Proxy MAC-IP Advertisement in EVPN

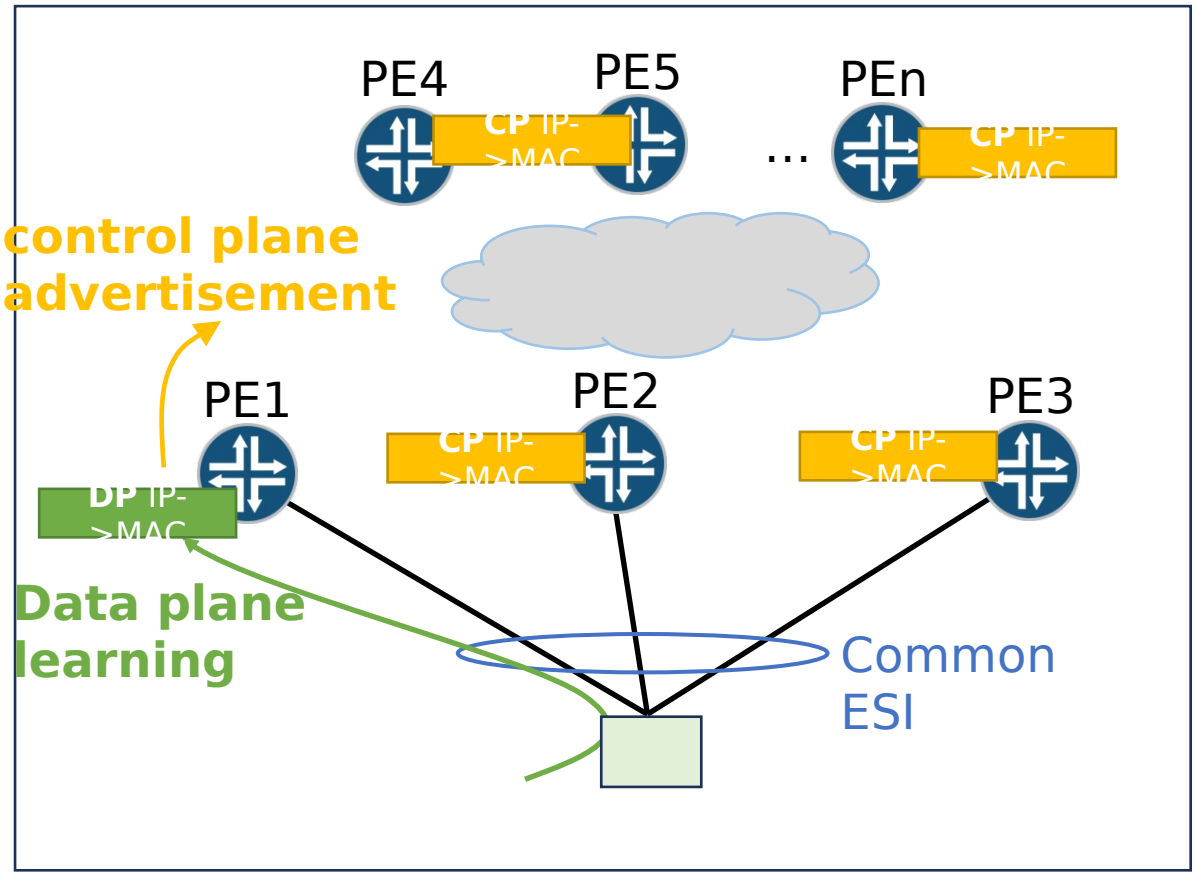
draft-rbickhart-evpn-mac-ip-proxy-adv-03

Ryan Bickhart (Juniper)
Wen Lin (Juniper)
John Drake (Juniper)
Jorge Rabadan (Nokia)
Alton Lo (Arista)
Patrice Brissette (Cisco)

