EVPN Anycast Multi-Homing

draft-rabnag-bess-evpn-anycast-aliasing-04

Jorge Rabadan (Nokia)
Kiran Nagaraj (Nokia)
Alex Nichol (Arista)
Ali Sajassi (Cisco)
Wen Lin (HPE)
Jeff Tantsura (Nvidia)

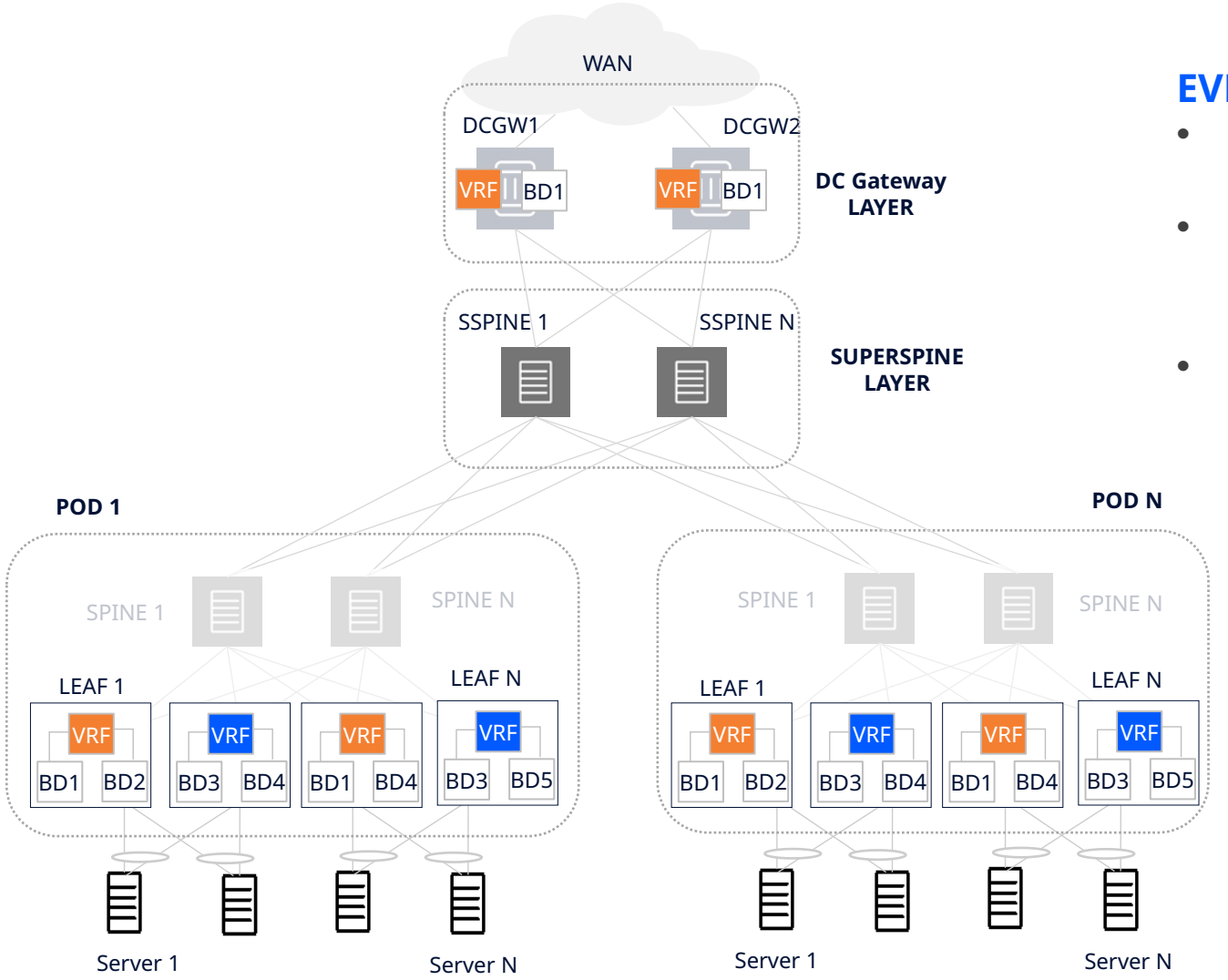
IETF123, Jul 2025
Madrid

Agenda

- Problem Statement and Solution Refresh
- Changes since last presented version (rev 02 – 04)
- Next Steps

Problem Statement

draft-rabnag-bess-evpn-anycast-aliasing-04



EVPN All-Active Multi-Homing in very large DCs

- Excessive control-plane noise caused by A-D per EVI routes carrying redundant information.
- ECMP complexity and processing overhead: underlay and overlay ECMP sets are programmed and updated independently during failures.
- Inefficient underlay forwarding during leaf-spine failures.

