BFD for Multipoint Networks over Point-to-Multi-Point MPLS LSP

draft-mirsky-mpls-p2mp-bfd

Greg Mirsky

MPLS WG Interim, May 2020

Motivation

- p2mp MPLS LSP is here to stay
- IETF published relevant RFCs – RFC 8562 and RFC 8563
- RFC 5884 BFD for MPLS LSP applies only to the case of p2p LSP
- RFC 7880 does not apply to the case of p2mp LSP

BFD for multipoint networks

- BFD for a multipoint network uses Demand mode, defined in RFC 5880, from the very start – no three-way handshake
- Only root transmits BFD control messages with its My Discriminator and Your Discriminator set to 0
- Leaf cannot demultiplex BFD sessions by Your Discriminator as in Asynchronous mode that is used in MPLS LSP per RFC 5884
- Leaf uses three-tuple <My Discriminator, IP Source address, Identity of the multipoint tree> to demultiplex BFD sessions
- RFC 8562 BFD for Multipoint Networks – egress (tail) detects a failure of the multicast tree (p2mp LSP). Ingress (head) doesn’t know the state of the tree
- RFC 8563 BFD for Multipoint Active Tails considers several mechanisms for the ingress LSR to learn about a failure of the p2mp LSP

IP/UDP Encapsulation

- Destination IP address randomly chosen:
  - from 127/8 range for IPv4
  - from ::ffff:127.0.0.0/104 range for IPv6
- Destination UDP port number 3784
- Source UDP port number from 49152 through 65535 range

Non-IP encapsulation

- Overhead of IP/UDP encapsulation, especially with IPv6, is significant
- Cannot use G-ACh type BFD Control, PW-ACH encapsulation (without IP/UDP Headers) 0x0007 defined in RFC 5885
- Request IANA to allocate the new G-ACh type Multipoint BFD Session without IP/UDP Headers
- Use Source Address TLV defined in RFC 7212

Bootstrapping BFD session over p2mp MPLS LSP

- LSP Ping with BFD Discriminator with Target FEC TLV from sub-TLVs defined in Section 3.1 RFC 6425
  - Since BFD over p2mp MPLS LSP is in Demand mode and an egress LSP does not periodically transmit BFD Control packets to the egress, it is RECOMMENDED that the LSP Ping with BFD Discriminator had Reply mode set to “Do Not Reply”
- BGP-BFD Attribute as defined in “Multicast VPN fast upstream failover”
- PIM-SM BFD Discriminator in PIM Hello message

P2MP BFD with Active Tail

- RFC 8563:
  - Head notification and tail solicitation with multipoint polling
    - Head occasionally transmits Poll sequence packet (BFD Control packet with P(Poll) bit set over p2mp LSP in addition to the periodic transmission of non-Poll BFD packets
    - The tail is expected to reply with F (Final) bit set over the unicast reverse path that is disjoint with the p2mp LSP (that is how an egress informs the ingress LSR of the detected failure)
    - If either multipoint Poll or the unicast Final lost, the ingress detects the defect but is not certain about the state of the p2mp LSP
  - Head notification with composite polling
    - The head's behavior is as described above. In addition, the ingress may send unicast Poll to a specified egress LSR, e.g., the one failed to respond to the multipoint Poll, over the forward unicast path (disjoint from p2mp LSP) (out-of-band for p2mp LSP)
    - Because this method uses the out-of-band probe, the ingress can better localize the failure and be aware of the state of p2mp LSP. It is not 100% certainty but still better than with only multipoint Polls.

