IETF 108, July 2020

draft-kompella-mpls-nffrr

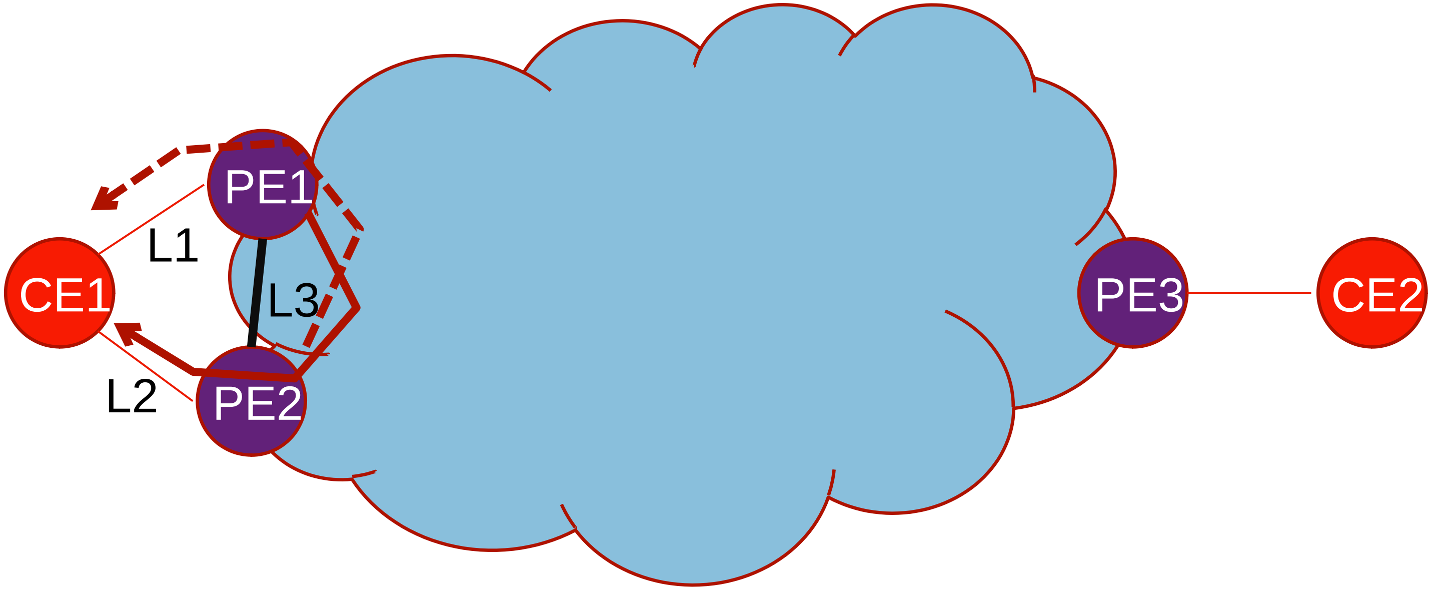
Kireeti Kompella, Wen Lin

PROBLEM STATEMENT

- Fast Reroute (FRR) offers high speed protection from link and/or node failures
    - FRR dramatically reduces packet loss on failure
    - FRR is thus a strong incentive to deploy MPLS
- However, on occasion, FRR may cause packet loss and/or congestion on some links
    - This can arise in several situations, primarily when there are multiple failures
- This draft describes some of these situations and proposes a solution using a new Special Purpose Label (SPL)

SITUATION 1: MULTI-HOMED CE

CE1, CE2 are clients of
an EVPN or an IPVPN



CE1 is dual-homed to PE1 and PE2 in “active-active” fashion.

