Arista VXLAN Bridging Cheat Sheet

by sh-arista via cheatography.com/93954/cs/20677/

Virtual eXtensible LAN

L2 ethernet frames tunneled across an L3 infrastructure. Why? Extends L2 boundaries, supports multi-pathing and load distribution, is an open standard, and is transparent to applications.

https://tools.ietf.org/html/rfc7348

Terminology

VTI

VXLAN Tunnel Interface - switchport linked to a UDP socket responsible for the encap/decap of the VXLAN header; IP interface of the VTEP; VLAN to VNI mapping; VTEP flood list for BUM traffic.

VNI

VXLAN Network Identifier - 24-bit number mapped to a VLAN to identify a network segment in the tunnel.

VTEP

VXLAN Tunnel End Point - the entry/exit point for the VXLAN overlay network; can be physical or SW virtual switch.

VXLAN Bridging

End hosts are communicating within the same VLAN and no gateway is needed.

VXLAN Routing

End hosts are communicating between VLANs and a gateway is needed for routing.

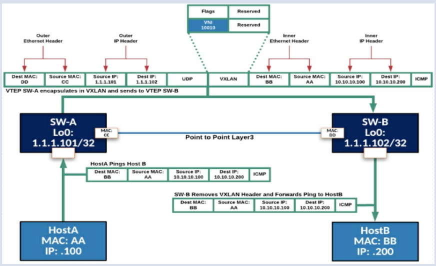
Configuration

```
SW-A (VTEP1)
Configure a loopback to serve as the L3 source
interface of the VXLAN Tunnel Interface (VTI):
interface loopback 1
  ip address 1.1.1.101/32
Create the VTI:
interface vxlan 1
Set the source interface to be the loopback just
created:
  vxlan source-interface loopback1
Set the destination UDP port (can be any unused UDP
port but needs to be consistent across all VTEPs:
```

Configuration (cont)

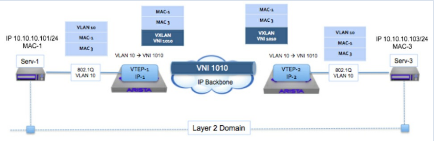
```
  vxlan udp-port 4789
Configure the VLAN to VNI mappings for any VLANs
that need to be extended:
  vxlan vlan 10 vni 10010
Configure the flood-set to include any VTEP IPs
that need to receive BUM traffic:
  vxlan flood vtep 1.1.1.102
Ensure routing is enabled for VXLAN to work:
ip routing
SW-B (VTEP2)
!
interface loopback 1
  ip address 1.1.1.102
!
interface vxlan 1
  vxlan source-interface loopback 1
  vxlan udp-port 4789
  vxlan vlan 10 vni 10010
  vxlan flood vtep 1.1.1.101
!
ip routing
!
```

Simple Topology and VXLAN Packet



Note the outer and inner header. There would be an ethernet outer header rewrite with every L3 hop in the L3 underlay. The outer IP header reflects the source and dest VTEP IPs. The inner header remains unchanged.

Packet Walk Through



The ingressing VTEP will map the VLAN to the VNI and encapsulate the packet with a VXLAN header with the VNI destined to the VTEP IP of the remote host found in "show vxlan address-table". Once the remote VTEP receives this packet, it decapsulates the packet and does the reverse VNI to VLAN mapping. The packet then switches per normal L2 (mac address table lookup).

MTU

```
vtep1#ping [VTEP IP] size 9214 df-bit
If using ECMP,
vtep1#ping [Uplink IP] size 9214 df-bit
```

The VXLAN header adds 50 bytes (54 bytes if outer L2 header includes dot1q tag), and the Do Not Fragment (DF) bit is set on the VXLAN encapsulated packet so ensure MTU is set correctly in the L3 underlay.

VXLAN Control Plane Options - in brief

IP Multicast	HER with Static Flood-Set	CloudVision eXchange (CVX)	EVPN

VXLAN provides the data-plane transport for any extended VLAN traffic. For control-plane traffic, there are several options based on scale, efficiencies, and other factors.

For BUM traffic, all of the above use HER, however the building of the flood lists, managing state, etc. will all be different depending on which option you choose.

See *Arista VXLAN Control Plane Options* for more details.