Head Notification with Multicast and Composite Polling

Unicast Poll/Final

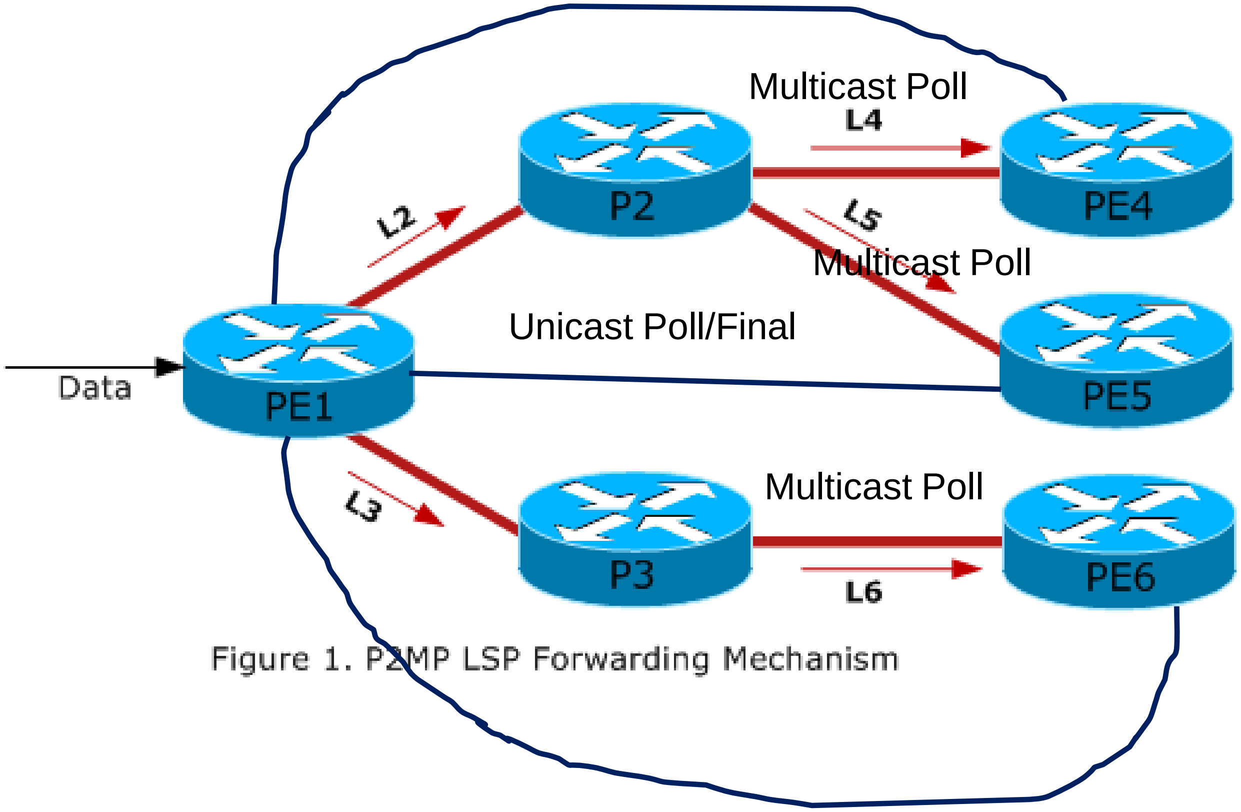
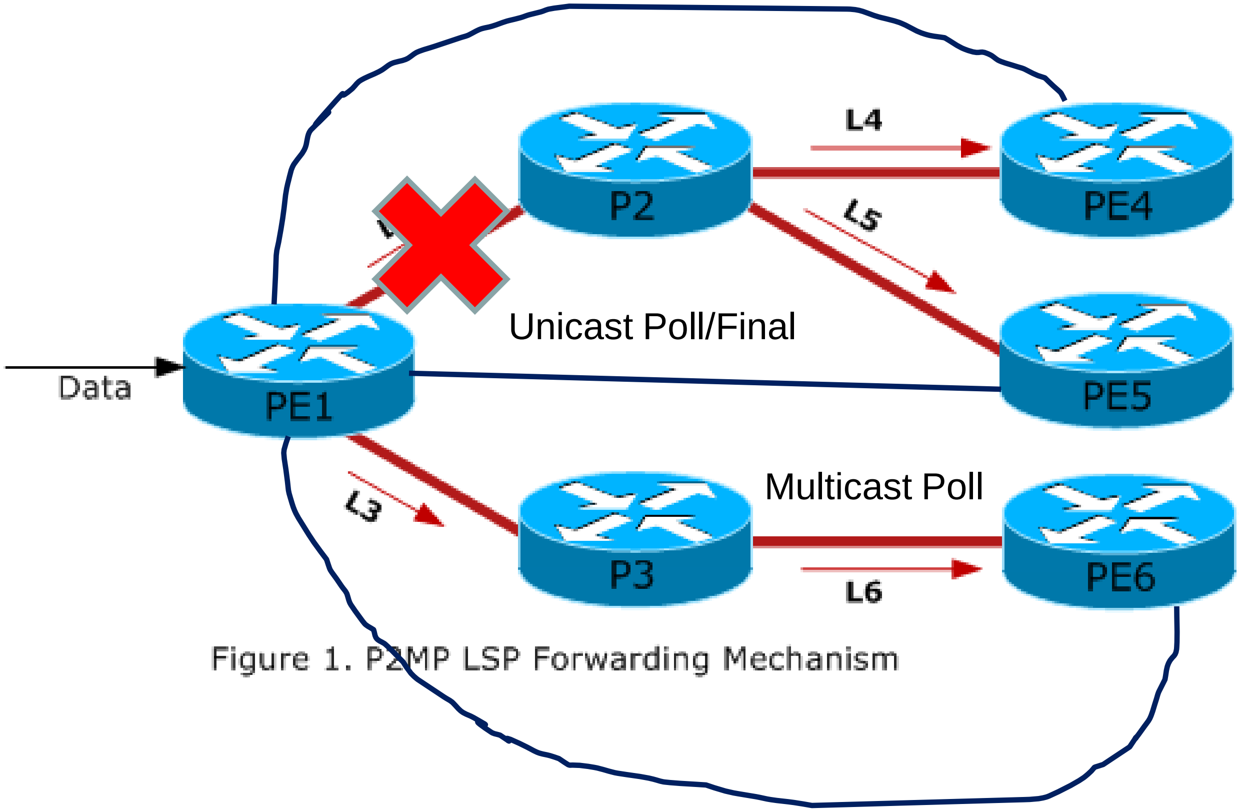


Figure 1. P2MP LSP Forwarding Mechanism

Unicast Poll/Final

Head Notification with Multicast and Composite Polling

Unicast Poll/Final



Unicast Poll/Final

Multicast Poll

Figure 1. P2MP LSP Forwarding Mechanism

Unicast Poll/Final

Head Notification Without Polling (Unsolicited, Event Triggered)

As suggested by the name, the ingress sends no Polls, but it is an egress LSR that, upon detecting a failure of p2mp LSP, transmits unicast Poll over the reverse unicast path with the Diag to signal the failure to the ingress LSR.