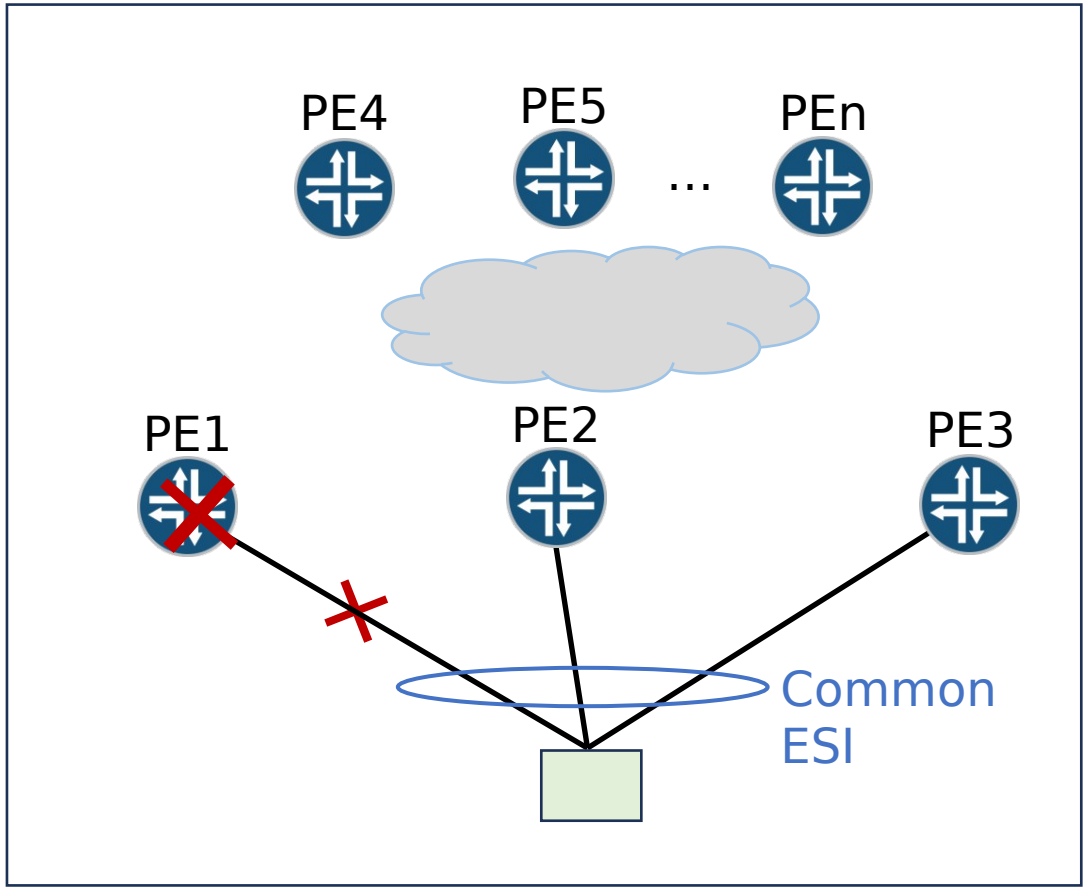
IETF 121, Mar 2024
Dublin, Ireland

MAC-IP Binding

- MAC-IP bindings in EVPNs may be locally learned in the data plane only on one of multihoming PEs



- IF PE1 suffers either node or egress link failures, the MAC-IP binding info is all lost from the EVPN PEs



