SRv6 in Verizon IT Data Centers

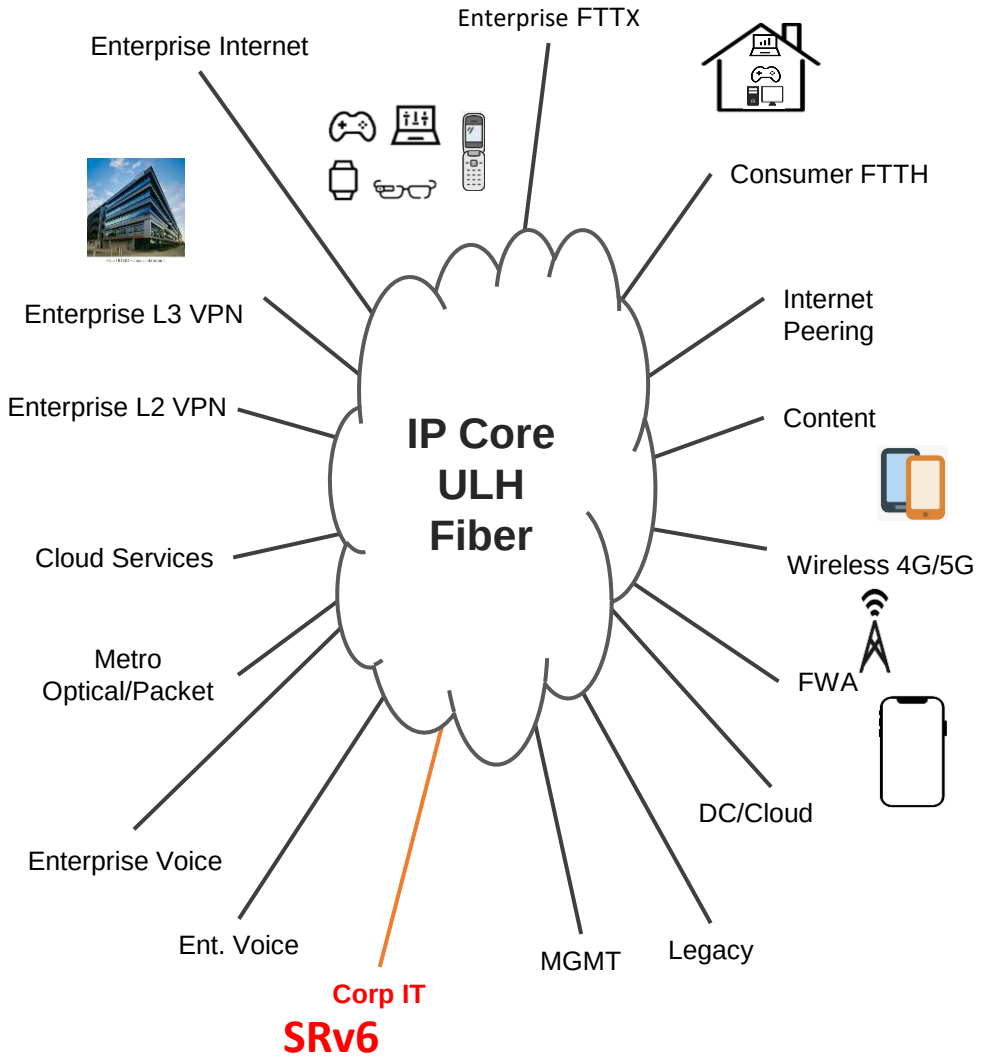
Gyan Mishra - Associate Fellow

Abstract:
In this presentation, Gyan will explore Verizon's IT Multi-POD data center architecture, which leverages SRv6 Next-C-SID and organizes each POD as an independent Autonomous System (AS). The discussion will focus on deploying end-to-end SRv6 Next-C-SID across both North-South and East-West inter-domain traffic engineered paths.

