SRv6

Record-Speed Standardization and Deployment

Clarence Filsfils
cf@cisco.com

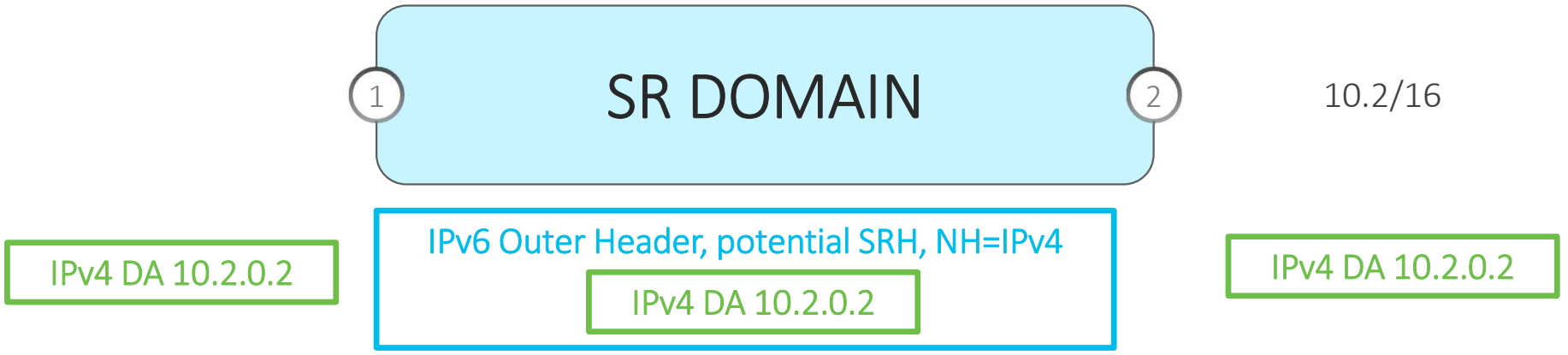
Agenda

- Introduction
- Deployment and Standardization Status
- The power of SRv6 uSID's
- uSID – Reminder
- uSID – Use-Cases
  - EPE, VPN, TILFA, uLoop, NFV, 5G
- Conclusion

segment-routing.net

Introduction

Transparent Service



- Customer packet is encapsulated from ingress to egress of the SR Domain
- SRv6 is applied to the OUTER header
- The inner packet is untouched

SRv6 Network Programming - RFC8986

- The End-to-End Policy is encoded as a Network Program
  - The first instruction is in the outer DA
  - The remaining instructions are in the SRH
- An instruction (a SID) may be bound to any behavior
  - TILFA FRR and uLoop Avoidance
  - Traffic Engineering: internal to the domain and across peering links
  - L2/L3 VPN's
  - NFV
  - Any HW custom behavior: P4 on Silicon1
  - Any SW custom behavior: Container orchestrated by Kubernetes

Powerful Service Creation

- Any service can be encoded as an ordered list of instructions: e.g.
  - Low-latency Slice
  - & VPN
  - & Service Chaining of various NFV’s distributed in regional and core DC’s
  - & Absolute Loss Measurement

Stateless Fabric

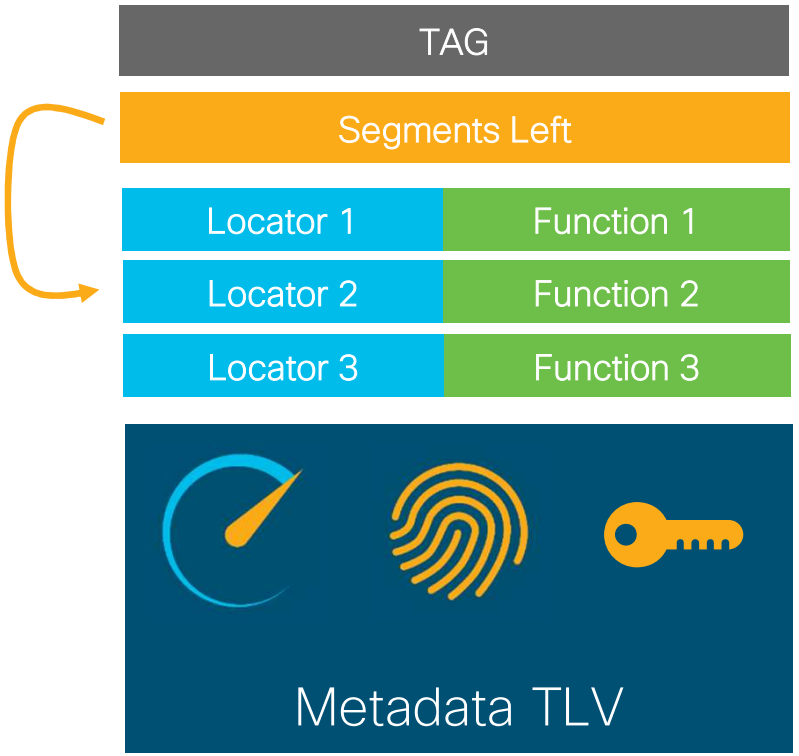
- The state (network program) is in the header
- The state is not in the fabric

De-Facto SDN Architecture

- The SDN controller programs the host/edge with the end-to-end network program
- The Application programs the network

SR Extension Header

- SRv6 is a native extension of IPv6
  - RFC 8754
  - As foreseen 25 years ago by RFC2460
- SRH contains an ordered list of SID's



SRv6 Provides the Required Reach

- Reach required by 5G
  - IP is on the UE
  - IP is in the socket
- Reached required by Hyper-Scale DC
  - IP scales to billions of containers

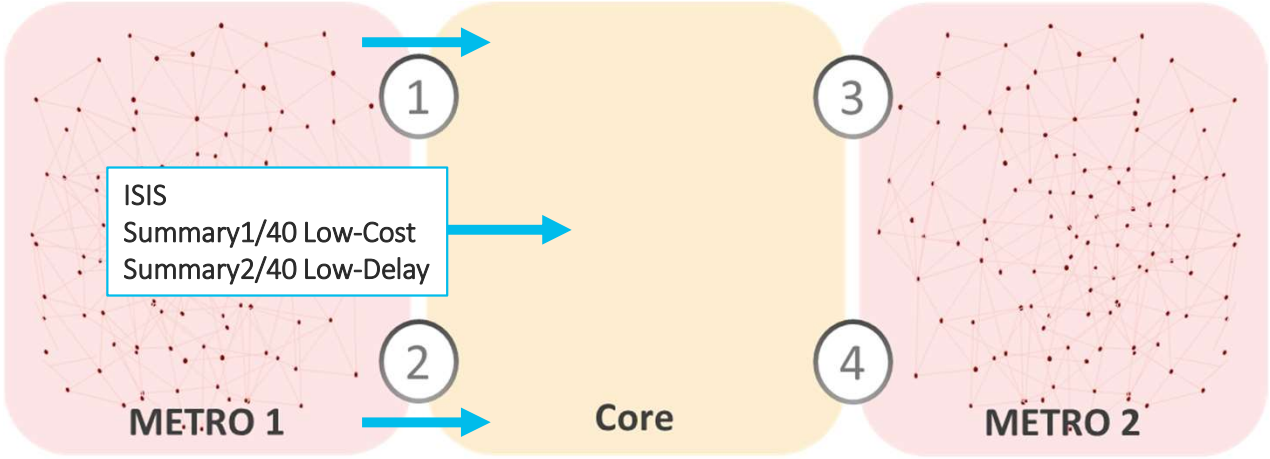
SRv6 Provides a Unified Dataplane

- One single dataplane natively supported by all IP nodes:
  - UE
  - Container
  - Host
  - Cell Site
  - Access
  - Metro
  - Core
  - DC

Seamless Deployment

- Seamless forwarding through IPv6 transit nodes
- Most use-cases do not need an SRH
  - DA SID contains up to 6 micro-instructions (uSID's)

Prefix Summarization



- 1000 times less IGP routes than with MPLS
- No BGP inter-AS Option A/B/C
- End-to-End Flex Algo Continuity

