データセンターネットワークで SRv6って本当に使えますか？

JANOG55@京都

Cisco Systems G.K.
Teppei Kamata

Jan 23th, 2025

自己紹介

- 名前：鎌田徹平
- 所属：シスコシステムズ合同会社
- 役職：Solutions Engineer
- Service provider様向けのPre-sales SE職
- Interop Tokyo ShowNet NOC team member
- JANOGでは40, 43, 50, 53で登壇

本日の内容

- データセンターへのSRv6の適用事例が海外を中心に出てきています。
- 日本はSRv6先進国と言われている状況の中、なぜデータセンターでの適用事例が少ないのか(私の知る限り一社のみ?) 、会場で議論させていただければと思い申し込みさせていただきました。
- 本発表では、以下のポイントについて具体的な事例と共に説明します :
  - データセンターにおけるSRv6 uSIDの具体的な適用方法とメリット: 実際のデータセンター環境での適用事例を紹介し、メリットについて説明します
  - コンテナネットワークにおけるホストルーティングの実装: LLDPベースのホストルーティングの実装とCNI(Cillium)実装についてデモを交えて説明します
  - そのうえでSRv6をデータセンターネットワークに適用するうえでの課題について議論させてください

- SRv6自体の説明はこれまでのJANOGセッションなどをご参照ください。

Outperform MPLS/VxLAN

Outperform MPLS - Daniel Voyer (Bell Canada)

- Native Optimum Slicing
  - SLID is encoded in Flow Label
- HW Linerate Push: 3 times better
  - J2 uSID linerate push: 30 uSIDs >> 10 MPLS Labels
- HW Counter and FIB consumption: 4 times better
  - uSID requires 4 times less counters and FIB entries than MPLS
- Routing scale: 20 times better
  - uSID supports summarization. MPLS requires host routes.
- Lookup efficiency: 2 to 3 times better
  - uSID can process 2 to 3 SIDs in a single lookup (LPM nature)
- Load-balancing: optimum and deterministic
  - uSID provides HW friendly entropy (fixed offset, shallow)

Bell SRv6 uSID Deployment
Paris 2022

Outperforms VxLAN – Gyan Mishra (Verizon)

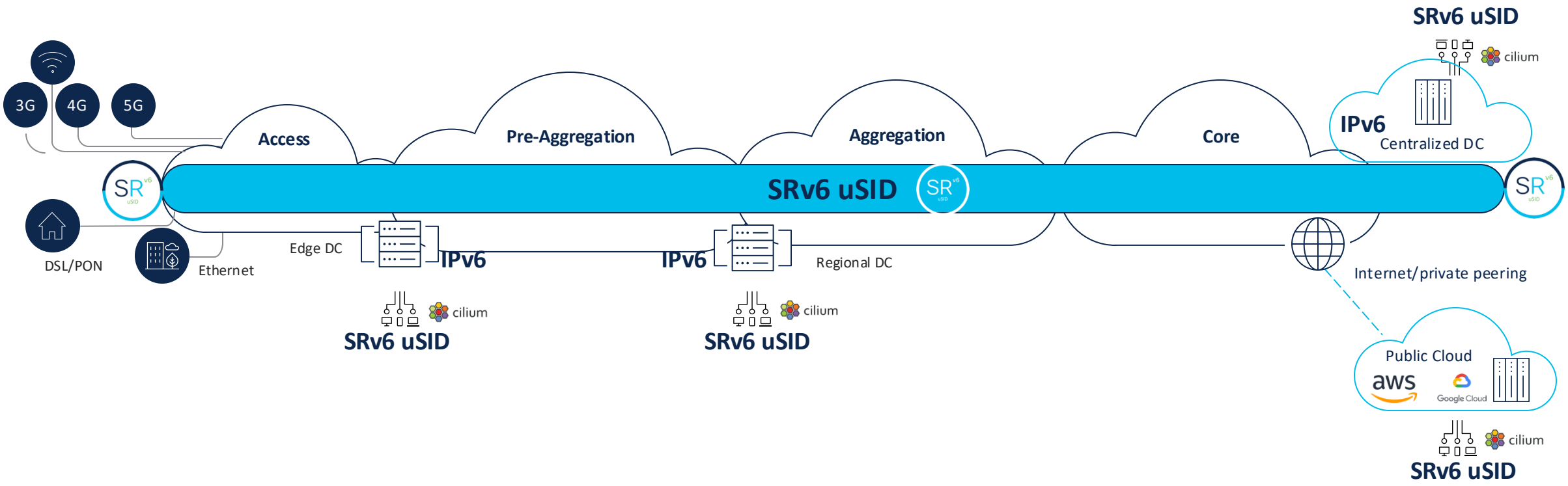
- Seamless Host support for Network Programming
  - 6 uSID's in outer DA: RFC2460 IPinIP with opaque DA
- TE in the DC
  - elephant flows exist, asymmetric fabrics exist, TE is needed
- TE in the Metro/Core from the host
  - An SRv6 uSID DC allows for the application to control the network program in the metro/core without complex DPI and protocol conversion at the DC boundary,
- uSID DC provides lower MTU overhead (~5%)
  - Lower MTU overhead means lower DC cost
- Vendor, Merchant and SONIC/SAI maturity
  - uSID support across DC vendor (Cisco), Merchant (Cisco, Broadcom, Marvell), Sonic/Sai (Alibaba deployment)

SRv6 uSID DC Use-Case
Paris 2023

Worse, the Combination of Alternatives is Painful

- Boundary Gateways:
  - DPI
  - Protocol Conversion
- Consequence:
  - Low Scale
  - High Cost
  - Low Reliability

Simplify Telco Cloud and Cloud networking



Any Service over IP without any shims

Unified Solution with Better Reliability

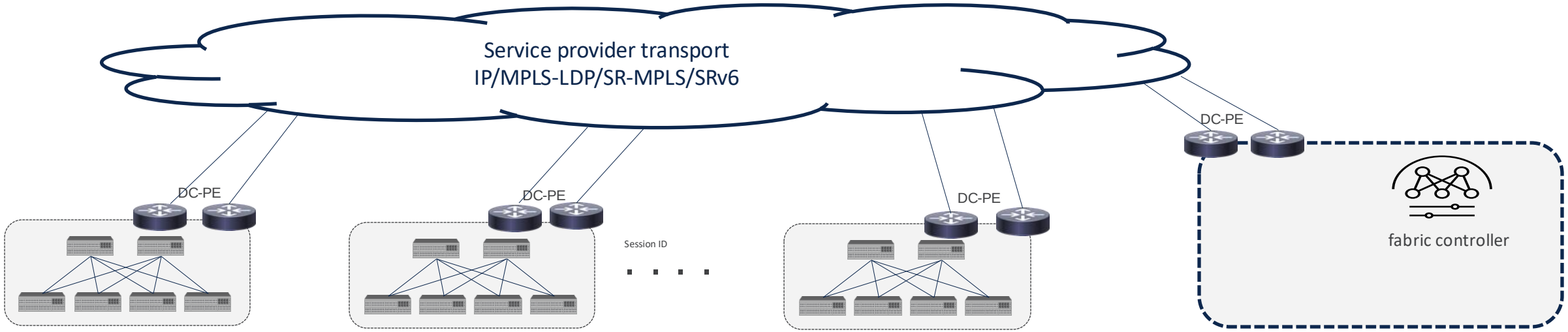
Seamless Brownfield Deployment

Native Host and Cloud

Embedded Security

Telco Cloud distributed DC - Present mode

- A smaller VXLAN fabric is required for relative bigger Edge DC sites
- VXLAN is only used within the DCs
- DC to transport handoff is required for all external communication (VXLAN to VRF mapping)
- FC is centrally hosted in one of the edge fabric or in a central DC can manage all edge DC sites



SRv6 from Cell Site to Internet thru Aggregation, Metro, Core & DC

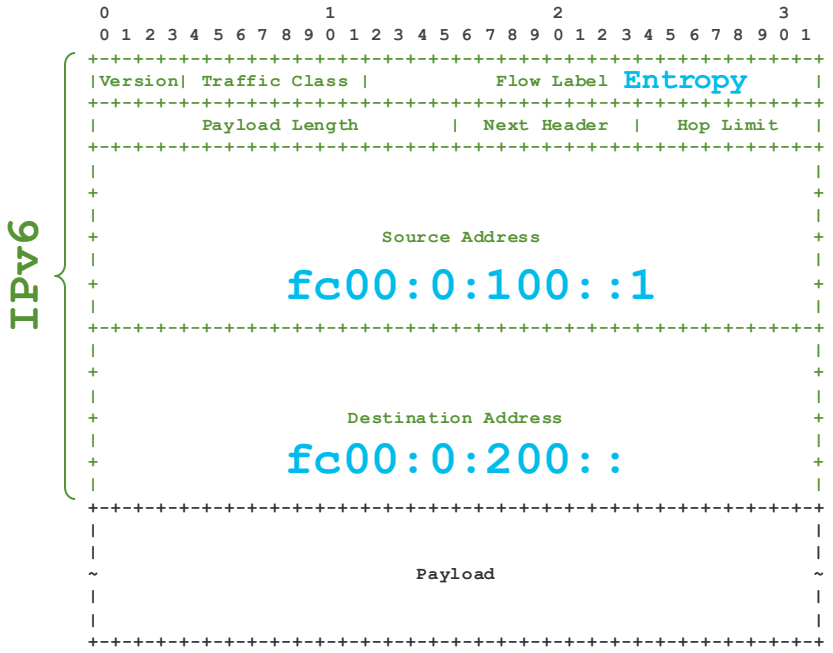
- SRv6 uSID is the sole solution that is end-to-end across segments
  - MPLS failed in the DC
  - VxLAN is a no-go for the WAN/Metro
  - NSH is an IETF failure
- What are the benefits?
  - VxLAN/MPLS gateways are eliminated
  - NSH is eliminated
  - TE/FRR/VPN/Service Chaining is naturally integrated
  - Fulfills the DC Traffic Engineering requirements
    - > Inter DC TE with policy starting from DC
    - > Intra DC Traffic Engineering