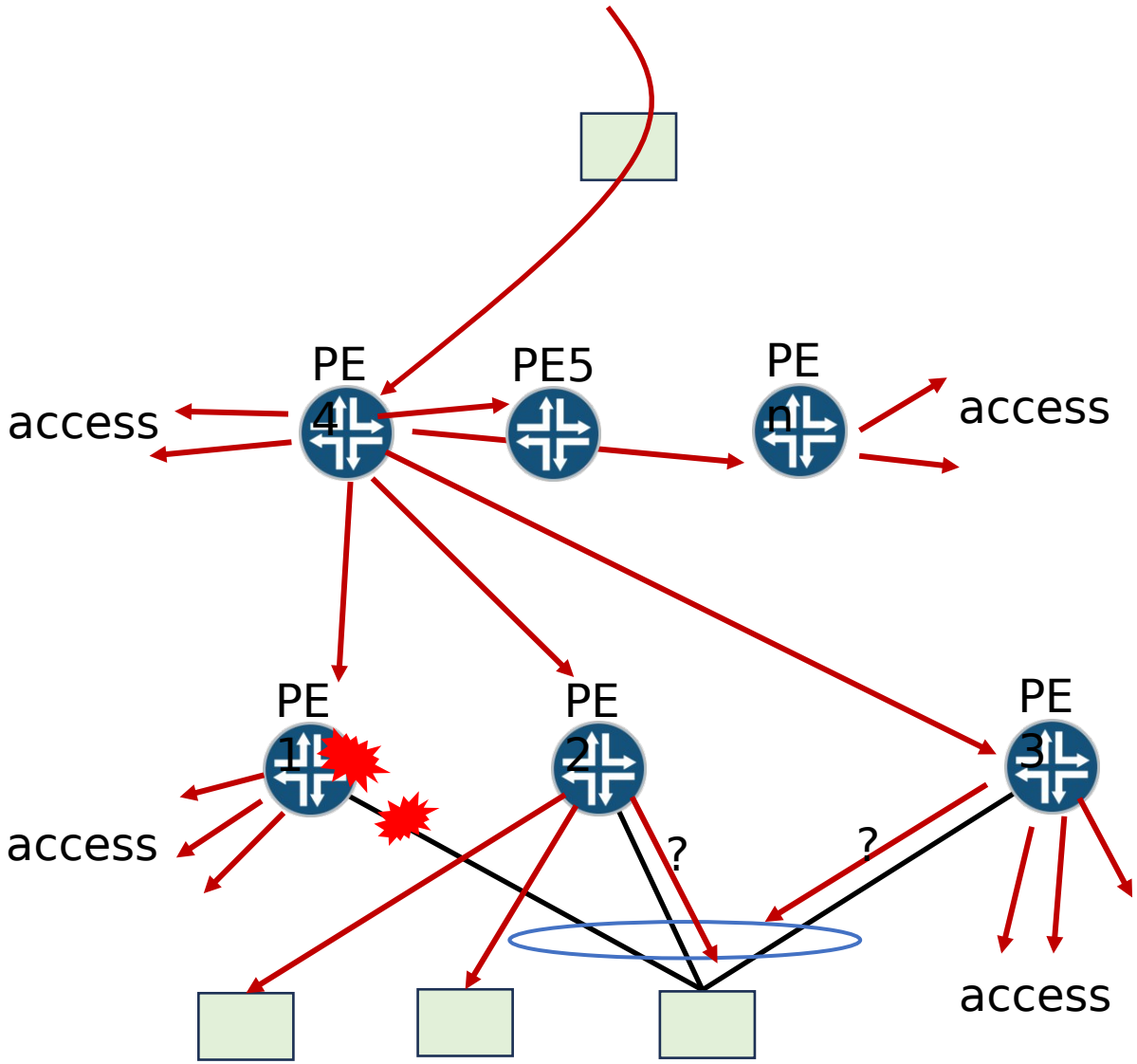
Problem to Address

Traffic that is not in-flight before MAC-IP is re-learned:

- Excessive flooding
- Waste of bandwidth
- Potential traffic blackholing

How to address this?

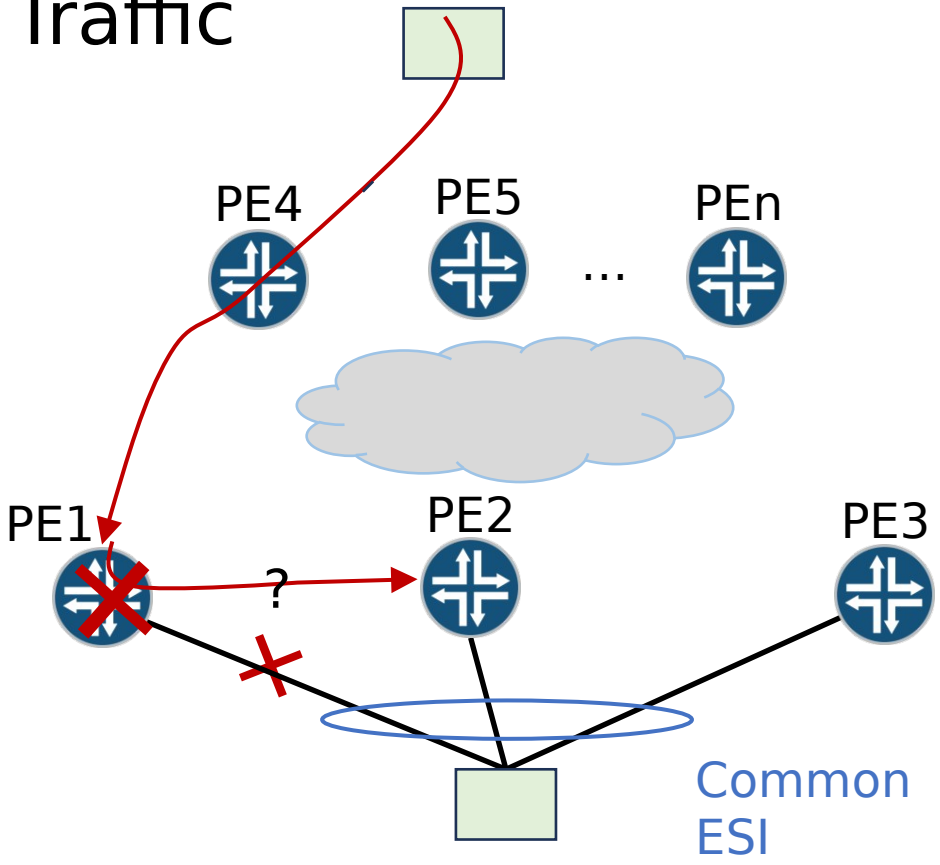
- Deal with it globally – on all EVPN PEs?
  - Control plane involvement
  - Potential Race condition on EVPN PEs?
- Deal with it “locally” – only on its peer multihomed PEs?
  - Make it transparent to the rest of PEs
  - Host is local to MH PEs, simpler.