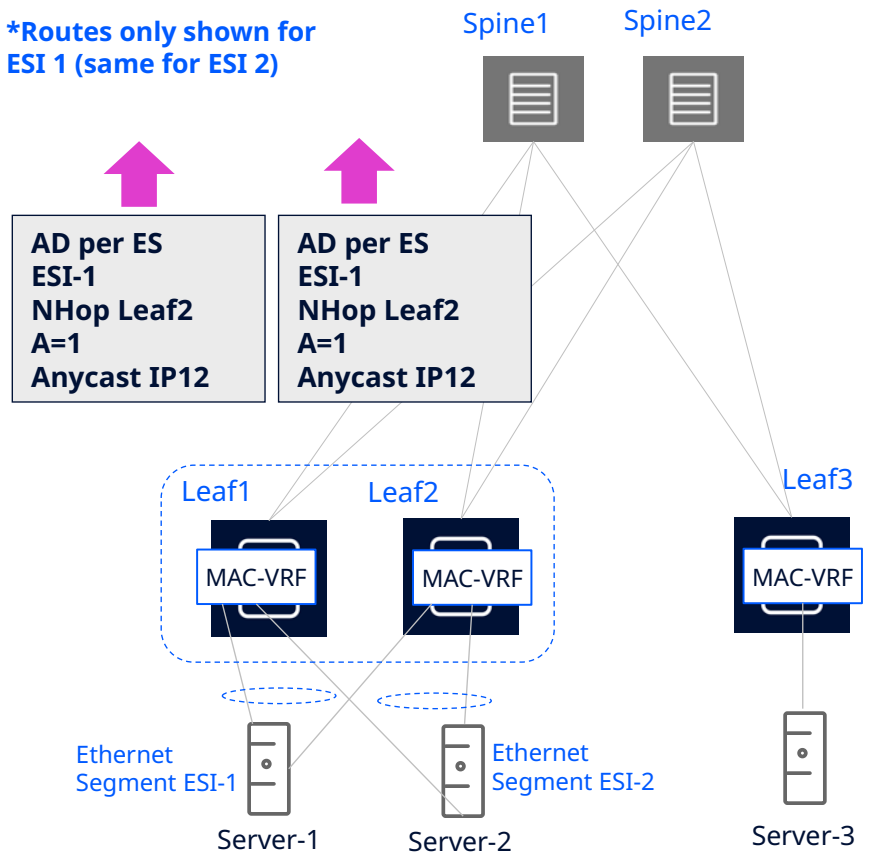
Proprietary MC-LAG use Anycast VTEPs

To address the above challenges, but:

- do not interoperate and
- do not support advanced EVPN Multihoming features

Anycast Multi-Homing Solution

draft-rabnag-bess-evpn-anycast-aliasing-04



Egress Leaf new behavior (Leaf1/Leaf2)

No changes to:

- DF election
- ESI advertisement in EVPN routes
- IMET routes processing and BUM forwarding

Ingress VNI:

- Both Leaf1 and Leaf2 must use the same ingress VNI for the network instance

Control Plane:

- Suppress AD per-EVI routes
- Advertise AD per-ES routes with: A=1 flag (ESI label ext community), Anycast VTEP in BGP TEA
- MAC/IP routes: ESI included, “unicast” VTEP as next-hop

Data Plane:

- VXLAN traffic sourced from unicast VTEP
- Anycast VTEP as source IP used only for FRR traffic

Ingress Leaf new behavior (Leaf3)