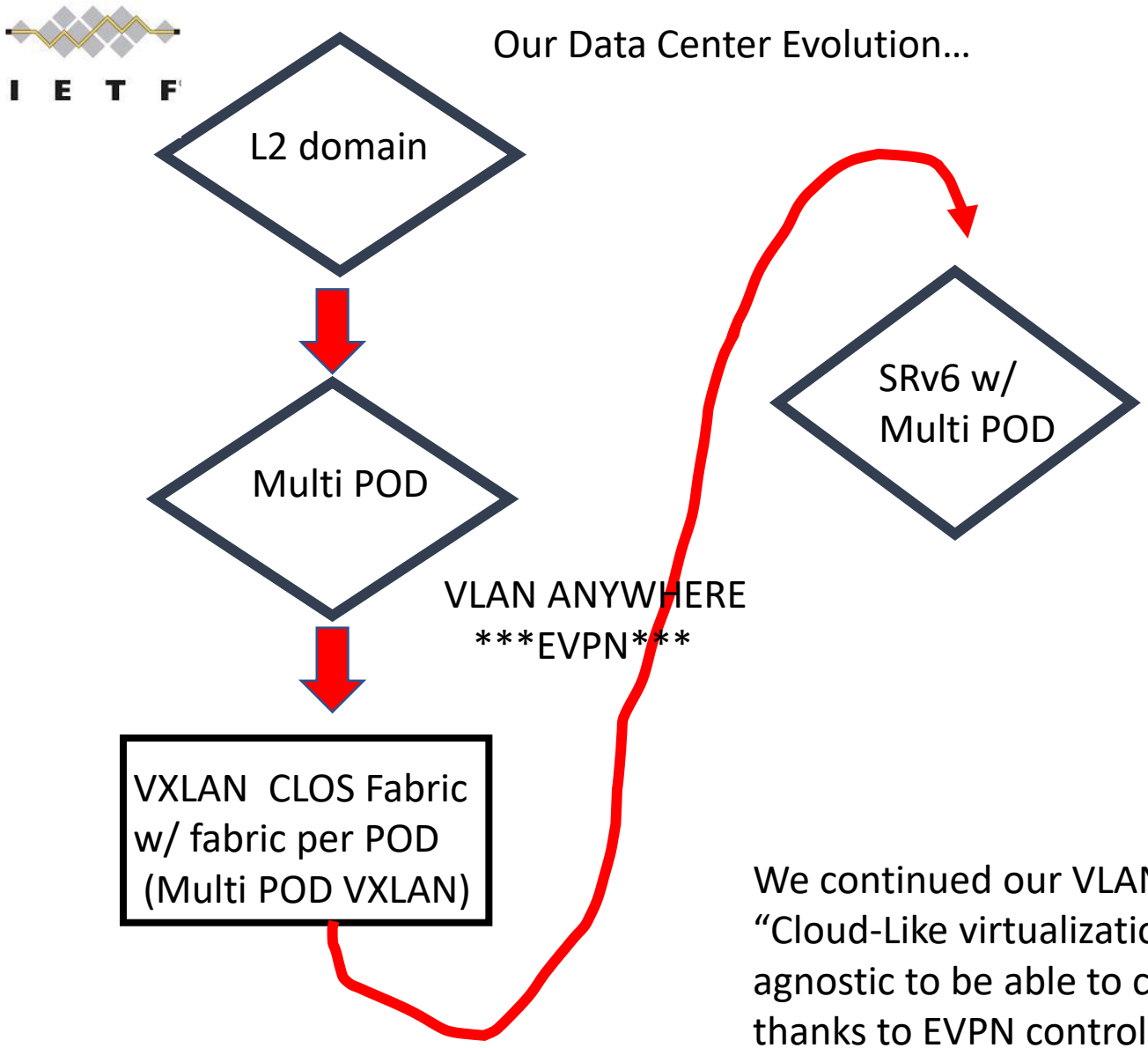
IETF 121 - Dublin

Verizon Networks

- Verizon comprises numerous large and small networks, each with its own administrative domain
- These networks are interconnected to deliver a cohesive service offering that meets customer expectations
- Utilizes various traffic engineering protocols, including RSVP-TE, SR-MPLS, and SRv6, to manage its network traffic
- Caters to consumers, businesses, and government entities, handling both wireless and wireline traffic across a vast range of services



Verizon ⇔ IT Data Center Story & Journey



We thought of maybe eliminating the Multi POD architecture, however the cost to keep the POD model was minimal with additional Spine (4) & BL (2) – 6 nodes per fabric was nothing & keeping the DC carved into individual AS per POD for fault isolation, resiliency, robustness & Public Cloud like AZ-Availability Zone capability, that it made sense to continue that model forward as we transition to SRv6.

We needed TE capabilities that VXLAN did not provide & an SDN controller based architecture that can optimize our forwarding plane based on desired bandwidth, latency and other constraints auto-magically ⇔ **"Our very own AIOps capability for self healing networks!!"**