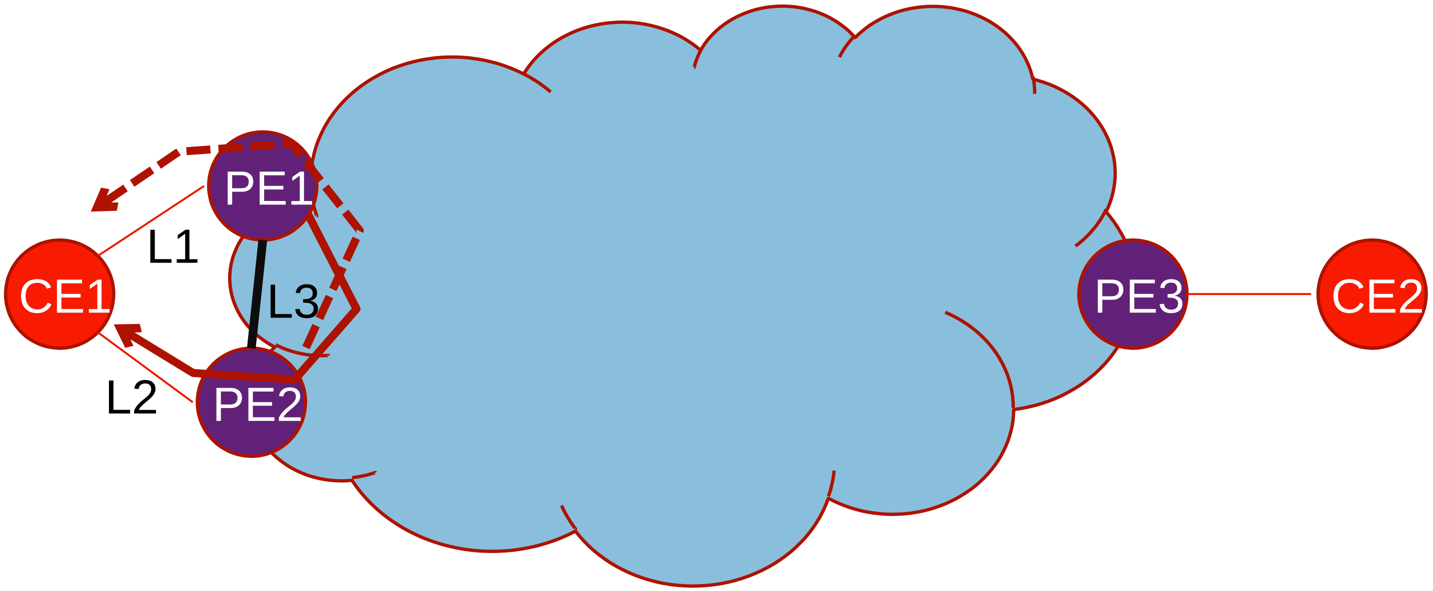
If link L1 fails, PE1 can quickly detect and recover using a pre-signaled backup path via PE2. Similarly, if link L2 fails, PE2 can do the same.