This Draft is Not About…

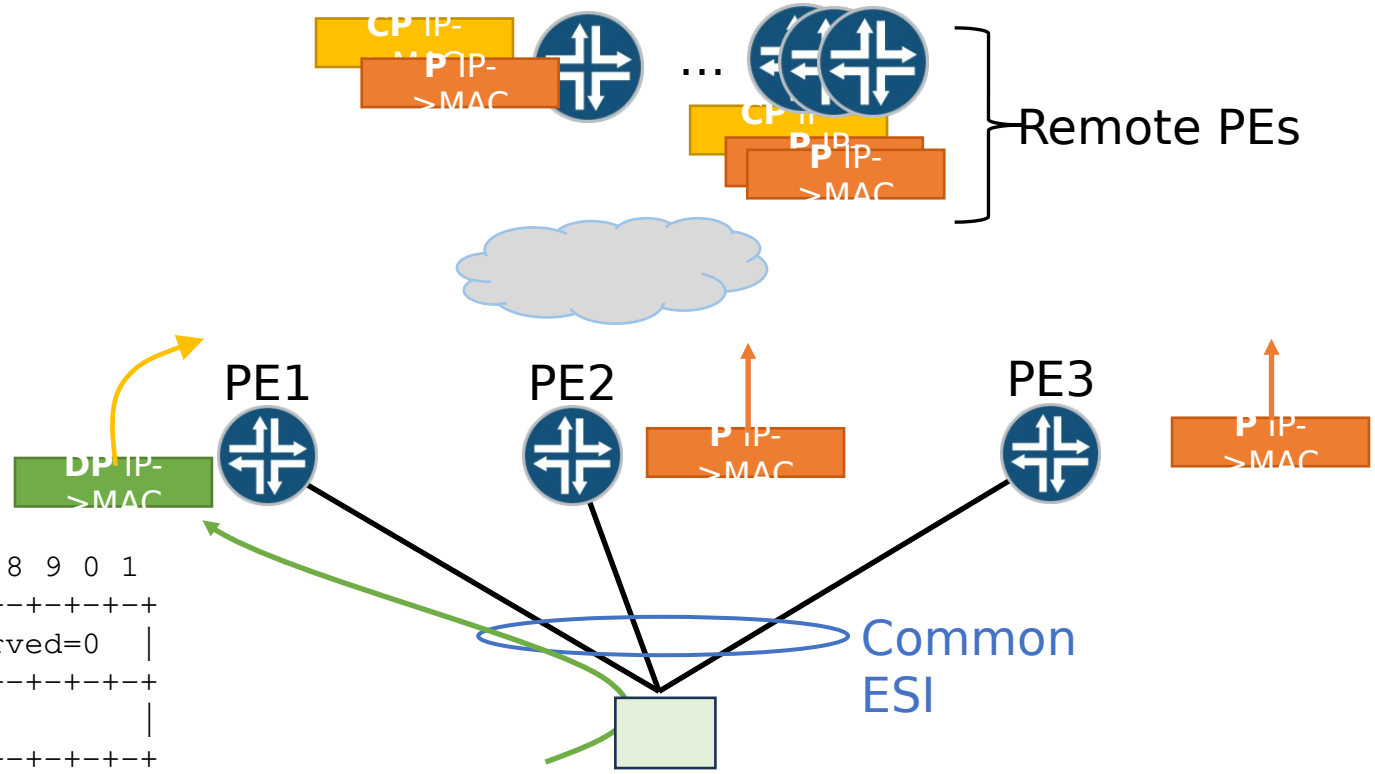
Addressing Fast Reroute for In-flight Traffic

- Egress link failure: Egress Link Protection behaves as is on PE1
- Node failure: lack of FRR support for EVPN node failure today, still an area to conquer.