Troubleshooting

show interface vxlan 1

> Should be "up"; correctly reflect configured VLAN-to-VNI mappings; confirm control plane (multicast, HER, CVX, EVPN)

show vlan

> Ensure extended VLANs show active on the "Vx1" interface

Troubleshooting (cont)

show mac address-table

> The L2 forwarding table should show that mac addresses are either learned locally or from across the VXLAN overlay - "Vx1"; if we are not learning MACs from another VTEP confirm flood list and L3 reachability between VTEPs

show vxlan address-table

> Shows the VXLAN MAC info, including the Host MAC, remote VTEP IP, and MAC moves.

show vxlan vtep

> Displays the remote VTEPs discovered by the local VTEP

show ip route

> All VTEP IPs should have L3 reachability (ping to confirm)

show vxlan counters software

> See block for details

tcpdump

> #bash tcpdump -nei <intf> port 4789

Optional Configuration

```
When all VTEPs carry same VLANs:
interface vxlan 1
  vxlan flood vtep <remote-vtep-ip> <remote-vtep-ip>
When VTEPs carry a subset of VLANs:
interface vxlan1
  vxlan vlan <X> flood vtep <remote-vtep-ip> <remote-vtep-ip>
    vxlan vlan <Y> flood vtep <remote-vtep-ip> <remote-vtep-ip>
VNI can be displayed or entered as dotted notation:
interface vxlan1
vxlan vni notation dotted
```

https://www.arista.com/en/um-eos/eos-section-22-3-vxlan-configuration

Things to Note

*Every VTEP's VTI IP address (vxlan source-interface loopback) needs to be reachable from every other VTEP. Advertise these in the underlay routing protocol and confirm pings sourced from this VTEP IP can reach all other VTEP IPs.

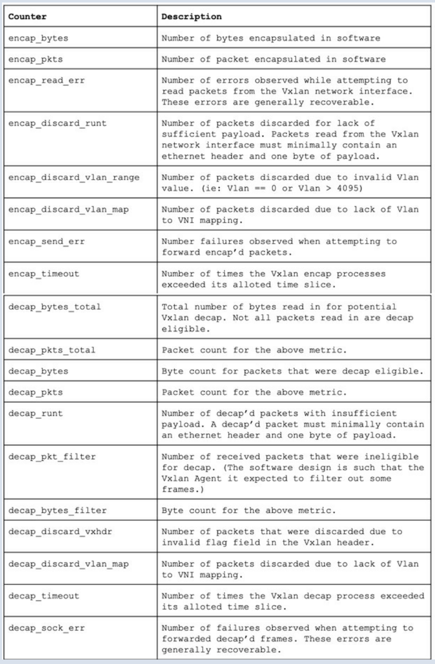
*The default UDP destination port is 4789. If this is changed, make sure it is changed on all the VTEPs.

*When using static HER, make sure that the flood lists match on all VTEPs within a VXLAN domain.

VXLAN Bridging + MLAG

Easy! Just mirror all VXLAN config to both MLAG peers. This provides for seamless failover should something happen to a peer. As both peers are presenting as one logical VTEP, they will share the same Loopback 1 IP address as well as VTI configuration.

show vxlan counters software



Counter	Description
encap_bytes	Number of bytes encapsulated in software
encap_pkts	Number of packet encapsulated in software
encap_read_err	Number of errors observed while attempting to read packets from the Vxlan network interface. These errors are generally recoverable.
encap_discard_runt	Number of packets discarded for lack of sufficient payload. Packets read from the Vxlan network interface must minimally contain an ethernet header and one byte of payload.
encap_discard_vlan_range	Number of packets discarded due to invalid Vlan value. (ie: Vlan == 0 or Vlan > 4095)
encap_discard_vlan_map	Number of packets discarded due to lack of Vlan to VNI mapping.
encap_send_err	Number failures observed when attempting to forward encap'd packets.
encap_timeout	Number of times the Vxlan encap processes exceeded its alloted time slice.
decap_bytes_total	Total number of bytes read in for potential Vxlan decap. Not all packets read in are decap eligible.
decap_pkts_total	Packet count for the above metric.
decap_bytes	Byte count for packets that were decap eligible.
decap_pkts	Packet count for the above metric.
decap_runt	Number of decap'd packets with insufficient payload. A decap'd packet must minimally contain an ethernet header and one byte of payload.
decap_pkt_filter	Number of received packets that were ineligible for decap. (The software design is such that the Vxlan Agent it expected to filter out some frames.)
decap_bytes_filter	Byte count for the above metric.
decap_discard_vxhdr	Number of packets that were discarded due to invalid flag field in the Vxlan header.
decap_discard_vlan_map	Number of packets discarded due to lack of Vlan to VNI mapping.
decap_timeout	Number of times the Vxlan decap process exceeded its alloted time slice.
decap_sock_err	Number of failures observed when attempting to forwarded decap'd frames. These errors are generally recoverable.

If you are seeing issues and have further questions on these counters, please reach out to Arista TAC:
https://www.arista.com/en/support/customer-support
Or search/post questions on the public forum:
https://eos.arista.com/