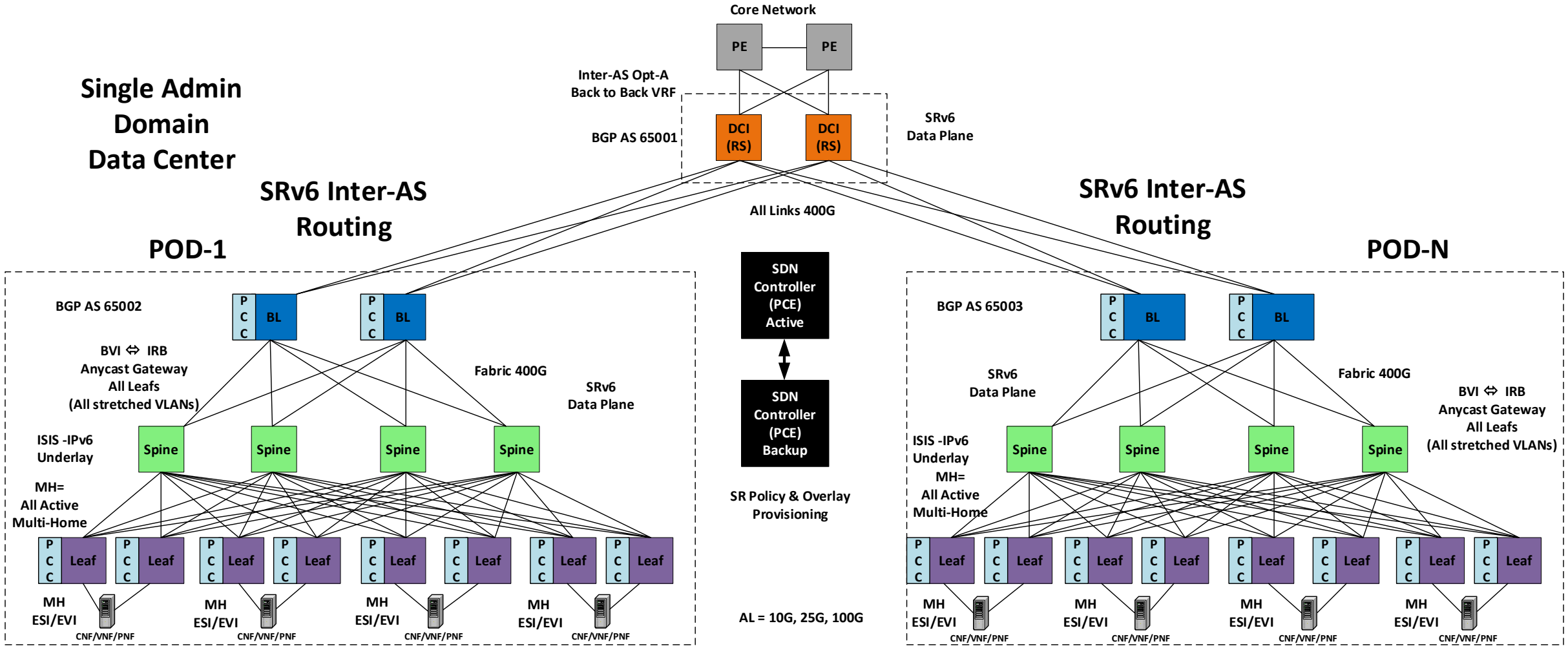
We continued our VLAN Anywhere sprawl which gave us the "Cloud-Like virtualization" layer for compute to be network agnostic to be able to connect anywhere in the fabric → many thanks to EVPN control plane which continues to thrive with SRv6!

Why SRv6 NEXT-C-SID ?

- **Data Plane Extensibility:** IPv6 data plane extensibility with IPv6 extension headers
- **SRv6 Data Plane Stitching:** All nodes need not be SRv6 enabled along E2E path.
  Data plane stitching with Vanilla IPv6 nodes (All nodes need not be SRv6 capable as with MPLS LSP)
- **Next-C-SID** - 5 hops of steering end to end without an SRH using a single Next SID container **Simplicity**
- **DC fabric to host extensibility:** SR DC fabric can easily be extended to host with IP data plane. (No special encapsulation required as with MPLS, VXLAN etc)
- **Host OS SRv6 Next-C-SID Support:**
  - ❑ Linux kernel SRv6 Next-C-SID support since 2022
  - ❑ VPP (Vector Packet Processing)
  - ❑ eBPF/Cilium
  - ❑ Router-in-container (Router instance placed between container VM/workload itself & server nic
- **Cloud Native Computing:** Future extensibility to cloud native architecture such as Kubernetes, Redhat portfolio - RHEL (Red Hat Enterprise Linux), Openshift, Openstack
- **Advanced Traffic Engineering:** PCE/SDN Centralized controller based architecture for End-to-End steering using “data lake” forensics from BGP-LS data & other telemetry sources for dynamic steering optimizations based on a variety of constraints such as latency, bandwidth, SRLG etc providing AIOps self healing, self-driving networking capabilities
- **Protocol Elimination**: All the MPLS Carrier Grade Services L2 VPN EVPN E-LINE, E-LAN, E-TREE, L3 VPN without MPLS, RSVP-TE and still able to provide Advanced Traffic Engineering capabilities
- **Simplified Configuration & provisioning:** Simple 3 step process, 1-Configure Locators, 2-Configure ISIS extension, 3-BGP

SRv6 Multi POD Data Center Use Case

SRv6 Multi POD DC Use Case (NEXT-C-SID)