Optimum Load Balancing

- The entropy of the inner packet is present in the outer Flow Label

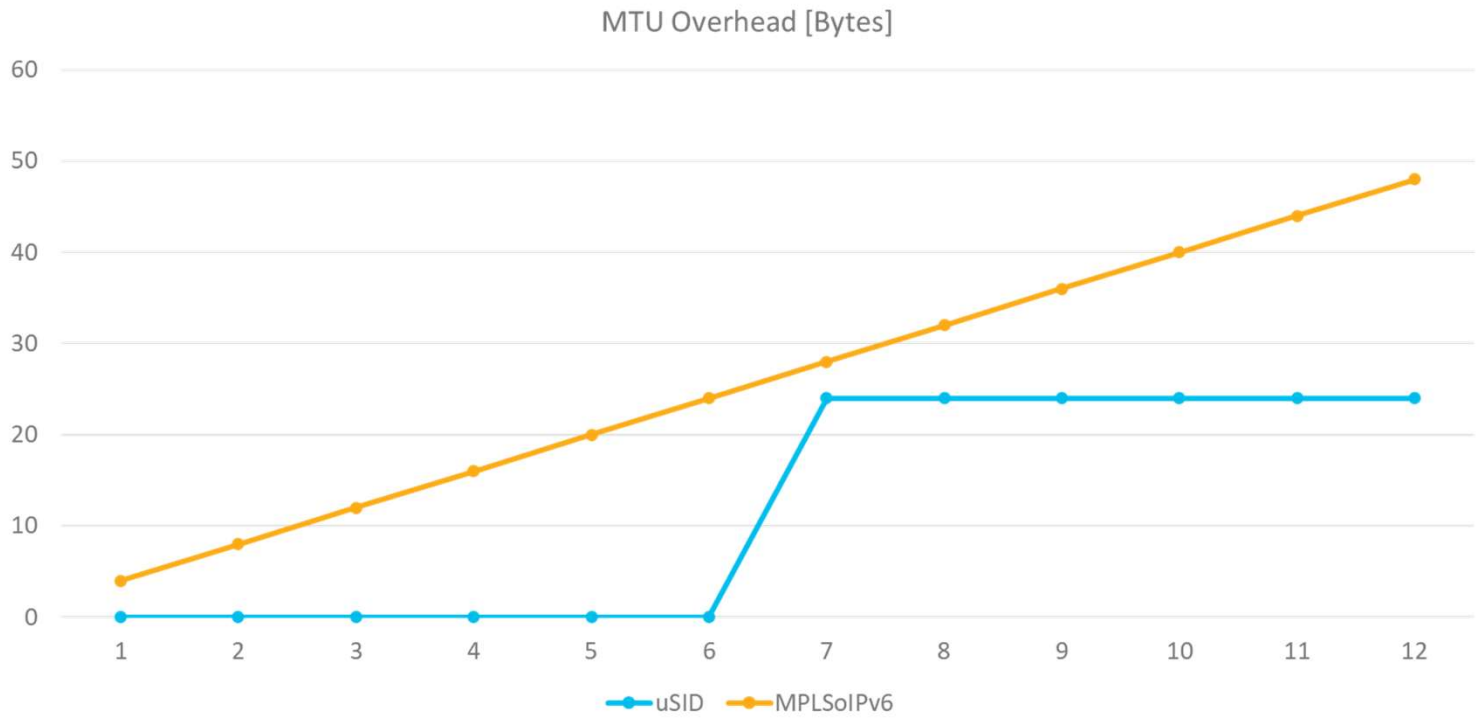
Negligeable IPv6 address space consumption

- Iliad < 1 billionth of private IPv6 space (FC/8)
- SBB < 1 millionth of current SBB public space

Less is More

- No MPLS dataplane
  - 66% less dataplane entries and counters
- No MPLS to the host
  - Bringing MPLS to the host is very complex: i.e. there is no label in a socket
- No VxLAN gateway at the DC edge
  - SRv6 end-to-end
  - Massive economical benefit
- No UDP trick for LB
- No RSVP-TE for TE/FRR
- No NSH for NFV

Lowest MTU Overhead



One of the many benefits of SRv6 Micro Instructions

Optimized for Hardware Efficiency

- Legacy merchant: J1
  - 8-year-old, 4 technology generations old (28nm)
  - 10’s of millions of SRv6 subscribers at linerate
- Silicon1 P100
  - World’s Highest Performance Routing device
    - > World's most efficient fixed router (24x800G)
    - > World’s first 28.8T (36x800G) line card (with the G100 as a fabric)
  - Linerate with more than 20 SRv6 micro-instructions
  - This is likely twice more than ever required

Deployment, Eco-System and Standardization

Record-Speed Deployment

- ~100M SRv6 subscribers with 2.5 years of commercial service
- ~100 deployments, with ~10 public reports

SoftBank	Indosat	China Unicom	MTN Uganda
Iliad	China Telecom	China Bank	Noia
Rakuten	Bell Canada	Cernet2	Line

Rich Eco-System

- 25 HW implementations
  - Cisco Systems, Nokia, Arrcus, Kaloom, NoviFlow, Huawei, ZTE, Juniper
  - Broadcom, Barefoot, Intel SmartNIC, Marvell, Mellanox,
  - Spirent, Ixia
  - Multiple Interop Reports
- 11 open-source platforms/ Applications
  - Linux, FD.io VPP, P4, iptables, nftables, snort, SERA, ExaBGP, GoBGP, GoBMP, Contiv-VPP

Mature Standardization

- Proposed Standard
  - RFC 8402 SR Architecture
  - RFC 8754 SRv6 DataPlane
  - RFC 8986 SRv6 Network Programming
  - RFC xxxx SRv6 ISIS Extension
- Last step to Proposed Standard RFC
  - BGP
  - BGP-LS
  - Flex Algo
  - Policy Architecture
  - OAM

The Power of SRv6 uSIDs

Perfect SRv6 Integration

- uSID reuses SRH (RFC8754) without any change
- uSID applies the SRv6 Network Programming (RFC8986)

Perfect IPv6 Integration

- uSID is an ingenuous leverage of CIDR (RFC7608)

"CIDR rules, even within an SR domain. For that reason, the fact that the bottom 64 bits in the "address" look funny or change is simply irrelevant. They are invisible to routing (which is done based on the prefix)…."

Brian Carpenter, former IETF and IAB chair

https://mailarchive.ietf.org/arch/msg/ipv6/37bH3Ag7jaNgKwnZY_mKIxWRwdM/

Extreme Hardware Friendliness

- Simply built on “longest match lookup”
  - The basic capability of any NPU
  - The basic nature of routing in the Internet (CIDR)
- One single lookup can resolve multiple uSID’s at once
  - No other technology can do this
- uSID’s of various sizes can be combined
  - always with one single longest-match lookup

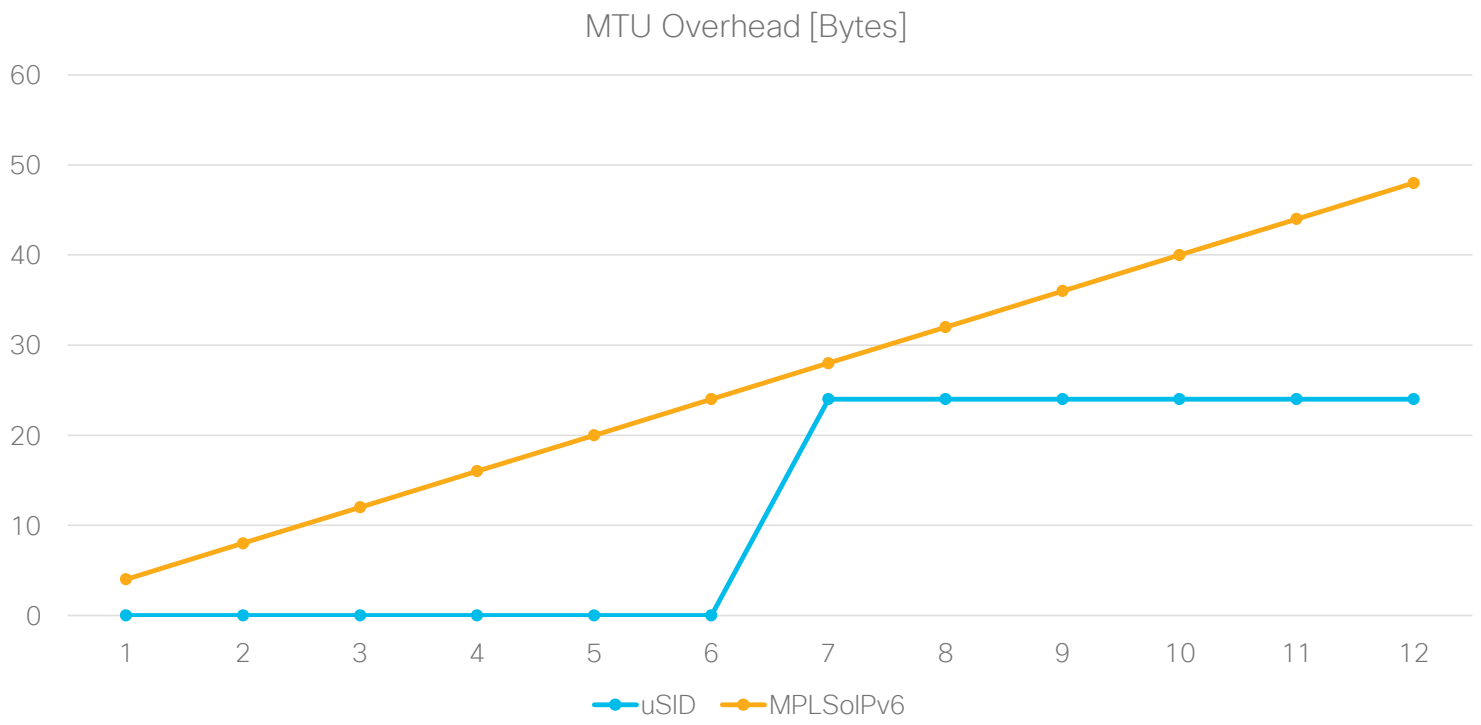
Any instruction can be bound to a uSID

- Slicing: Min Cost, Min Delay
- Disjointness
- Traffic Engineering
- TILFA / uLoop
- L2VPN, L3VPN, EPE Peering Optimization
- NFV
- Measurement (Absolute packet loss)
- Custom: HW P4, SW Container…

Scalable

- Deployed Solution already supports 15 million fabric nodes
  - More can be added in the future
- Solution supports billions of services per node
  - E.g. a PW aggregation PE can easily support 512k PW’s

