

draft-boutros-bess-elan-services-over-sr-01

IETF 109 Online

November 19, 2020

Siva Sivabalan (Ciena)

Jim Uttaro (AT&T)

Daniel Voyer (Bell Canada)

Ben Win (Comcast)

Himanshu Shah (Ciena)

Luay Jalil (Verizon)

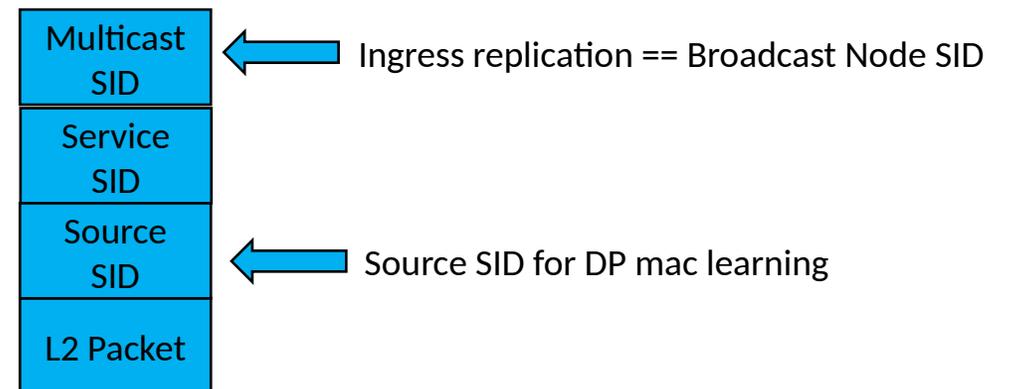
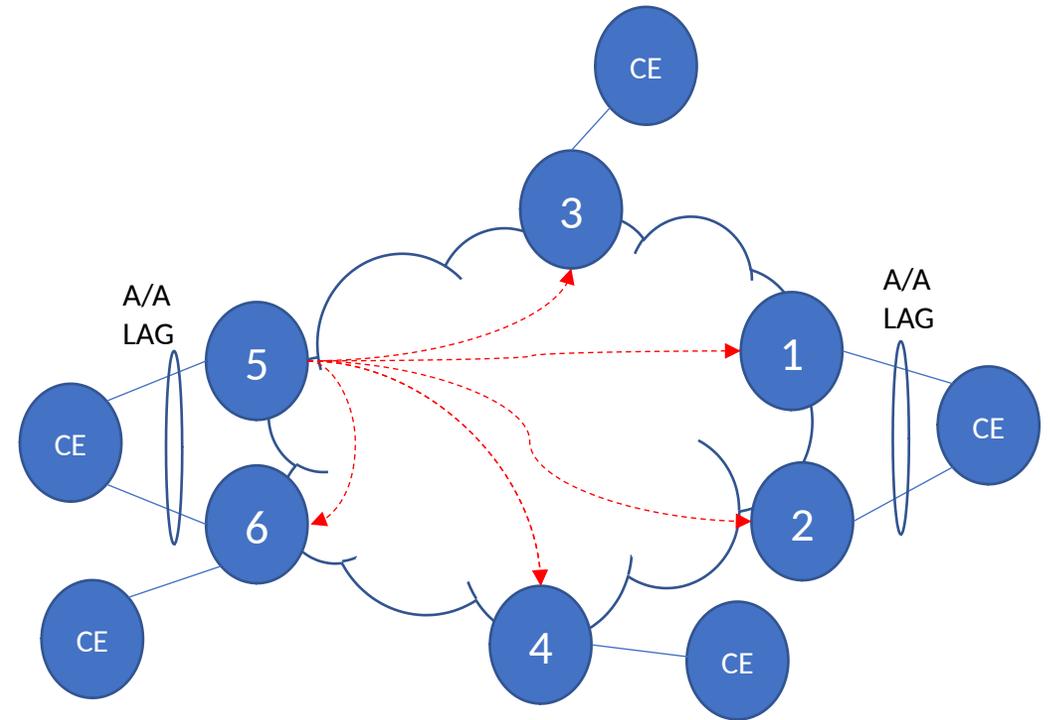
Sami Boutros (Ciena)

