1102 VXLAN        e01a.ea3d.5188   forward   dynamic
1102 CPU          e01a.ea5b.dc5b   forward   static
1102 port1.0.2    eccd.6d20.c0df   forward   dynamic
1112 CPU          e01a.ea5b.dc5b   forward   static
1122 CPU          e01a.ea5b.dc5b   forward   static
1132 CPU          e01a.ea5b.dc5b   forward   static
1142 CPU          e01a.ea5b.dc5b   forward   static
```

The **show vxlan** command displays: resource usage, remote VTEPs with their current underlay nexthop (if resolved), VNIs with associated VTEPs and member ports.

```
SBx908NG2#show vxlan
Resource Usage:
---------------
VTEP Usage                        3/127
VNI Usage                         5/1000
VTEP multiplied by VNI Usage      15/8192
Hardware Virtual Port Usage       128/16384

Remote VTEPs (2)
------------------
VTEP HOST_103 (192.168.1.103)
  VNIs (2): 2000, 2001
  Nexthops:
    IP Address      Mac Address      Interface        Port
    172.16.102.1    e01a.ea53.2c9d   vlan102          port1.0.1

VTEP HOST_109 (192.168.1.109)
  VNIs (2): 2000, 2001
  Nexthops:
    IP Address      Mac Address      Interface        Port
    172.16.102.1    e01a.ea53.2c9d   vlan102          port1.0.1

VXLANs (2)
------------
VNI 2000   Access VID 1102
  Local prefix 192.168.1.102/32
  Remote VTEPs (2): 192.168.1.103, 192.168.1.109
  Vlan Member Ports:
    Interface    Tagged?
    port1.0.2    Yes

VNI 2001   Access VID 1112
  Local prefix 192.168.1.102/32
  Remote VTEPs (2): 192.168.1.103, 192.168.1.109
  Vlan Member Ports:
    Interface    Tagged?
    port1.0.2    Yes
```

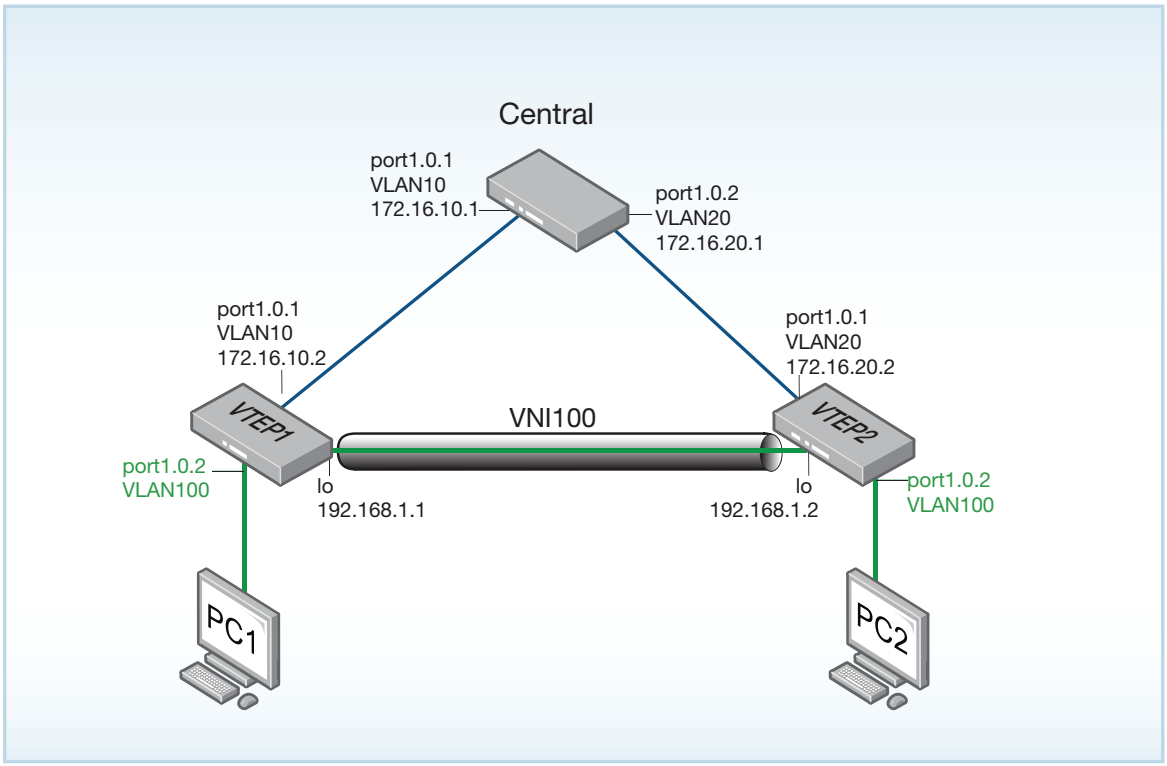
```
x530#show vxlan
Resource Usage:
---------------
VTEP Usage                        3
VNI Usage                         5
Hardware Virtual Port Usage       15/896