Lowest MTU Overhead



Impressive uSID Roadmap

- TILFA
- uLoop Avoidance
- Flex-Algo: Low-Cost/Low-Delay Slicing
  - With VPN Automated Steering into Flex-Algo
- Performance Monitoring: Link Latency (for Low-Delay slice)
- Seamless Inter-Domain SRv6 with summarizing and slicing
- L3VPN (IPv4 and IPv6), IPv4 Internet, IPv6 Internet, PW
- VPN GW to interconnect “new” SRv6-VPN buildup with legacy VPN

Impressive 5G Deployments

News Release

Rakuten Mobile Advances Its Network for 5G and IoT Services with Cisco

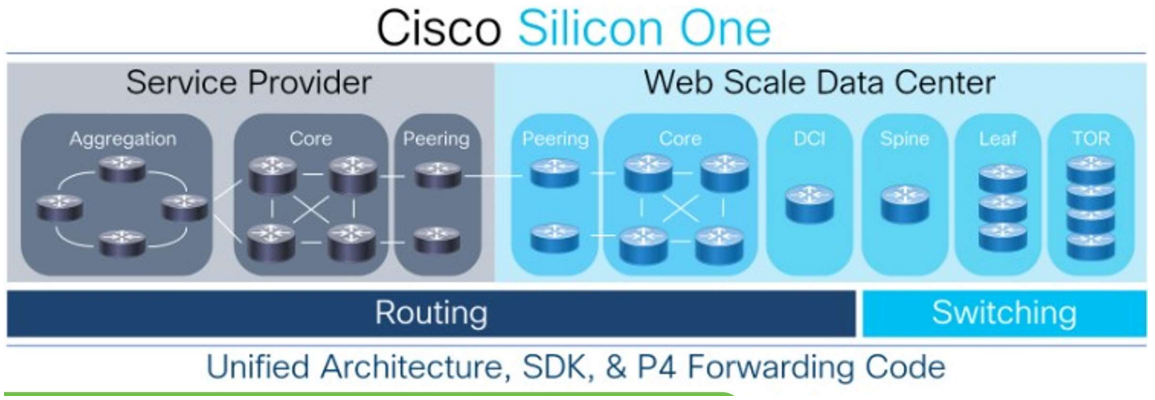
June 25, 2021

TOKYO, JAPAN, June 25, 2021 – Cisco and Rakuten Mobile, Inc. today announced a major milestone for Rakuten Mobile's network infrastructure in support of efforts to build a better, more inclusive internet for the future.

Rakuten Mobile operates the world first's fully cloud-native mobile network. It launched 4G service in Japan in April 2020, and launched 5G non-standalone (NSA) services in September 2020 in record time. With four million subscribers today, Rakuten Mobile continues to advance and scale its network to support new demands driven by the growth of remote and mobile workers.

With the implementation of Segment Routing over IPv6 (SRv6) and Cisco Routed Optical Networking, Rakuten Mobile plans to expand its capabilities to support enterprise customers with 5G and IoT services. To support its future 5G SA services with network slicing capabilities, Rakuten Mobile will introduce SRv6 micro-segments, an extension to the SRv6 network programming model that is key to addressing multi-domain 5G deployments

Impressive SDN Deployments



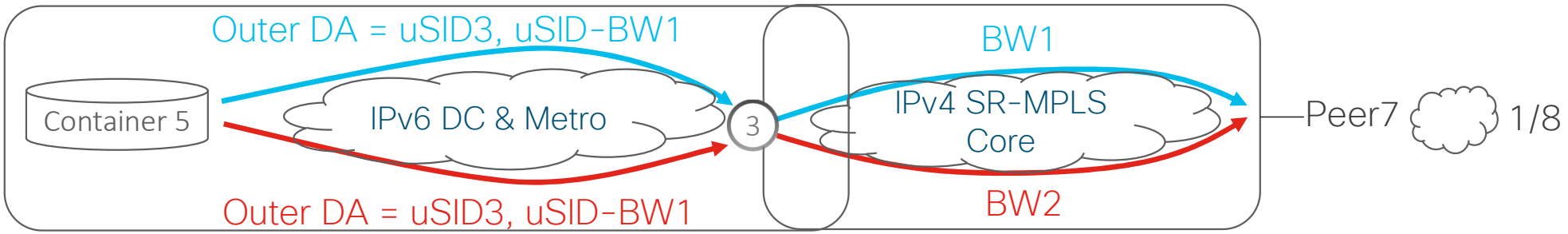
"Embracing the path towards network softwarization and "in-network" or ubiquitous computing, SRv6 Network Programming and its uSID implementation brings software like programmability and agility at the protocol level, with a broad set of implementations.

Silicon1 and its P4 SDK brings it a step further by extending its reach at the ASIC level for massive scale, speed and cost efficiency. "

Daniel Bernier

Impressive Host-Based Deployment leveraging legacy SR-MPLS Services

Major Cloud Player



"Host/Container/GW do IPinIP encap based on application(outer DA = SID), service anchor point(edge) to decap and execute the SID function accordingly. Basically it's endpoint+network programming approach, leveraing the strong SRv6 eco-system across hardware, chip and OS"

Architect

SRv6 uSID - Reminder

SRv6 Network Programming

- SRv6 Network Programming (RFC8986)
  - The source encodes any end-to-end program as an ordered list of instructions
  - The first instruction is in the outer DA
  - The remaining instructions are in the SRH
- An instruction is called a SID
- A Container SID may contain up to 6 micro-instructions called uSID's
  - The IETF term is "NEXT-CSID"

Container of 6 uSID's

FC00:0000:1111:2222:3333:4444:5555:6666

Min-Cost Block

- We recommend to allocate uSID's from Private IPv6 Space (FC/8)
  - FC/8: unroutable outside the domain, hence more secure
  - Public address is also possible
- We recommend /32 Block
  - /16 and /48 are also possible
- In this presentation, we will use
  - FC00:0000/32 for Min-Cost Slice: shortest path based on minimum ISIS cost
  - FC00:0008/32 for Min-Delay Slice: shortest path based on minimum latency

Container of 6 uSID's

FC00:0000:1111:2222:3333:4444:5555:6666

Min-Cost Block | uSID1 | uSID2 | uSID3 | uSID4 | uSID5 | uSID6

- uSID "default size" is 4 nibbles (16 bits) ":WXYZ:"
  - uSID's of different length can be mixed
  - 32-bits uSID's are used for ultra-scale service
- Up to 6 uSID's in the outer DA
- A uSID program reads left to right
- Intuitively: within the Min-Cost Slice, first go to 1111 then 2222 then …

Less than 6 uSID's in the outer DA

Outer DA: FC00:0000:1111:2222:3333:4444:0000:0000

uSID1 uSID2 uSID3 uSID4 EoC EoC

- Unused uSID's in the micro program are filled with ":0000:"
- ":0000:" means "End of Container" (EoC)

If more than 6 uSID's are required

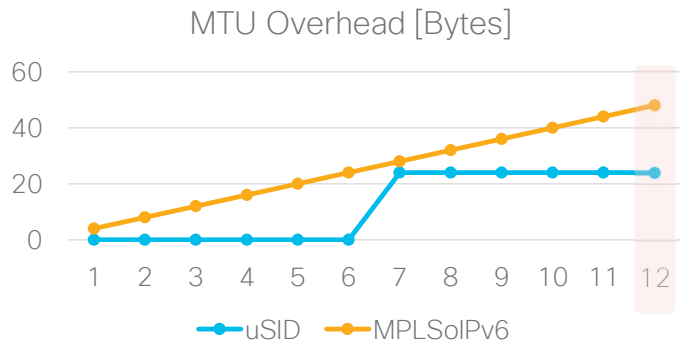
Outer DA: FC00:0000:0001:0002:0003:0004:0005:0006

uSID1 uSID2 uSID3 uSID4 uSID5 uSID6

Outer SRH: FC00:0000:0007:0008:0009:0010:0011:0012

uSID7 uSID8 uSID9 uSID10 uSID11 uSID12

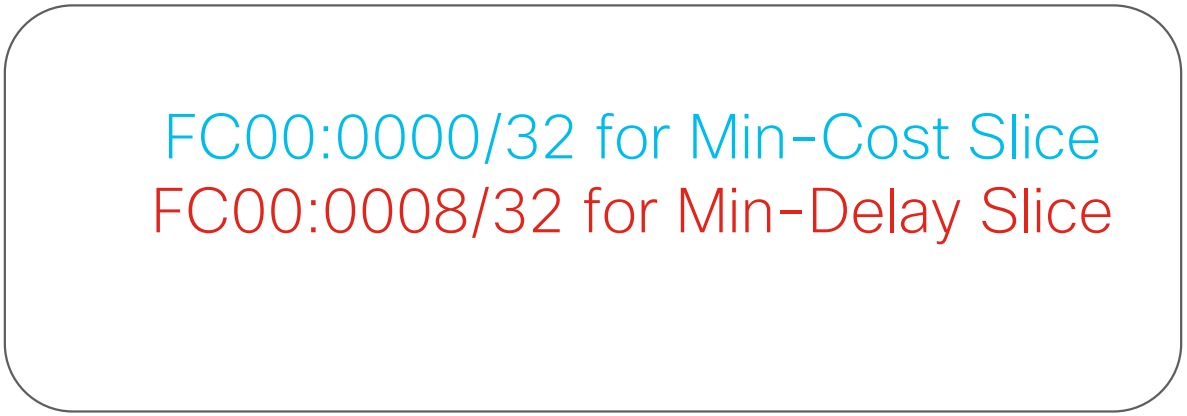
- 12 uSID's with an outer SRH holding one single additional uSID container
  - 6 in the DA, 6 in the SRH
  - With solely 24-bytes of MTU overhead
  - 50% less overhead than MPLS in IPv6 (12*4 bytes)