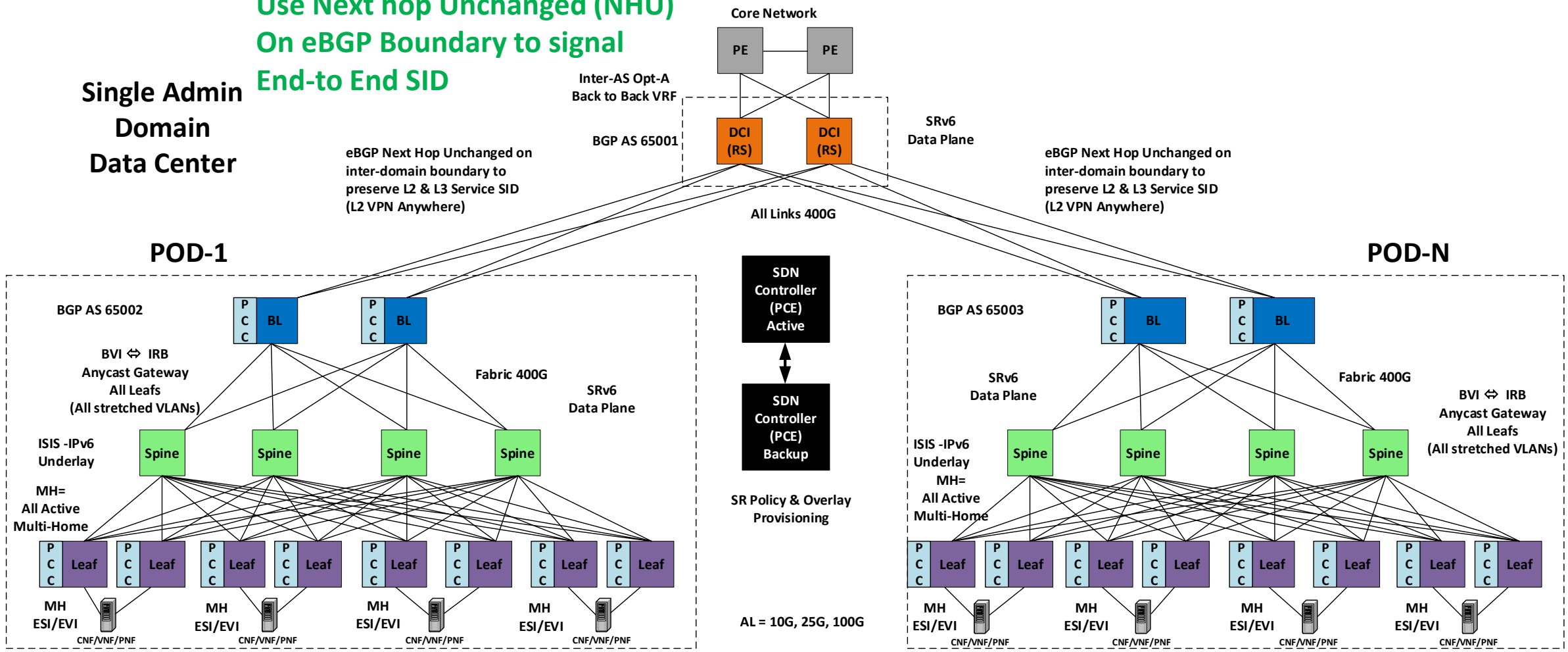


SRv6 Multi POD Data Center Solution

SRv6 Multi POD DC (NEXT-C-SID) Solution

Operational Guidance:
Use Next hop Unchanged (NHU)
On eBGP Boundary to signal
End-to End SID

Single Admin Domain Data Center



RFC 9252 SRv6 BGP Overlay Services

SRv6 BGP Overlay Services Specification is clear on Inter-Domain Routing

Finding related to SRv6 uSID Inter Domain routing

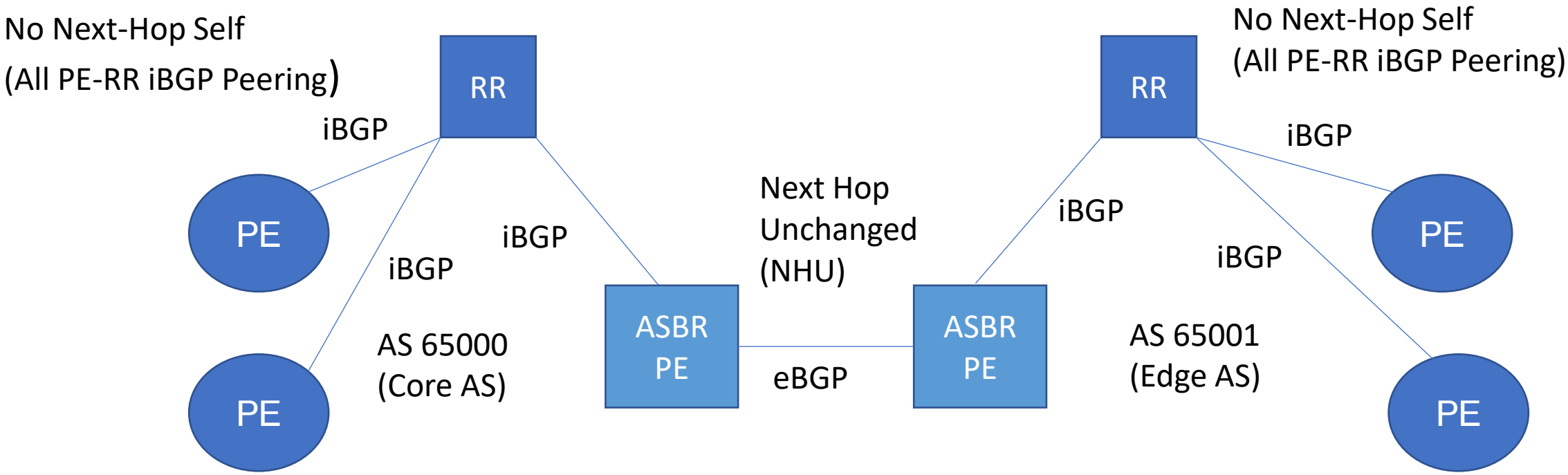
RFC 9252 bottom of section 2:

A BGP speaker receiving a route containing the BGP Prefix-SID attribute with one or more SRv6 Service TLVs observes the following rules when advertising the received route to other peers:
•If the BGP next hop is unchanged during the advertisement, the SRv6 Service TLVs, including any unrecognized Types of Sub-TLV and Sub-Sub-TLV, **SHOULD** be propagated further. In addition, all Reserved fields in the TLV, Sub-TLV, or Sub-Sub-TLV **MUST** be propagated unchanged.

