SRv6 uSID

Ahmed Abdelsalam, Mounir Mohamed

Cisco Systems

MENOG 23
Middle East Network Operators Group

Riyadh, Saudi Arabia

segment-routing.net

- Discover the latest news on Segment Routing (link)
- Highlights of the SRv6 Conference in Paris, April 2023 (link)
- Highlights of the SRv6 workshop in Tokyo, Japan (link)
- Cisco Knowledge Network "SRv6 Standardization Deployed at Scale" (link)

SRv6 uSID Reality check

SRv6 is Proposed Standard

Architecture	• SR Architecture – RFC 8402 • SRTE Policy Architecture – RFC 9256
Data Plane	• SRv6 Network Programming – RFC 8986 • IPv6 SR header – RFC 8754
Control Plane	• SRv6 BGP Services – RFC 9252 • SRv6 ISIS – RFC 9352 • SR Flex-Algo – RFC 9350
Operation & Management	• SRv6 OAM – RFC 9259 • Performance Management – RFC 5357

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, P4, OPEN Compute Project®, Kubernetes, SONiC, FRRouting

Smart NIC

Intel, Mellanox Technologies

Open-Source Applications

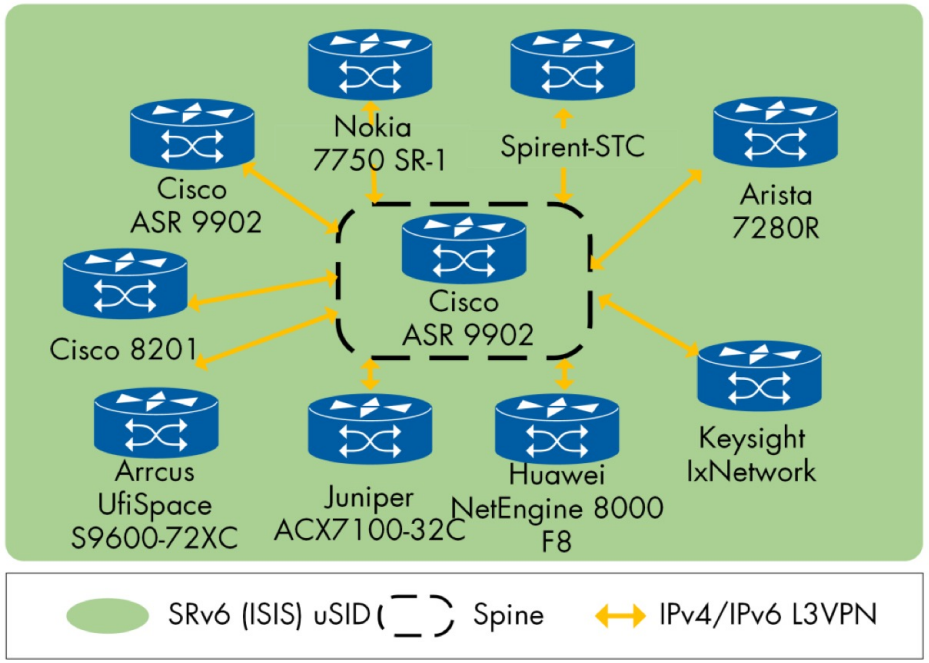
Pyroute2, IPTABLES SERA, GoBGP, exa BGP, eBPF, cilium, TCPDUMP, WIRESHARK, SNORT, GoBMP

Partners

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

EANTC 2023 testing

- Successful multi-vendor interoperability test for SRv6 uSID over 11 implementations.
- BGP-Based overlay services over SRv6 **uSID** (RFC9252 as extension for RFC4363, and RFC7432, including L3VPN, EVPN VPWS, LAN, RT5) full VPN overlay services over SRv6 and uSID
- Full BGP-based overlay services over uSID with TI-FLA, UPA, and SR-TE.
- Multiple Silicon families from BRCM, Cisco, Huawei, Juniper, Nokia



EANTC

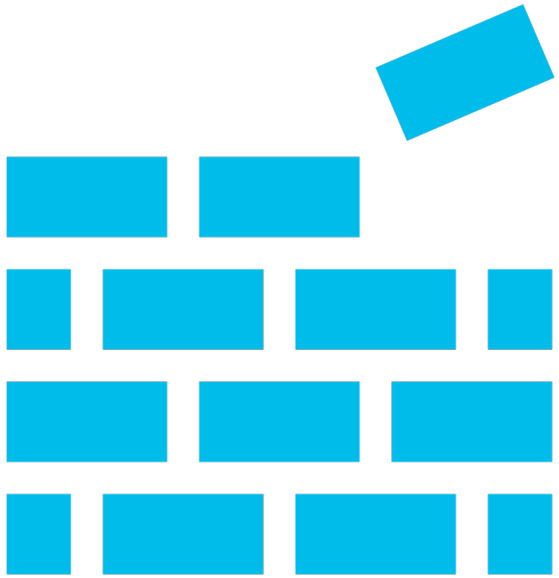
Multi-Vendor MPLS SDN
Interoperability Test Report
2023

https://eantc.de/fileadmin/eantc/downloads/events/2023/EANTC-InteropTest2023-TestReport.pdf