Destination IP address – IP address of the Multipoint Head (either from Source IP Address or Source Address TLV)

UDP Destination port – 4784 per RFC 5883 Multi-hop BFD

Your Discriminator is set to My Discriminator value associated with the BFD session (in the received BFD Control packets from the ingress)

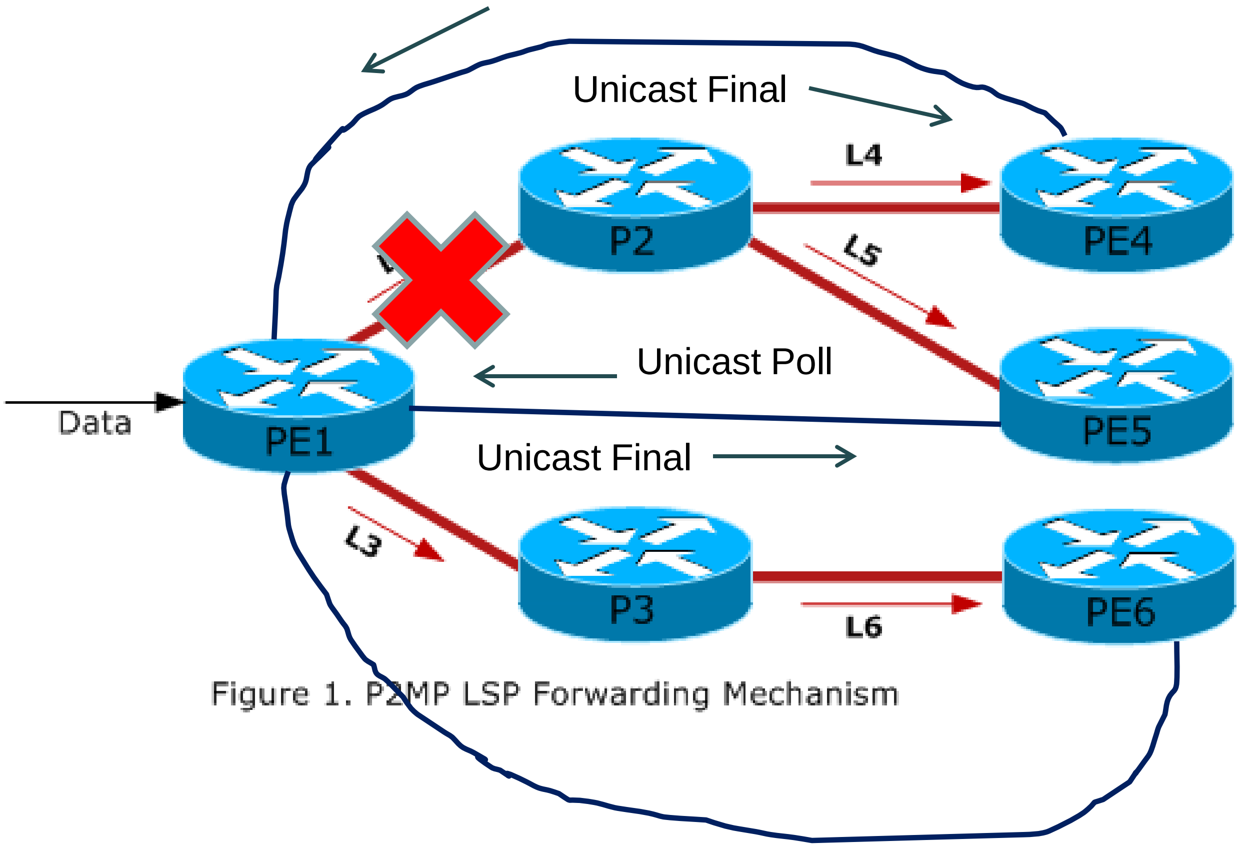
Poll bit is set

Sta (Status) – Down

Diag - Control Detection Time Expired value

That Poll packet transmitted periodically (one per second) until either the failure clears or the Final packet from the ingress LSR received.

Head Notification Without Polling



Unicast Poll

Unicast Final

Unicast Poll

Unicast Final

Figure 1. P2MP LSP Forwarding Mechanism

Unicast Poll

Next steps

- Your comments, suggestions, questions always welcome and greatly appreciated
- WG adoption