•If the BGP next hop is changed, the TLVs, Sub-TLVs, and Sub-Sub-TLVs **SHOULD** be updated with the locally allocated SRv6 SID information. Any received Sub-TLVs and Sub-Sub-TLVs that are unrecognized **MUST** be removed.

On all eBGP peering connections the next hop must be set to “unchanged”

INTER DOMAIN SRv6 uSID

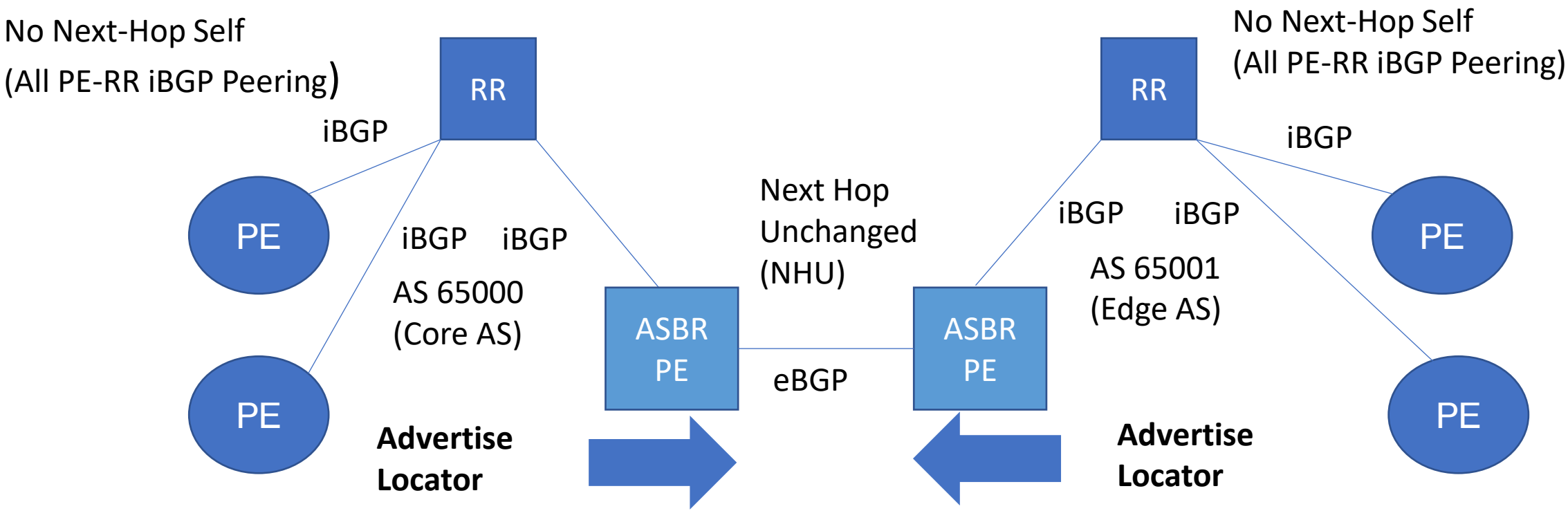


Rule for Inter Domain Peering

- iBGP -No Next-Hop Self
- eBGP –Next Hop Unchanged

(Requirement to Preserve L2 VPN & L3 VPN Service SID across INTER-AS Boundary)