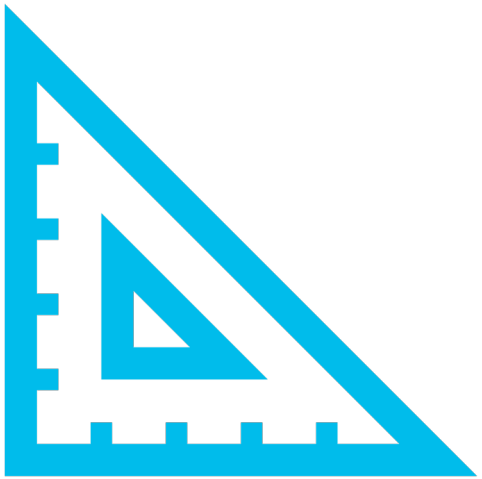
Simplicity Always Prevails

Benefits to operators

Integrated Solution



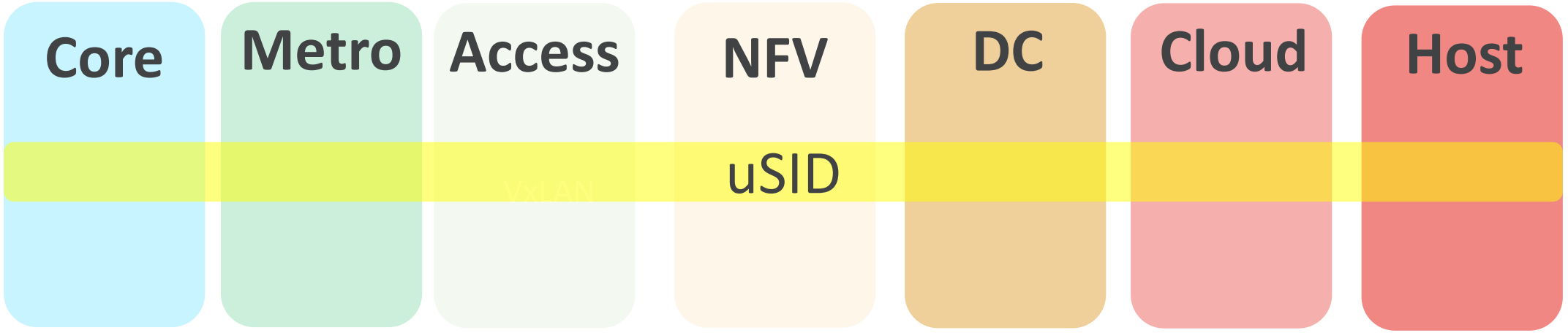
Creation



Measure

Analyze

Unified Solution across domains



Dan Voyer
Bell Canada
Paris 2022

Dan Bernier
Bell & NoviFlow
Paris 2022

Gyan Mishra
Verizon
Paris 2023

Dan Bernier
Bell Canada
Paris 2023

Unified Core Metro Access DC Cloud IP solution

Outperforms per-domain custom shim (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)