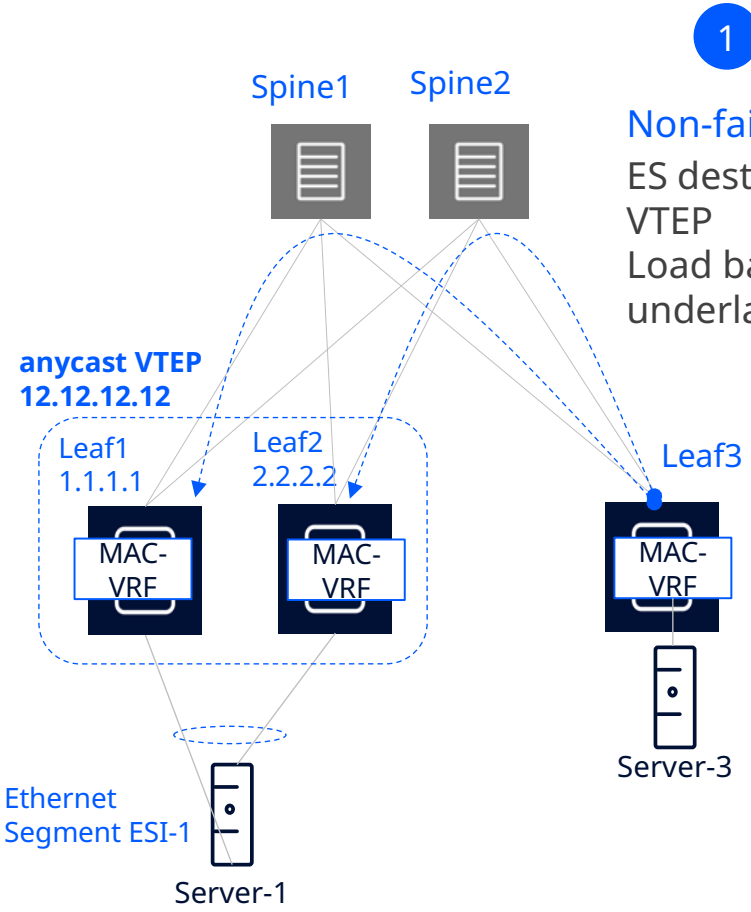
Control Plane Behavior

- ES destination:
  - creation based on AD per-ES routes with A=1
  - Single VTEP used (the anycast VTEP from AD per-ES)
  - VNI derived from the MAC/IP route
- MAC entries:
  - Installed against the ES destination as usual
- Mass withdrawal per ES supported (upon reception of the last MP_UNREACH_NLRI A-D per ES for the ESI)
- OPTIONAL: when only one egress NVE remains active in the ES, the ingress NVE reprograms the ES destination from using the Anycast VTEP to using the unicast VTEP of the remaining egress NVE

No Data Plane changes

Behavior During Node Failure

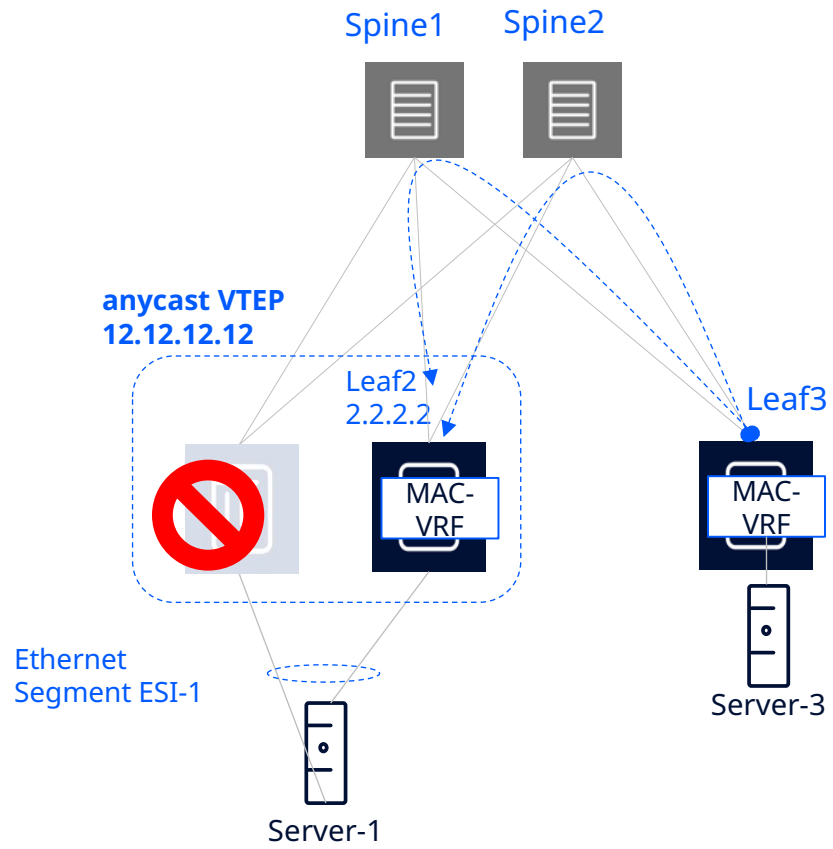
draft-rabnag-bess-evpn-anycast-aliasing-04