Solution – Make It Transparent to the Remote PEs

- Upon receiving the control plane learned MAC-IP, peer MH PEs also originate their own advertisements for type-2 route: MAC-IP marked with a proxy indication



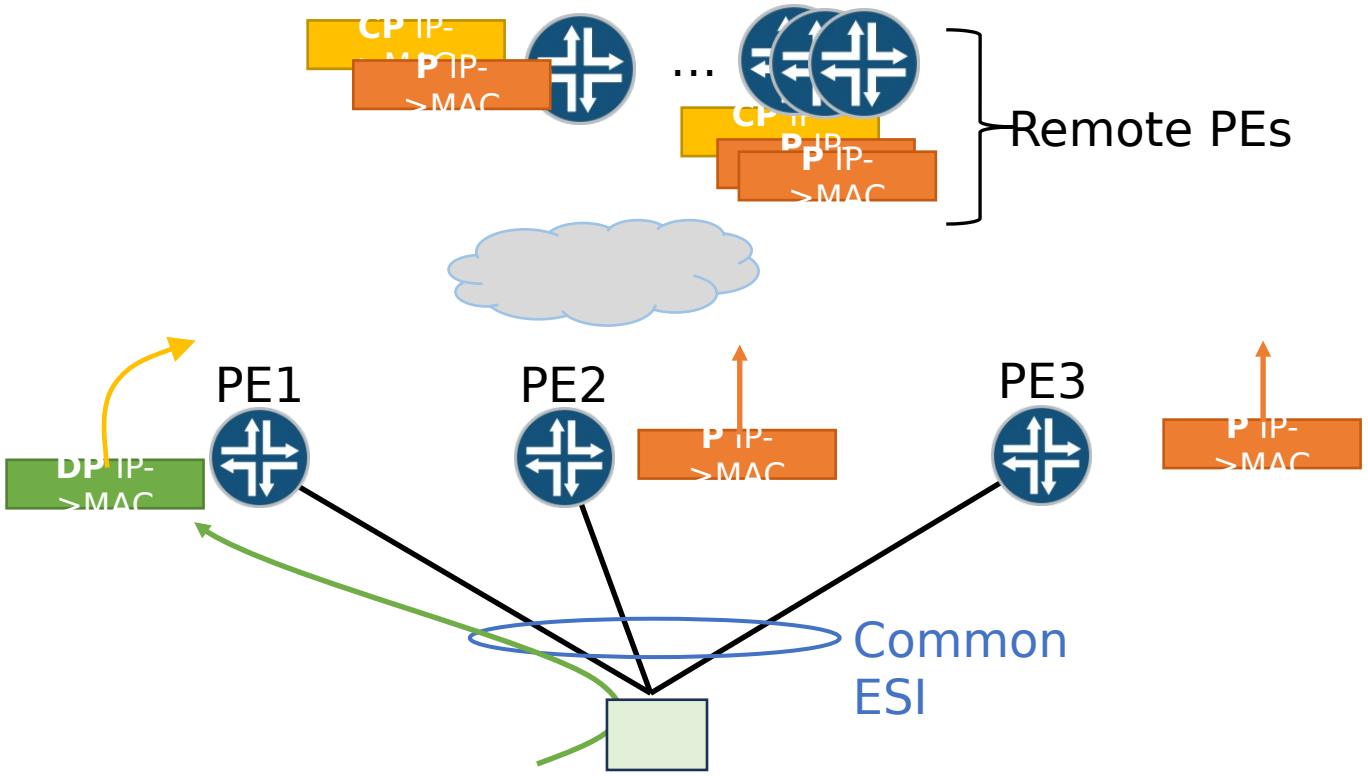
- EVPN ARP/ND Extended Community:

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=0x06     | Sub-Type=0x08 |Reserved |P|O|R|   Reserved=0  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                              Reserved=0                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Bit Name: P -> Proxy MAC-IP advertisement

