Virtual Hub-and-Spoke in BGP EVPNs

draft-keyupate-bess-evpn-virtual-hub-00.txt

K. Patel, A. Sajassi, J. Drake, Z. Zhang, W. Henderickx

IETF 98, March 2017

Chicago

History

- Presented at IETF 94, November 2015
- Added a co-author, Jeffrey Zhang, for this rev.
- Added a new section on handling of broadcast and multicast traffic (contributed by Jeffrey)
- Change the name from
  - “draft-keyupate-evpn-virtual-hub-00.txt” to
  - “draft-keyupate-bess-evpn-virtual-hub-00.txt”

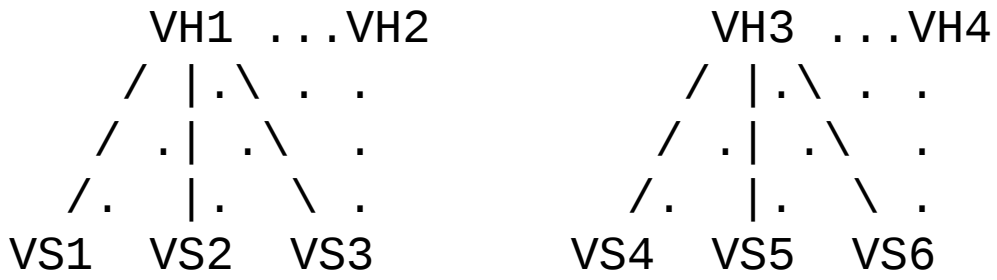
Background

- EVPN route scale in DCs is typically achieved by
  - Leaf nodes learn and store local DC routes (MAC and IP addresses)
  - Any remote DC routes are stored on Gateway nodes
  - Leaf nodes are installed with default route pointing towards Gateway nodes
- This draft further optimizes route learning such that
  - Leaf nodes learn and store routes (MAC and IP addresses) for local sites only
  - Leaf nodes do not store routes advertised by other Leaf nodes residing in the same DC

Background – Cont.

- This draft extends RFC7024 for EVPN address family
  - Rules for generating and processing of unknown MAC Route
  - Modifications to Aliasing, Split Horizon and Mass Withdraws
  - Forwarding considerations for Bridging mode and IRB mode
  - Rules for ARP Suppression
- What was not covered, was the rules/procedures for broadcast/multicast traffic

Handling of BUM Traffic



- Virtual Spoke 1 (VS1), VS2, VS3 are associated with Virtual Hub 1 (VH1) and VH2. VS4, VS5, VS6 are associated with VH3 and VH4
- For BUM traffic, we want a single copy to be sent from the VS to one of its associated VHs and then that VH send the BUM traffic to all other VSs that need to received this traffic (either directly or indirectly)

Split Horizon

- When VH1 relays traffic from VS1,
  - For Ingress-Replication (IR), it must not send traffic back to VS1
  - For P2MP tunnel, it must indicate VS1 as source
- In case of IR, when IP unicast tunnel is used, the outer IP SA identifies the source VS
- In case of IR, when MPLS unicast tunnel is used, downstream assigned label by VH1 is used to identify the source VS
- In case of P2MP MPLS tunnel, upstream assigned label by VH1 is used to identify the source VS
- PE Distinguisher (PED) Label Attribute is used as both upstream and downstream assigned label
  - It is advertised along with IMET route by VH1
  - It is used to identify both the source PE and the specific EVI/BD

Route Advertisement

- Just like any other routes, IMET routes from V-hubs are advertised with RT-VH and RT-EVI so they are imported by associated V-spokes and all V-hubs
- IMET routes from V-spokes are advertised with RT-EVI so they are imported by all V-hubs
- If a V-hub uses mLDP P2MP tunnel to send/relay traffic, only its associated V-spokes and all V-hubs will see the V-hub's IMET route and join the tunnel
  - Another V-hub needs to relay traffic to its associated V-spoke
  - That V-hub uses (*,*) S-PMSI AD route to advertise to its cluster

Designated Forwarder in a Cluster

- If there are multiple V-hubs in a cluster, a V-spoke choses one
- If the receiving V-hub uses mLDP to relay traffic, then V-hubs in other clusters further relay the traffic
- But only one V-hub in each cluster can do so
- Thus DF election is needed among V-hubs in a cluster

Traffic Forwarding Rules

- Traffic from a V-hub's local ACs is forwarded using tunnel announced in IMET route
- For mLDP, traffic is relayed by V-hubs of other clusters to their associated V-spoke
- Traffic received by a V-hub from a V-spoke, it needs to relay to other PEs using the tunnel announced in IMET route.
- In case of IR, the source V-spoke identified by incoming label or source IP address, is excluded from replication list
- In case of P2MP tunnel, the popped incoming label is imposed again to identify the source V-spoke

Next Step

- More discussions among interested partitas
- Finalize the new routes
- Clarify that this approach is incremental on top of HRW draft – to avoid too many permutations
- Beef-up backward compatibility section for both mechanisms