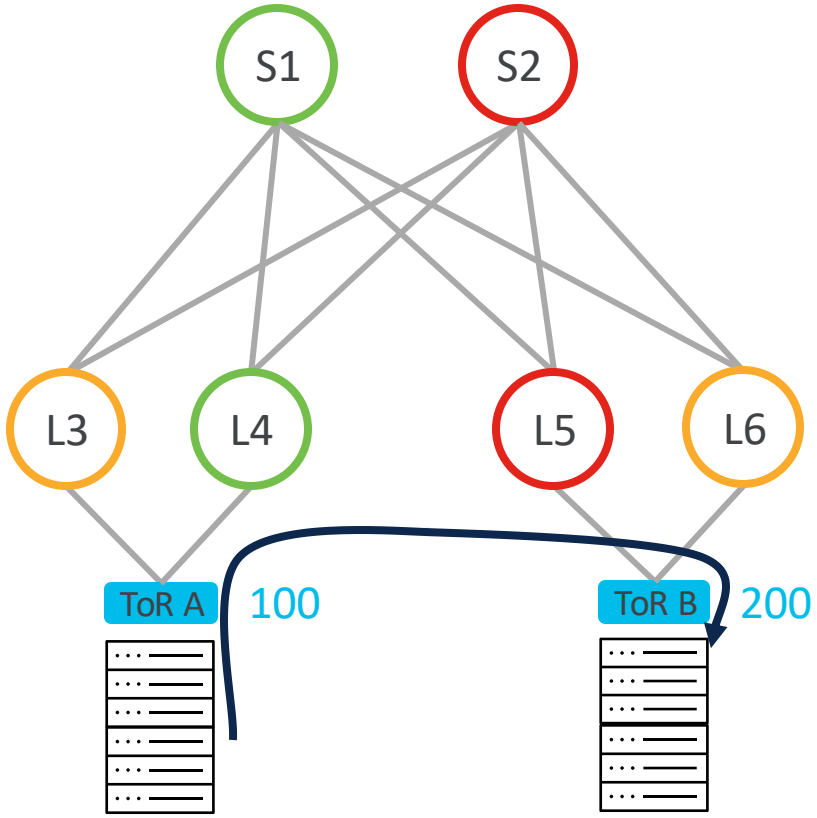
Simple IPv6 encapsulation

- Most of the time uSID is inexistent and the packet is just pure IPv6
  - Base solution: IP
  - Add uSID for VPN, TE, FRR, NFV where needed
  - Add IPM for measurement where needed
- When SRv6 uSID is used, very rarely an SRH is present
  - Simple IPv6 encapsulation fits 6 uSIDs (no SRH)
  - After 20 years of MPLS, most MPLS use-cases require less than 6 labels

Simple IPv6 encapsulation

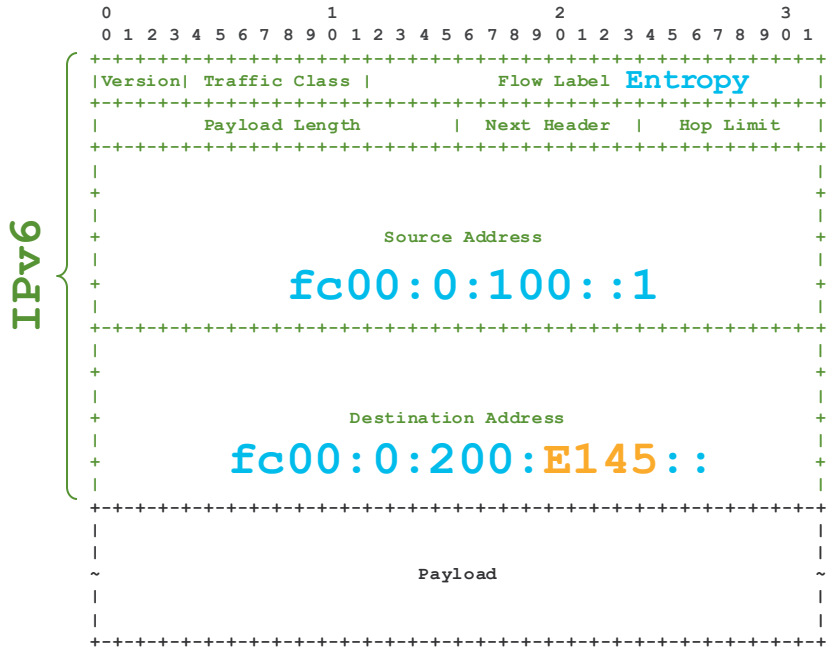
- Best effort reachability from 100 to 200 only requires plain IPv6

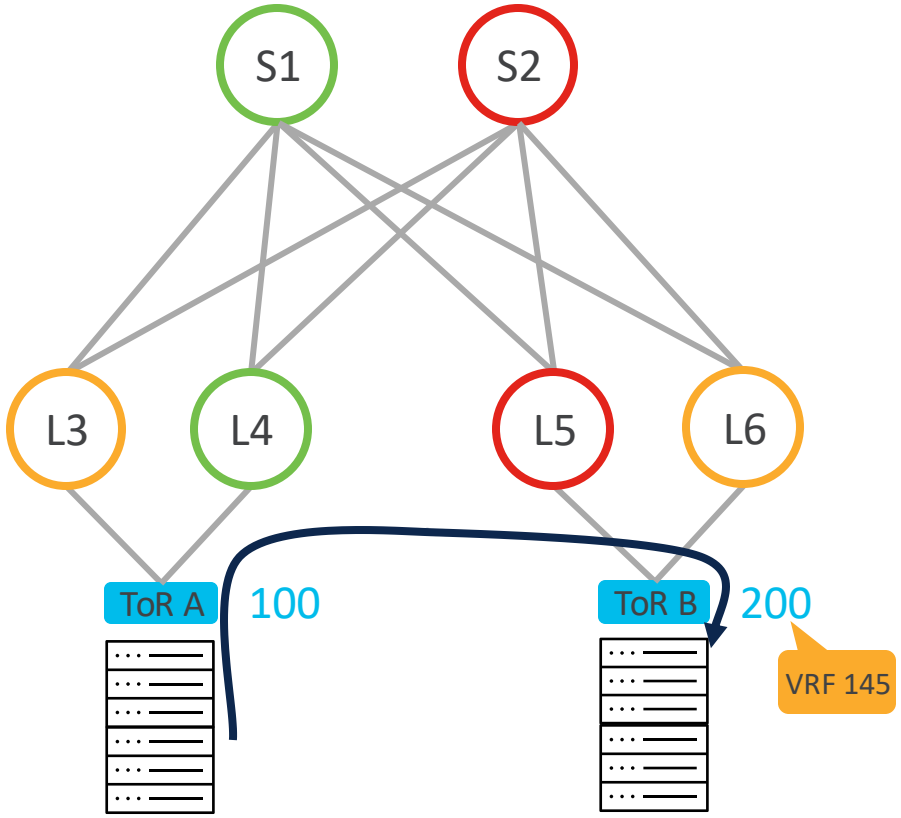




Simple IPv6 encapsulation

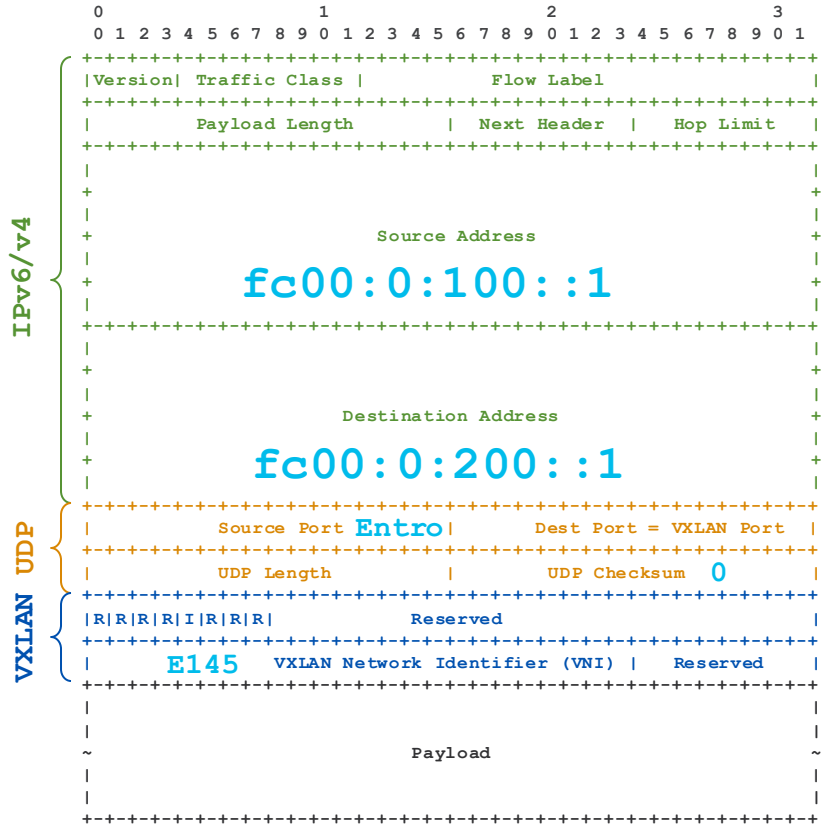
- Best effort reachability from 100 to 200 only requires plain IPv6
- If we need to create a VPN service, we only add a single uSID