1

Non-failure state

ES destination points at anycast VTEP
Load balancing based on underlay ECMP-set



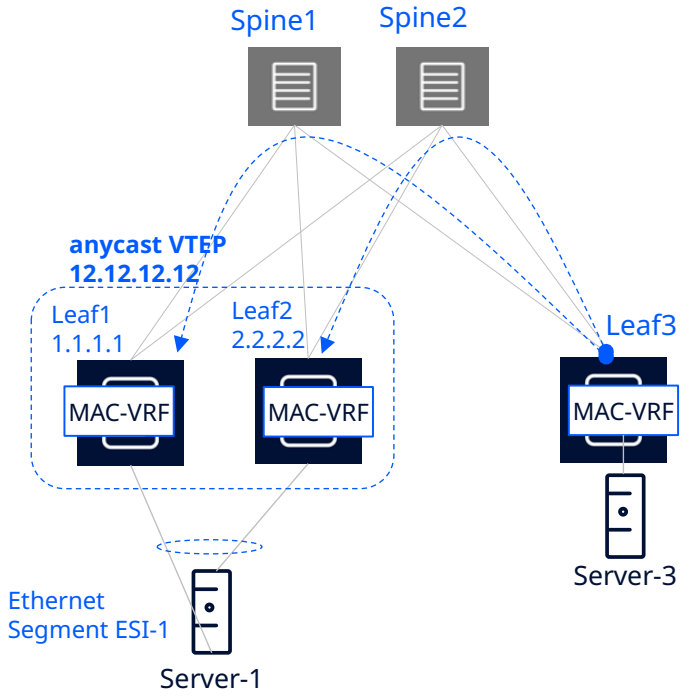
2

Node failure state

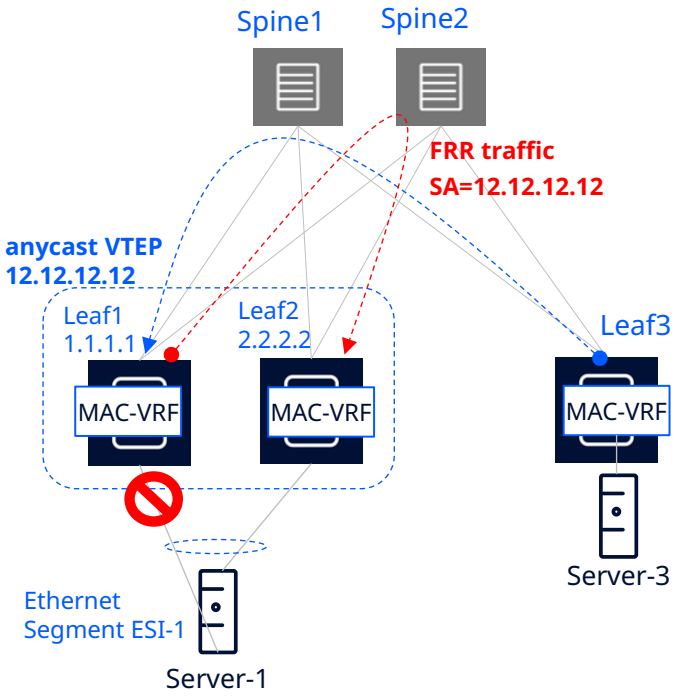
In-flight packets immediately rerouted to Leaf2
Upon receiving the AD per ES MP_NLRI_UNREACH from Leaf1, Leaf3 (optionally) reprograms the ES destination to use the “unicast” VTEP of Leaf2