https://www.segment-routing.net/conferences/MPLS-WC-2022-Daniel-Voyer/

Bell SRv6 uSID Deployment
Paris 2022

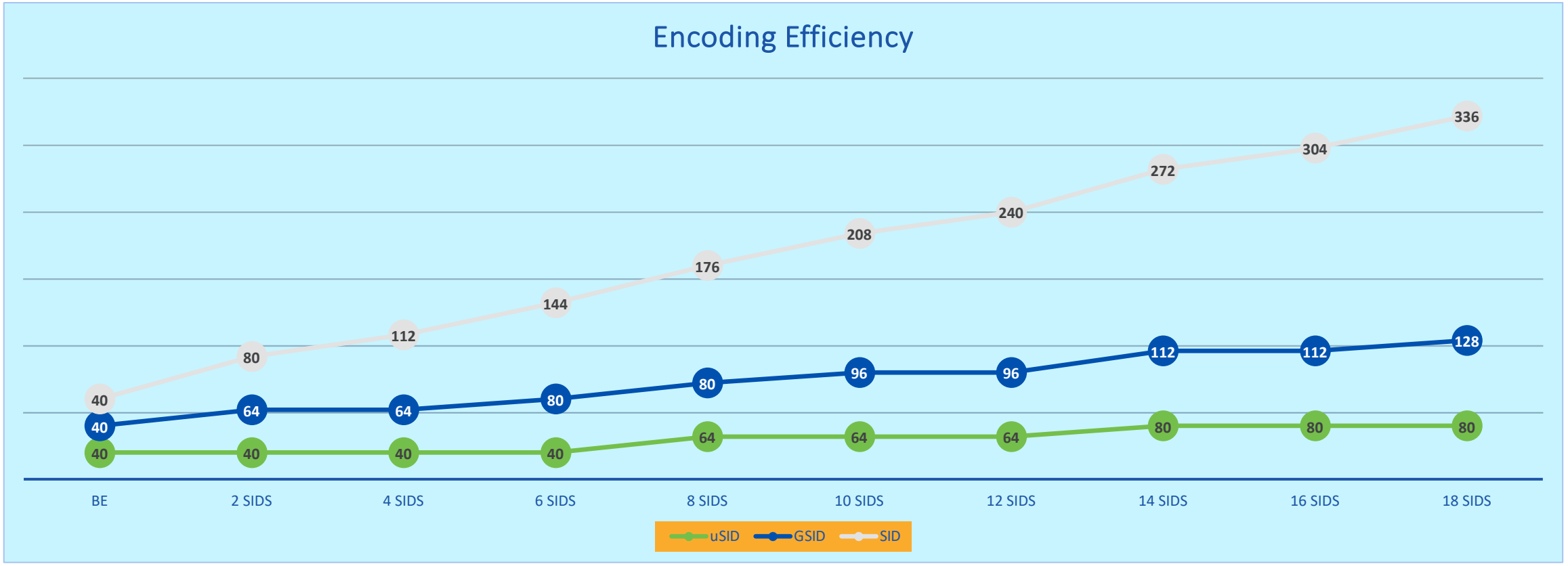
Outperforms VxLAN – Gyan Mishra (Verizon)

- 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)

https://www.segment-routing.net/conferences/Paris23-Verizon-Gyan-Mishra/

SRv6 uSID DC Use-Case
Paris 2023

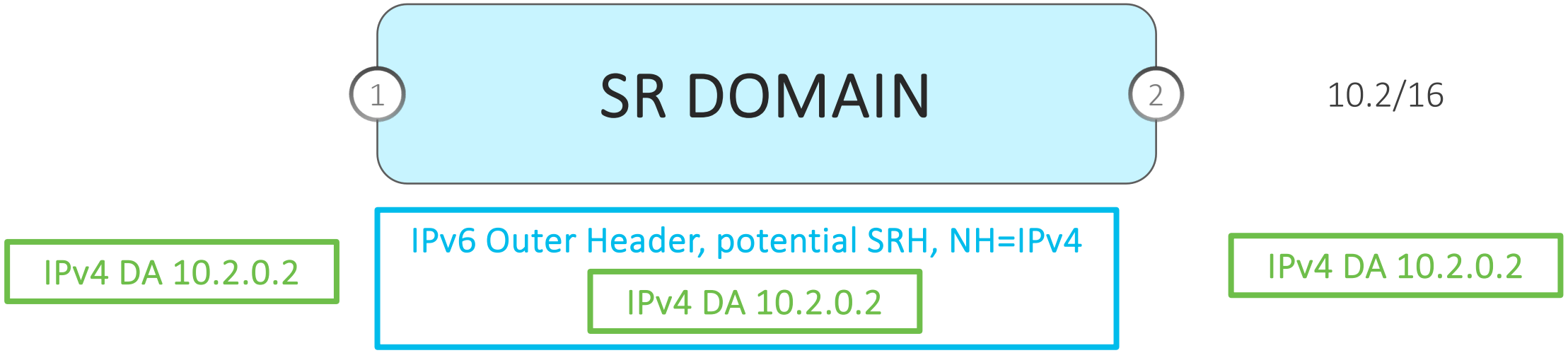
SRv6 uSID Encoding Efficiency



- Better Encoding Efficiency = Lower MTU overhead = Fiber Cost Discount

SRv6 uSID - Reminder

Transparent Service



- The SRv6 uSID program is placed in the outer IPv6 header
  - DA holds up to 6 uSID instructions
  - SRH is Rare: Most use-cases need less than 6 instructions
- The inner packet is untouched
  - Customer packet is encapsulated from ingress to egress of the SR Domain

SRv6 uSID Program

- The End-to-End Policy is encoded as a List of uSID instructions
  - The first 6 uSID instructions in the outer DA
  - The remaining uSID instructions are in the SRH (rarely needed)
- An uSID instruction 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 program in HW
  - Any SW custom behavior: Container orchestrated by Kubernetes

Powerful Service Creation

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

Container of 6 uSID’s

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

Min-Cost Block

- The /32 block can come from any space
  - Global “registered” IPv6
  - “Local” FC/7
- We recommend FC/7:
  - unroutable outside the domain
- 2 Blocks are typically deployed. One per Flex-Algo: Cost vs Latency
  - FC00:0000/32 for Best-Effort ISIS Flex-Algo
  - FC00:0008/32 for Low-Latency ISIS Flex-Algo

