MVPN/EVPN
Tunnel Aggregation with Common Labels

Zhaohui Zhang (Juniper)
Eric Rosen (Juniper)
Wen Lin (Juniper)
Zhenbin Li (Huawei)
P2MP Tunnel Aggregation

- A single P2MP tunnel used for multiple VPNs/BDs
- An ingress PE imposes a VPN/BD-identifying label followed by tunnel label
  - Per MVPN/EVPN specifications, the VPN/BD-label is upstream allocated from the ingress PE’s label space, and is advertised in the corresponding PMSI/IMET route
- An egress PE maintains context label tables – one per Ingress PE, with the VPN/BD-labels signaled from the ingress PE
  - Tunnel label of an incoming packet identifies the context label table in which the inner VPN/BD label is looked up
EVPN Multi-homing ES

- EVPN multi-homing split-horizon procedure
  - When an ingress PE sends BUM packets from an multi-homed ES using a P2MP tunnel, it imposes an ES-identifying label to indicate the source ES so that receiving PEs will not send packets out of ACs attached to the source ES
  - This is another form of tunnel aggregation
    - A P2MP tunnel, even if used only for a single BD, is used for traffic from multiple Eses
    - The ES-identifying label is upstream allocated, just like VPN/BD-identifying labels
Scaling Issue

• Serious scaling issue with the upstream allocated labels
  • X VPNs/BDs with (Y+1) PEs in each \( (X \times Y) \) labels on each PE
    • 1000 VPNs each with 1001 PEs \( \rightarrow \) 1M labels on each PE
  • Problem has not surfaced before, likely because P2MP tunnel aggregation has not been deployed
  • BIER transport is an inherent aggregation tunnel and is getting deployed
  • This applies to MP2MP tunnels as well
Solution

- PEs coordinate their label allocation:
  - From a common label pool carved out of the downstream-allocation label space
    - No longer “upstream allocated”
      - This simplifies forwarding
    - Referred to as “Domain-wide Common Block”
      - Much like SRGB
  - All uses the same label for the same VPN/BD/ES
    - This reduces the number of labels needed
      - X labels for X VPNs/BDs/Eses
DCB Not Large Enough?

• Use a separate label space
  • Different from the downstream-allocation space
  • But still:
    • Shared across all PEs
    • Same label used by all for the same VPN/BD/ES

• This separate label space is identified by a label from the DCB
  • Label stack: <tunnel label, label-space-identifying-DCB label, VPN/BD/ES-identifying label>
Signaling

- If VPN/BD label is from the DCB, a C-bit in the flags field of the PTA field of the PMSI/IMET route is set.
- If VPN/BD label is from a separate common label space, the PMSI/IMET route carries a Context Label Space ID Extended Community:
  - Transitive Opaque EC: \(<\text{ID Type}, \text{ID Value}>\)
  - \(<0, \text{DCB Label}>\)
- ESI label must be from the same label pool/space as the BD label pool/space:
  - No additional signaling is needed.
Summary

• Instead of upstream allocated VPN/BD/ES-identifying labels, use labels from a Domain-wide Common Block
  • A common pool from all PE’s downstream-allocation label space
  • Simpler forwarding and better scaling
• If the DCB is not enough, use a separate label space shared among all PEs
• Additional details in the draft on tunnel segmentation
Next Steps

• Seeking Comments
• Will seek adoption after further polishing