An SRv6 uSID’s may be bound to any behavior

- VPN
- TILFA/uLoop
- TE
- EPE
- NFV
- Any customer behavior
  - HW: P4
  - SW: Container

SR Domain



- Each slice gets a /32 uSID block

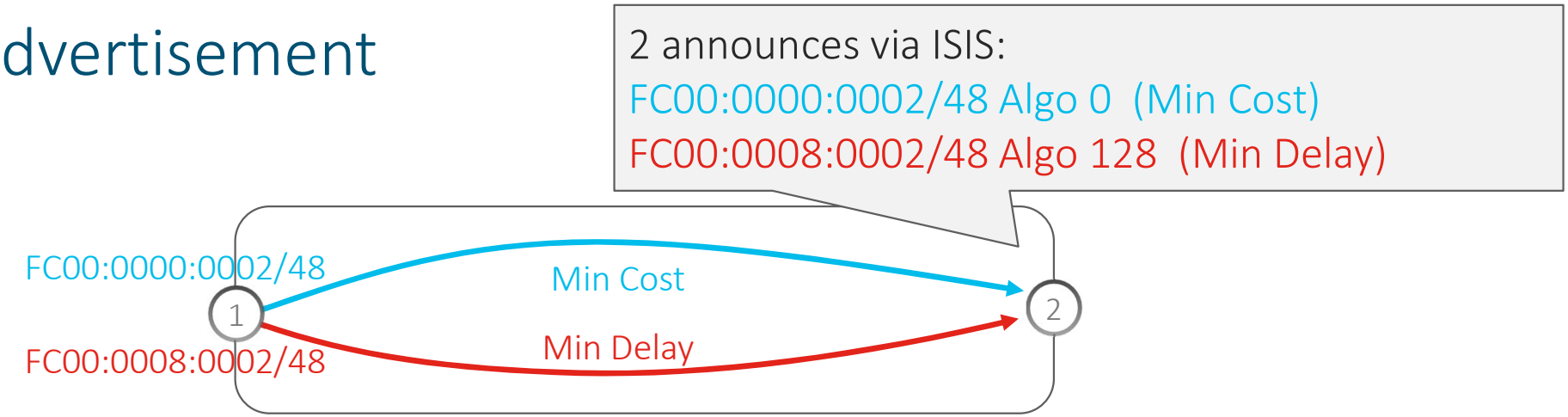
SR Node

FC00:0000.0002/48 Min-Cost uN of 2

FC00:0008.0002/48 Min-Delay uN of 2

- A node gets a Shortest-Path Endpoint uSID (uN) from each slice
- A uN is a /48 off the /32 of the related slice
- Classic Prefix-Based Routing (CIDR)

ISIS Advertisement



- Classic IP Routing
  - Flex-Algo based routing to a /48

BGP Advertisement

2 announces via BGP:
RD9:10.2.0/24, RT9, via 2,
with SID: FC00:0000:0002:F009::

- Intuitive uSID program:
  - Within the Min-Cost Slice (FC00:0000)
  - Follow the shortest-path to 2 (0002)
  - Execute VPN9 Decaps at 2 (F009)
- Seamless Deployment
  - Any transit node routes on a classic /48
- Hardware Efficiency
  - Egress PE 2 processes multiple uSID's with a single /64 lookup
  - FC00:0000:0002:F009/64

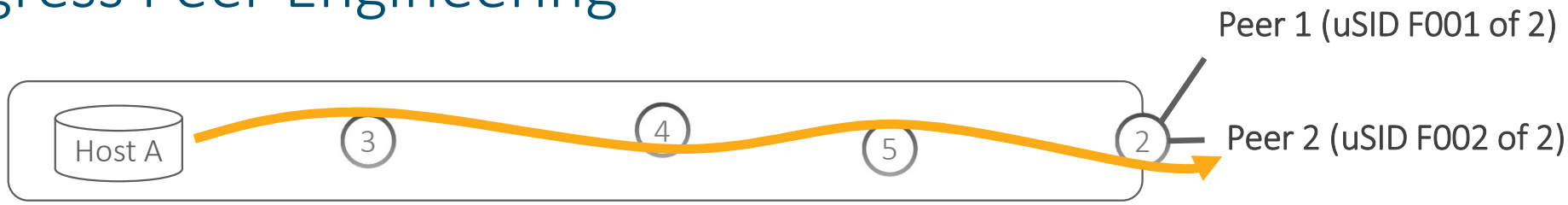
BGP Advertisement per Slice

2 announces via BGP:
RD9:10.2.8/24, RT9, via 2,
with SID: FC00:0008:0002:F009::

- Intuitive uSID program:
  - Within the Min-Cost Slice (FC00:0008)
  - Follow the shortest-path to 2 (0002)
  - Execute VPN9 Decaps at 2 (F009)
- Seamless Deployment
  - Any transit node routes on a classic /48
- Hardware Efficiency
  - Egress PE 2 processes multiple uSID's with a single /64 lookup
  - FC00:0008:0002:F009/64

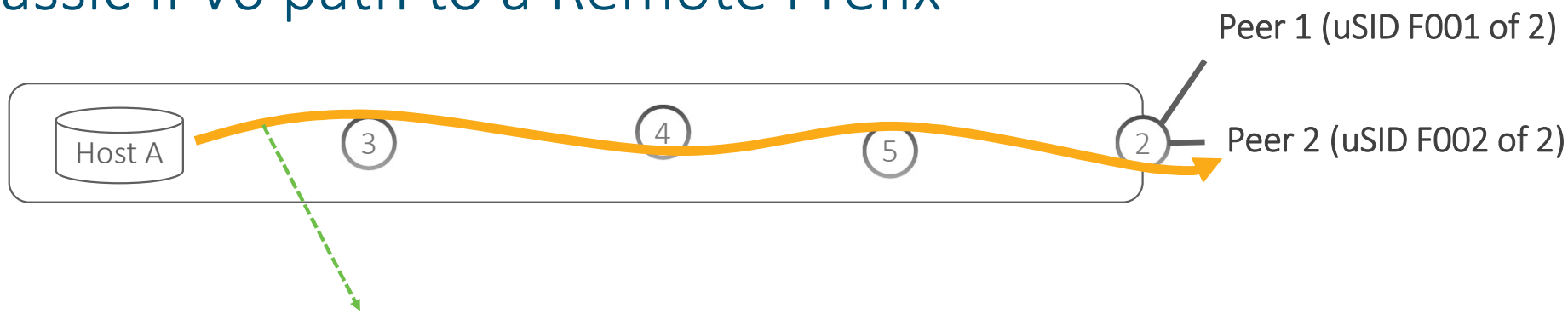
SRv6 uSID
EPE Use-Case

Egress Peer Engineering



- Well-known use-case described by Facebook, Google and Alibaba (Sig17)
- SDN programs Host A to encapsulate its egress traffic via
  - Intra-Domain TE: 3 then 4 then 5 then 2
  - Inter-Domain TE: via Peer2
- Implemented with a single outer DA
  - FC00:0000:0003:0004:0005:0002:F002::
- Intuitive reading:
  - Within Min-Cost slice, go to 3, then 4, then 5, then 2 and @2 take peering F002