⟶ Protection path from PE1 via PE2 to CE1

- - → Protection path from PE2 via PE1 to CE1

} VPN FRR

SITUATION 1: MULTI-HOMED CE

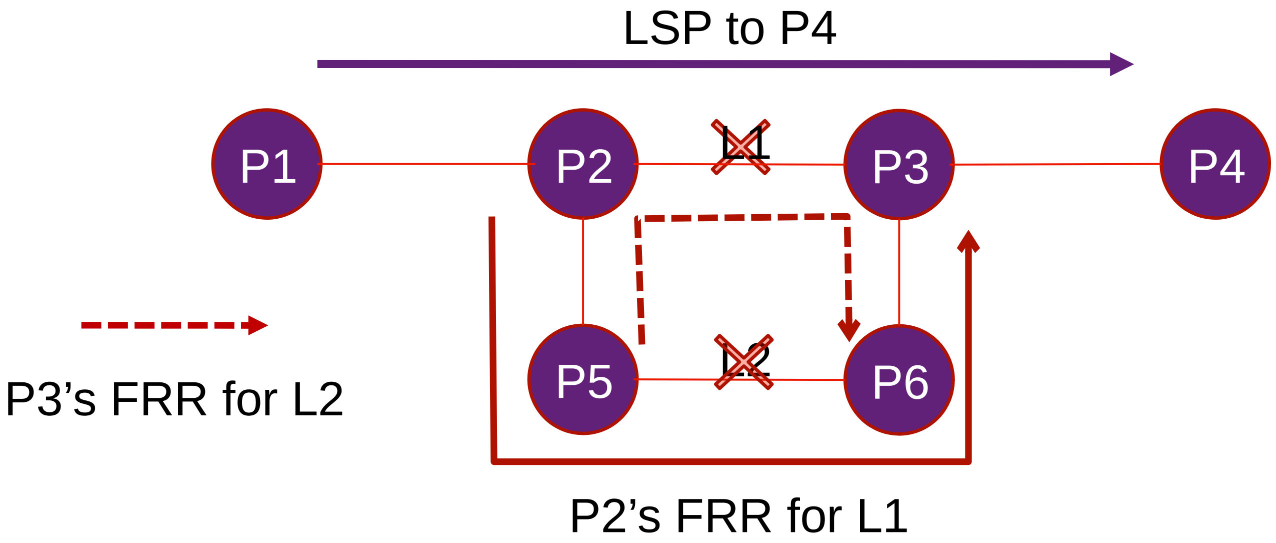


However, if CE1 fails, PE1 likely sees this as a link failure, and activates FRR. The packets to CE1 via L1 then take the backup path via PE2 to CE1.

But PE2 also sees a link failure of L2 and does the same: sends the packets back to CE1 on the backup path via PE1.

Thus, each packet loops between PE1 and PE2 until TTL expires, congesting link L3. Thus, in this case, FRR actually hurts rather than helps. This situation can continue indefinitely.

SITUATION 2: MPLS TRANSPORT

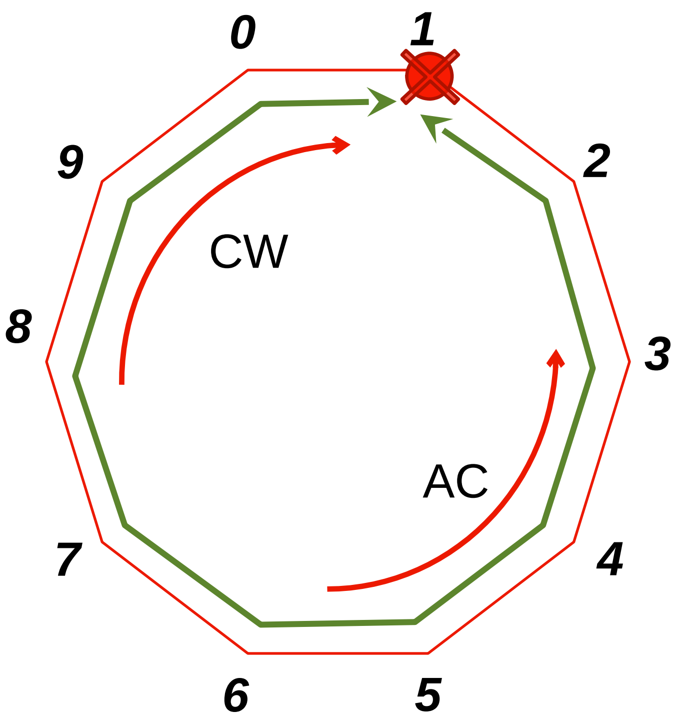


This is “standard” FRR for links, using RSVP-TE or xLFA.

P2 activates FRR if link L1 fails; P5 does the same if link L2 fails

However, if both links L1 and L2 fail, either because of fate-sharing or happenstance, both P2 and P5 activate FRR, causing congestion on the link between them until the LSP is rerouted

SITUATION 3: RMR NODE FAILURE



In RMR, protection is done by creating two counter-rotating LSPs to each egress node.

If node 5 sends traffic to node 1, it can choose to send it clockwise or anticlockwise.