INTER DOMAIN SRv6 uSID ROUTING SIMPLICITY



Rule for Inter Domain Peering

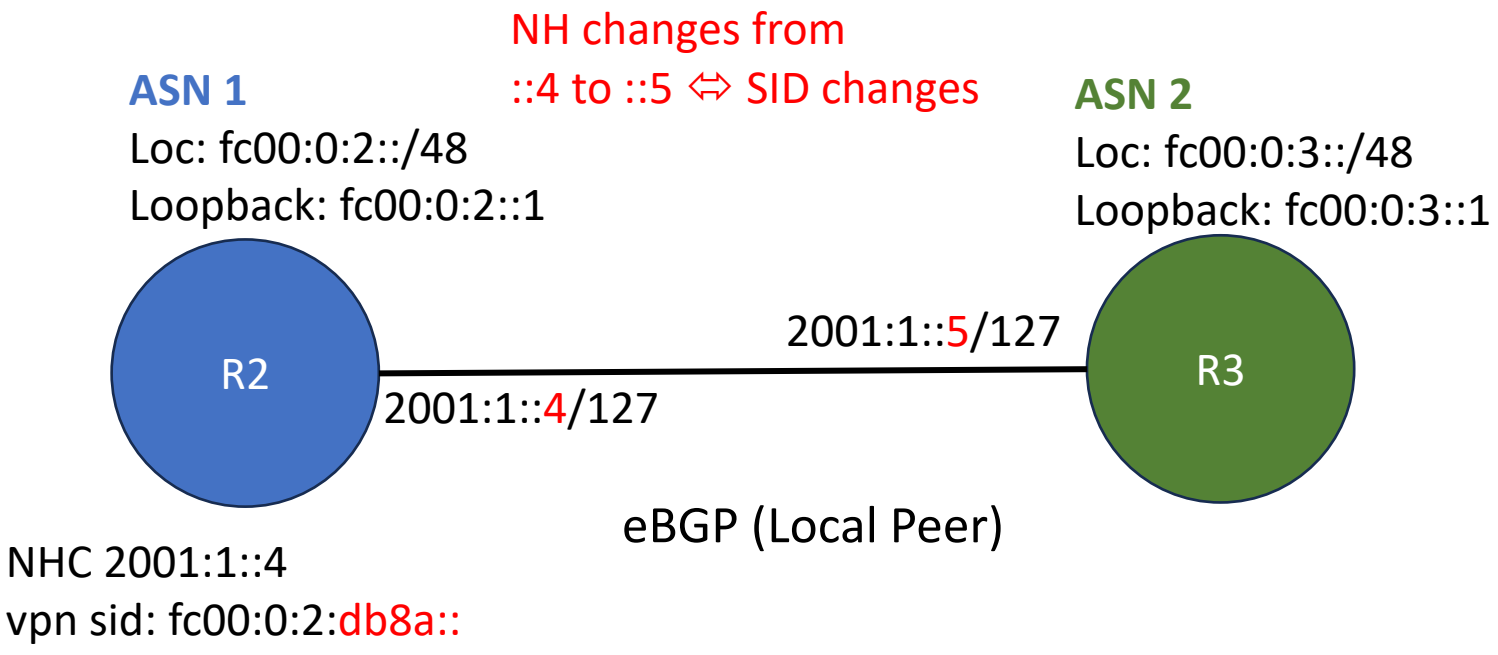
- Locator Reachability (That's it!!)

This allows host endpoints to provide static steering capabilities without PCE across any SR Algo cross domain

eBGP direct peering (NHC = NH Changed) – (Local Peer)

- **NHC = Next Hop Changed is default Behavior on eBGP peering**
- When the Next hop changes the label changes
- MPLS Service label (L2 VPN / L3 VPN = SRv6 Service SID
- When the Label changes the SRv6 SID changes

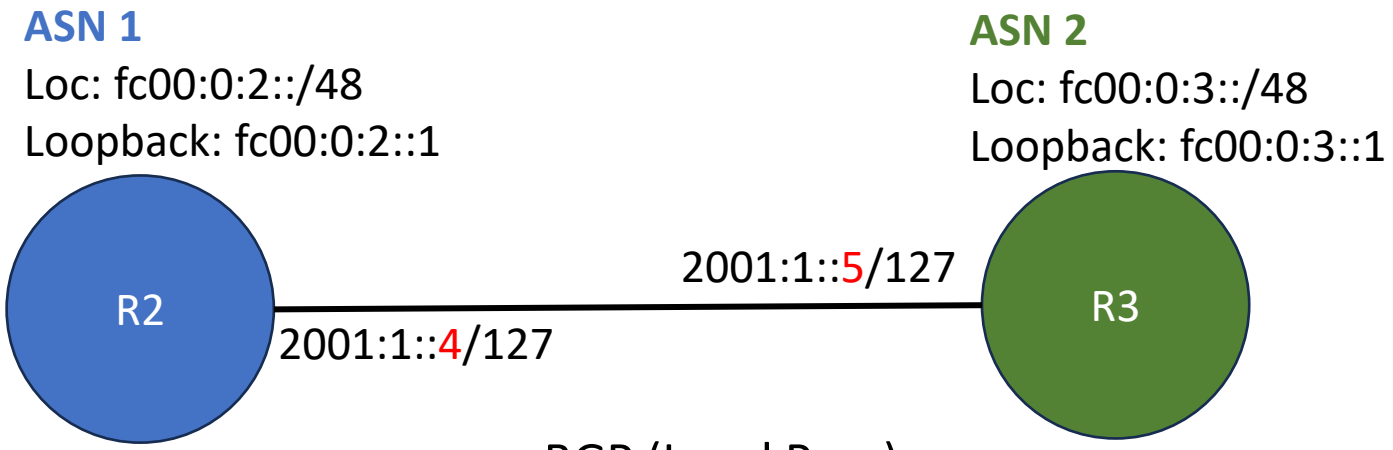
Operational Guidance:
eBGP boundary is Missing signaling end-to-end SID”



IETF

eBGP direct peering (NHU=Next Hop Unchanged) – (LOCAL Peer)

- **NHU = Next Hop Unchanged**
- When the Next hop is Unchanged the MPLS Label is preserved
- MPLS Service label (L2 VPN / L3 VPN = SRv6 Service SID
- When the MPLS label is preserved the SRv6 Service SID is preserved
- **eBGP boundary is now signaling end-to-end SID"**