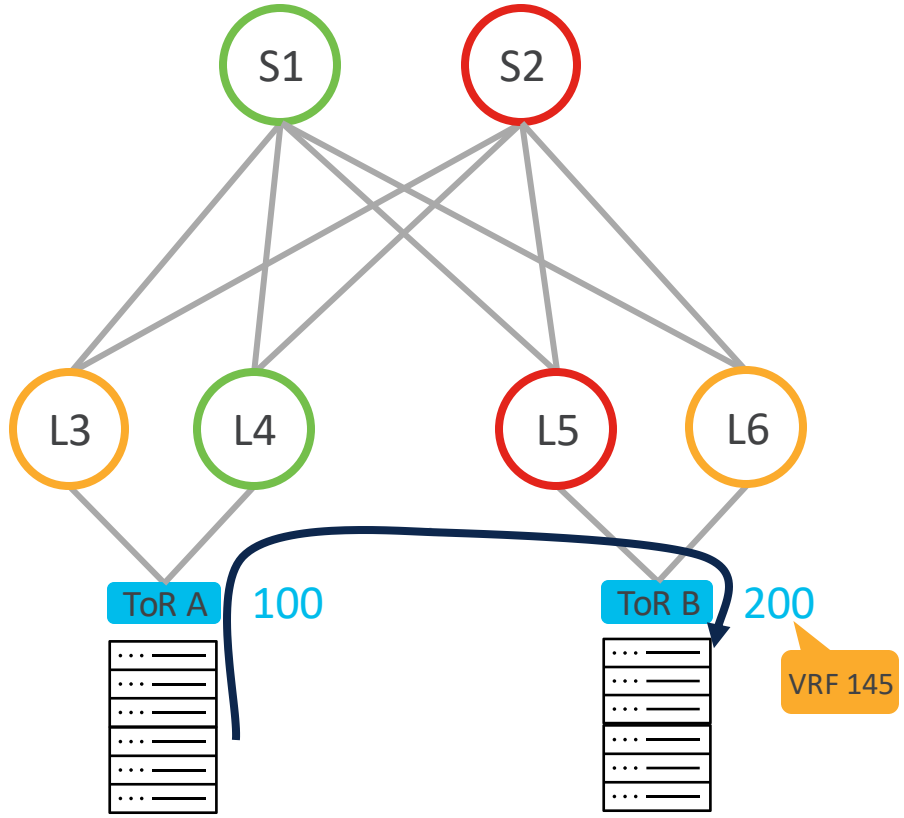




Non-Simple Encapsulation

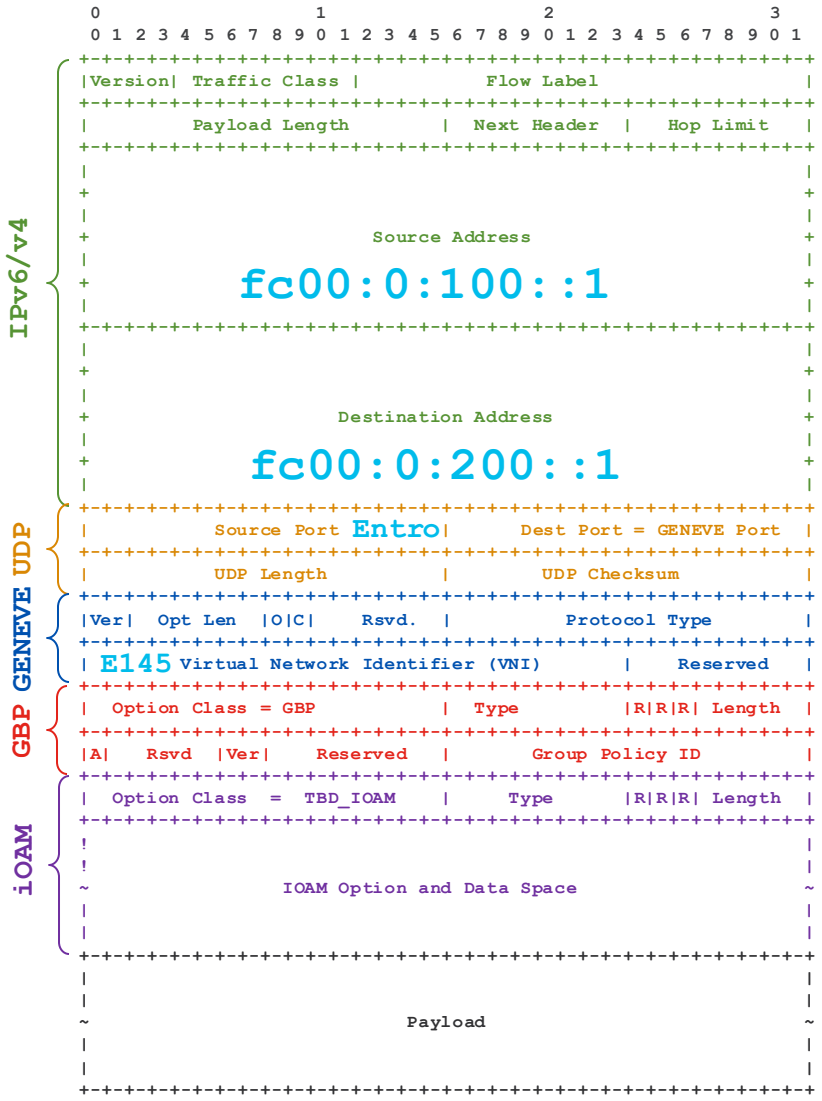
- To compare it with VXLAN….

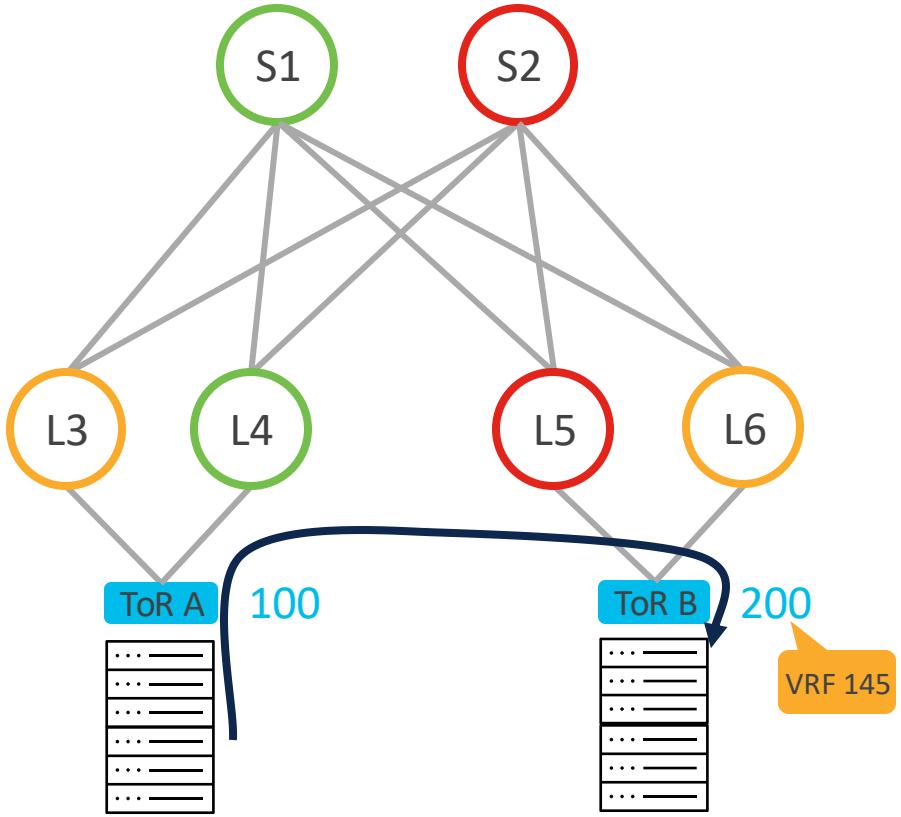




Non-Simple Encapsulation

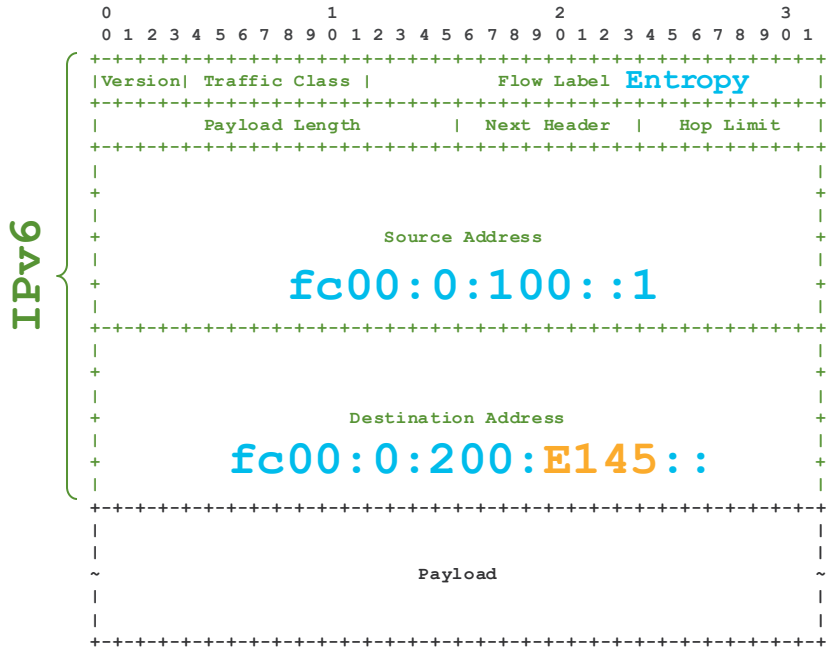
- GENEVE makes it **even more complex** with additional Options (GENEVE Extension Headers)

