EVPN Anycast Multi-Homing

draft-rabnag-bess-evpn-anycast-aliasing-02

Jorge Rabadan (Nokia)
Kiran Nagaraj (Nokia)
Alex Nichol (Arista)
Ali Sajassi (Cisco)

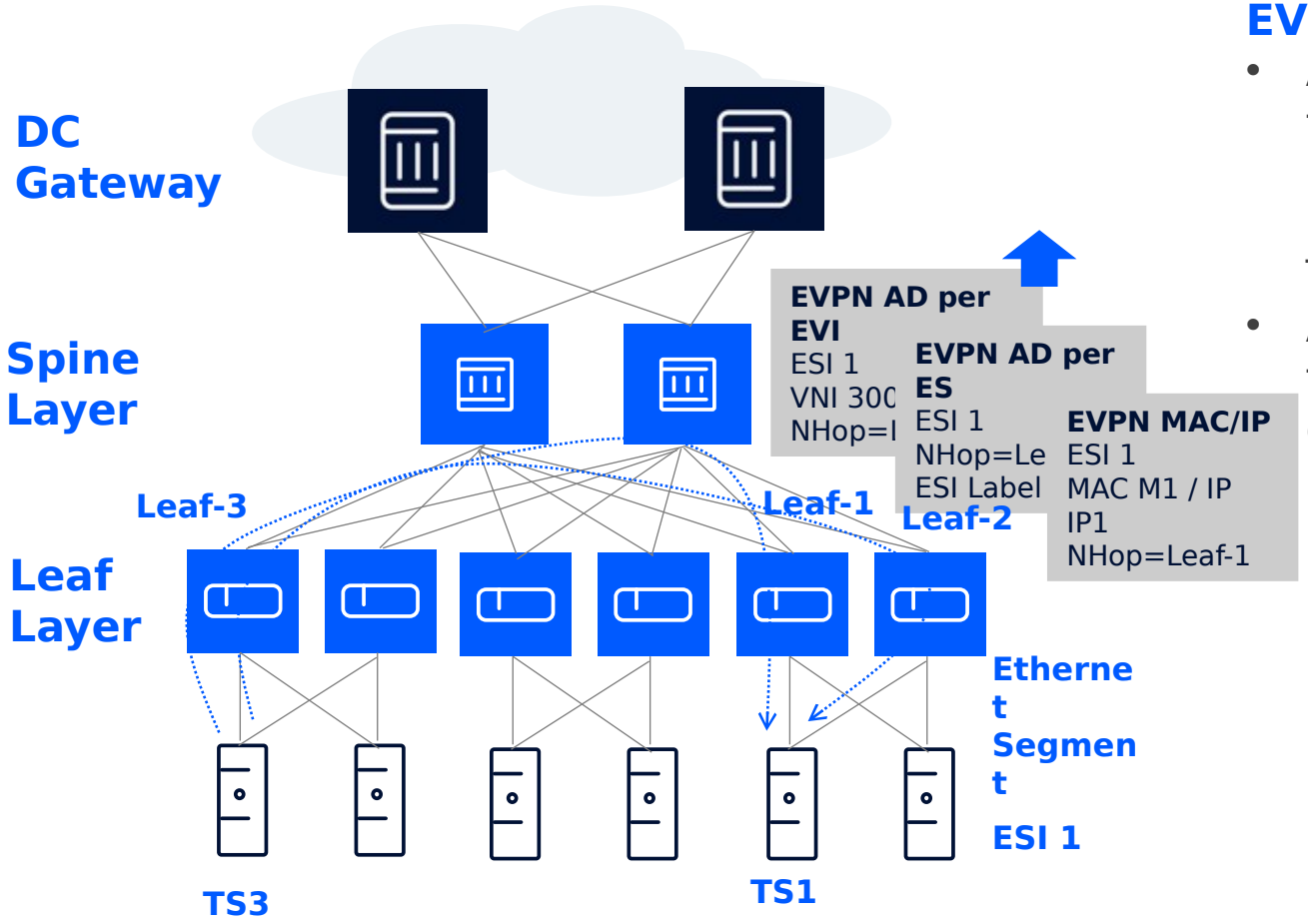
IETF121, Nov 2024
Dublin

Agenda

- Problem Statement refresh
- Changes since rev 00
- Next Steps

Problem Statement

draft-rabnag-bess-evpn-anycast-aliasing-02



EVPN Multi-Homing

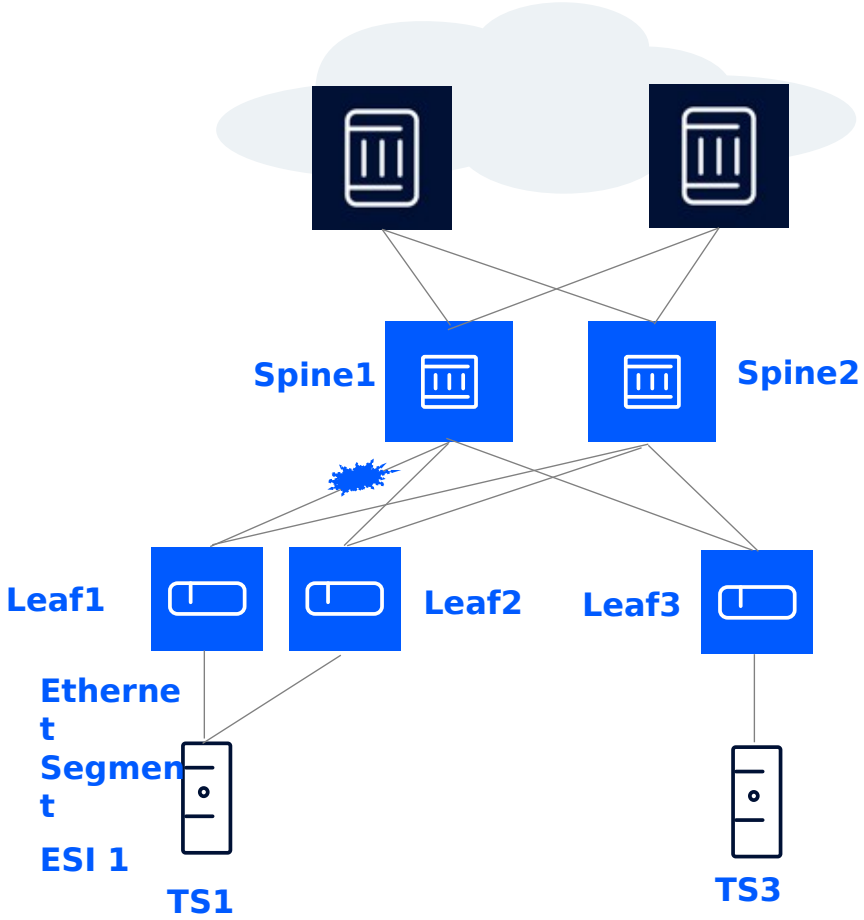
- Aliasing is an all-active multi-homing procedure that makes sure unicast traffic from TS3 to TS1 is per-flow load balanced to Leaf-1/Leaf-2 irrespective of the MAC/IP Advertisement route for TS1 not being advertised from both Leaf nodes
- Aliasing on Leaf-3 creates an “overlay ECMP-set” for ESI 1 in addition to the “underlay ECMP-set” to each VTEP of the ES

Challenges in very Large DCs

- Control Plane Scale
- Convergence and Processing overhead
- Inefficient forwarding during a failure in the underlay

Problem Statement: Inefficient forwarding during failures

draft-rabnag-bess-evpn-anycast-aliasing-02



Failure Example

- Leaf1-to-Spine1 fails
- In-flight VXLAN packets from Leaf3 with destination VTEP Leaf1 arrive at Spine1 and are rerouted via e.g., Leaf2->Spine2->Leaf1->TS1, while they could go directly via Leaf2->TS1.
- After the underlay routing protocol converges, all VXLAN packets with destination VTEP Leaf1 are correctly sent to Spine2 and Leaf3 removes Spine1 from the underlay ECMP-set for Leaf L1.