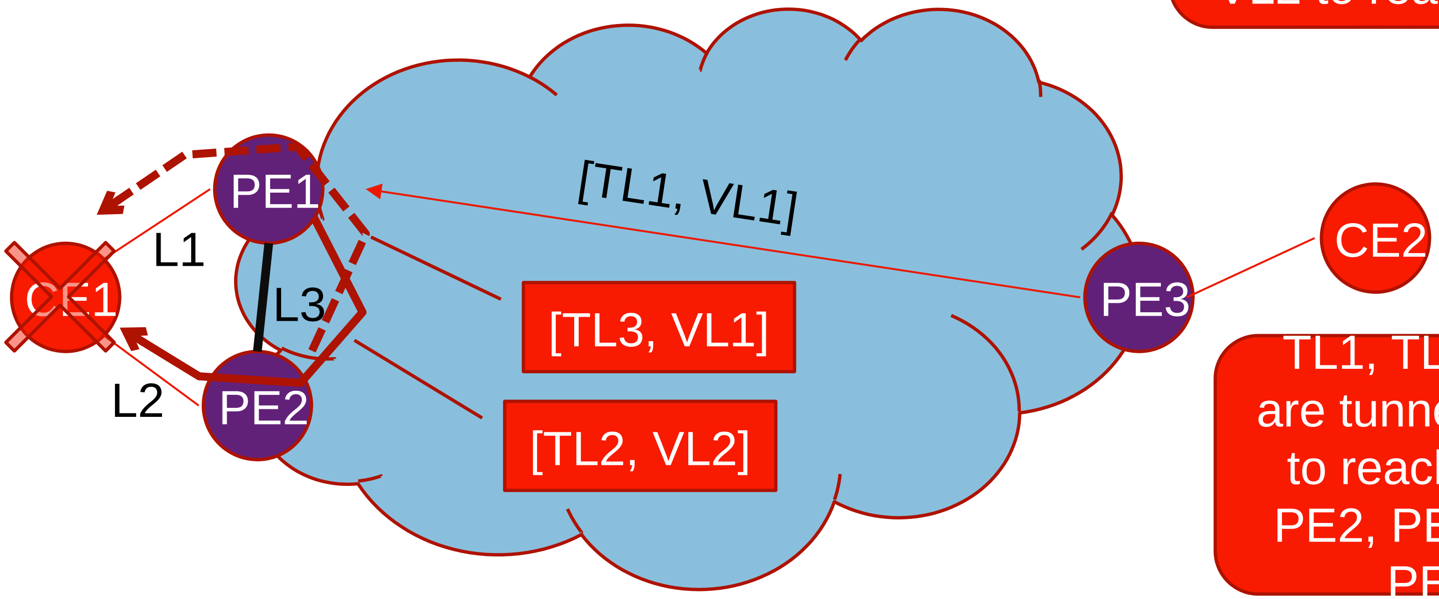
If it chooses CW, and the link between node 7 and node 8 fails, node 7 sends the traffic AC to node 1.

However, if node 1 (the egress) fails, node 0 will protect by sending traffic AC; node 2 will protect by sending traffic CW, and traffic will loop forever

Some mitigation strategies are suggested in the RMR draft, but no good solution.