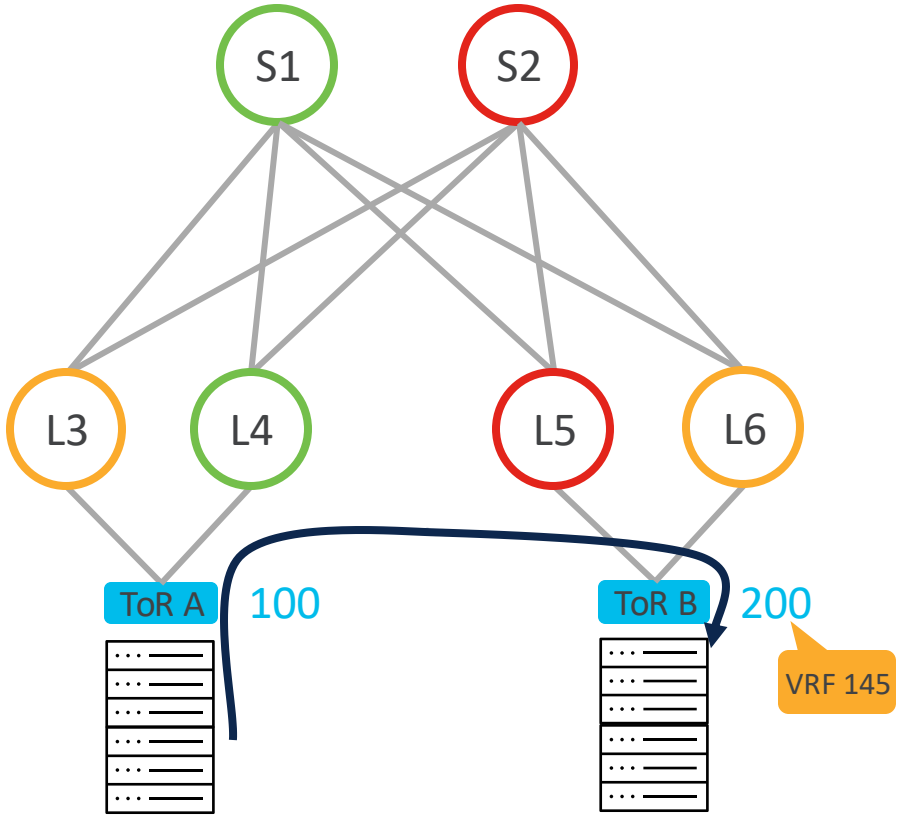




Simple IPv6 encapsulation

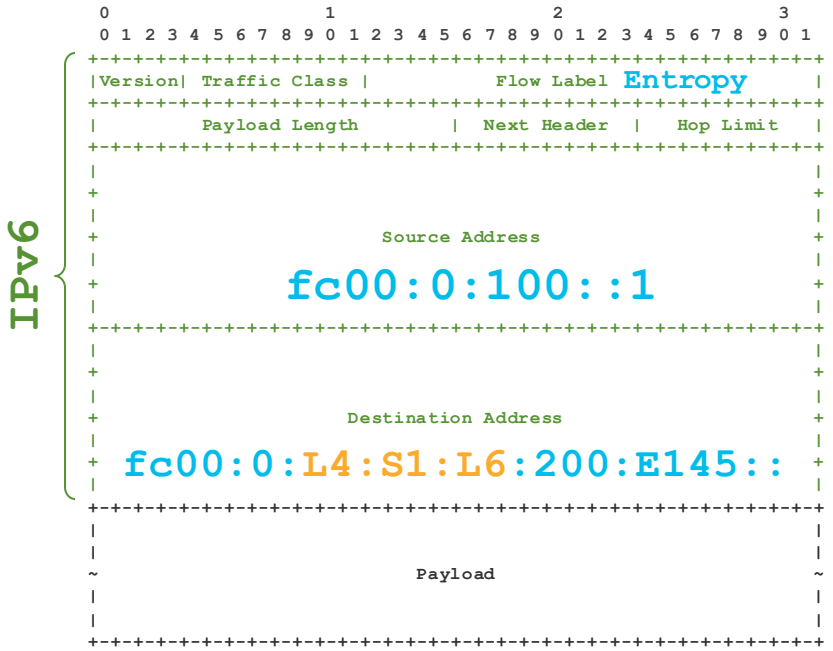
- Best effort reachability from 100 to 200 only requires plain IPv6
- If we need to create a VPN service, we only add a single uSID

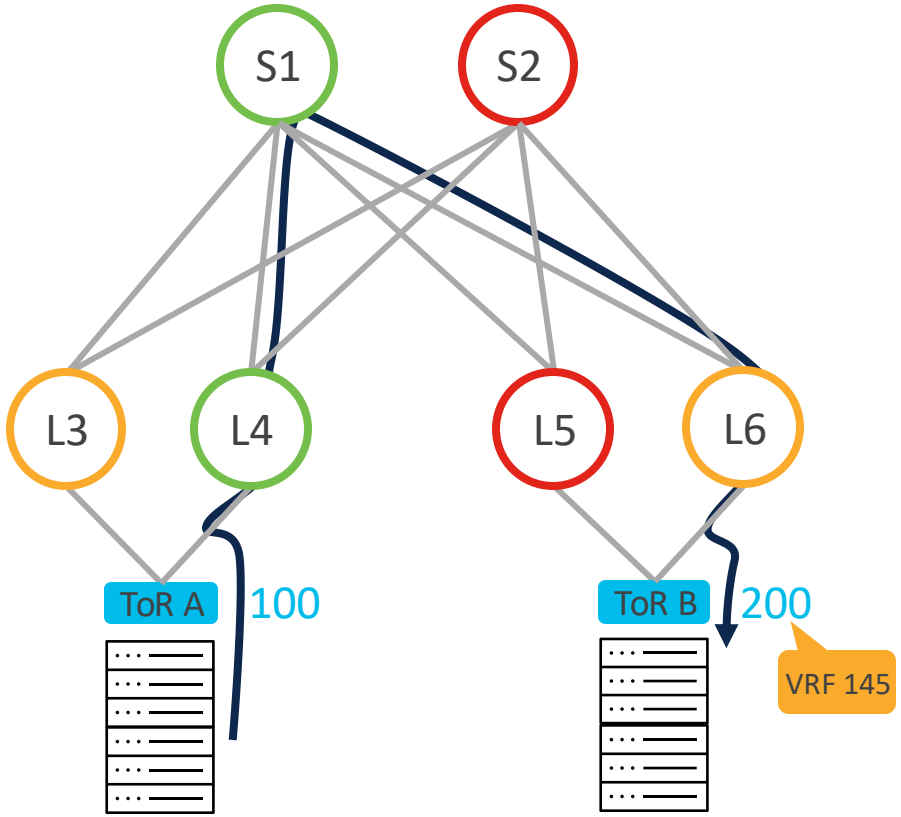




Simple IPv6 encapsulation

- Best effort reachability from 100 to 200 only requires plain IPv6
- If we need to create a VPN service, we only add a single uSID
- If we want to enforce a Traffic Engineered path <4, 1, 6> we add three uSIDs:





- Apple-to-apple comparison:
  - IPv6 uSID provides more functionality AND less overhead than VXLAN

Service Programming (Chaining)

Applicability Examples

Network Services

Firewall
Load-Balancer

Overlay

Multi-Could

As-a-Service

Network as a Service

Solution

Services are expressed with segments (Loc::Fun:arg)

- Flexible
- Scalable
- Stateless

Benefits

Simplicity and automation

- No static configuration required
- No Traffic bottlenecks
- No Dedicated encapsulation header (NSH) that require states to be maintained for each service chain

Packets are steered through a sequence of services on their way to the server

Traditional Approach: Services are placed on the traffic route

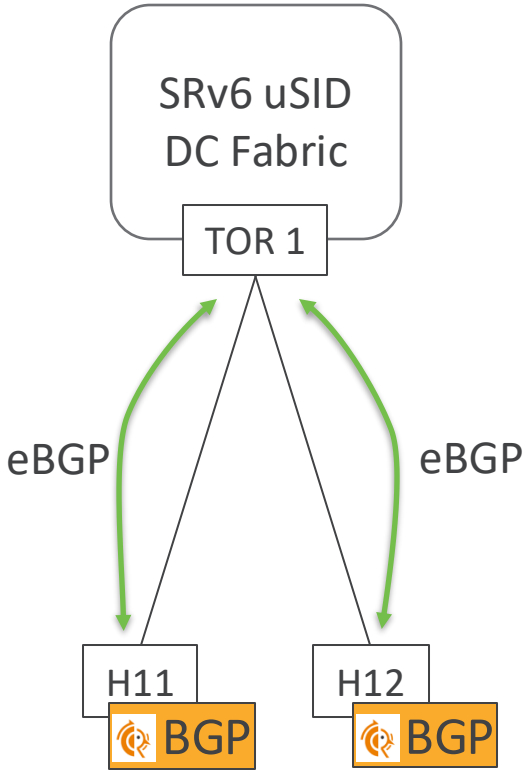
SFC with NSH requires dedicated header + maintaining of states

SFC with SRv6; Services are expressed with segments

Bringing SR simplification to the DC LHR

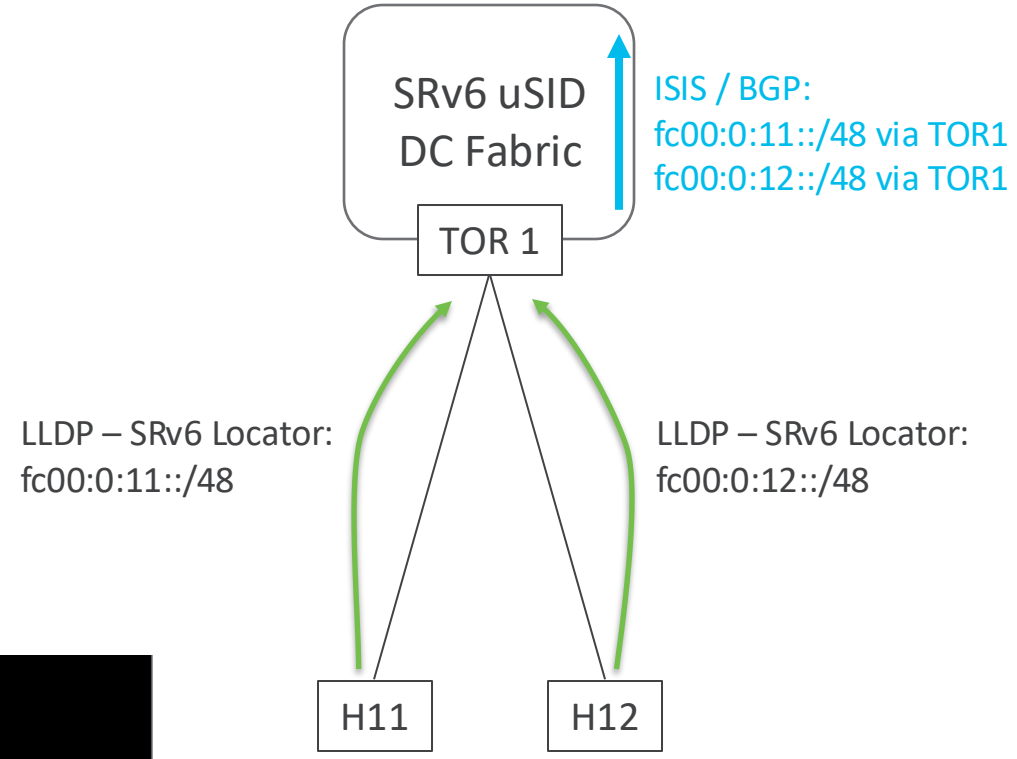
Compute attachment to the DC network

- Use-case: services terminated at the host
  - 5G deployments with NFVs
  - Security appliances (e.g., firewall)
  - Container networking (e.g., Calico)
- How do we perform the routing today?
  - FRR stack on the Host with a BGP peering session Host<->ToR
  - Operationally complex, and expensive!
- We can do better!



