draft-sajassi-l2vpn-evpn-overlay-01.txt

A. Sajassi (Cisco), S. Salam (Cisco), Nabil Bitar (Verizon), W. Henderickx (Alcatel-Lucent)

IETF 85, November 2012

Atlanta

Objectives

- This draft describes how E-VPN can be used as an Network Virtualization Overlay (NVO) solution
- It considers two main scenarios:
  - NVE residing in the hypervisor, and
  - NVE resides in a ToR device
- and explores different tunnel encapsulation options for E-VPN over IP under them
  - MPLS over GRE
  - VXLAN and NVGRE

VXLAN/NVGRE Encapsulation

- VXLAN encapsulation is based on UDP and uses 8-byte header on top of UDP
- Provides a 24-bit virtual network (or segment) identifier
- Typically there is a one-to-one mapping between the segment-id and the tenant VLAN ID
- Typically tenant VLAN ID is not included in encapsulation – e.g., analogous to VLAN-based service in E-VPN
- If it does, segment-id is still used at the egress VTEP to identify a bridge domain - e.g., analogous to VLAN bundle service in E-VPN

NVGRE Encapsulation

- Encapsulation is based on GRE
- It mandates inclusion of GRE key field
- Just like VXLAN, a 24-bit ID is used to identify a segment (called Virtual Subnet ID)
- There is a one-to-one mapping between VSID and the tenant VLAN ID
  - This maps to VLAN-based service in E-VPN

What does the above mean for E-VPN BGP routes ?

- It means that Ethernet Tag ID can be used as a segment-id
- This is not coincidence, E-VPN was designed with a 24-bit Ethernet tag in mind (for I-SID in 802.1ah)
- No need to use MPLS label field – it can be set to null or omitted
- Although there is no change to E-VPN BGP routes, E-VPN procedures are impacted

MPLS over GRE Encapsulation

- E-VPN MPLS client layer can be transported over IP PSN tunnel transparently
  - There is no impact to EVPN routes
  - There is no impact to EVPN procedures and associated data-plane functions
- GRE key can be used to provide a 32-bit entropy field
- Load balancing can be supported by new core routers built to support NVGRE encap
- Load balancing cannot be efficiently supported by existing routers that don’t support GRE key for ECMP

VXLAN/NVGRE Encapsulation

- Depending on where NVE resides, either a subset or the full set of E-VPN routes & procedures are needed
- Depending on the location of NVE (e.g., TOR), there are changes to E-VPN procedures and data-plane functions as will be described next

NVE Residing in Hypervisor

- Requires only a subset of E-VPN routes & attributes
  - MAC Advertisement Route
  - Inclusive Multicast Route
  - MAC Mobility Extended Community Attribute
  - Default GW Extended Community Attribute

NVE Residing in Hypervisor – Cont.

- Requires only a subset of E-VPN Procedures
  - Local learning of MAC addresses
  - Advertising locally learned MAC addresses in BGP
  - Performing remote learning using BGP
  - Discovering other NVEs and constructing multicast tunnels
  - Handling MAC address mobility

NVE Residing in ToR Switch

- It assumes servers are multi-homed to ToR switches operating either in active/active or active/standby modes
- If servers are single-homed to ToR, then the scenario becomes similar to that of "NVE residing on hypervisor"
- Requires the entire set of E-VPN BGP routes and attributes

NVE Residing in ToR Switch – Cont.

- Requires the entire set of E-VPN multi-homing procedures
  - Multi-homed Ethernet Segment Auto-discovery
  - Fast Convergence and Mass Withdraw
  - Split-Horizon
  - Aliasing and Repair-Path
  - DF election

NVE Residing in ToR Switch – Cont.

- Modification to E-VPN procedures on what type of multi-homing is used: active/active versus active/ standby
- For active/standby, if repair-path functionality is not used, then no changes to E-VPN procedures is needed else need to modify
    - Aliasing and Repair-Path
- For active/active, we need to modify
    - Split-Horizon
    - Aliasing and Repair-Path

Modifications to Split-Horizon

- In E-VPN, and MPLS label is used for SH filtering to support active/active MH
- In VXLAN/NVGRE encap, we cannot add such label and other means are needed
- Several options are being considered:
  - Assign and IP address for each site (ESI)
  - Use client source MAC to perform SH filtering
  - Use source PE address along with local switching

Modifications to Aliasing and Backup-Path

- In E-VPN, Ethernet AD route is advertised by a multi-homed PE with a VPN label used to load-balance traffic between PEs, even when a given MAC is learnt by only a single PE
- For VXLAN/NVGRE, instead of MPLS label, we can advertise IP address of the PE along with Ethernet AD route
- Remote PEs resolve and ESI (site ID) to a list of IP addresses corresponding to the tunnel endpoints connected to that multi-homed site

Summary

- MPLS over GRE uses E-VPN MPLS client layer as is and thus no impact to BGP routes and procedures
- VXLAN/NVGRE can impact E-VPN procedures if NVE reside in ToR and if it runs active/active multi-homing or active/standby multi-homing with backup-path feature