SITUATION WITHOUT NFFRR

PE1 advertises VL1 to reach CE1; PE2 advertises VL2 to reach CE1



TL1, TL2, TL3 are tunnel labels to reach PE1, PE2, PE1 from PE2

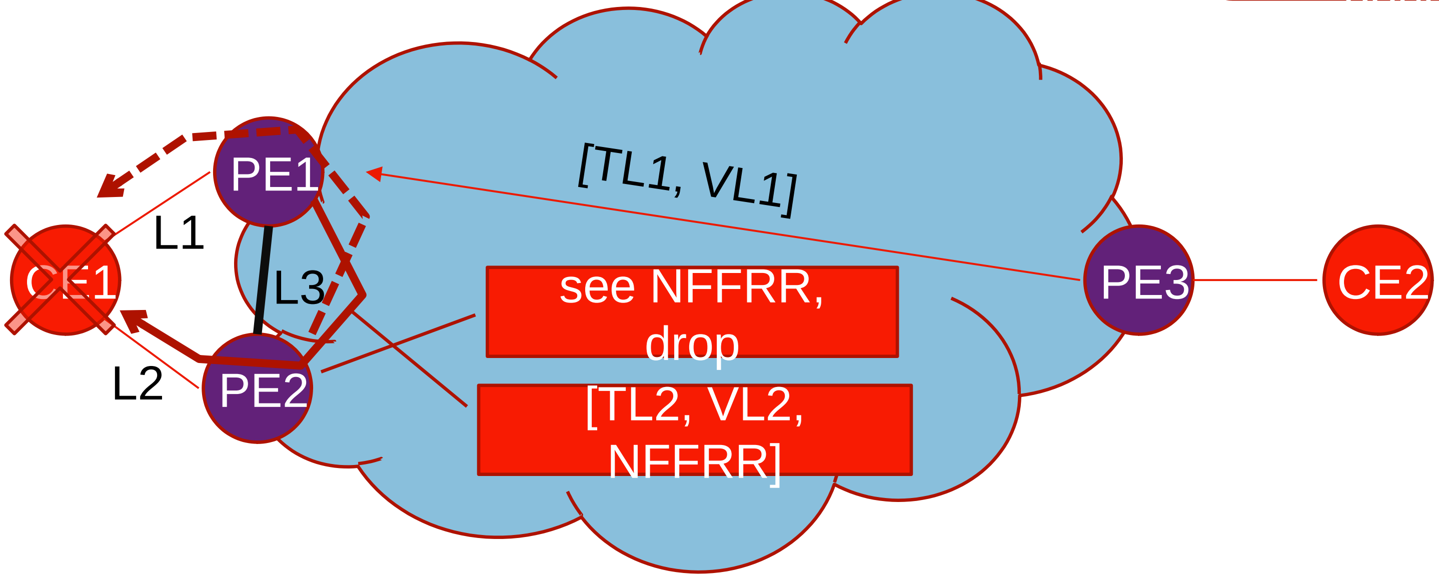
PE3 sends a packet to CE1 via PE1, with label stack [TL1, VL1]

PE1 pops the tunnel label (or PHP applies), sees VL1 and tries to send the packet to CE1 via L1. PE1 sees that L1 is down, sends the packet via PE2 with label stack [TL2, VL2]

PE2 behaves similarly, sending the packet back to PE1. Loop!

SOLUTION: WITH NFFRR SPL

NFFRR: an SPL to indicate FRR has been done



see NFFRR, drop

[TL2, VL2, NFFRR]