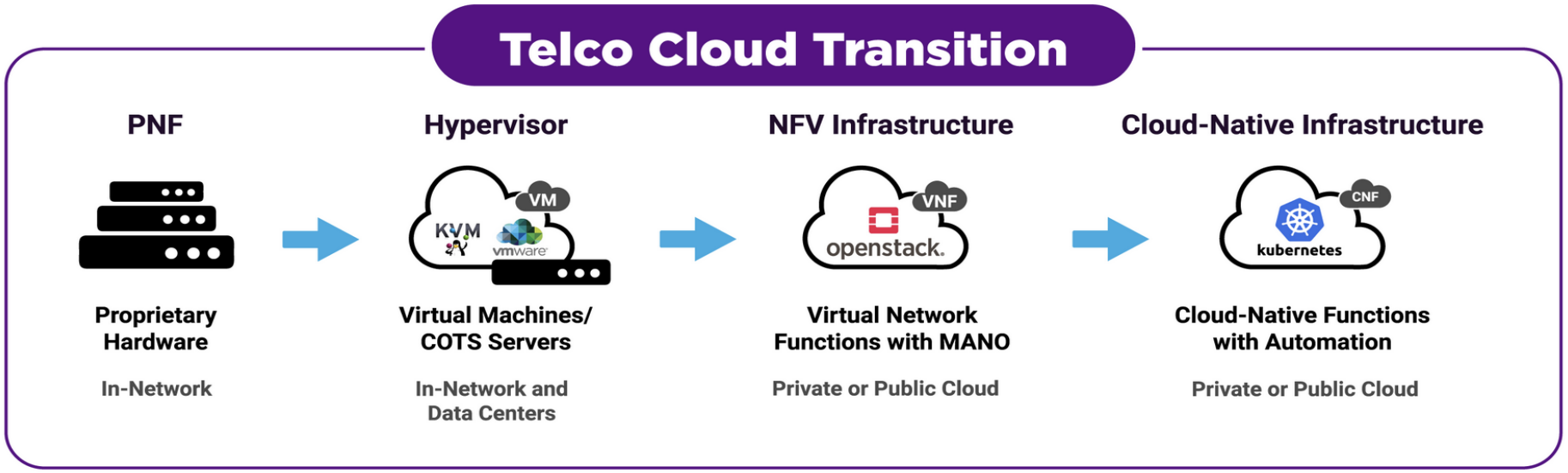
Lightweight Host Routing

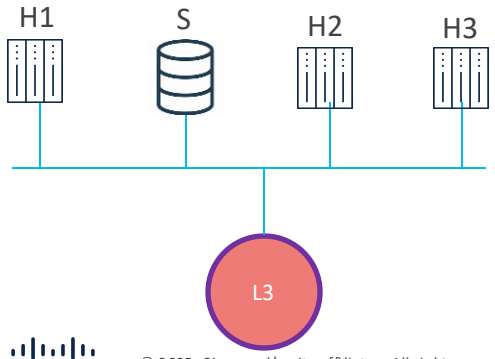
- Lightweight solution based on LLDP
  - Linux include its SRv6 Locator (IP Prefix) within the LLDP advertisements
  - TOR redistributes the prefix into BGP/ISIS
- Simpler solution:
  - Provides reachability (routing) up to the host
  - Provides visibility into the container (workload IP address)
  - Provides liveness detection (built-into LLDP)
- Lightweight:
  - No need to run BGP stack on the host

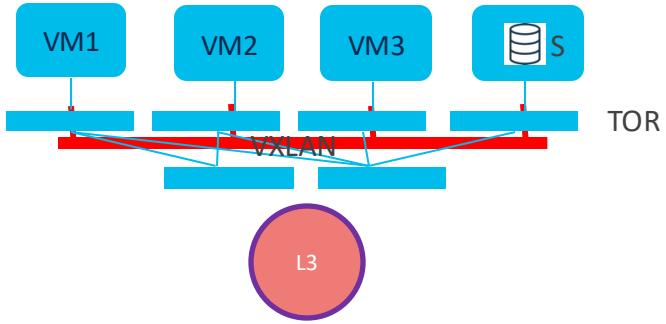


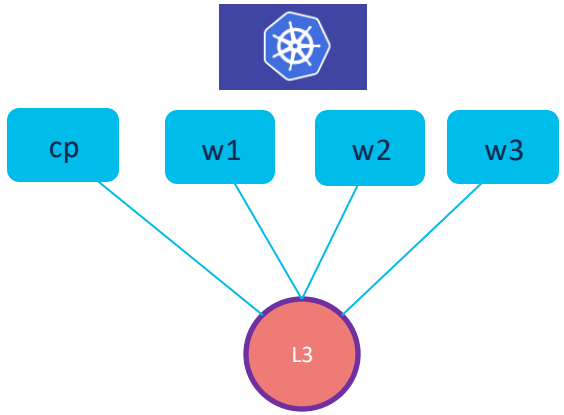

```
RP/0/RP0/CPU0:l7#sh lldp srv6
Mon Oct  2 12:31:04.214 UTC

Locator                                  Holdtime        Mgmt Address                    Interface
-------------------------------------------------------------------------------------------------------------------
---
fcbb:bb00:1818::/48                      120             7::18                           TenGigE0/0/0/3

Total entries: 1

RP/0/RP0/CPU0:l7#
```