Classic IPv6 path to a Remote Prefix

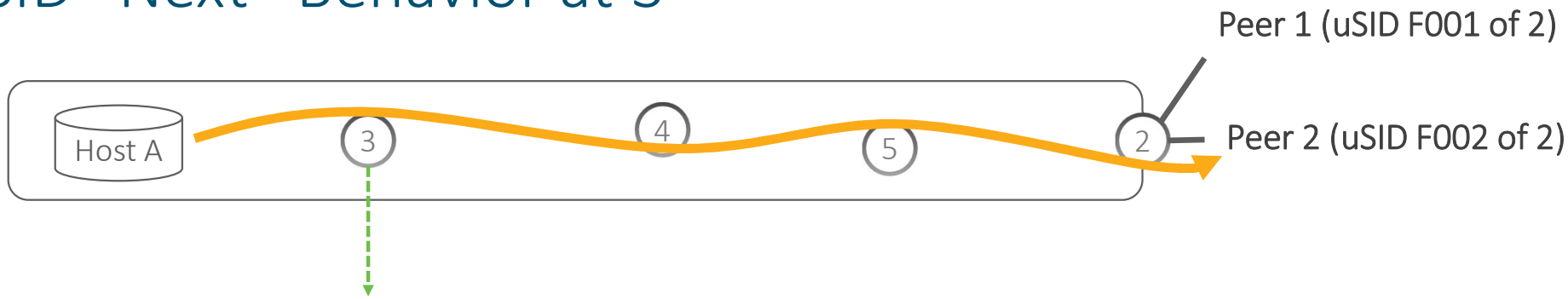


DA = FC00:0000:0003:0004:0005:0002:F002:0000

➔ Follows remote prefix FC00:0000:0003/48

Simple application of 25-year-old CIDR: RFC4632 and RFC7608

uSID “Next” Behavior at 3



DA = FC00:0000:0003:0004:0005:0002:F002:0000

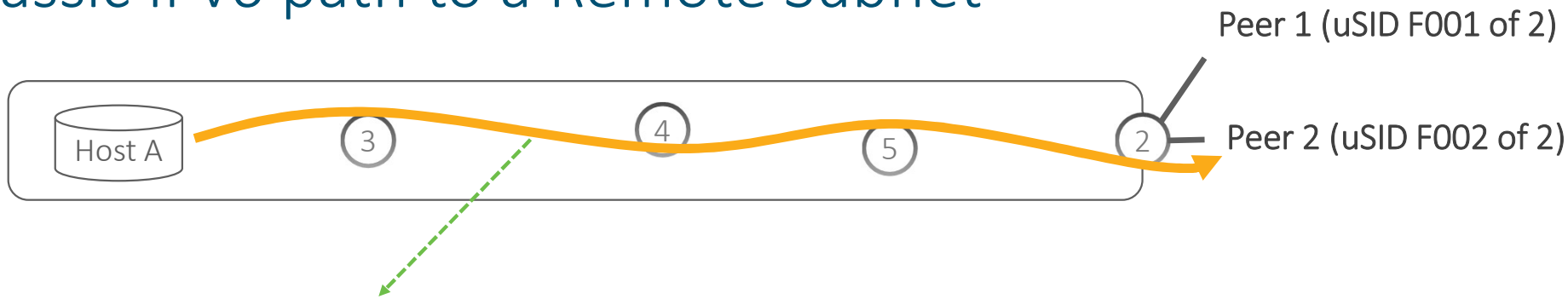
➔ Matches local SID FC00:0000:0003/48

➔ Apply “Next” Behavior

➔ DA becomes FC00:0000:0004:0005:0002:F002:0000:0000

Simple application of SRv6 Network Programming: RFC8986

Classic IPv6 path to a Remote Subnet

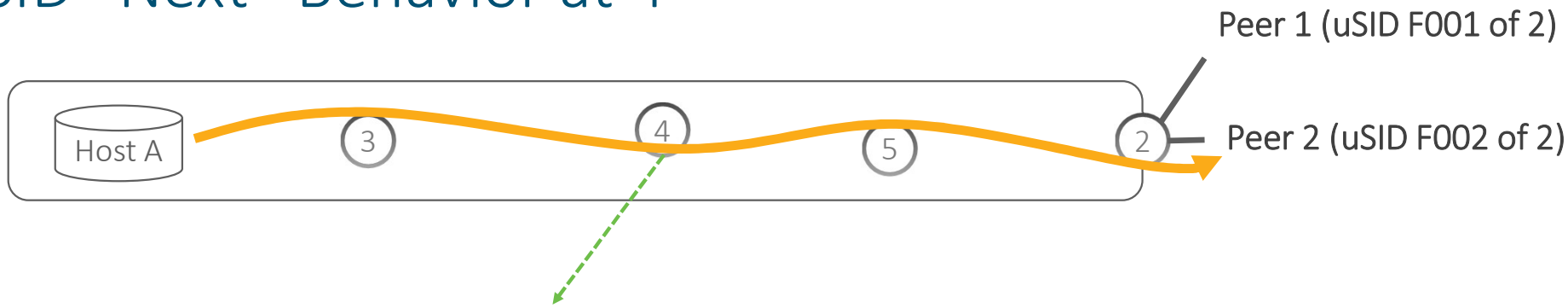


DA = FC00:0000:0004:0005:0002:F002:0000:0000

➔ Follows remote prefix FC00:0000:0004/48

Simple application of 25-year-old CIDR: RFC4632 and RFC7608

uSID “Next” Behavior at 4



DA = FC00:0000:0004:0005:0002:F002:0000:0000

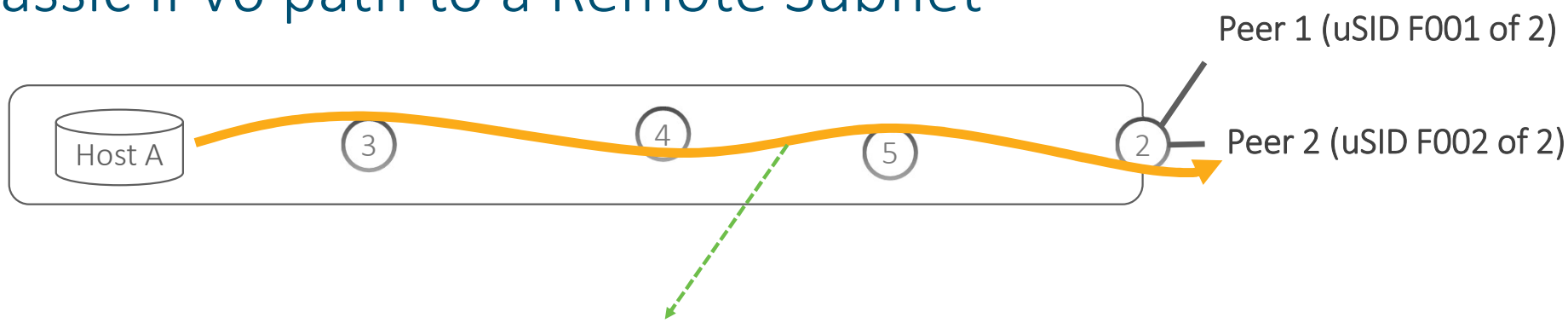
➔ Matches local SID FC00:0000:0004/48

➔ Apply “Next” Behavior

➔ DA becomes FC00:0000:0005:0002:F002:0000:0000:0000

Simple application of SRv6 Network Programming: RFC8986

Classic IPv6 path to a Remote Subnet

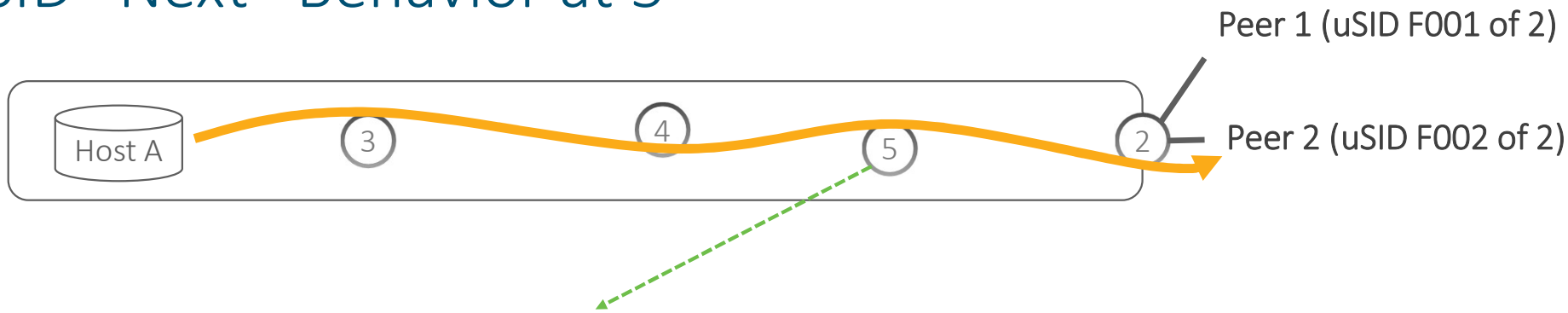


DA = FC00:0000:0005:0002:F002:0000:0000:0000

➔ Follows FC00:0000:0005/48

Simple application of 25-year-old CIDR: RFC4632 and RFC7608

uSID “Next” Behavior at 5



DA = FC00:0000:0005:0002:F002:0000:0000:0000

➔ Matches local SID FC00:0000:0005/48

➔ Apply “Next” Behavior

➔ DA becomes FC00:0000:0002:F002:0000:0000:0000:0000

Simple application of SRv6 Network Programming: RFC8986

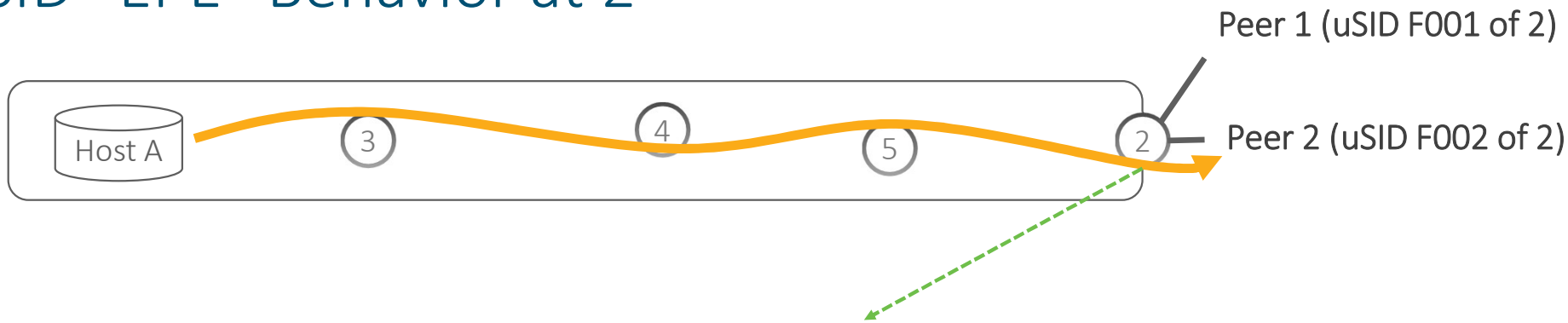
Classic IPv6 path to a Remote Subnet

DA = FC00:0000:0002:F002:0000:0000:0000

➔ Follows FC00:0000:0002/48

Simple application of 25-year-old CIDR: RFC4632 and RFC7608

uSID “EPE” Behavior at 2



DA = FC00:0000:0002:F002:0000:0000:0000:0000

➔ Matches local SID FC00:0000:0002:F002/64

➔ Apply “EPE” Behavior