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

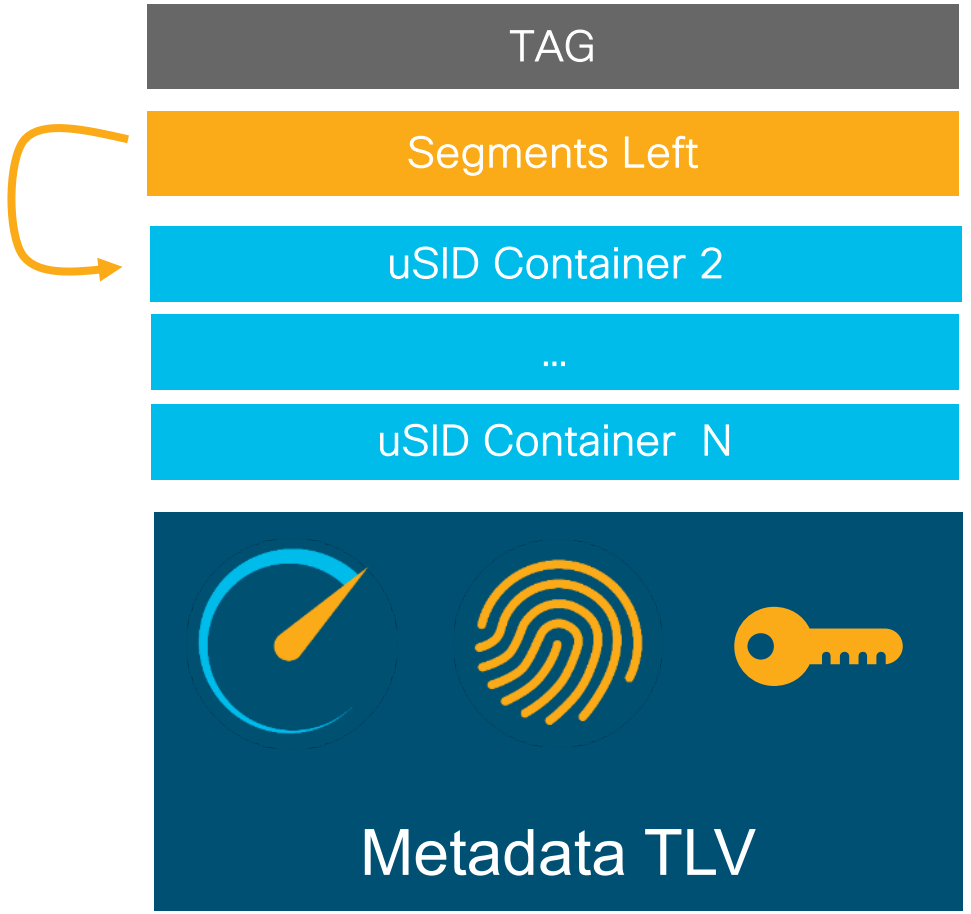
Outer SRH: FC00:0000:0007:0008:0009:0010:0011:0012

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

SRH

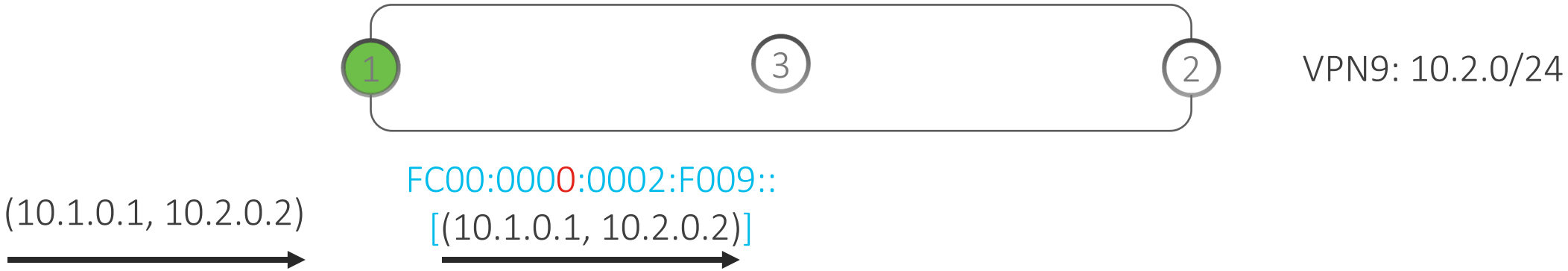
- SRv6 is a native extension of IPv6
  - RFC 8754
  - As foreseen 25 years ago by RFC2460
- SRH contains an ordered list of uSID Containers
  - DA : uSID Container 1
  - SRH: remaining uSID Containers (2…N)