Potential for Simplification

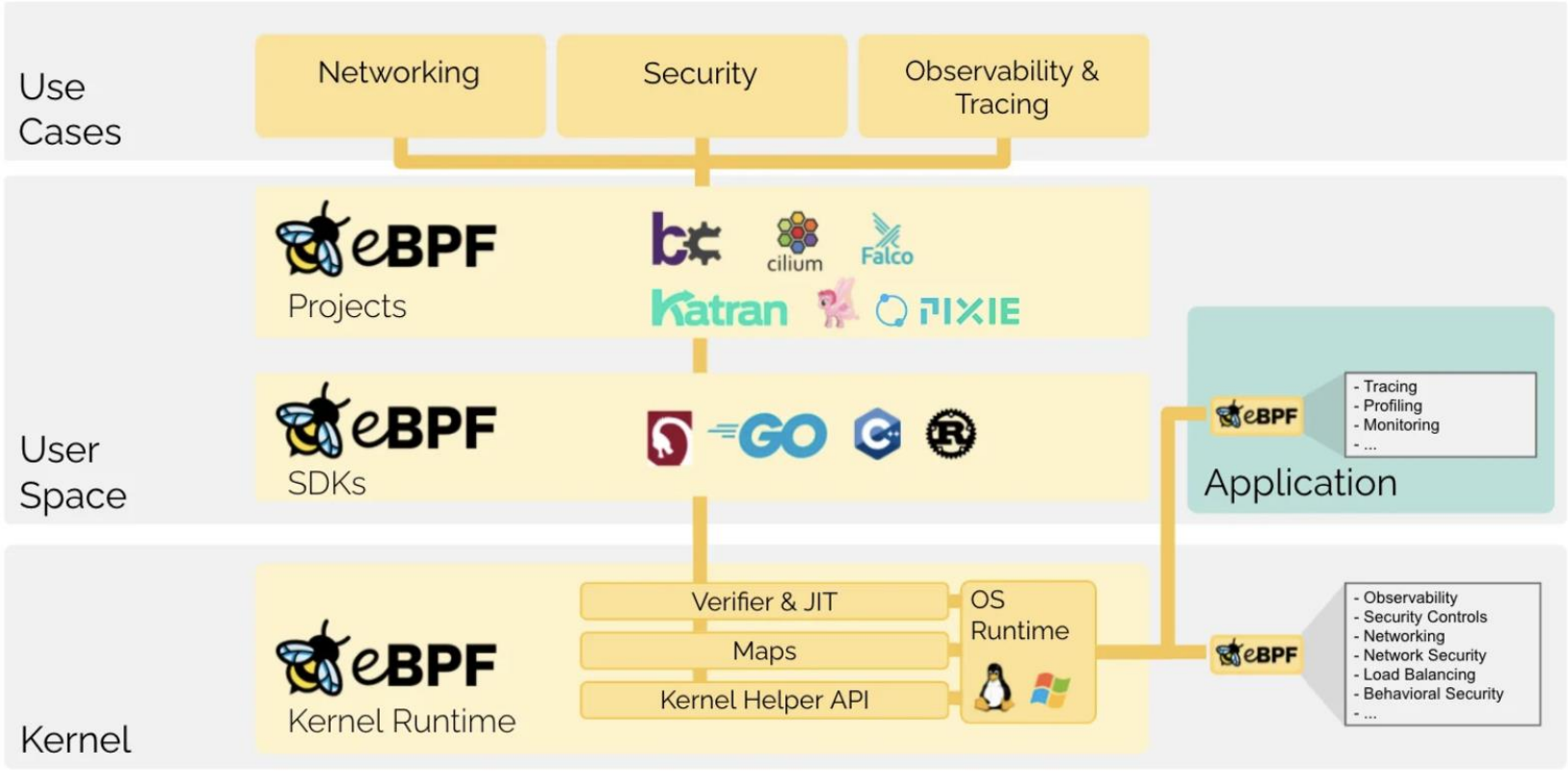




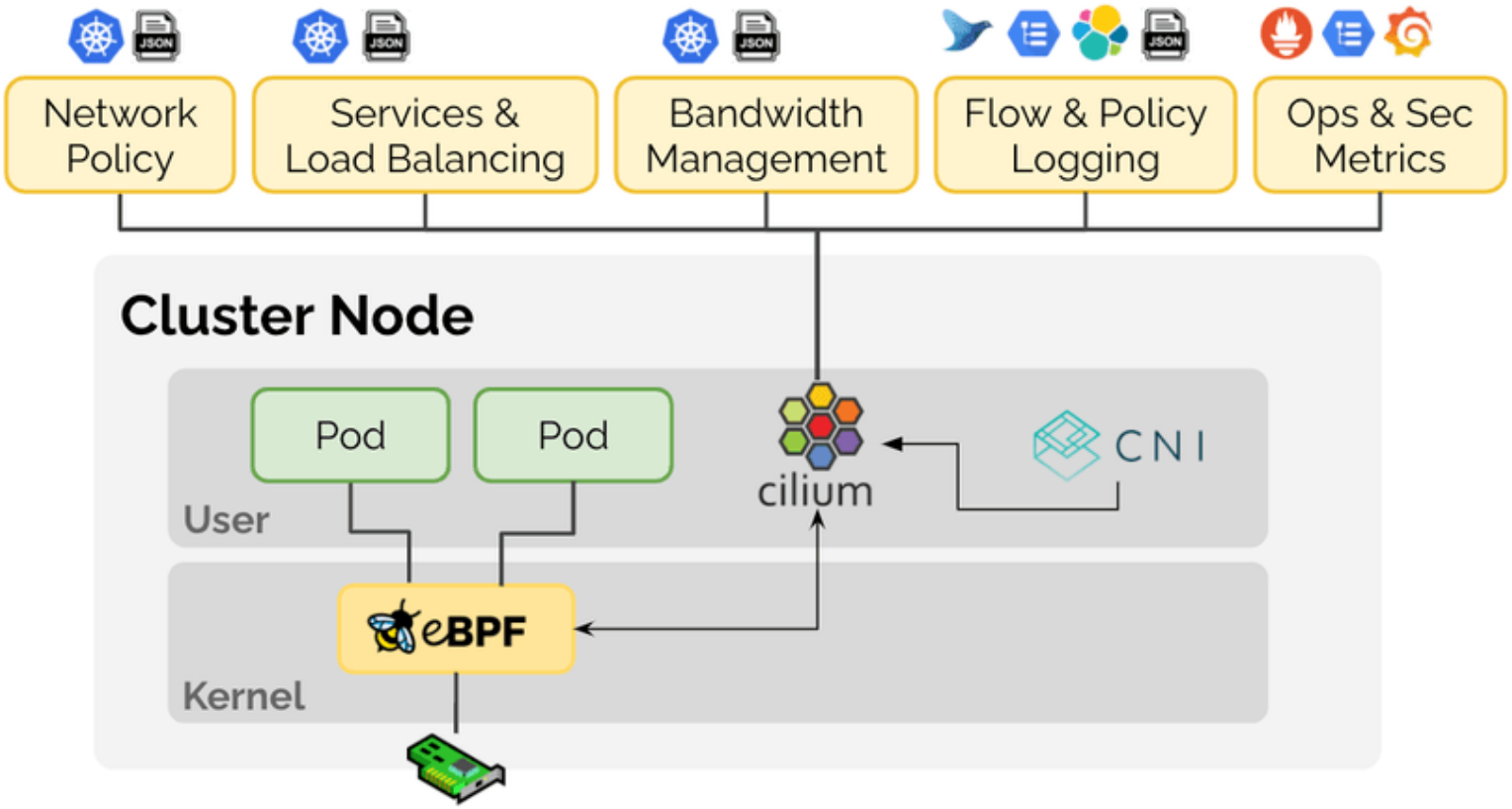




eBPF

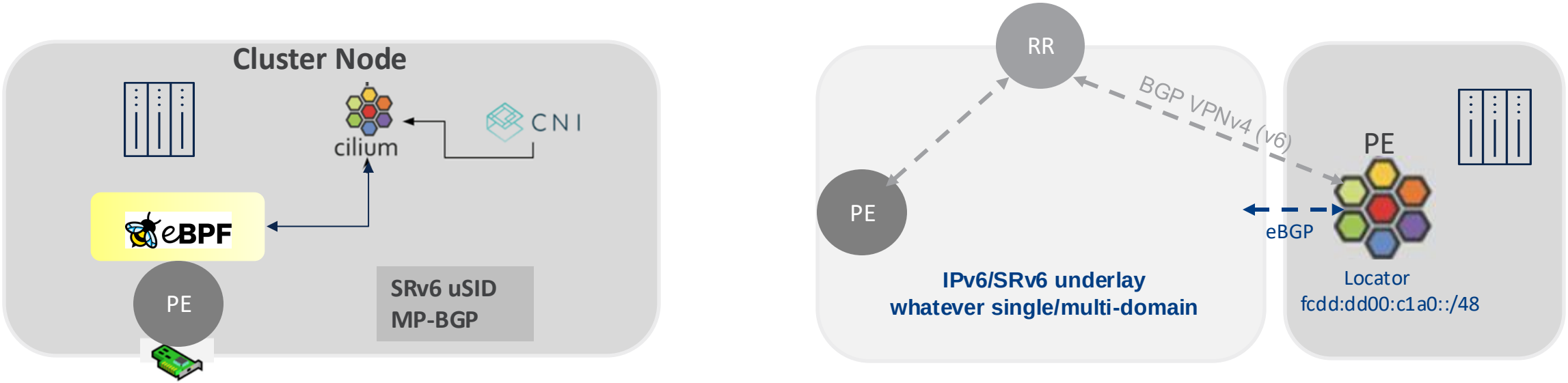


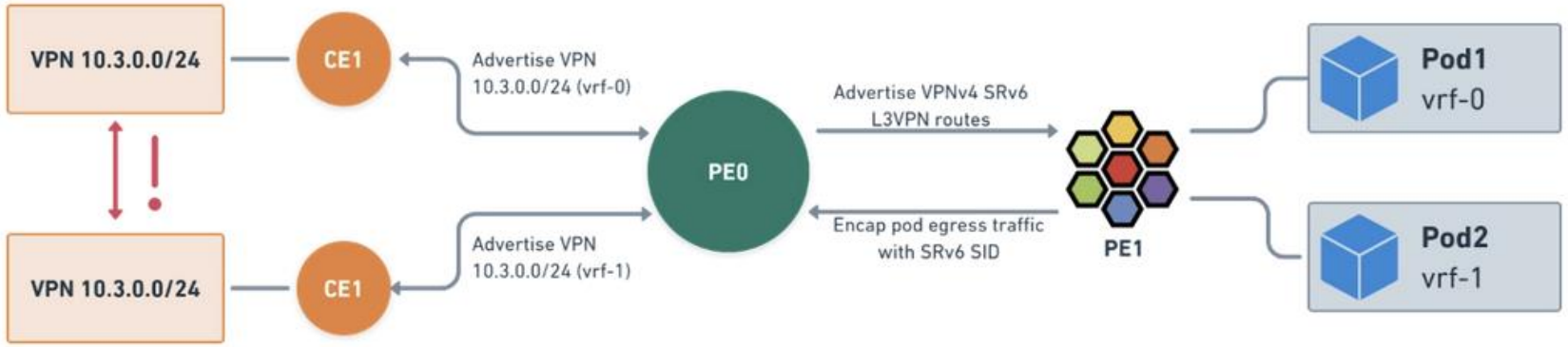
Cilium



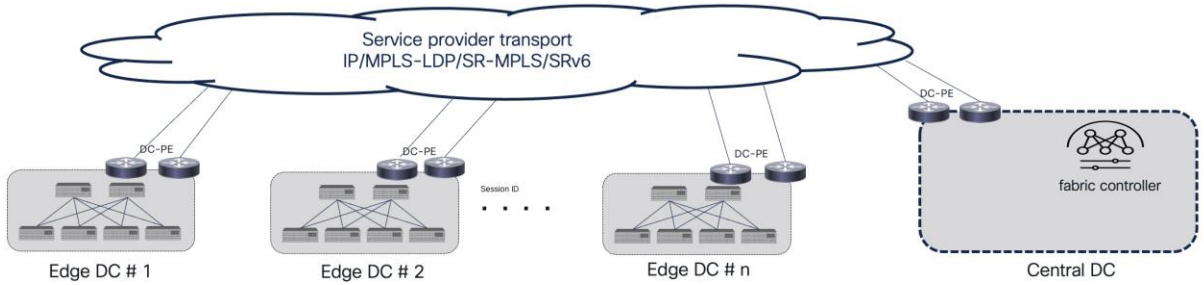
Networking
Security
Visibility

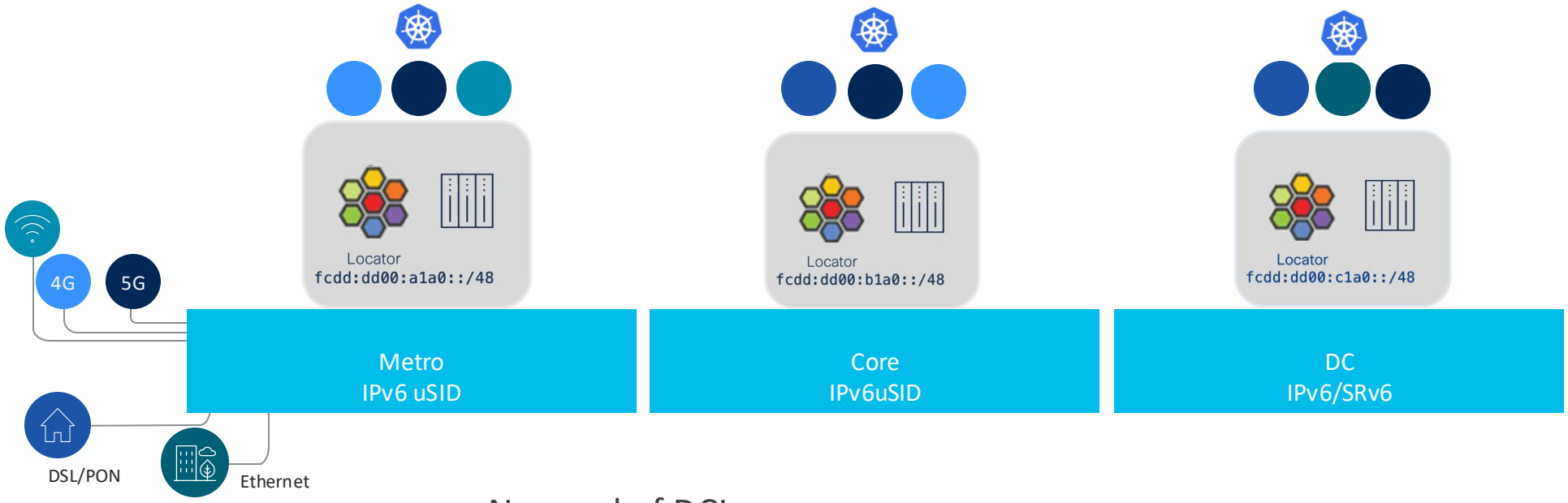
With SRv6 the compute node is becoming a PE





Benefits

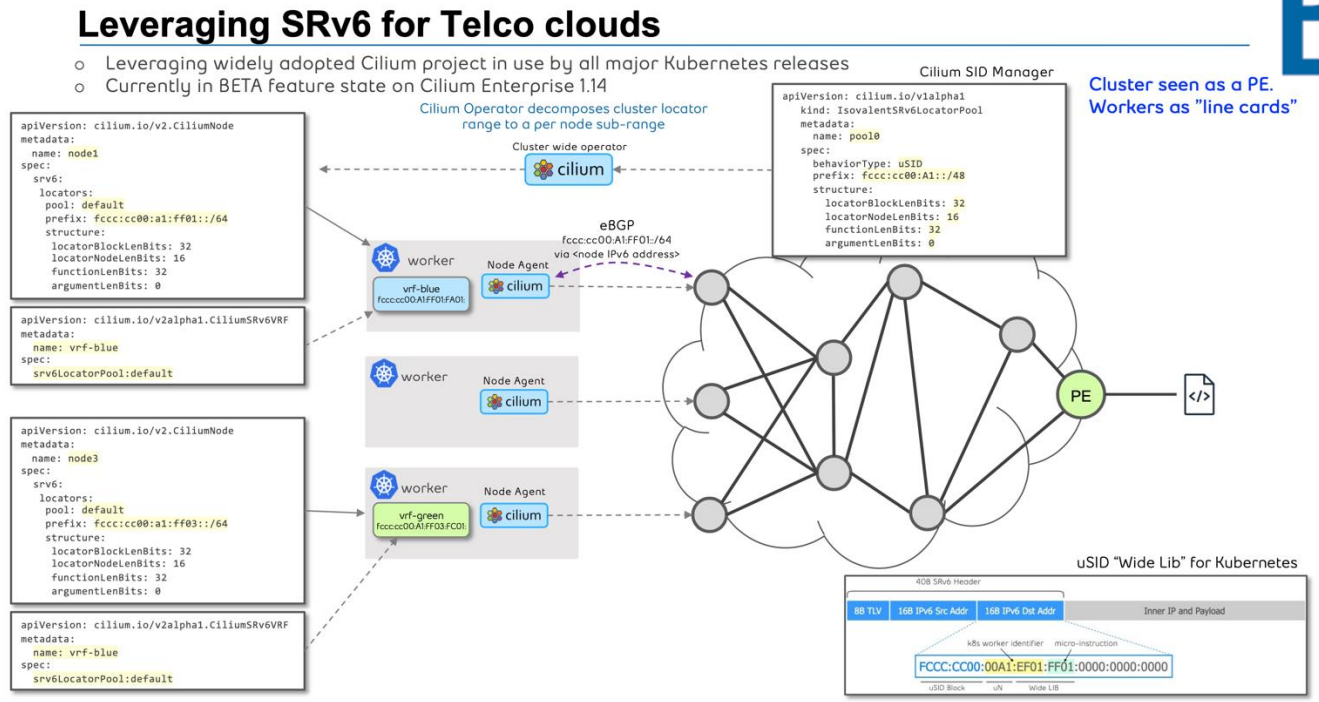


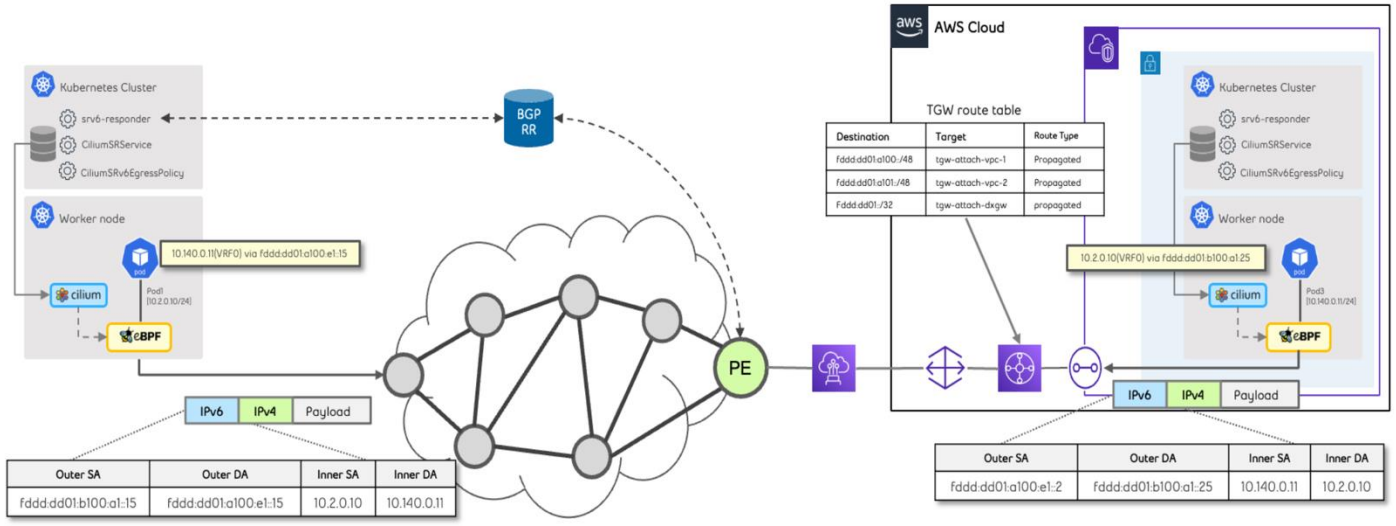


- No need of DCI
- No need of underlay L2 VXLAN
- No need of DC fabric in small sites
- Simplified networking via end 2 end VRF to the host
- Embedded security policies in Cilium
- Minimize service touch points

Bell Canada*

- One of the first SRv6 deployments
- Implementing Cilium SRv6 for Telco Cloud
- Working with AWS and Google to extend it to public cloud
- Dan Bernier's public recordings
  - Presentation at CNCF
  - Presentation at MPLS SD&AI