➔ Decaps the inner packet and forward it to Peer2

Simple application of SRv6 Network Programming: RFC8986

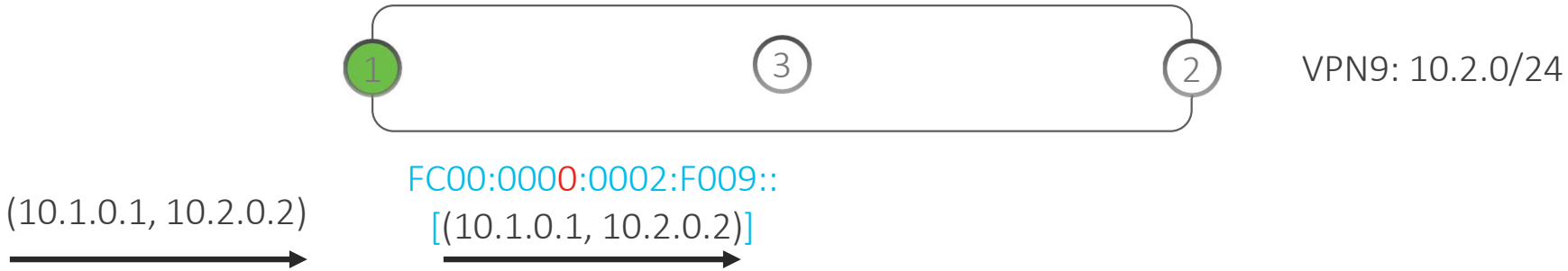
Benefit: Seamless Deployment & Legacy Platform Leverage

- Transit nodes perform classic IPv6 forwarding
  - No upgrade (legacy is fine)
- Hosts encapsulate in IPv6 with an opaque outer DA provided by SDN
  - No upgrade (legacy is fine)
- Similar constructs may be used on the Cell Site Router

SRv6 uSID
VPN & Slice Use-Cases

VPN over Min-Cost 5G Slice - Ingress PE

Commercial Deployment



- iPE1 learns via BGP that 10.2.0/24 in VPN9 is reachable via SID FC00:0000:0002:F009
- iPE1 encapsulates with outer DA = FC00:0000:0002:F009
- Intuitive reading: FC00:0000:0002:F009
  - Within Min-Cost slice, take shortest-path to 2 where VPN-Decaps into VRF9 is implemented

VPN over Min-Cost 5G Slice – Transit P

Commercial Deployment

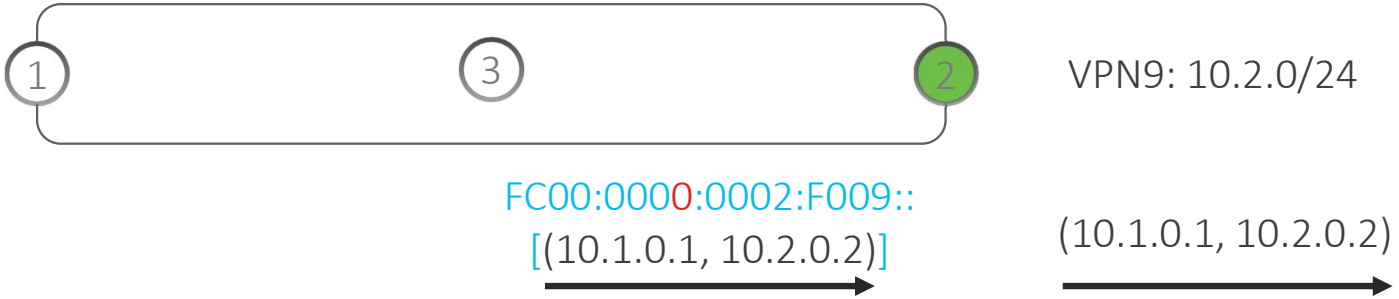
VPN9: 10.2.0/24

- Transit Node 3 forwards along remote prefix FC00:0000:0002/48
  - ISIS Shortest-Path with Algo 0 (Min Cost)

Simple application of 25-year-old CIDR: RFC4632 and RFC7608

VPN over Min-Cost 5G Slice – Egress PE

Commercial Deployment

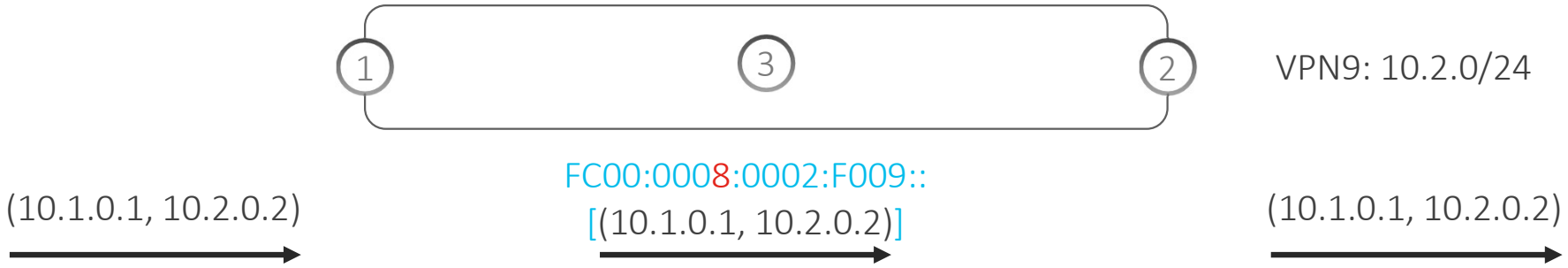


- ePE2 matches local SID FC00:0000:0002:F009/64
- ePE2 Applies “VPN Decaps” Behavior into VRF9

Simple application of SRv6 Network Programming: RFC8986

VPN over Min-Delay 5G Slice

Commercial Deployment

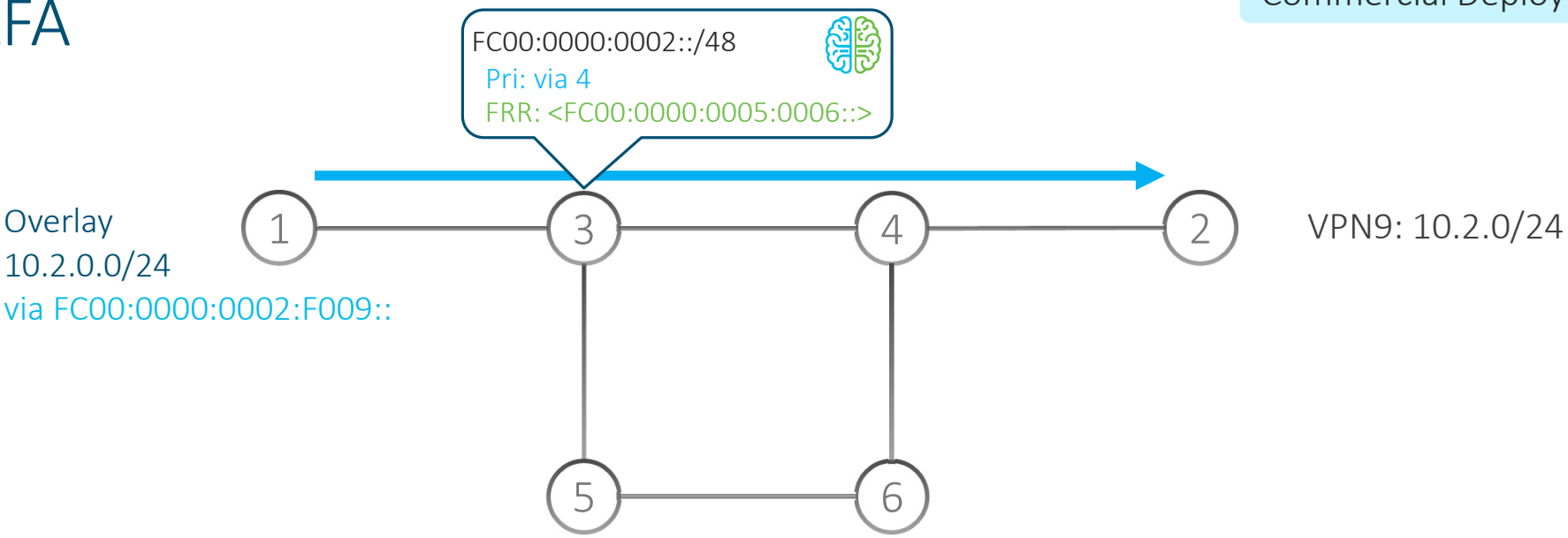


- iPE1 learns via BGP that 10.2.0/24 in VPN9 is reachable via SID FC00:0008:0002:F009
- @1: encapsulates with outer DA = FC00:0008:0002:F009
- @3: forwards based on remote prefix FC00:0008:0002/48 along Min-Delay shortest path
- @2: matches local SID FC00:0008:0002:F009/64 and applies “VPN Decaps” Behavior
- Intuitive reading: FC00:0008:0002:F009
  - Within Min-Delay slice, take shortest-path to 2 where VPN-Decaps into VRF9 is implemented