Behavior During Link Failure and FRR

draft-rabnag-bess-evpn-anycast-aliasing-04



1 Non-failure state



2

Failure FRR State for flows sent to Leaf1

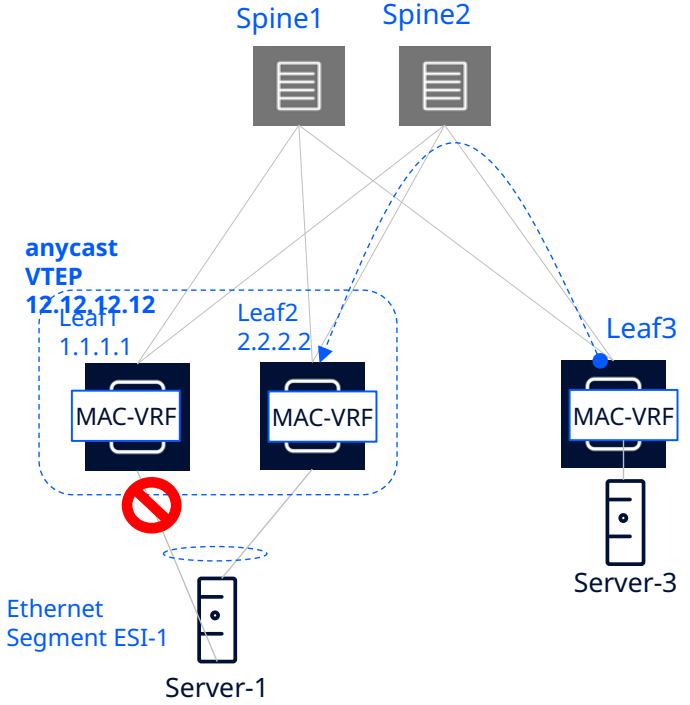
In-Flight Packet Handling

Packets decapsulated; if MAC DA points to ES peer, re-encapsulated with:

- **SA:** anycast VTEP
- **DA:** ES peer VTEP

FRR packets on Leaf2:

- Identified by anycast VTEP source address
- Never forwarded back into a tunnel



3

Failure non-FRR state

Upon receiving the AD per ES MP_NLRI_UNREACH from Leaf1, Leaf3 (optionally) reprograms the ES destination to use the “unicast” VTEP of Leaf2

NOTE: this is only possible if there are only two egress Leaf nodes.

Changes in rev 04 compared to previous versions

draft-rabnag-bess-evpn-anycast-aliasing-04

Rev 3 changes

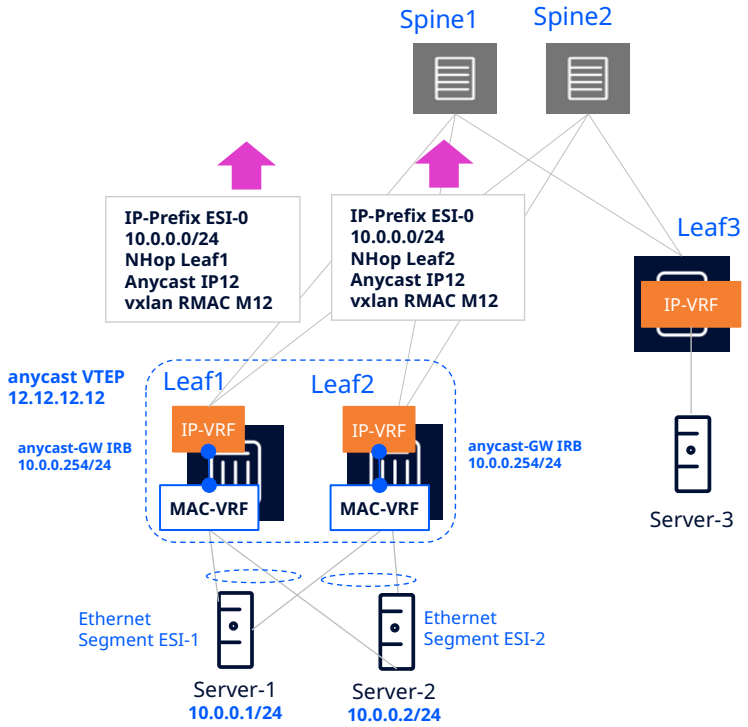
- Two new co-authors joined: Wen and Jeff
- Anycast VTEP encoding in A-D per ES routes - Use of BGP Tunnel Encapsulation Attribute [RFC9012] Tunnel Egress Endpoint Sub-TLV (code point 6)