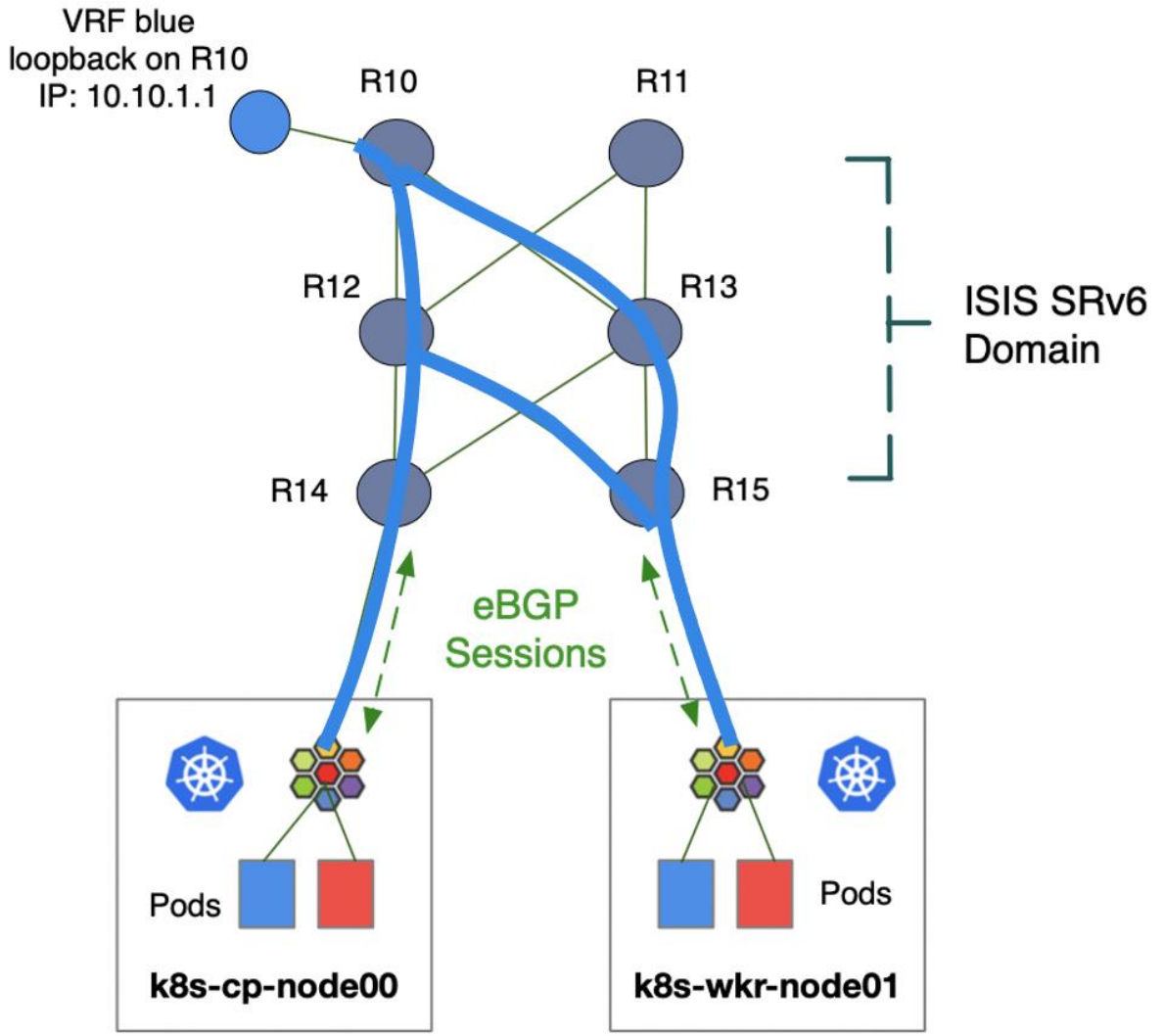


*Courtesy of Dan Bernier, Dir Technology Strategy, Bell Canada

Demo

- デモトポロジ

- R10-15のXRdとCilium間を
  eBGPでつないでSRv6 VPNの
  通信をお見せします



Example output

```
cisco@k8s-cp-node00:~/cilium-srv6/cilium/cilium$ kubectl get
sidmanager
k8s-cp-node00 -o yaml
apiVersion: isovalent.com/v1alpha1
kind: IsovalentSRv6SIDManager
metadata:
generation: 1
name: k8s-cp-node00
resourceVersion: "27756"
uid: 4220c57d-478d-4764-92c9-d050e4a53a9a
spec:
locatorAllocations:
- locators:
- behaviorType: uSID
prefix: fc00:0:15b::/48 <------- control plane node locator
structure:
argumentLenBits: 0
functionLenBits: 16
locatorBlockLenBits: 32
locatorNodeLenBits: 16
poolRef: pool0
```

```
status:
sidAllocations:
- poolRef: pool0
sids:
- behavior: uDT4 <----------- uSID L3VPN IPv4 table lookup
behaviorType: uSID
metadata: blue
owner: srv6-manager
sid:
addr: 'fc00:0:15b:e46b::’   <---- uSID locator+function entry for
structure:                        control plane node VRF blue
argumentLenBits: 0
functionLenBits: 16
locatorBlockLenBits: 32
locatorNodeLenBits: 16
```

Availability

Rich SRv6 uSID Ecosystem

Network Equipment Manufacturers

Cisco, Huawei, Nokia, Ciena, Juniper Networks, ZTE, Arrcus, Arista, H3C

Merchant Silicon

Broadcom, Marvell, Barefoot Networks (Intel)

SR v6 Segment Routing

Open-Source Networking Stacks

Linux, FD.io (VPP), P4, OCP, OPEN Compute Project, Kubernetes, SONiC, FRRouting

Smart NIC

Intel, Mellanox Technologies

Open-Source Applications

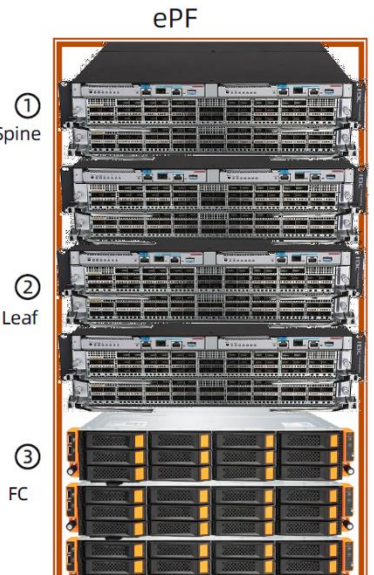
Pyroute2, IPTABLES SERA, GoBGP, exa, BGP, eBPF, cilium, TCPDUMP, WIRESHARK, SNORT, GoBMP

Partners

kaloom, Trend Micro, ENEA Qosmos Division, NoviFlow, ixia A Keysight Business, Spirent Promise. Assured.

IPv6 uSID in SONiC/SAI

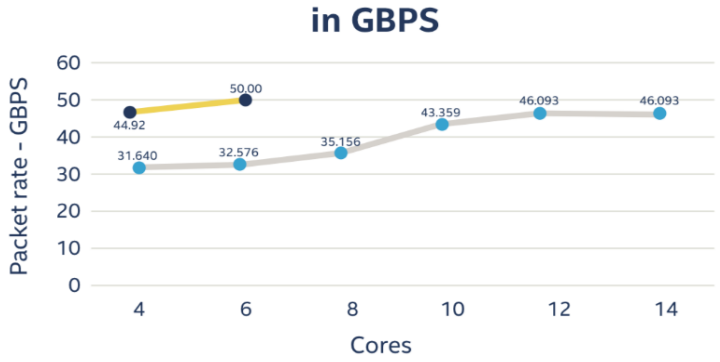
- Collaboration with Alibaba to enable IPv6 uSID whitebox running SONiC
- Full-stack solution: SAI Adaptor + SONiC + FRR
  - SAI: Industry collaboration (CSCO, BCOM, Intel) to define the SRv6 SAI header
    - SAI has full SRv6 uSID support + IPM
  - SONiC/FRR: Today we support an SRv6 fabric with BGP services + ISIS-SRv6 + IPM



Announced@Apsara Conference 2022

SRv6 acceleration on the Host

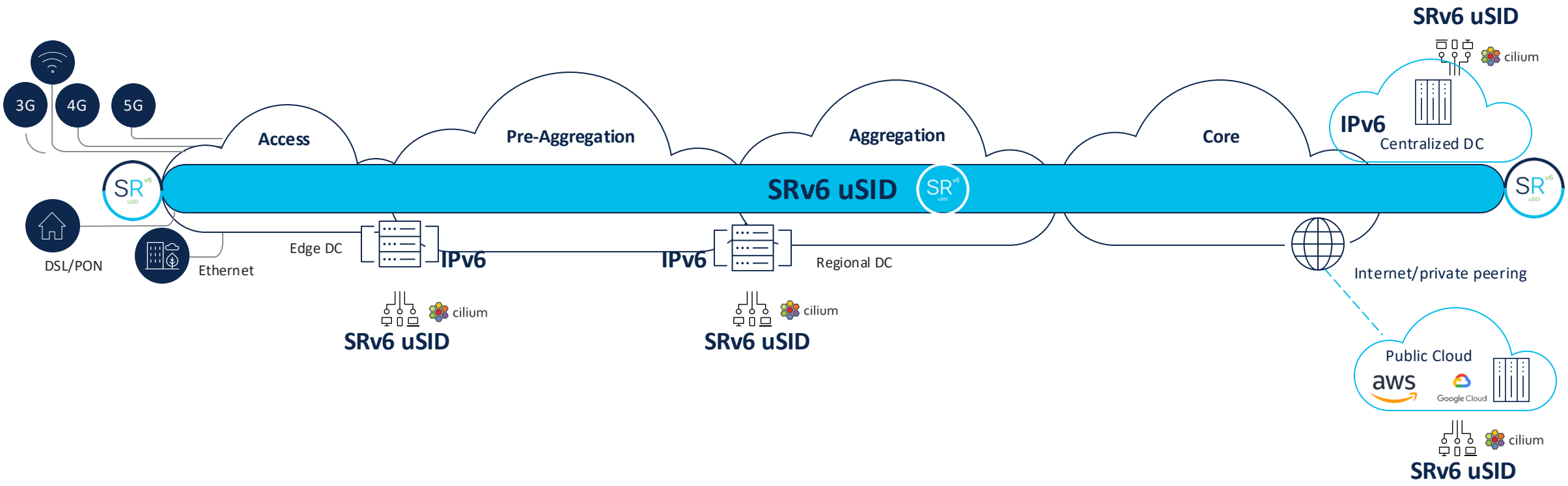
- SmartNIC:
  - Provides a clean separation between IT and Network
    - On top of performance gain by offloading networking and security
  - Intel and Mellanox
- Host stack:
  - Linux Kernel and FD.io VPP for opensource networking stacks
    - Service chaining to external firewalls and load-balancers
  - Host routing stacks: FRRouting, GoBGP, ExaBGP
  - Kubernetes CNIs (Cilium and eBPF)
    - Steering capabilities between Kubernetes PODs and clusters



https://www.intel.com/content/dam/www/programmable/us/en/pdfs/literature/solution-sheets/sb-accelerate-srv6-processing.pdf

Conclusion

Conclusions



Any Service over IP without any shims

Unified Solution with Better Reliability

Lower equipment and operational cost

Native Host and Cloud

Embedded Security

議論したいこと

- なぜ日本のデータセンターへのSRv6の適用事例が少ないのか?
- SRv6のデータセンターへの適用に向けて足りないものは何か?
- Service chainingとかはやりやすい？でも実際に必要ですか？

- 例えば、一般的なデータセンタースイッチでSRv6サポートされたら使いますか?
  - すでにSRv6をWAN側でも使っている人たちにはメリットがありそう？
  - データセンター単体で考えてもメリットありますか？

- 標準化団体にフィードバックしたい内容とかありますか？


CISCO