VXLAN, Enhancements, and Network Integration

Apricot 2014 - Malaysia

Eddie Parra
Principal Engineer, Juniper Networks
Router Business Unit (RBU)
eparra@juniper.net

Legal Disclaimer: This statement of product direction sets forth Juniper Networks’ current intention, and is subject to change at any time without notice. No purchases are contingent upon Juniper Networks delivering any feature or functionality depicted on this statement.
## VARIOUS ENCAPSULATION METHODS

### Overlays
- **VXLAN**
  - Cumulus, Arista, Broadcom, Cisco, VMware, Citrix, Red Hat
- **NVGRE**
  - Microsoft, Arista, HP, Broadcom, Juniper
- **STT**
  - Nicira, Rackspace, eBay, Yahoo!
- **Geneve**
  - VMware, Microsoft, Red Hat, Intel

### Fabrics
- **TRILL**
  - Intel, Cisco, Brocade
- **IEEE 802.1aq**
  - Huawei, ALU
- **FabricPath**
  - Cisco
- **VCS**
  - Brocade
- **Qfabric**
  - Juniper

### Other
- **GRE**
  - Ethernet-over-GRE
- **IP-IP**
- **MPLS**
  - MPLS over GRE
  - MPLS over UDP
- **L2TP**
- **GTP-U**
- …etc
### VXLAN PLATFORM AND VENDOR SUPPORT

#### Broadcom Trident 2 (aka “T2”) Platforms

- **QFX5100-48S (1RU)**
  - 48x10 GbE
  - 6x40 GbE

- **QFX5100-96S (2RU)**
  - 96x10 GbE
  - 8x40 GbE

- **QFX5100-24Q**
  - 24x40 GbE
  - 2 x Modules:
    - 8x10 or 4x40 GbE

#### Other T2 Platform Vendors

- Cumulus Networks
- Arista
- Dell
- Cisco
- HP

---

**Juniper MX-Series and EX9200**
VXLAN ENCAPSULATION AND TERMINOLOGY

VXLAN Encapsulation

VTEP

Host-A

Router-A

Router-B

Host-B

VXLAN Tunnel End Point (VTEP)

VXLAN Network Identifier (VNI)

VXLAN Segment

Terminology

1

2

3

VXLAN Tunnel End Point (VTEP)

VXLAN Network Identifier (VNI)

VXLAN Segment
VIRTUAL EXTENSIBLE LOCAL AREA NETWORK (VXLAN)

- **Encapsulation Overview**
  - Layer 2 Overlay scheme over Layer 3 network
    - Designed for VM-to-VM communication in mind
    - VXLAN should be transparent to end hosts
  - Provide L2 segmentation ability > 4096 VLANs
    - 24 bit VXLAN Network Identifier (VNI)
      - 16M VXLAN segments

- **Forwarding Overview**
  - Data-Plane based learning and forwarding
    - VXLAN relies on Data-Plane learning of associated host MAC addresses to VTEP IP’s through source learning
    - Similar to Layer 2 with flood and learn

---

**VXLAN Encapsulation**

<table>
<thead>
<tr>
<th>Outer MAC DA</th>
<th>Outer MAC SA</th>
<th>Optional Outer 802.1Q</th>
<th>Outer IP DA</th>
<th>Outer IP SA</th>
<th>Outer UDP</th>
<th>VXLAN ID (24 Bits)</th>
<th>Inner MAC DA</th>
<th>Inner MAC SA</th>
<th>Optional Inner 802.1Q</th>
<th>Original Ethernet Payload</th>
<th>FCS</th>
</tr>
</thead>
</table>

**Original Ethernet Frame**
1) Host-A sends an ARP for Host-B.
2) Router-A looks up the VNI association for Host-B.
3) There is no entry and the ARP is VXLAN encapsulated and sent out to the IP multicast group per that VNI.
4) Router-B receives the Multicast packet, verifies the validity of the VNI, and learns the inner source MAC of Host-A.
5) Host-B receives the ARP and responds.
6) Router-B looks up the VNI associated for Host-A, and VXLAN unicasts to Router-A.
7) Router-A receives the unicast packet, verifies the validity of the VNI, and learns the inner source MAC of Host-B.
VXLAN INTEGRATION WITH EXISTING SERVICES

- **Overview**
  - Terminate (aka “Stitch”) VXLAN segments into existing network services, such as L3VPN, VPLS and E-VPN
    - Use routing/switch instances as centralized anchor points within a geography

- **Integration Areas**
  - Data Center Interconnect (DCI)
  - Virtual Provide Cloud Gateway
  - Access to Edge
    - MBH, Business, Residential, Wholesale
    - Subtending nodes
INTER-VXLAN ROUTING

Use Cases:

- Inter-Connecting
  - VXLAN Segments
  - L2 - VLANS
  - L3 – IRB
  - L2VPN / L3VPN
  - VPLS / E-VPN

- Augment Merchant Silicon with In-House Silicon
  - Example: Trident-2 does not support the ability to route packets into VXLAN tunnels and vice versa based on payload IP header.

- Controlled VTEP Broadcast Replication
BROADCAST DOMAIN REPRESENTATION
Enhancements:

- Broadcast replication using VXLAN Unicast
- Endpoints are statically defined
- In-line Data Plane learning and forwarding functions the same

Use Cases:

- No IP Multicast support between VTEPs
- A static point-to-point deployment, whereby a given VNI only has two VTEPs
- VXLAN communication must be secure using a mechanism that does not support IP Multicast
**CONTROL MODES**

**Data Plane Based**
- VXLAN IETF Draft based
- Multicast for L2-BUM traffic
  - Or Unicast BUM replication

**Control Plane Based**
- P2P tunnels built by the controller
  - Juniper Contrail or VMware NSX
  - OVSBD (or NETCONF)
- Controller MAC Learning
- Can be combined with Data Plane Control
DAYONE GUIDE: VXLAN CASE STUDIES

- Day One Guide
  - Native VXLAN with Multicast
    - PIM/OSPFv2
  - Unicast Only VXLAN
    - No Multicast
  - Inter-VXLAN Routing
    - Network Service Integration
  - VXLAN over IPSec Transport
    - IPsec Tunnel Mode

By Eddie Parra & Russell Kelly

Tentatively Scheduled for May, 2014
SUMMARY

- VXLAN Consideration
  - Think beyond VXLAN’s design use cases
  - Use platform diversity to your advantage
    - Economics, Power, Space, …etc

- JUNOS VXLAN Support
  - Target Release: JUNOS 14.1
    - May timeframe
  - Account teams can provide beta images
    - Feel free to email me accordingly
THANK YOU...
REFERENCES

Standards

VXLAN: A Framework for Overlay Virtualized L2 Networks over L3 Networks

Generic Overlay OAM and Datapath Failure Detection
http://www.ietf.org/id/draft-jain-nvo3-overlay-oam-01.txt

The Open vSwitch Database (OVSDB) Management Protocol