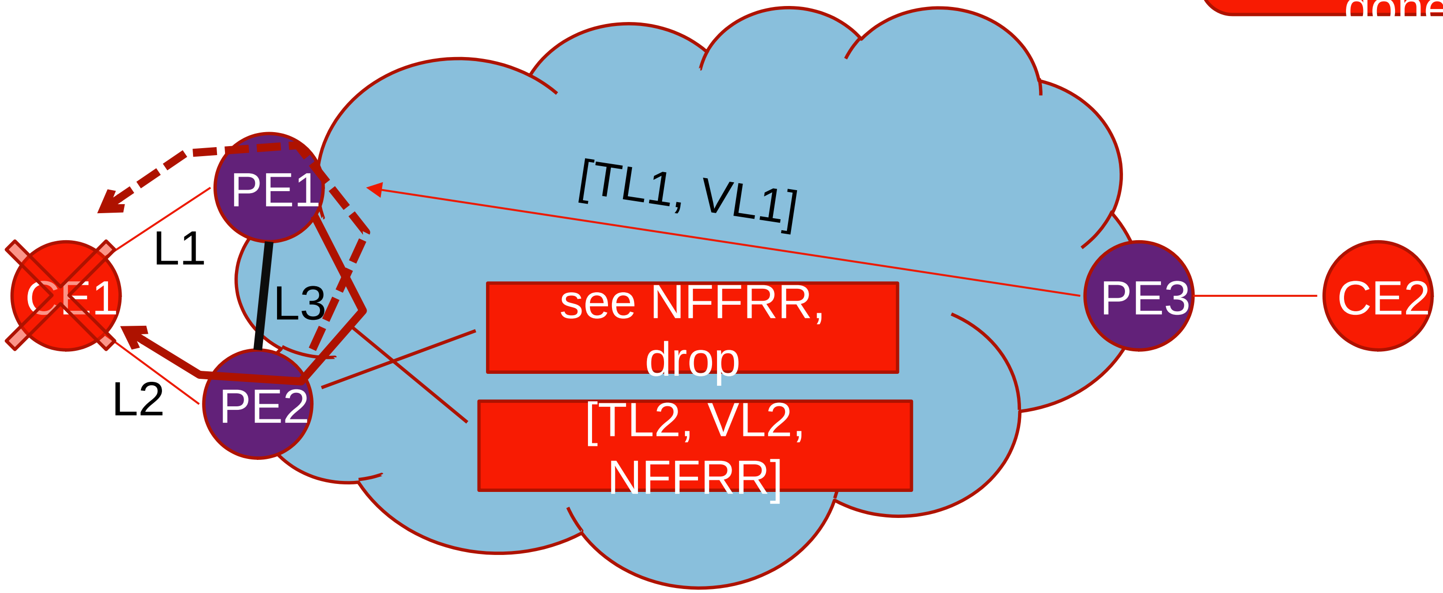
PE3 sends a packet to CE1 via PE1, with label stack [TL1, VL1]

PE1 pops the tunnel label (or PHP applies), sees VL1 and tries to send the packet to CE1 via L1. PE1 sees that L1 is down, sends the packet via PE2 with label stack [TL2, VL2 , NFFRR]

With PHP, PE2 sees [VL2, NFFRR], pops VL2 and notes dest is CE1 via L2, notes and pops NFFRR. Since L2 is down, PE2 wants to do FRR, but since NFFRR was present, PE2 drops the packet.

SOLUTION: WITH ANOTHER ALLOCATED

NFFRR: an SPL to indicate FRR has been done



PE3 sends a packet to CE1 via PE1, with label stack [TL1, VL1]

PE1 pops the tunnel label (or PHP applies), sees VL1 and tries to send the packet to CE1 via L1. PE1 sees that L1 is down, sends the packet via PE2 with label stack [TL2, VL2 , NFFRR]

With PHP, PE2 sees [VL2, NFFRR], pops VL2 and notes dest is CE1 via L2, notes and pops NFFRR. Since L2 is down, PE2 wants to do FRR, but since NFFRR was present, PE2 drops the packet.

LABEL FOR NFFRR

- Allocate an SPL for indicating NFFRR
  - Why SPL, not ESPL?

    If the protection path uses a label stack (as in SPRING), then one would potentially have to insert the NFFRR indication with each label in the stack. With ESPL, this would be expensive
  - Make the solution generic for different multihomed service (IP-VPN or EVPN, etc) or MPLS based FRR: when NFFRR label is present, no further FRR.
  - Control plane signals NFFRR capability, but no need for other extensions to signal NFFRR in each individual protocol space.
- Request an early allocation (label value 8).
  - To prototype the solution, and to play with different implementations

NEXT STEPS

- Discussion of this draft on the list
- Companion documents to signal the ability to process NFFRR (BGP and IGP)
- Examination of techniques and algorithms to better understand under what circumstances to use NFFRR
  - Pessimistic use of NFFRR may prevent FRR and lead to unnecessary packet loss …
  - … whereas optimistic non-use may lead to worse problems