SRv6 uSID
TI-LFA and uLoop Use-Cases

TI-LFA

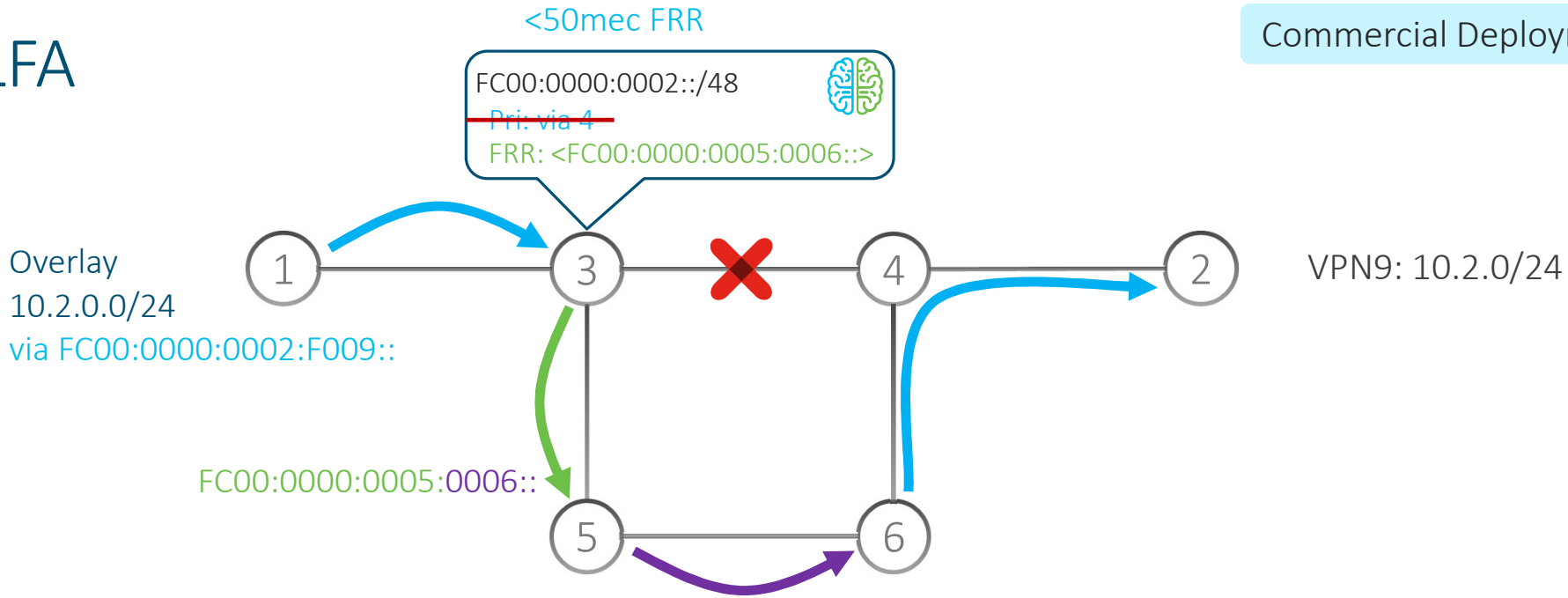
Commercial Deployment



- 50msec Protection upon local link, node or SRLG failure
- Simple to operate and understand
  - automatically computed by the router's IGP process
  - 100% coverage across any topology
  - predictable (backup = post-convergence)
- Optimum backup path
  - leverages the post-convergence path
  - avoid any intermediate flap via alternate path
- Incremental deployment
- Distributed and Automated Intelligence

TI-LFA

Commercial Deployment



- Upon failure of its outgoing interface to destination FC00:0000:0002::/48, 3 does
  - Add an SRH with the received DA FC00:0000:0002:F009::
  - Set DA to FC00:0000:0005:0006::
- Intuitive Reading:
  - FC00:0000:0005::/48 leads the packet along the shortest path to 5
  - FC00:0000:0006:0000::/64 leads the packet along the shortest path to 6 and finishes the micro-program in the DA
  - The next SID is then loaded from the SRH into the DA
  - The packet resumes its path to 2 with original DA = FC00:0000:0002:F009::

uLoop

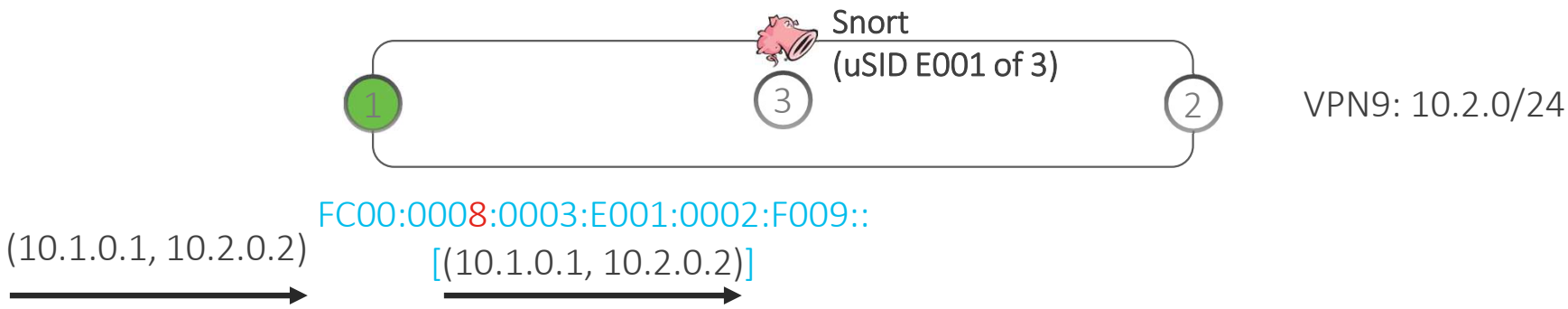
Commercial Deployment

- Straightforward Implementation with SRv6 uSID
  - Same as TILFA

SRv6 uSID
NFV Use-Cases

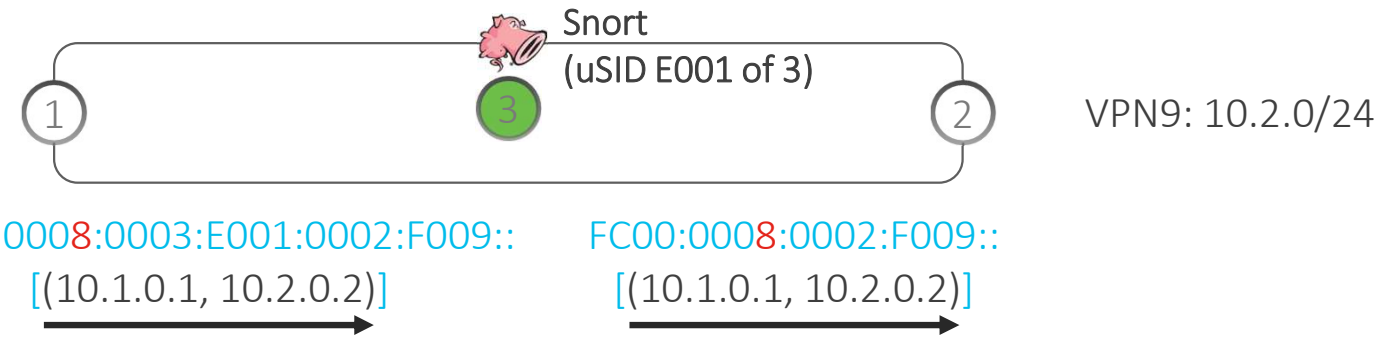
Commercial Deployment

Firewall NFV in Min-Delay 5G Slice with VPN - Ingress PE



- iPE1 learns via BGP that 10.2.0.0/24 in VPN9 is reachable via SID FC00:0008:0002:F009 and Firewall SLA
- iPE1 encapsulates with outer DA = FC00:0008:0300:E001:0002:F009
- Intuitive reading: FC00:0008:0300:E001:0002:F009
  - Within Min-Delay slice, take shortest-path to 3 and apply Snort policy (E001); then take shortest-path to 2 where VPN-Decaps into VRF9 is implemented