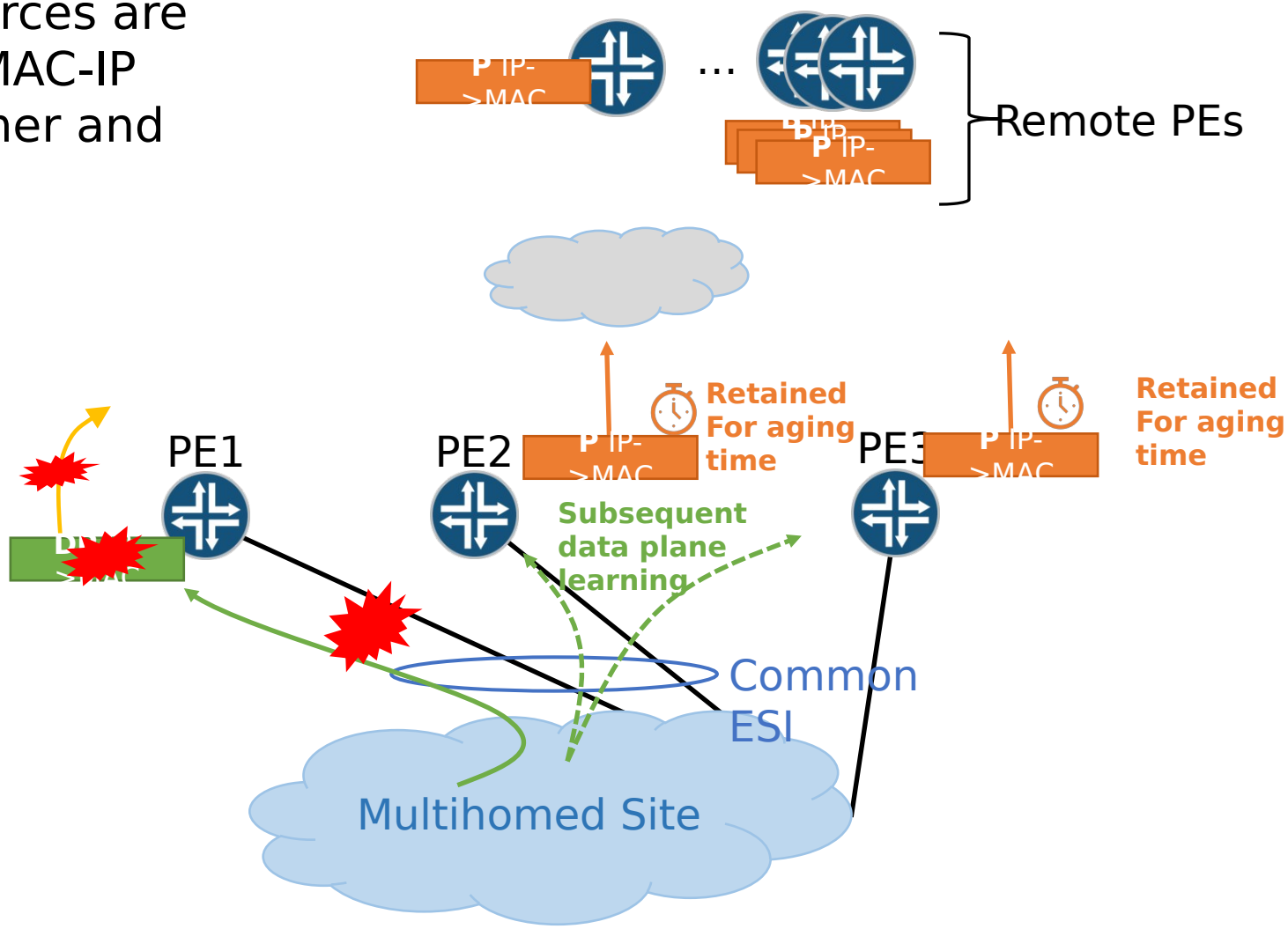
Solution – Remote PE Behavior

- Remote PEs receive the proxy MAC->IP advertisements and give them standard treatment



Solution - Failure Handling and MAC-IP relearning

- Upon failure, non-proxy MAC-IP sources are withdrawn, PE2/ PE3 keep proxy MAC-IP advertisement, but start an age timer and trigger ARP requests
- If MAC-IP learned from non-proxy source before aging timer expires, cancel the aging timer
- Otherwise, when aging timer expires, withdraw proxy MAC-IP advertisement



Summary for the Solution

- Close the gap between the link / node failure and the subsequent relearning of the MAC->IP on one of the remaining multihoming PEs.
  - Avoid unnecessary churn in the network
  - Remote PEs operate as usual, no need for upgrade

Other facts:

- Extra MAC+IP EVPN Type 2 advertisements and processing in the control plane – could be a scale concern for control plane
- It does not increase data plane footprint

Current State and Next Step

- Known to be implemented by multiple vendors, FRR and SONIC, deployed in the fields for a couple of years.
- Would like to seek WG adoption at this point
- Will continue to seek comments