IP/LDP Fast-Reroute
Using Maximally Redundant Trees
draft-ietf-rtgwg-mrt-frr-architecture-01

Alia Atlas, Robert Kebler,
Ijsbrand Wijnands,
Gábor Enyedi, András Császár,

IETF 83, Paris, France
Overview

• Took multicast-related work from draft-atlas-rtgwg-mrt-frr-architecture and created separate draft.

• Covers:
  – Global Protection 1+1 (aka multicast live-live)
  – Multicast fast-reroute
    • PLR Replication
    • Alternate Trees
Global Protection 1+1

• MRT provides two maximally disjoint trees.
• MRMTs (maximally redundant multicast trees) can be created via PIM or mLDP signaling specifying the appropriate MT-ID.
• Traffic Self-Identification important to handle cut-links/cut-nodes
  – mLDP traffic always has different labels per MRMT
  – PIM recommended to use different G or S per MRMT
Convergence for MRMT

• On topology change, both Blue MRT and Red MRT *change*.
  – Not possible to compute maximally redundant tree to an existing one (in general)

• Two options to handle:
  – Make-before-break on each MRMT
  – Ordered Convergence – still under discussion
    • Receivers repair broken MRMT
    • Then update unbroken MRMT
Inter-area/inter-level behavior for MRMT

- Need to protect against ABR/LBR failure.
- Approach A: exactly 2 ABR/LBR between two areas
  - BR1 receiving join for MRMT determines whether MT-ID needs to be changed (Blue to Red or vice versa) to avoid BR2 in upstream area/level.
  - For mLDP, control-plane changes to MT-ID is all that is needed
  - For PIM, if different (S,G) for Blue MRMT vs. Red MRMT, then traffic rewriting is needed by BR.
Example: Red to Blue Change Needed

(a) Area 0
Red Next-Hops to S
BR1's is BR2
BR2's is B
B's is S

Blue Next-Hops to S
BR1's is A
BR2's is BR1
A's is S

(b) Area 10
Y's Red next-hop: BR1
Y's Blue next-hop: BR2
Approach B: BR Stream Selection

• Works for any number of BRs
• When BR receives a join from downstream area, BR joins both Blue and Red MRMTs in upstream area.
• BR uses stream-selection to pick which traffic to forward to downstream area.
  – For PIM, different (S,G) means traffic rewriting.
• Each area/level is independently protected
Multicast Fast-Reroute: Differences from Unicast

• Final destinations unknown to PLR and may be large, so can only repair to next-hop or next-next-hops
• If failure not KNOWN to be node, then need to repair to both next-hops and next-next-hops
• If failure not KNOWN to be link & node-protection desired, then need to repair to both next-hops and next-next-hops.
• Updating multicast state can take much longer than unicast convergence.
• For PLR replication, PLR and MP cannot predict which interface alternate traffic will arrive at the MP on.
MP decides whether to accept alternate traffic

• If link/node failure can’t be told apart, a next-next-hop MP may receive two copies of traffic
  – Primary traffic from UMH
  – Alternate traffic

• *Because* of 100% unicast alternate coverage:
  – If RPF interface (for PIM) or links from UMH are up, then MP can assume primary traffic will flow
  – Otherwise, accept and use alternate traffic

• MP switches behavior based on link state – not received traffic (more secure).

• MP must do make-before-break so it continues to accept alternate traffic until its new primary UMH is sending traffic.
Multicast FRR: PLR-Replication

- PLR learns MPs to replicate to
  - PLR-driven: failure-point proxies info for next-next-hops
  - MP-driven: failure-point tells next-next-hops the PLR and each MP requests protection from PLR.

- PLR replicates traffic, encapsulates it with label or IP to MP.

- Traffic is forwarded to MP using unicast forwarding.
  - Route might be alternate or new primary
Multicast FRR: Alternate Trees

- Motivation: PLR replication can cause lots of traffic replication on links
- Create alternate-tree per (PLR, FP, S, G)
- Signal backup-joins to Blue UMH or Red UMH based on computation
- Allows use of native multicast – but does add multicast state
- Traffic must self-identify as to which alternate-tree it is in.
  - MPLS labels for mLDP and PIM
  - IP-in-IP possible for PIM – but need to deal with G assignment.
- Always forward alternate traffic on alternate tree
- MP also determines whether to accept alternate traffic and forward onto primary multicast tree.
Bypass Alternate Trees

• Motivation: IPTV – many different G for same S
  – Reduce alternate-tree state
  – Bypass shown to scale well for RSVP-TE FRR
• Downside: Alternate Traffic can go to MPs that don’t subscribe to that G
• For PIM, top level encap is the same and (S,G) underneath is globally understood.
• For mLDP, requires upstream-assigned labels for inner label.
  – Probably targeted-LDP between MP and PLR so PLR can distribute.
• Adds some complexity – but can substantially reduce state (e.g. 1000 G can share same bypass alternate-tree).
Summary

• Draft has significantly more details than previous sections.
• Trying to address multiple use-cases.
• Looking for comments and review.
• Have heard significant interest.
• Plan to evolve and have a more complete/stable version for next IETF.