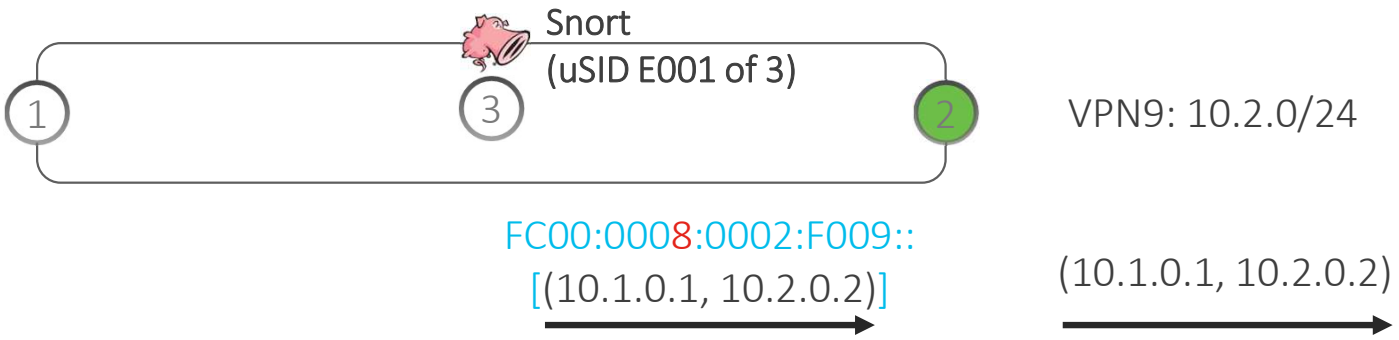
Firewall NFV in Min-Delay 5G Slice with VPN - NFV



- Node 3 matches local SID FC00:0008:0003:E001::/64
- Node 3 applies “Snort policy” to the packet
- Node 3 activates the next uSID and forwards over ISIS Shortest-Path with Algo 128 (Min delay)

Simple application of SRv6 Network Programming: RFC8986
Any behavior can be bound to a SID

Firewall NFV in Min-Delay 5G Slice with VPN – Egress PE



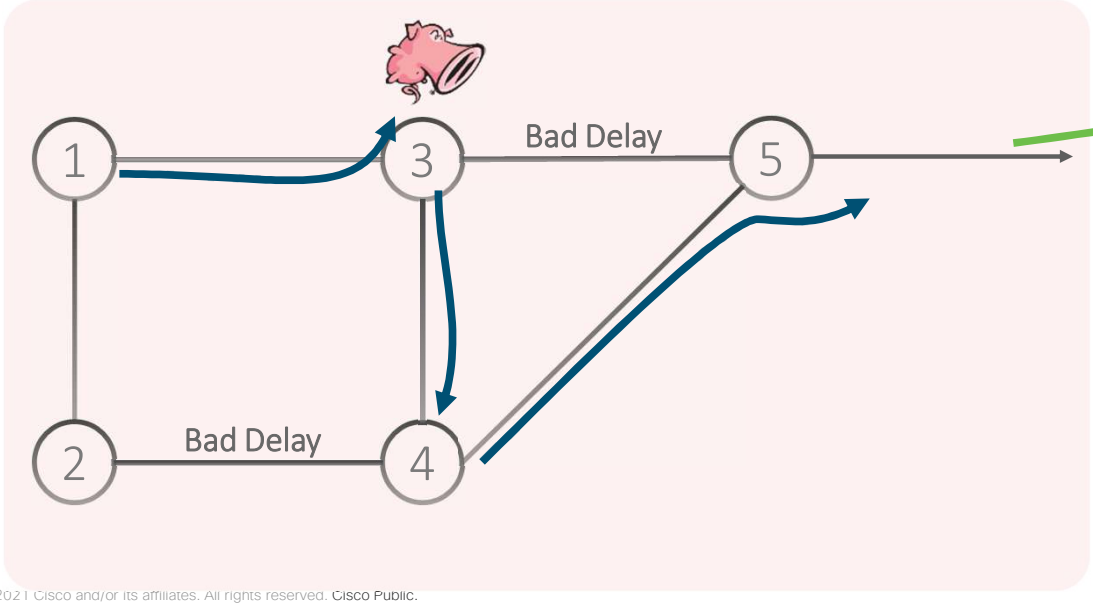
- ePE2 matches local SID FC00:0008:0002:F009/64
- ePE2 Applies “VPN Decaps” Behavior into VRF9

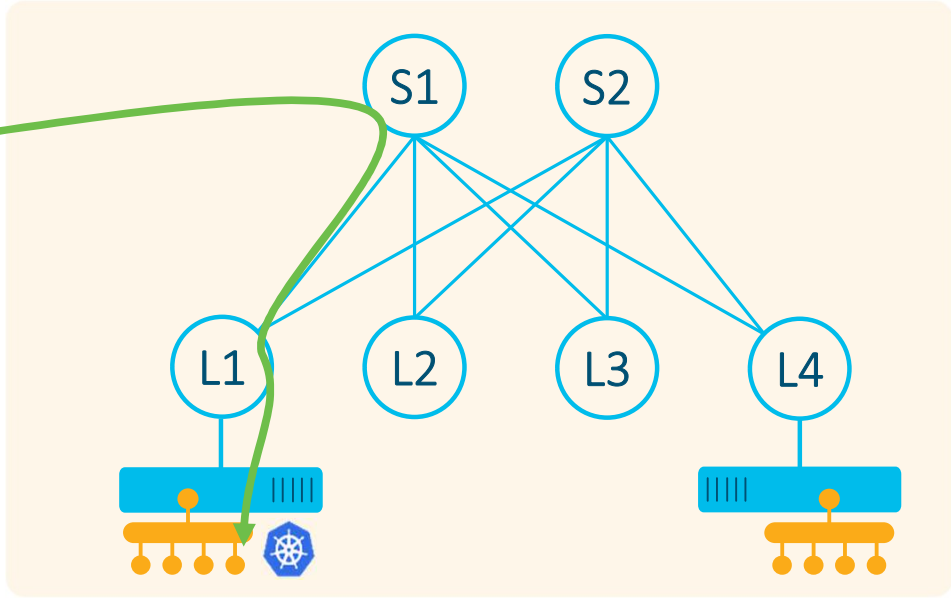
Simple application of SRv6 Network Programming: RFC8986

SRv6 uSID in WAN+MSDC

- Hyper-scale by combining uSIDs with 128b SIDs
  - Container-scale ready

Network Program: < FC00:0:0300:0400:0500 , 2001:db8:1234::K8s >





SRv6 uSID
5G Use-Cases

5G Use-Case

- Combination of all the previous use-cases
  - TILFA
  - uLoop
  - EPE
  - Slicing and VPN
  - NFV

Conclusion

Simplicity Always Prevails

~~LDP~~

~~RSVP-TE~~

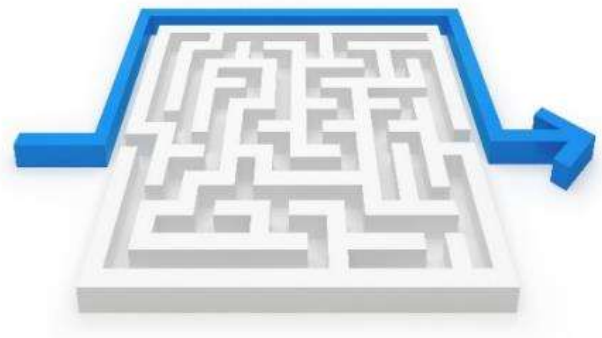
~~Inter-AS Option~~

~~A/B/C~~

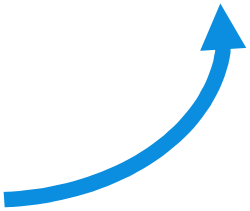
~~MPLS~~

~~UDP/VxLAN~~

~~NSH~~



Furthermore with more scale and functionality



Deployment and Eco-system at Record Speed

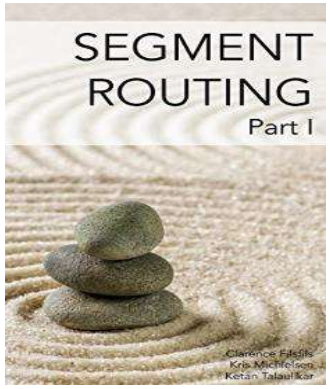
- ~100M Commercial Subscribers supported by SRv6 services
- ~100 SRv6 Deployments
- Mature SRv6 Industry
- SRv6 is Proposed Standard: RFC8402, RFC8754, RFC8986
- Mature SRv6 Open Source
- Best MTU Overhead
- Seamless Deployment
- Any program can be expressed

Much more to come

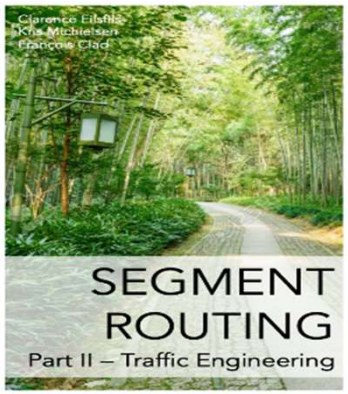
- It is only the beginning ☺

candy18 © 123RF.com

Stay up-to-date



amzn.com/B01I58LSUO



amazon.com/dp/B07N13RDM9

SRv6 Part III Coming by end CY21

twitter.com/SegmentRouting

segment-routing.net

facebook.com/SegmentRouting/

linkedin.com/groups/8266623

ask-segment-routing@cisco.com


CISCO

Appndix

uSID is a Spring WG Document

- *"the rough consensus of the working group is that we should adopt this document"*. Spring Chair
  - 83% of the WG voted for the adoption
  - 81% asserted that CSID is a single SRv6 based Data Plane
- *"we do not object to C-SID behavior work continuing in SPRING, we simply need a … separate 6MAN document to clarify and categorize SRv6 SIDs"*. 6MAN Area Director
  - Suresh Krishnan has volunteered editing this draft