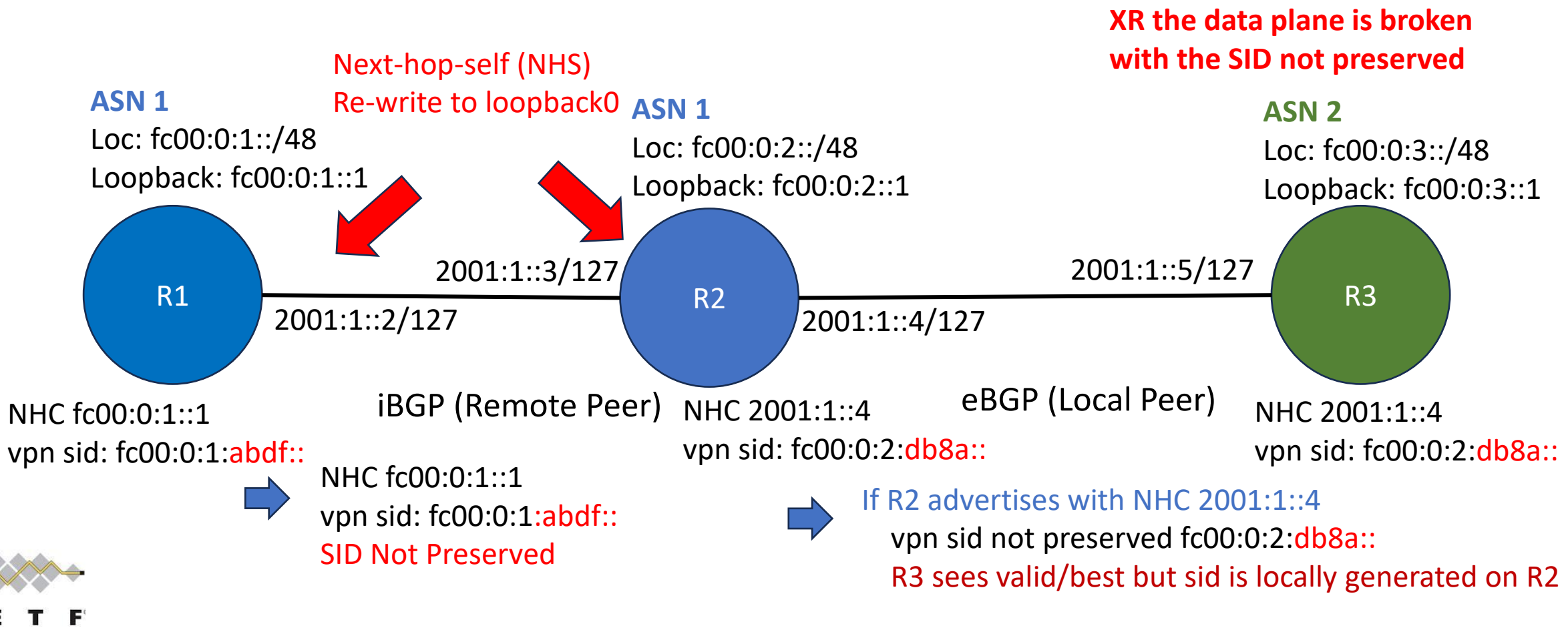
If R2 advertises with NHU fc00:0:2::1
vpn sid is preserved: fc00:0:1:e005::
R3 sees valid/best NH is *accessible*

***XR** BGP knob fixes the inaccessible "IGNORE CONNECTED CHECK"*
neighbor w.x.y.z
ignore-connected-check

Operational Guidance:
eBGP boundary is now signaling end-to-end SID"