SRv6 uSID
VPN & Slice Use-Cases

VPN over Min-Cost 5G Slice - Ingress PE

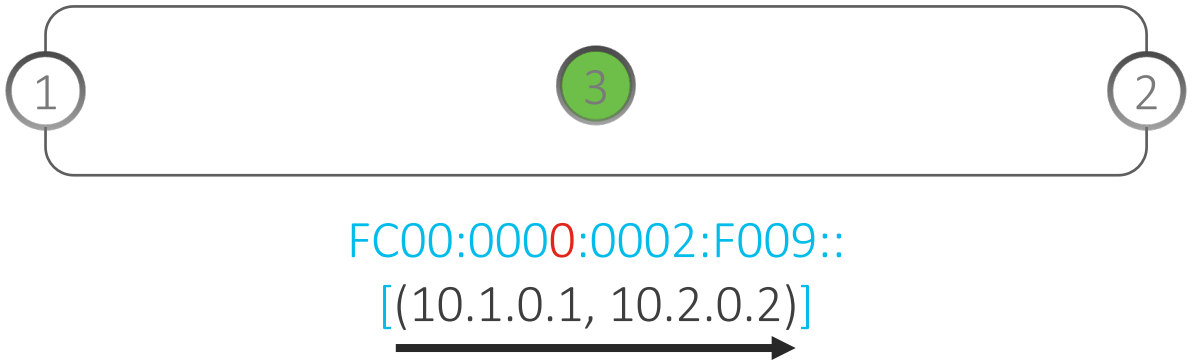
Deployed use-case



- 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

Deployed use-case

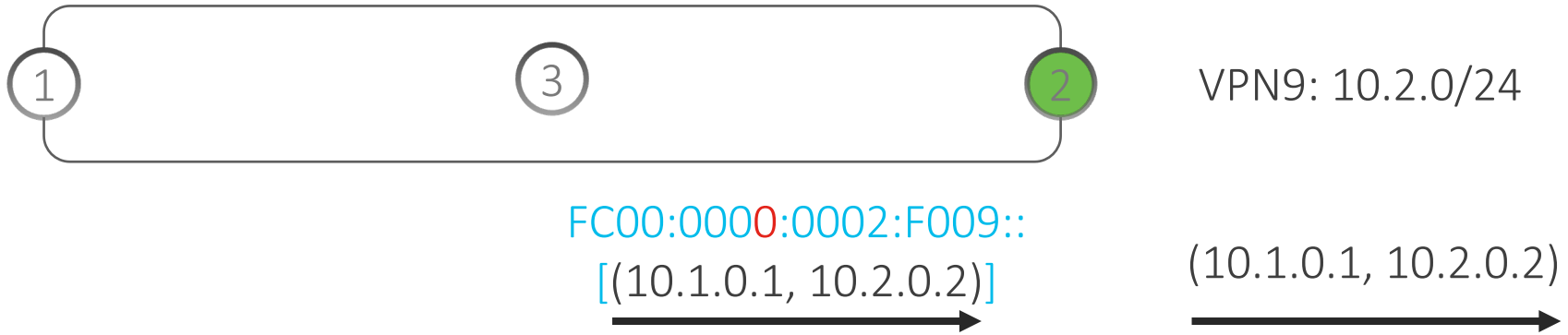


- Transit Node 3 forwards along remote prefix FC00:0000:0002/48
  - ISIS Shortest-Path with Algo 0 (Min Cost)

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

VPN over Min-Cost 5G Slice – Egress PE

Deployed use-case

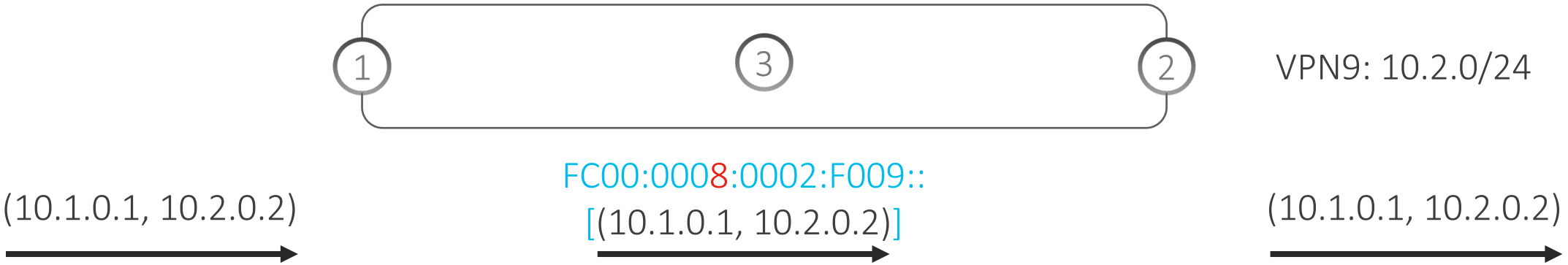


- ePE2 matches local SID FC00:0000:0002:F009/64
- ePE2 Applies “VPN Decaps” Behavior into VRF9