Remote VTEPs (2)
------------------
VTEP HOST_103 (192.168.1.103)
  VNIs (2): 2000, 2001
  Nexthops:
    IP Address      Mac Address      Interface        Port
    172.16.102.1    e01a.ea53.2c9d   vlan102          port1.0.1

VTEP HOST_109 (192.168.1.109)
  VNIs (2): 2000, 2001
  Nexthops:
    IP Address      Mac Address      Interface        Port
    172.16.102.1    e01a.ea53.2c9d   vlan102          port1.0.1

VXLANs (2)
------------
VNI 2000   Access VID 1102
  Local prefix 192.168.1.102/32
  Remote VTEPs (2): 192.168.1.103, 192.168.1.109
  Vlan Member Ports:
    Interface    Tagged?
    port1.0.2    Yes

VNI 2001   Access VID 1112
  Local prefix 192.168.1.102/32
  Remote VTEPs (2): 192.168.1.103, 192.168.1.109
  Vlan Member Ports:
    Interface    Tagged?
    port1.0.2    Yes
```

Configuration example

This simple network has a VXLAN tunnel (VNI100) between VTEP1 and VTEP2.

You can see that the VXLAN tunnel forms a single Layer 2 broadcast domain (VLAN100) between VTEP1 and VTEP2. The configuration for VTEP1, VTEP2, and Central is provided on the following pages.

Network diagram



VTEP1

```
platform jumboframe
platform acls-to-vlanclassifiers more-vlan-classifiers tunnel-support
platform inter-chip-header 4-word
!
mls qos enable
!
vlan database
 vlan 10,100 state enable
!
access-list hardware ACL_DROP_VXLAN
 deny mac any any tunnel-terminated
!
interface port1.0.1
 switchport
 switchport mode access
 switchport access vlan 10
!
interface port1.0.2
 switchport
 switchport mode access
 switchport access vlan 100
 access-group ACL_DROP_VXLAN
!
interface lo
 ip address 192.168.1.1/32
!
interface vlan10
 mtu 1550
 ip address 172.16.10.2/24
!
interface vlan100
 no ip igmp snooping
 no ipv6 mld snooping
!
router ospf
 network 172.16.10.0/24 area 0
 network 192.168.1.1/32 area 0
!
ping-poll 1
 ip 172.16.10.1
 active
!
nvo vxlan
 source-interface lo
 remote-vtep VTEP_2 192.168.1.2
 map-access vlan 100 vni 100
```

VTEP2

```
platform jumboframe
platform acls-to-vlanclassifiers more-vlan-classifiers tunnel-support
platform inter-chip-header 4-word
!
mls qos enable
!
vlan database
 vlan 20,100 state enable
!
access-list hardware ACL_DROP_VXLAN
 deny mac any any tunnel-terminated
!
interface port1.0.1
 switchport
 switchport mode access
 switchport access vlan 20
!
interface port1.0.2
 switchport
 switchport mode access
 switchport access vlan 100
 access-group ACL_DROP_VXLAN
!
interface lo
 ip address 192.168.1.2/32
!
interface vlan20
 mtu 1550
 ip address 172.16.20.2/24
!
interface vlan100
 no ip igmp snooping
 no ipv6 mld snooping
!
router ospf
 network 172.16.20.0/24 area 0
 network 192.168.1.2/32 area 0
!
ping-poll 1
 ip 172.16.20.1
 active
!
nvo vxlan
 source-interface lo
 remote-vtep VTEP_1 192.168.1.1
 map-access vlan 100 vni 100
```

Central

```
platform jumboframe
!
vlan database
 vlan 10,20 state enable
!
interface port1.0.1
 switchport
 switchport mode access
 switchport access vlan 10
!
interface port1.0.2
 switchport
 switchport mode access
 switchport access vlan 20
!
interface vlan10
 mtu 1550
 ip address 172.16.10.1/24
!
interface vlan20
 mtu 1550
 ip address 172.16.20.1/24
!
router ospf
 network 172.16.10.0/24 area 0
 network 172.16.20.0/24 area 0
```

C613-22128-00 REV C

Allied Telesis NETWORK SMARTER

North America Headquarters | 19800 North Creek Parkway | Suite 100 | Bothell | WA 98011 | USA | T: +1 800 424 4284 | F: +1 425 481 3895
Asia-Pacific Headquarters | 11 Tai Seng Link | Singapore | 534182 | T: +65 6383 3832 | F: +65 6383 3830
EMEA & CSA Operations | Incheonweg 7 | 1437 EK Rozenburg | The Netherlands | T: +31 20 7950020 | F: +31 20 7950021

alliedtelesis.com