PW History

- Historically, a Pseudo-Wire (PW) is a P2P service between 2 endpoints.
 - A **PW** context in both control and data-plane (i.e. the MPLS label) represents **both the service and the service endpoint**.
 - This led to **scale issues** specially with ELAN service, where a 10,000 service distributed over 100 endpoints would need 100K+ PW labels (service IDs) per endpoint.
 - PW(s) follow L2 semantics had **no Active/Active** redundancy.
- SR-Optimized ELAN with data-plane MAC learning:
 - **Improves** the **scale** issue (e.g., 10,000 services will be presented by only 10,000 Service SIDs regardless of how many endpoints participate in the service).
 - **Maintains** the PW **P2P semantics** between 2 endpoints by presenting the endpoint by another SID under the service SID in the SID list.
 - Solves the **Active/Active Redundancy** and multipathing using Segment Routing **anycast** SIDs.

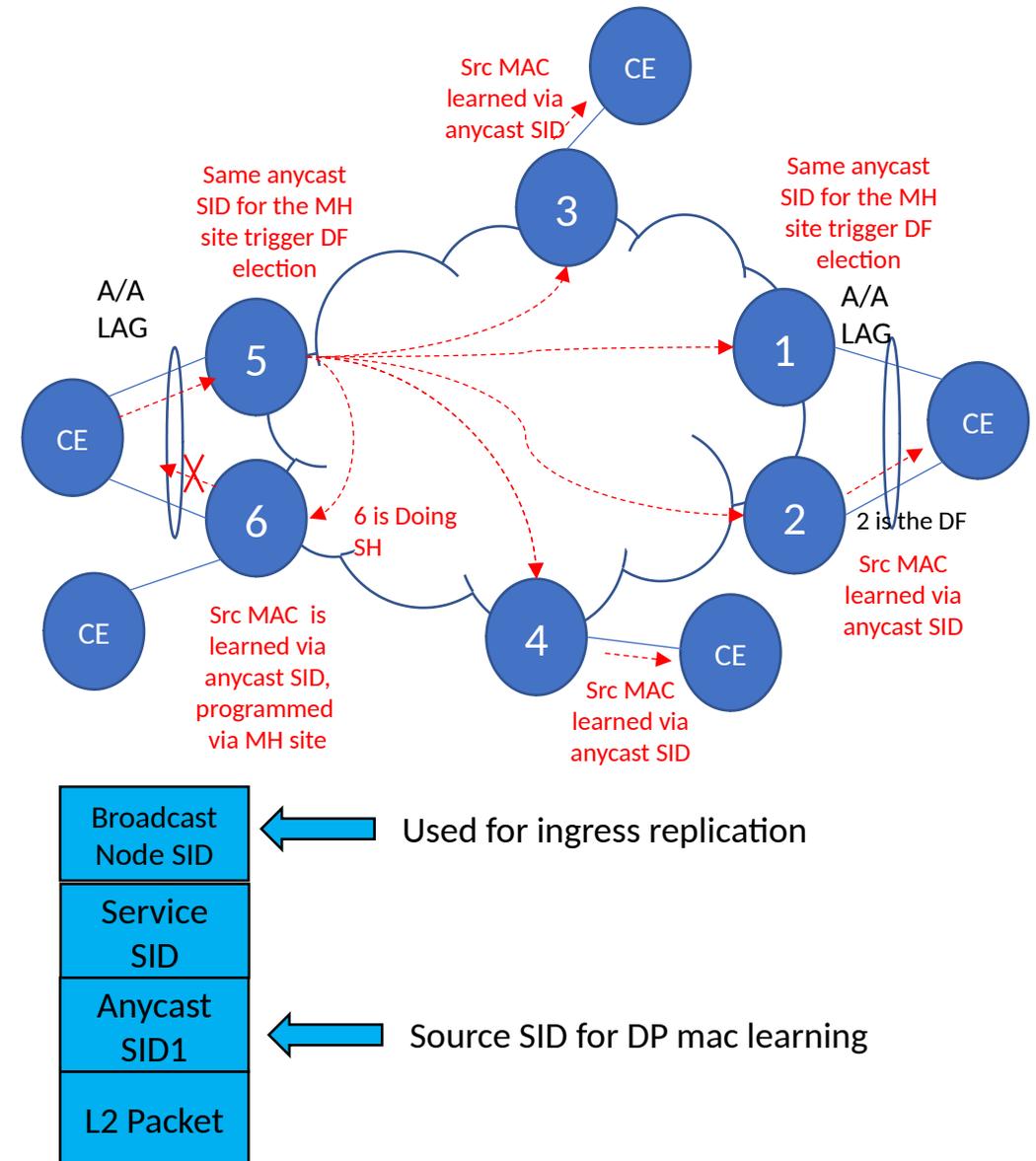
SR-Optimized ELAN Service discovery

- ELAN **Service SID advertised by BGP** for service auto discovery:
 - Such that a single route contains a bitmap of all service SIDs as well as the Broadcast Node SID (for BUM traffic) associated with the advertising node.
- Upon receiving BGP update, a node can **discover** the service(s) hosted on the advertising node and hence can build P2MP flooding trees for L2 BUM traffic.
- P2MP flooding trees can be built for a given service or a group of services (**aggregate inclusive**)
- **Ingress replication per service** can be performed using broadcast SID.



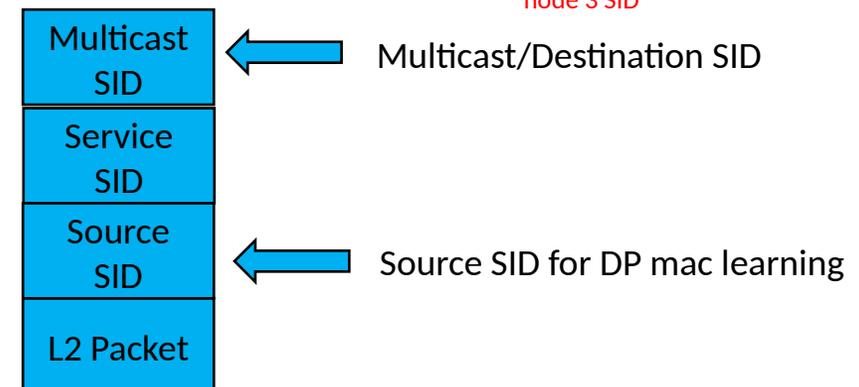
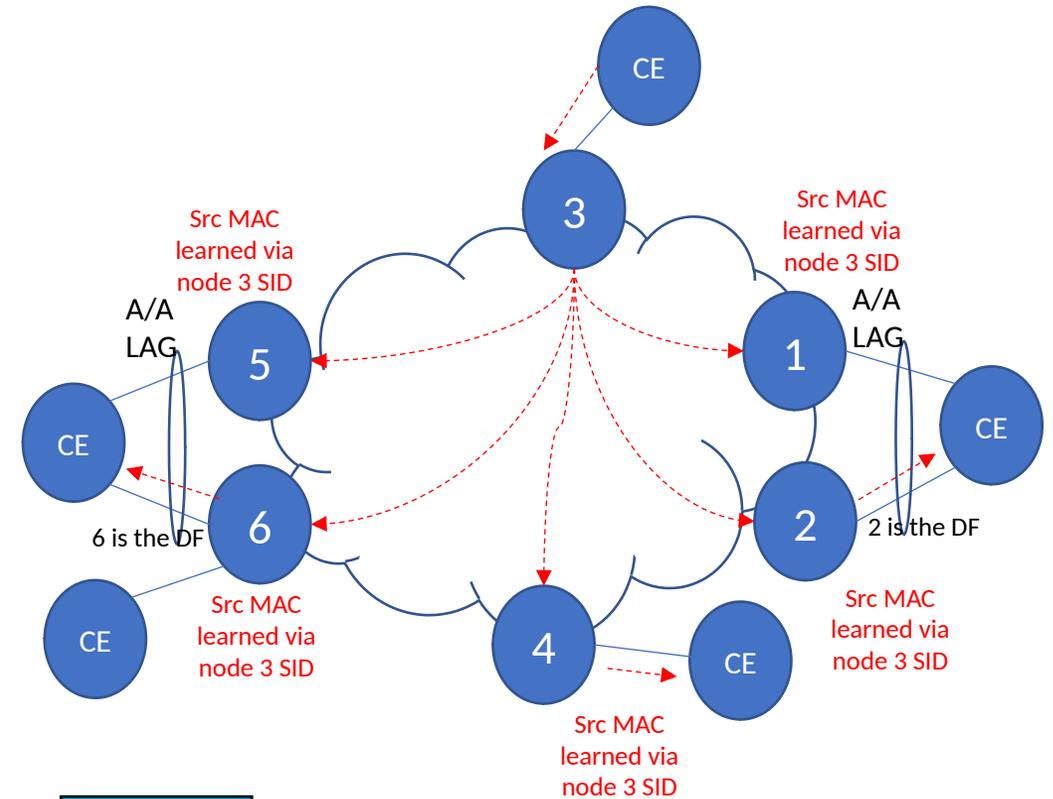
SR-Optimized ELAN service A/A Redundancy