Simple application of SRv6 Network Programming: RFC8986

VPN over Min-Delay 5G Slice

Deployed use-case

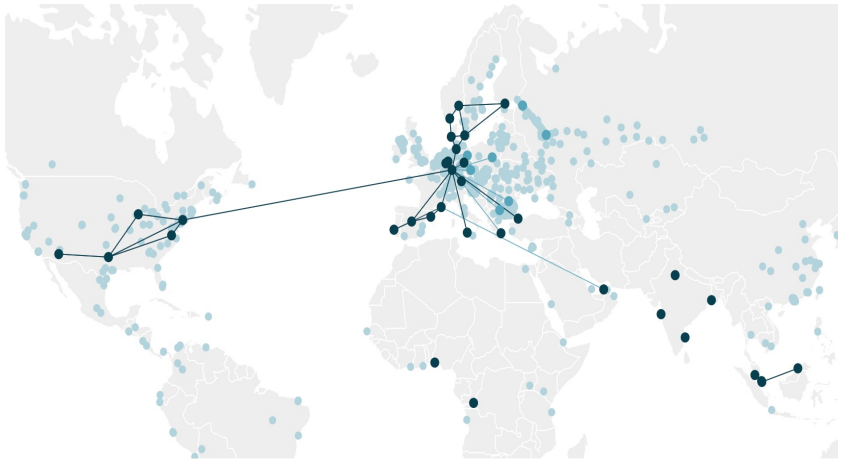


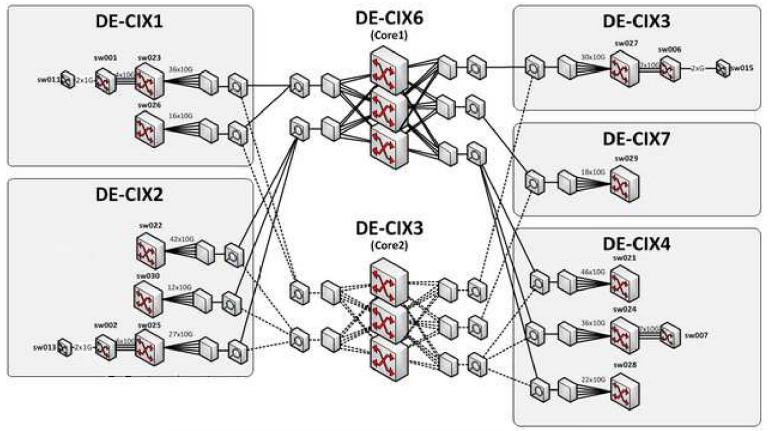
- iPE1 learns via BGP that 10.2.0/24 in VPN9 is reachable via SID FC00:0008:0002:F009
- @1: encapsulates with outer DA = FC00:0008:0002:F009
- @3: forwards based on remote prefix FC00:0008:0002/48 along Min-Delay shortest path
- @2: matches local SID FC00:0008:0002:F009/64 and applies “VPN Decaps” Behavior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

Peering and Exchange use-cases

SRv6 for Exchange use-case

- IXPs expands their global coverages to offer remote peering services with additional VLANs via EVPN overlayed over MPLS with RSVP-TE.
  - IXPs moved from VPLS to EVPN due to VPLS complexity
  - Not all new locations interconnected to IXP BB
  - Different IX sites using different mechanisms (VXLAN/MPLS)
- SRv6 enables IXPs:
  - Manage all IXP/DC sites using a single transport without interworking function (VXLAN to MPLS)
  - Native IP based ECMP (No labels/Entropy)
  - Latency based traffic-steering with simplified stack = simplified core platforms
  - Provides global remote peering without dedicated links (over internet)
  - Operation and troubleshooting simplicity
  - Native measurement capabilities (SRv6 PM , SRv6 Path Tracing)
  - Much More
    - > Automate 50ms convergence
    - > Allows to deliver constraint path (Shortest path/Low latency/..)

