VXLAN Design and Deployment

Aftab Siddiqui
Internet Society
siddiqui@isoc.org
Fundamentals

• Virtual eXtensible LAN (VXLAN) – defined under RFC7348, co-authored by: - Cumulus Networks, Arista, Broadcom, Cisco, Red Hat, VMware and Intel

• VXLAN overlays a virtualized Layer 2 on top of a Layer 3 network allowing the extension of Layer 2.
  • “MAC in IP” Encapsulation
  • Layer 2 multi-point tunneling over IP UDP
  • 50Bytes Header

• Enables Layer 2 interconnection across Layer 3 boundaries
  • Transparent to the physical IP network
  • Provides Layer 2 scale across the Layer 3 IP fabric
Terminology

• **Virtual Tunnel End-point (VTEP).**
  - The VTEP acts as the entry point for connecting hosts into the VXLAN overlay network.
  - The task of the VTEP is to encap/decap with the appropriate VXLAN header.
  - It can be a software virtual switch or a physical switch.

• **Virtual Network Identifier (VNI)**
  - A 24-bit field added within the VXLAN header.
  - Identifies the Layer 2 segment of the encapsulated Ethernet frame
  - 16 million options as compare to 4000 in VLAN
Why VXLAN

- **Supported in Wide Variety of Platforms**
  - Broadcom: Trident II/II+, Tomahawk
  - Arista Networks, Cisco Systems, Dell, Juniper Networks
  - White Label: Bring Your Own Switch (BYOS) & Purchase Operating System (Cumulus, OcNOS & Others to Follow)
  - Widely Deployed in Many Networks

- **Supports Point to Point & Multi-Point**

- **UDP Header:**
  - Hashing Works Just Like Any Other UDP Traffic

- **Wide Vendor Support**
VXLAN Packet

Source: Cisco
VXLAN Packet

Frame 11: 209 bytes on wire (1672 bits), 209 bytes captured (1672 bits) on interface 0
Ethernet II, Src: Dell_01:4b:30 (14:18:77:01:4b:30), Dst: Dell_01:5d:30 (14:18:77:01:5d:30)
Internet Protocol Version 4, Src: 1.1.1.1, Dst: 2.2.2.2
User Datagram Protocol, Src Port: 37103 (37103), Dst Port: 4789 (4789)
  Source Port: 37103
  Destination Port: 4789
  Length: 175
  Checksum: 0x0000 (none)
  [Stream index: 0]
Virtual eXtensible Local Area Network
  Flags: 0x0000, VXLAN Network ID (VNI)
  Group Policy ID: 0
  VXLAN Network Identifier (VNI): 2000
  Reserved: 0
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.2
Internet Control Message Protocol
VXLAN – with QinQ
VXLAN Packet [QinQ]

**Before Encapsulation**
- Frame 2: 122 bytes on wire (976 bits), 122 bytes captured (976 bits) on interface 0
- 802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 600
- 802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 500
- Internet Protocol Version 4, Src: 10.10.10.2, Dst: 10.10.10.1
- Internet Control Message Protocol

**After Encapsulation**
- Frame 3: 168 bytes on wire (1344 bits), 168 bytes captured (1344 bits) on interface 0
- Internet Protocol Version 4, Src: 30.0.0.1, Dst: 30.0.0.2
- User Datagram Protocol, Src Port: 38190 (38190), Dst Port: 4789 (4789)
- Virtual eXtensible Local Area Network
- 802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 500
- Internet Protocol Version 4, Src: 10.10.10.1, Dst: 10.10.10.2
- Internet Control Message Protocol
Use Cases:

VXLAN – Data Center Interconnect (DCI)

• To interconnect Data Centers across geographically disperse sites
  • Layer 2 connectivity between sites, providing VM mobility between sites
  • Within the DC for server migration between PODs, for integrating new infrastructure
Use Cases:
VXLAN – Cloud Provider

• Provider looking to support multiple customers over a shared L3 infrastructure.
  • Wants the flexibility to deploy tenant resources across racks.
  • Layer 2 (VXLAN bridging) required to stitch the tenant’s resources/appliances together across racks.
  • VRF can be used to segregate tenant routing.
  • Fabric VTEP thus only required to provide layer 2 VXLAN bridging service
VXLAN – Not for everything

• Works fine:
  • Routing protocols
  • VLAN
  • QinQ
  • Multicast

• Doesn’t Work
  • LLDP (works locally only)
  • LACP (works locally only)
  • CoS Preservation
  • 802.1x
Summary

• **Operationally Simple Ethernet Line Services**
  • Eliminate Protocol Dependency & Complexity
  • Expand Vendor Choice

• **Practical & Economical Deployments**
  • Deploy Services Over Legacy Networks or IP Only Networks
  • Deploy Services Over Other Providers Networks

• **Future Use Cases:**
  • Remote TAP Transport
  • Hub and Spoke ELINE Services
  • More Mobile Management Applications to deliver Self Services
Any Questions?