- Anycast SID per Ether Segment (ES) is configured on all nodes attached to the MH site and advertised by the nodes connected to a Multi-Home site.
- Each node attached to the MH site advertises the **same anycast SID** to allow other nodes to discover the group membership and **perform DF election**.
- **Aliasing/Multi-pathing** achieved using the same mechanisms used for anycast SID.
- For example, **node 5 learns a MAC address from the CE** in dataplane and **floods the BUM data packet** to all other nodes including node 6. Thus, a receiver can learn the **MAC as reachable via the anycast SID** configured on node 5 and node 6 in “data-plane”
- **Node 6 applies Split Horizon** and hence does not send the received data packet back to the MH CE but **programs the MAC as reachable via the MH CE**.



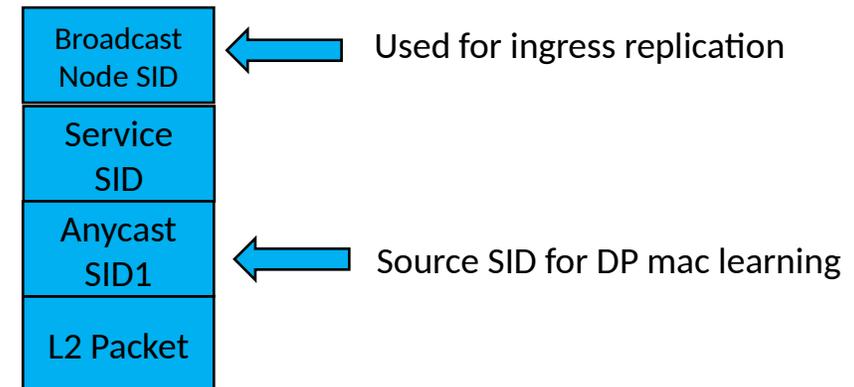
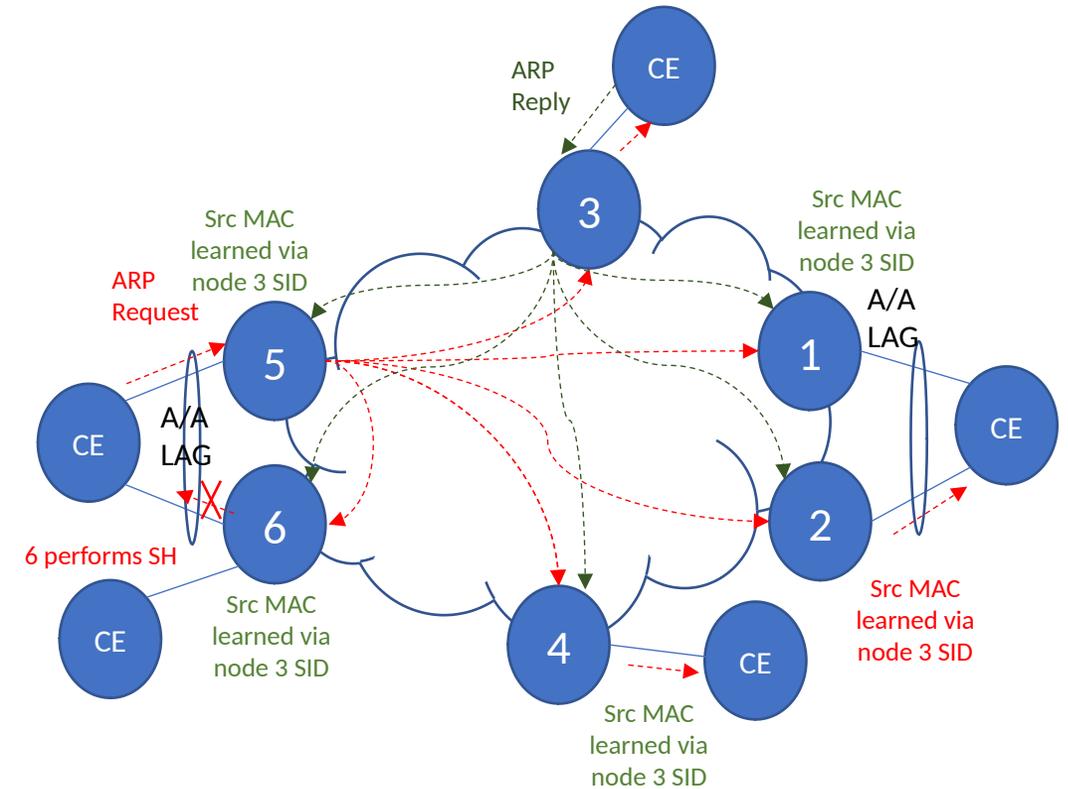
SR-Optimized ELAN service Data-Plane MAC learning

- MAC learned in data-plane **against** the **source node SID**, encapsulated under the service SID in the L2 forwarded packets.
- **Node 3** learns the CE **MAC** address and **floods** the BUM packet to **all nodes** configured with the **same service SID**.
- Node 1, 2, 4, 5 and 6 **learn** the **MAC** as **reachable** via the source **node SID** hosted on **node 3**.