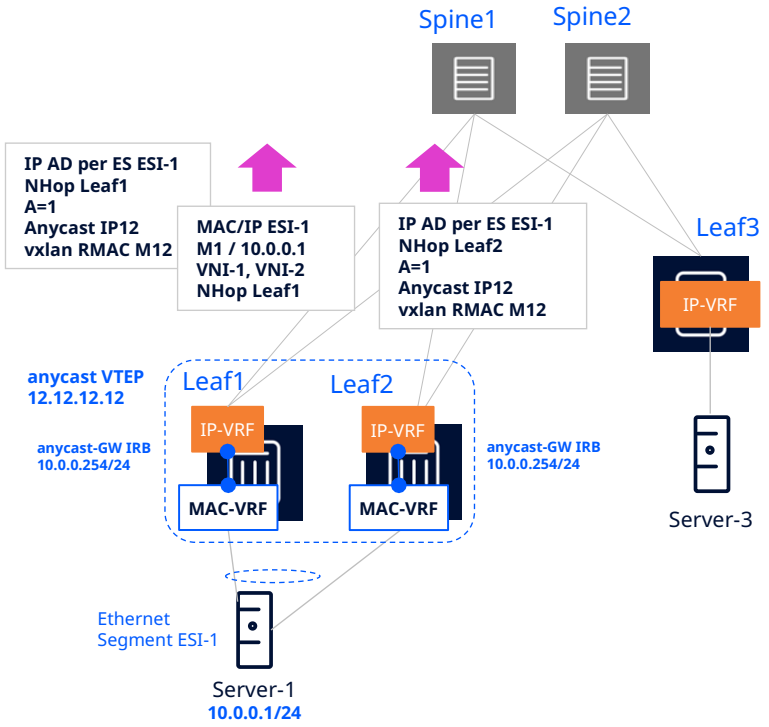
Rev 4 changes

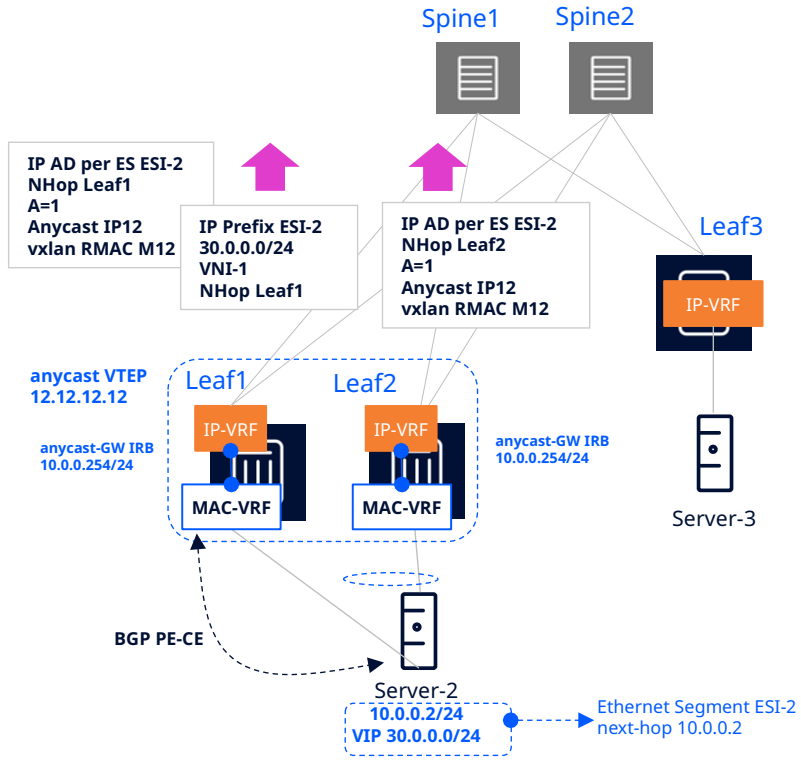
- Fixed typos, updated references and improved readability
- Added “Applicability of Anycast Multi-Homing to Inter-Subnet Forwarding” section (see next slide)

Applicability of Anycast Multi-Homing to Inter-Subnet Forwarding

draft-rabnag-bess-evpn-anycast-aliasing-04







1

Anycast Multi-homed IP Prefixes (RFC9136)

IP Prefix routes advertised with the Anycast IP in the BGP Tunnel Encapsulation Attribute (IFL and IFF unnumbered models) Router's MAC need to match across Egress Leaf nodes in the redundancy group

2

Anycast Multi-Homing and IP Aliasing for host routes

(draft-ietf-bess-evpn-ip-aliasing sect 1.1 and 1.2)

IP AD per ES route advertised with Anycast IP and RMAC
IP AD per EVI routes are suppressed
EVPN host route programmed against anycast VTEP

3

Anycast Multi-Homing and IP Aliasing IP Prefix routes

(draft-ietf-bess-evpn-ip-aliasing sect 1.3)

IP AD per ES route advertised with Anycast IP and RMAC
IP AD per EVI routes are suppressed
EVPN IP Prefix route programmed against anycast VTEP

Next Steps

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

- **Feedback** from WG
- The Authors Solicit **WG Adoption**

Thank you