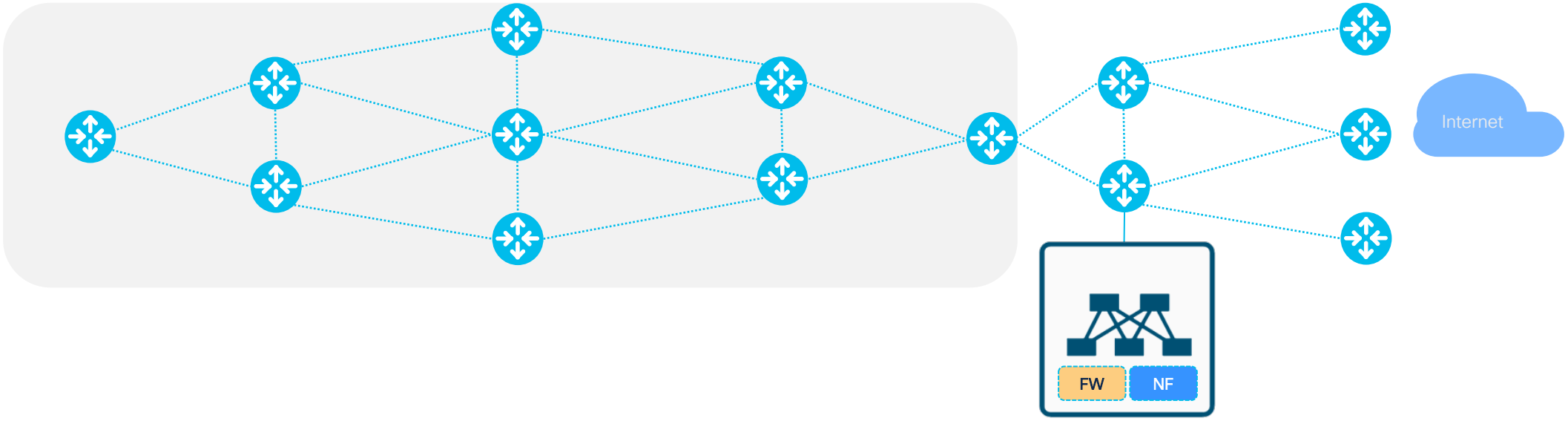




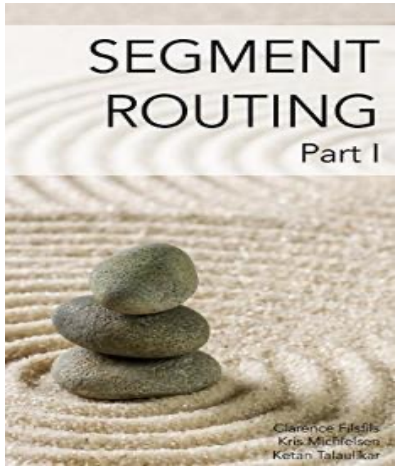
https://www.de-cix.net/en/services/globepeer-remot

SRv6 for Peering use-cases

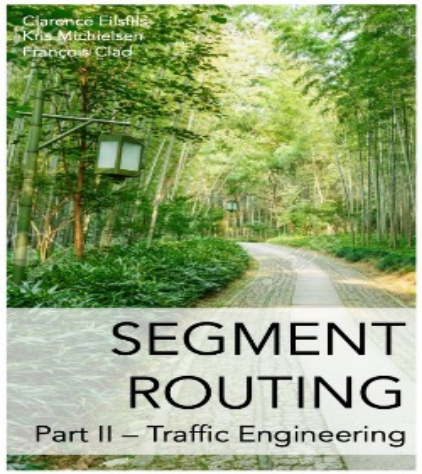
- Secured Business-internet service
  - In this use-case the service requires two instructions: “go to NF via Low-Delay Slice and apply FW” and “go to Egress PE via Low-Delay Slice and forward in INT-FW-SV.
  - The Ingress IBR maps the received customer packet to the policy and encapsulates this packet in an outer IPv6 header with two instructions. The first instruction (D:0:0:2:NF::) is encoded in the DA of the outer header. The second instruction (D:0:0:3:V9::) is encoded in the SRH that has been added in the outer packet header.