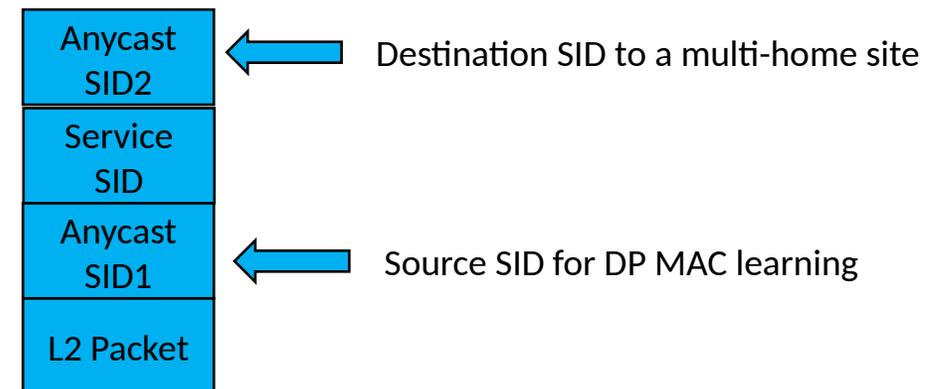
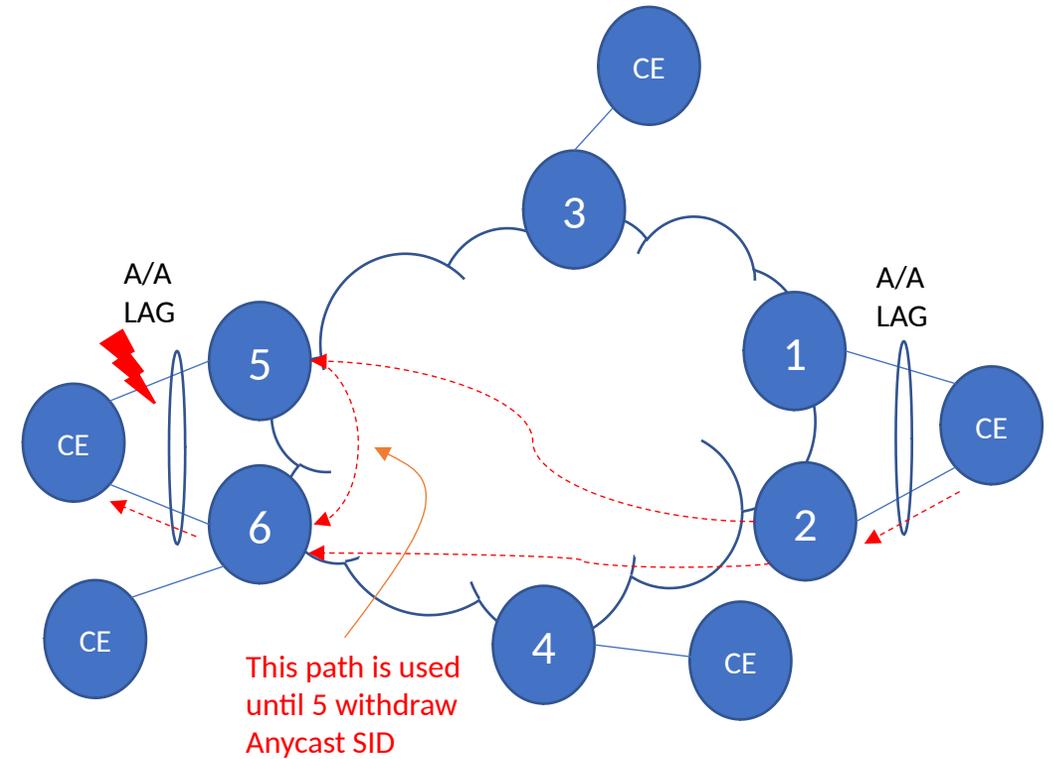
SR-Optimized ELAN service ARP suppression

- **Cleaning** ARP packet **requests** and **replies** can be used to learn IP/MAC binding for ARP suppression.
- ARP **replies** are **unicast** however **flooding** ARP **replies** can allow all nodes to learn the MAC/IP bindings for the destinations too.