IETF

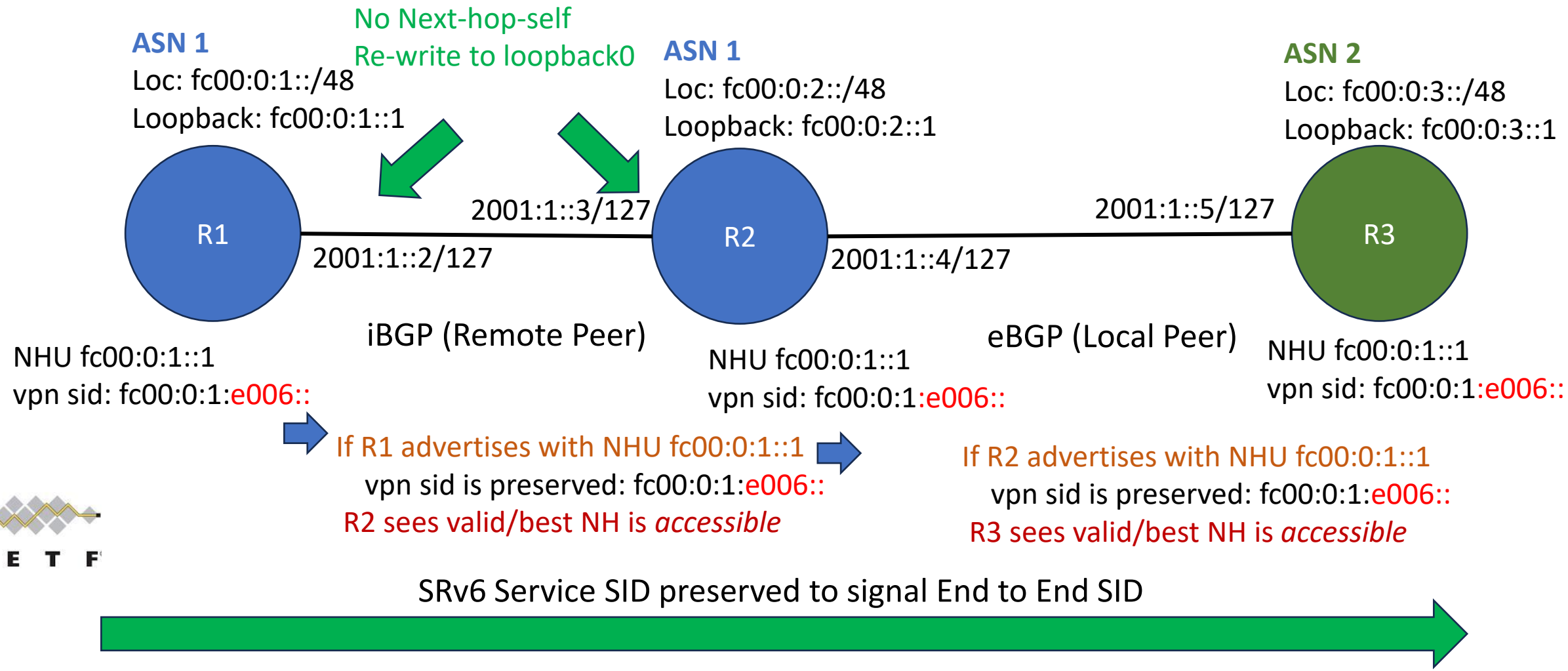
eBGP direct peering (NHC=Next Hop Changed) – (Remote Peers)



IETF

eBGP direct peering (NHU) – (Remote PE)

***XR** BGP knob fixes the inaccessible "IGNORE CONNECTED CHECK"*
neighbor w.x.y.z
ignore-connected-check



I have submitted this draft below in SRv6Ops Please review & comment…

SRv6 Inter Domain Routing
draft-mishra-srv6ops-inter-domain-routing-01

Authors:
Gyan Mishra – Verizon Inc
(Associate Fellow)
Bruce McDougall – Cisco Inc

Final Summary & Takeaways

Verizon operational use case takeaways:

- Multi-POD design with AS per POD is possible with SRv6 where the goal remains as it has in the past and is still very relevant to reduce the size of the fault domain for higher scalability & resiliency.
- Multi-POD design is very possible with SRv6 for end to end East-West flows and North-South Inter domain traffic engineered paths.

SRv6 specification takeaways:

- Operational Guidance to always use Next-Hop-Unchanged (NHU) on all eBGP boundaries on both sides of the eBGP peering to signal “End-to-End SID” for SID propagation End-to-End in both directions.
- Operational Guidance to not use Next-Hop-Self on iBGP PE-RR peering so that the received service SID from an Ingress Domain is propagated “End-to-End” to an Egress Domain and vice versa in both directions.
- Operational Guidance is applicable to both SRv6 (Full SID) and SRv6 compression.
- There is NO issue or problem with the SRv6 specification or SRv6 compression specification.
- RFC 9252 SRv6 BGP service overlay clearly states in Section 2 that the next hop must be “Unchanged” for the propagation of the service sid caveat. We just need to get this documented for operator awareness.

Next Steps:

- Submit informational draft to Spring WG on the SRv6 Inter domain routing issue & resolution and so this is well documented for vendors & operator deployments.
- All vendors should document this solution on their website portal or in a whitepaper for operators

SRv6 Developer Consortium (Co-Founders Gyan Mishra & Bruce McDougall)

GOALS & OBJECTIVES: The goal of the SRv6 Developer consortium is the advancement & proliferation of SRv6 and the development of ground breaking new innovative use cases for SRv6

Segment-routing.net portal ⇔ Links to SRv6 Developer consortium resources for innovation

- Github icon at bottom of the page links to segment-routing "Github"
  https://github.com/segmentrouting/srv6-labs

- Slack channel icon at the bottom of the page links to segment-routing "Slack Channel"
  segment-routing-net.slack.com (Send email to "hayabusagsm@gmail.com" Gyan Mishra for access)

MPLS World Congress Links of SRv6 Inter Domain Routing use cases (Gyan Mishra):
https://www.segment-routing.net/conferences/Paris24-Verizon-Gyan-Mishra/

MPLS World Congress YouTube Channel Video of SRv6 Inter Domain Routing Series 1-9 Video's
https://www.youtube.com/@segment-routingnet

Gyan Mishra – Github & YouTube channel contains all the R&D work from MPLS WC 2023 & 2024
3-srv6-dc-case-studies

Bruce McDougall – Github contains R&D work on SONiC labs & Containerlab.dev labs
1-starter-topologies
2-use-case-topologies

Contact Gyan Mishra – hayabusagsm@gmail.com
For any Q&A


verizon