A Framework for Computed Multicast applied to SR-MPLS

draft-allan-pim-sr-mpls-multicast-framework-00

Dave Allan,
Jeff Tantsura,
Ian Duncan

IETF102

What is the draft about?

- Using computation to determine the routing of multicast segments in an MPLS based SR network, and how tunneling using node-SIDs can be used as part of multicast tree construction
  - Either distributed or centralized control models

- The draft describes
  - Terminology
  - Overall approach
  - Loose and Explicitly Routed multicast distribution trees
  - Algorithm
  - FIB installation procedures

Motivations/1

Reduce state!

- Multicast state can rapidly dwarf unicast state
- A quick comparison:
  - ND = network diameter
  - T = total number of multicast trees
  - L = average number of leaves per tree
  - For “flat” multicast trees (e.g. PIM or mLDP)
    - State ~= T x f(ND x L)
  - With the approach described in the draft
    - State = T x L x 2 (worst case)

Motivations/2

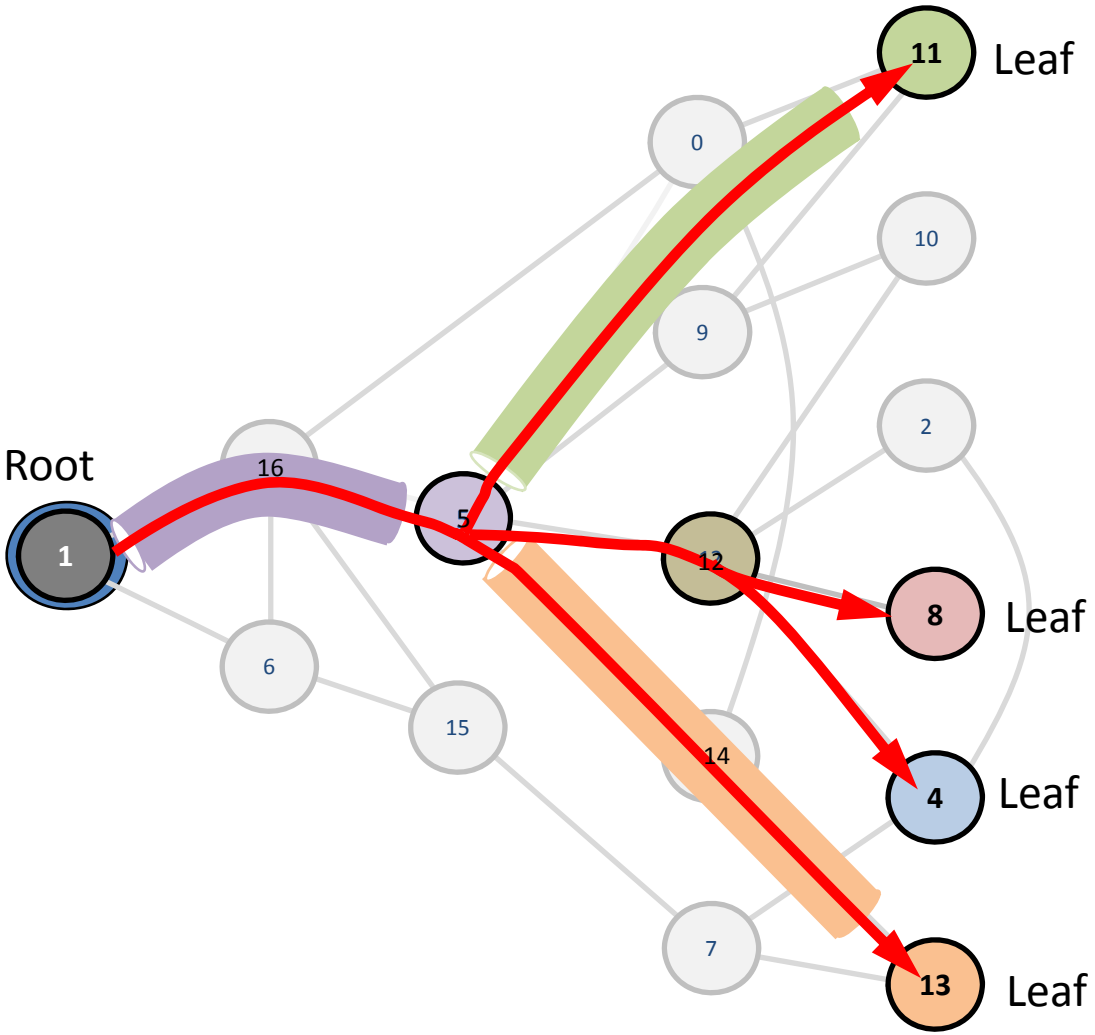
- Leverage the MPLS dataplane and SR as much as possible
  - Use the SR-MPLS data plane in ways PIM or mLDP "like" approaches cannot
  - Implement multicast where BIER not technically or economically feasible

Approach

- The draft describes an architecture whereby multicast trees are a hybrid of roots, leaves, and replication points interconnected with tunnels, with the routing of the tree determined entirely from information in the IGP

- This provides multiple benefits
  - Minimized messaging the converge the network
  - Reduced dataplane state
  - Reduces bandwidth requirements vs. straight IGP derived trees (PIM, mLDP, BIER)
  - Unicast convergence provides recovery for most failures

An example tree



Root – node 1
Leaves – nodes 4,8,11,13
Replication points – nodes 5 & 12

Multicast SID ‘x’

Node SID ‘5’

Node SID ‘11’

Node SID ‘13’

Required tree attributes

- The use of tunnels requires a minimum cost or near minimum cost multicast tree in order to be ECMP “friendly”
  - No duplication of packets on any link → no logical multicast
- An ECMP “friendly” tree construction algorithm is in the draft
- Serendipitously, it is also the source of improvements in bandwidth efficiency
  - It shifts replication points closer to the leaves

Loose and Explicitly Routed Trees

- A loose tree is composed of a single multicast segment (with a SID), where only the root and the leaves have been specified
- An explicitly routed tree is composed of a concatenation of multicast segments where the roots, waypoints and leaves have been specified
  - The routing of individual segments is still computed
  - The routing of an MDT can then be specified to an arbitrary level of granularity

Changes from the last time around

- This has been presented before
  - Last time was IETF 97
- Current draft
  - Updated terminology to align with current state of SR-MPLS
  - Editorial improvements
  - Motivations added
  - Improvements to the algorithm description
  - Offers some thoughts on SR-Controller operation

Next Steps

- Collect feedback
- Planned updates to the draft
  - Improvements to FIB installation procedures
  - Bring the draft up to date w.r.t. “MPLS friendliness”
- We will bring forth in future drafts:
  - The required IGP extensions
  - Interworking with existing mechanisms
- We will pursue standards track
  - So looking for PIM WG adoption

Questions?

Backup

- Existing implementations
  - SPRING charter focuses on no DP changes
- This does not require a DP change
  - Existing silicon can replicate into tunnels
  - ECMP at a replication node can be a control plane function
    - The action for a multicast SID is to replicate a packet to a set of interfaces, and there is a stack manipulation to be performed for each interface
      - This maps to a continue and push
    - The ECMP aspect is what interface is selected for the particular tree from the set of possible next hops for the node SID
      - In RFC 3031 terms an ILM → {NHLFE1, NHLFE2, etc.}