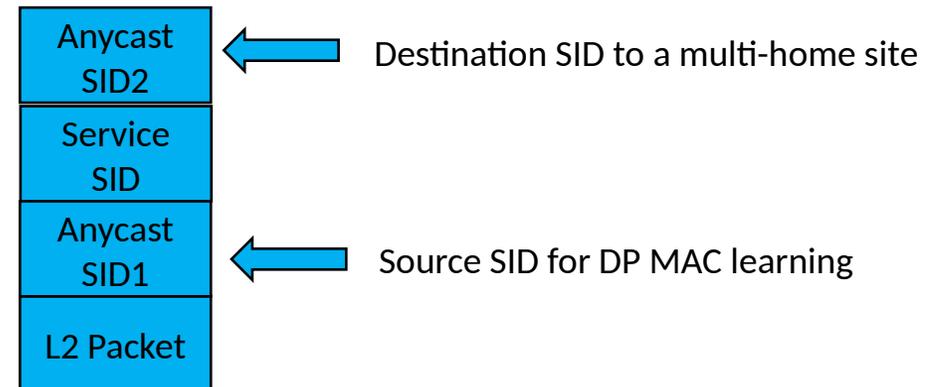
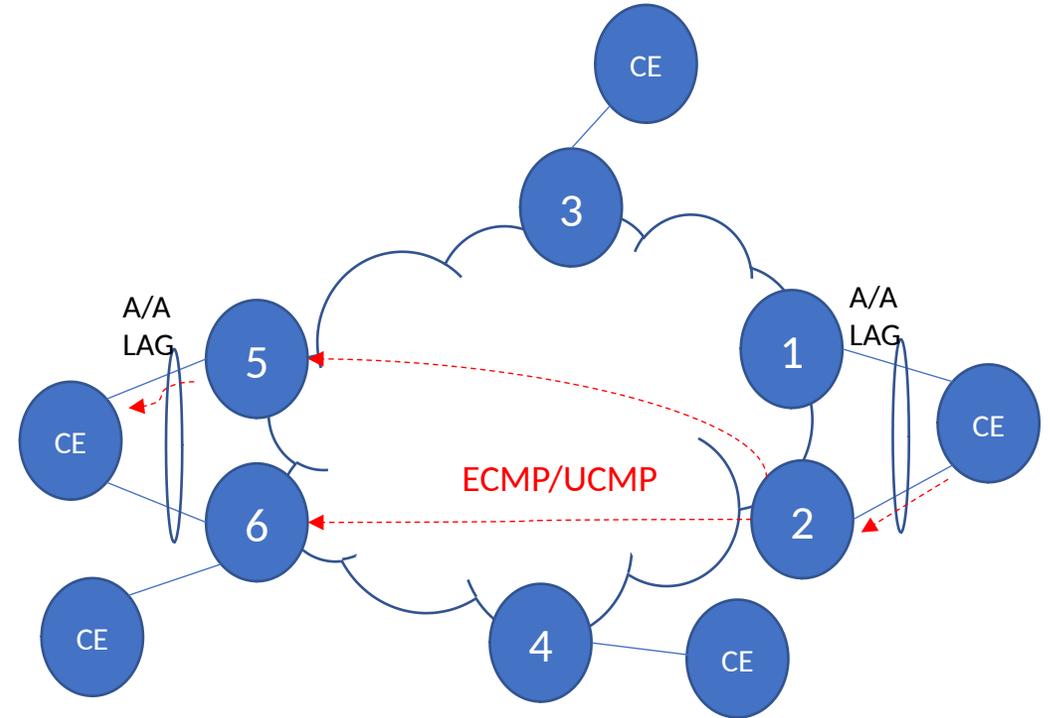
SR-Optimized ELAN service Mass withdrawal

- Node failure is learned via BGP/IGP, and there **is no need for additional MAC withdrawal mechanism.**
- On **link failure** the **node** can **withdraw** in IGP/BGP the **anycast SID** associated with the site so as not to receive any packets destined to the MH site.
- On **link failure** between node 5 and CE, **node 5** can **forward** the received L2 **packet from the core to node 6** (using anycast SID shared by both 5 and 6) **for fast convergence until it withdraws the anycast SID** associated with the MH site.



SR-Optimized ELAN service ECMP Multi-pathing

- Packets destined to the MH CE connected to node 5 and node 6 can be **load-balanced (ECMP/UCMP)** across the core given that the MAC addresses were **learned** via **anycast SID** hosted node 5 and 6.



Benefits of SR-Optimized ELAN

- Maintain **data-plane MAC learning benefits** such as **fast convergence**, **fast MAC move**, and **scale** through **conversational learning**.
- Bring the benefits of **A/A multihoming**, **multipathing**, and **ARP suppression**.
- **Simpler** and much better **control plane** scale over legacy PWs, by splitting the endpoint ID from the service ID and representing them by 2 SIDs in the SID segment list.
- **Leverage** the benefits of **Segment Routing anycast SID** for redundancy and **fast convergence**, and to discover nodes sharing the same anycast SID to perform DF election.
- **Eliminate** the need for any **overlay fast convergence mechanism**.

**Thank
You**