Changes in rev 02 compared to previous versions

draft-rabnag-bess-evpn-anycast-aliasing-02

Simplified procedures agreed among co-authors (rev 02)

- The goal is to match the functionality of MC-LAG and not the added complexity in previous versions
- A single anycast VTEP is used for all the Ethernet Segments attached to the same group of multi-homed PEs

Other procedures in previous versions moved to Annex (rev 02)

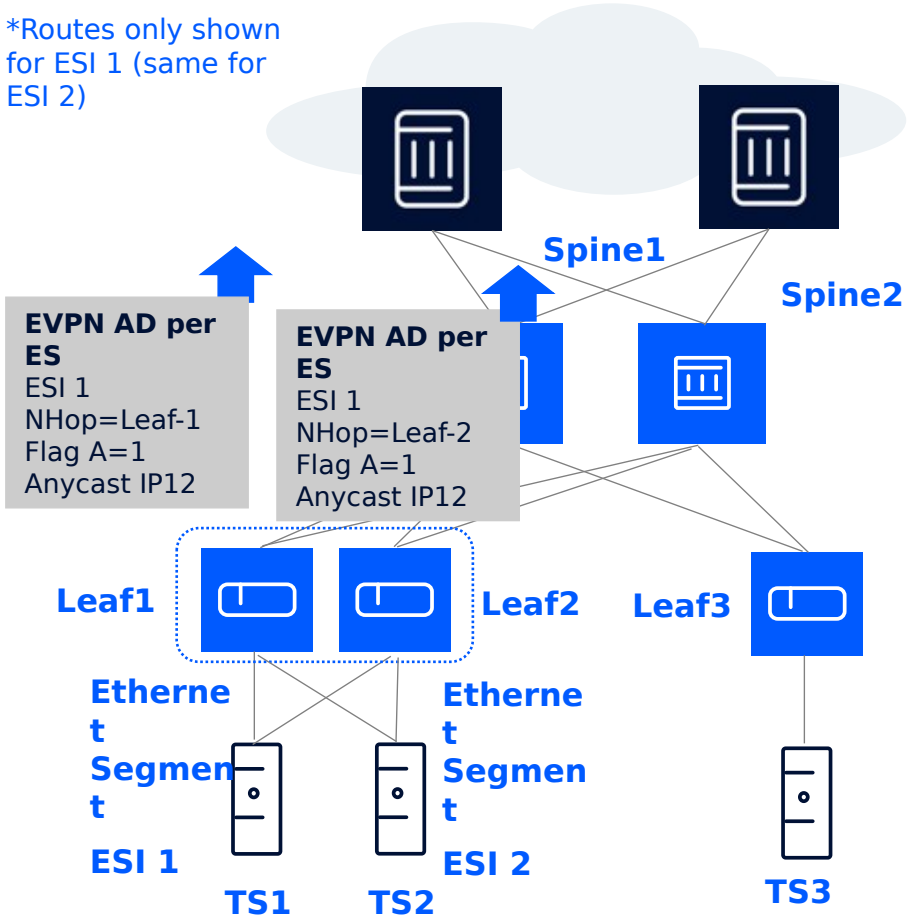
- Annex added for documentation purposes only, and it will be removed from the document before publication.
- Co-authors agreed that they introduce additional complexity and are therefore excluded, as they undermine the primary goal of using anycast VTEPs, which is to simplify EVPN operations.

Operational considerations section (added in rev 01, modified in rev 02) based on feedback

- Analyses how anycast VTEPs can help when eBGP is used in the underlay (RFC7938) and no allowas-in is used on the spines
- Warns about the operational considerations of using “anycast” multi-homing as opposed to “unicast” multi-homing

Anycast Multi-Homing Solution Procedures

draft-rabnag-bess-evpn-anycast-aliasing-02



What procedures are modified

- Only the Aliasing procedure on a per ES basis, based on the A flag
- Only if the ES NVEs use the same VNI/label

Egress NVEs (e.g. Leaf1 and Leaf2)

- NVEs configured for anycast MH mode, and with an anycast VTEP used for all ESes on the redundancy group (e.g. IP12)
- Anycast VTEP reachability advertised in the underlay
- A-D per EVI routes are suppressed for the ES
- In case of ES link failure on Leaf1, A-D per ES / ES route are withdrawn, and unicast traffic attracted to Leaf1 is FRR'ed to Leaf2

Ingress NVEs (e.g. Leaf3)

- Upon receiving MAC/IP route for e.g. TS1 with ESI1, programs TS1 with destination ESI1 (resolved to ESI1 anycast VTEP IP12)
- Incoming frames to TS1 encapsulated using IP DA = IP12 and use the underlay ECMP-set to IP12
- OPTIONAL: ES destination changed to unicast VTEP if only one egress Leaf remains active on the ES

Next Steps

draft-rabnag-bess-evpn-anycast-aliasing-02

- Capture more feedback from WG
- Add applicability to IP Aliasing and SRv6
- Solicit WG adoption

Thank you