Stay up-to-date



amzn.com/B01I58LSUO



amazon.com/dp/B07N13RDM9

SRv6 Part III Coming soon

twitter.com/SegmentRouting

segment-routing.net

facebook.com/SegmentRouting/

linkedin.com/groups/8266623

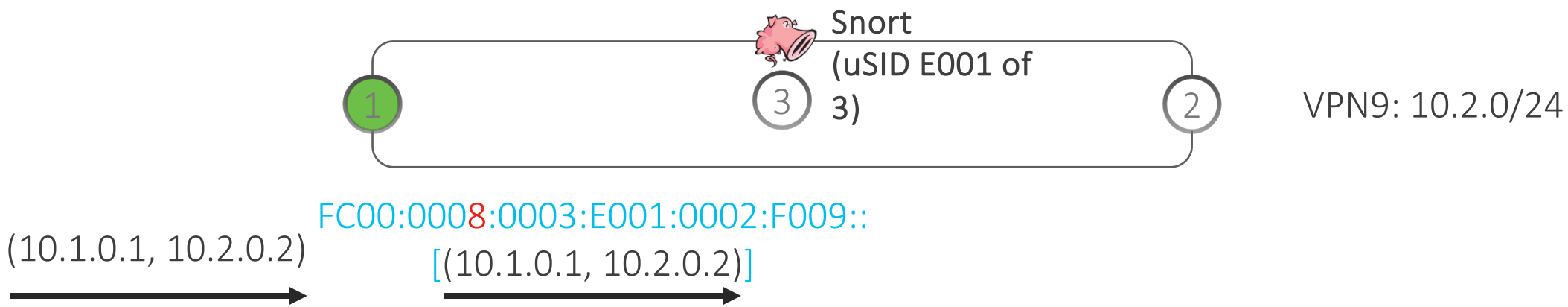
ask-segment-routing@cisco.com

The bridge to possible

SRv6 uSID
NFV Use-Cases

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

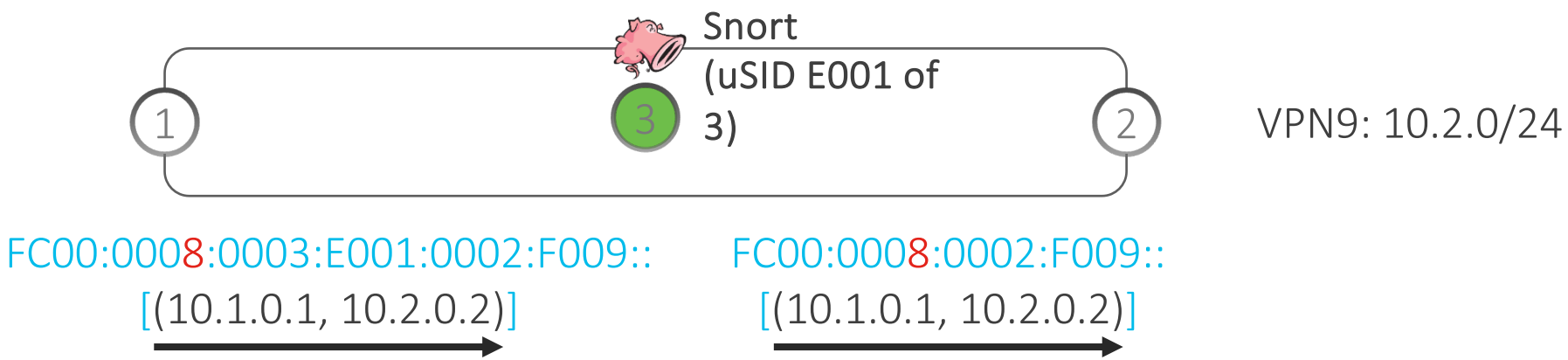
Deployed use-case



- iPE1 learns via BGP that 10.2.0/24 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

Deployed use-case

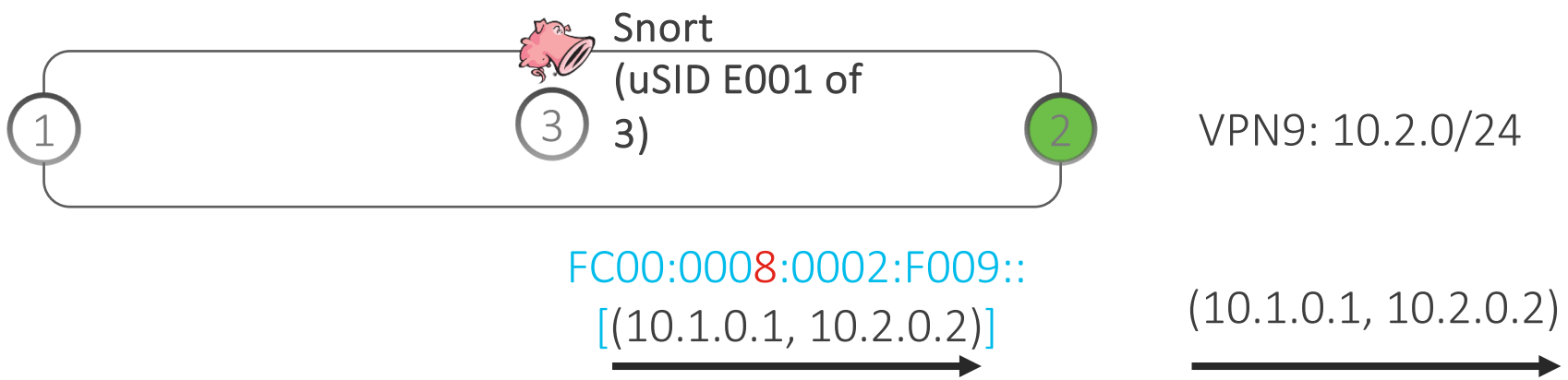


- Node 3 matches local SID FC00:0008:0003:E001::/64
- Node 3 applies “Snort policy” to the packet
- Node 3 activates the next uSID and forwards over ISIS Shortest-Path with Algo 128 (Min delay)

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

Deployed use-case



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

Simple application of SRv6 Network Programming: RFC8986