BGP Signaled Multicast
BGP-MCAST

draft-ietf-bess-bgp-multicast-02
Z. Zhang, L. Giuliano, K. Patel, I. Wijnands, M. Mishra, A. Gulko
draft-ietf-bess-bgp-multicast-controller-03
Z. Zhang, R. Raszuk, D. Pacella, A. Gulko

Presented by Z. Zhang
For BESS, IETF108
draft-ietf-bess-bgp-multicast-02

• Hop-by-hop signaling of IP multicast tree and mLDP tunnel
• Added RD to NLRI to support signaling in VRFs
  • Mainly to support signaling from controllers
  • In theory also as hop-by-hop replacement of BGP-MVPN signaling over the core
    • But not pursuing that
• Added inter-region support
  • Inline signaling through a region
    • Added Multicast RPF EC to address the problem of internal routers not having routes to source/root in case of BGP-LU
      • Similar to PIM RPF Vector and mLDP Recursive FEC
• Overlay signaling over a region
  • Similar to mLDP over targeted session
• Interworking with controller signaled multicast
  • Stitching for heterogeneous regions that use hop-by-hop or controller signaling
draft-ietf-bess-bgp-multicast-controller-03

• TEA enhancements
  • Upstream information is now encoded in TEA itself
  • MP2MP support
• Added multi-domain section
• Added SR P2MP support
TEA Enhancements

• Upstream information is now encoded in TEA itself
  • As a tunnel with a RPF sub-TLV
  • Upstream Router’s IP Address in NLRI is this router’s address
    • The Leaf A-D route is “from” downstream routers so the upstream router is this router

• Incoming Label Stack
  • Each tunnel in the TEA for MP2MP has an Incoming Label Stack for incoming traffic and regular Label Stack for outgoing traffic
  • The RPF tunnel in the TEA for P2MP only has an incoming Label Stack for incoming traffic, and other tunnels in the TEA for P2MP only have a regular Label Stack for outgoing traffic
Multi-domain Support: Different controllers for different domains

- Native IP Multicast
  - Ctrlr1 & Ctrlr2 independently signals to RTR B & C their respective downstream and upstream interface
- Labeled Multicast
  - Ctrlr1 & Ctrlr2 coordinate the label to use on the B-C link, or,
  - RTR C uses hop-by-hop signaling to B

- RTR B receives Leaf A-D routes from both Ctrlr1 & Ctrlr2. Normally only one of them is chosen as the best route and triggers forwarding state.
- In this case, RTR B is provisioned as a border router and must look for the routes from both controllers to stitch the two segments together
SR P2MP Support

• SR P2MP has been accepted by Spring/PIM WGs
  • Spring - Replication Segment
    • Building block - Replication state on individual nodes of a tree
  • PIM – SR P2MP policy for P2MP trees

• An MPLS SR P2MP tree is no different from mLDP/RSVP-TE P2MP tunnel in forwarding plane
  • Control plane differences: controller calculated and signaled
  • Signaling could be Netconf/PCEP/BGP
    • BGP signaling could be BGP SR-TE based or BGP-MCAST based
    • “Different ways to skin a cat”

• Only a new NLRI type is needed for BGP-MCAST based signaling
  • “Same way to skin a different cat”
Replication Segment

- A Replication Segment is identified by (Root-ID, Tree-ID, Node-ID), which is encoded in a new type (0x83) of S-PMSI route, which is in turn included in the Leaf A-D route signaled from the controller.
- Replication information is encoded in Tunnel Encapsulation Attribute.
  - In case of SR-MPLS, nothing new needed.
    - Nothing explicitly tied to SR except the SR P2MP tree and Replication Segment terms.
  - Optionally, TEA can have SR Policy tunnels.
    - This ties to SR explicitly.

```
+--------+-----------------------------------+
|        | Route Type - 4 (Leaf A-D)         |
|        | Length (1 octet)                  |
|        +-----------------------------------+
| L      | Route Type - 0x83 (SR P2MP S-PMSI) |
| E      | Length (1 octet)                  |
| A      +-----------------------------------+
| F      | RD (8 octets)                     |
| R      +-----------------------------------+
| O      | Root ID (4 or 16 octets)          |
| U      |                                   |
| T      |                                   |
| L      |                                   |
| E      |                                   |
| N      |                                   |
| T      |                                   |
| L      |                                   |
| E      |                                   |
| N      |                                   |
| T      |                                   |
| R      +-----------------------------------+
| I      | Upstream Router's IP Address      |
| K      |                                   |
| E      |                                   |
| Y      | Originating Router's IP Address   |
|        +-----------------------------------+
```
SR Policy Tunnel

• Originally defined to instantiate an SR P2P Policy
  • Specifying Binding SID and outgoing SID list for the SR P2P Policy

• When used in TEA for SR P2MP, it refers to a pre-installed SR P2P policy as a replication branch
  • Binding SID used to lookup the pre-installed outgoing SID list
  • One-SID SID list in the tunnel is the outgoing SID for the tree
SR P2MP Policy

- An SR P2MP Policy defines an SR P2MP Tree
  - Identified by (root-id, tree-id)
  - Includes a set of Candidate Paths (CPs) and a set of Leaf Nodes

- An SR P2MP tree’s CP is instantiated with Replication Segments for the root, leaf, and replication nodes for that CP stitched together

- An SR P2MP Policy is instantiated on the tree root by attaching a BGP Community Container to Leaf A-D routes for the root’s Replication Segments
  - CP priority
  - Optional Atom TLV
    - An IPv4/v6 Address List – for the set of leaves
    - An UTF-8 string – for policy name
Summary

• BGP signaled multicast for IP multicast, mLDP and SR P2MP is getting more and more mature
  • Draft work
  • POC implementation
• Needs to spell out more precise procedures
• Will present to IDR on TEA changes