



SRv6

Clarence Filsfils - Kris Michielsen - Pablo Camarillo - François Clad

Disclaimer

"Many of the products and features described herein remain in varying stages of development and will be offered on a when-and-if-available basis. This roadmap is subject to change at the sole discretion of Cisco, and Cisco will have no liability for delay in the delivery or failure to deliver any of the products or features set forth in this document."

Acknowledgements

- Ahmed Abdelsalam
- Clarence Filsfils
- François Clad
- Ianik Semco
- Jose Liste
- Kris Michielsen
- Pablo Camarillo

Industry at large backs up SR



Strong customer adoption WEB, SP, DC, Metro, Enterprise



De-facto SDN Architecture



Standardization **IFTF**



Multi-vendor Consensus



Open Source Linux, VPP

































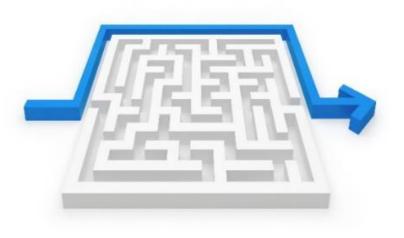








Simplicity Always Prevails



Furthermore with more scaleand functionality





SRv6 Eco-System



At record speed

- 9+ large-scale commercial deployments
 - Softbank, Iliad, China Telecom, LINE corporation,
 China Unicom, CERNET2, China Bank, MTN Uganda, NOIA Network, ...
- 25+ HW linerate implementations
 - Cisco Systems, Huawei, Juniper
 - Arrcus, Broadcom, Barefoot, Intel, Marvell, Mellanox, Kaloom
 - Spirent, Ixia
 - Multiple Interop Reports
- 11+ open-source platforms/ Applications
 - Linux, FD.io VPP, P4, iptables, nftables, snort, SERA, ExaBGP, GoBGP, GoBMP, Contiv-VPP, ...

Cisco Supports SoftBank on First Segment Routing IPv6 Deployment in Prep for 5G

<u>Link to PR - https://newsroom.cisco.com/press-release-content?type=webcontent&articleId=1969030</u>





Re: [spring] SPRING SRv6 Deployment Status draft

Sébastien Parisot <sparisot@free-mobile.fr> | Tue, 10 December 2019 09:34 UTC | Show header

Hi Satoru, Zafar,

I would like to provide an update to SRv6 deployment in Iliad's nationwide network in Italy.

As of the end of 2019, the SRv6 network consists of:

- 1000 Cisco NCS 5500 routers
- 1800 Iliad's Nodeboxes
- The network services 4.5 million mobile subscribers (as of Q3 2019)
- The network is carrying 300 Gbps of commercial traffic at peak hours
- It is expected to grow to more than 4000 Nodeboxes in 2020.

The following SRv6 features have been deployed:

- A Segment Routing Header based data plane
- End (PSP), End.X (PSP), End.DT4, T.Encaps.Red, T.Insert.Red functions
- BGP VPN SRv6 extensions
- ISIS SRv6 extensions
- SRH-based Topology Independent (TI-LFA) Fast Reroute mechanisms
- Support for ping and traceroute

Can you please update the SRv6 deployment draft accordingly?

Thanks, Sébastien

iliad disco

- Nationwide deployment in Italy
- 1000 Cisco NCS 5500
- 1800 Iliad Nodeboxes



SRv6 Network Programming Introduction

Segment Routing

- Source Routing
 - the topological and service (NFV) path is encoded in packet header
- Scalability
 - the network fabric does not hold any per-flow state for TE or NFV
- Simplicity
 - automation: TILFA sub-50msec FRR
 - protocol elimination: LDP, RSVP-TE, VxLAN, NSH, GTP, ...
- End-to-End
 - DC, Metro, WAN

Two dataplane instantiations



MPLS

leverage the mature MPLS HW with only SW upgrade



- 1 segment = 1 label
- a segment list = a label stack

Segment Routing



IPv6

- leverages RFC8200 provision for source routing extension header
- 1 segment = 1 address
 - a segment list = an address list in the SRH

IPv6 provides reachability



IPv4 limitations & work-arounds

- Limited address space
- × No engineered Load Balancing
- × No VPN
- × No Traffic Engineering
- × No Service Chaining

Data (L5,L6 & L7)

Socket header (L4)

IPv4 header (L3)

Ethernet (L2)

IPv4 limitations & work-arounds

- × Limited address space
- × No engineered Load Balancing
- × No VPN
- × No Traffic Engineering
- × No Service Chaining

Data (L5,L6 & L7)

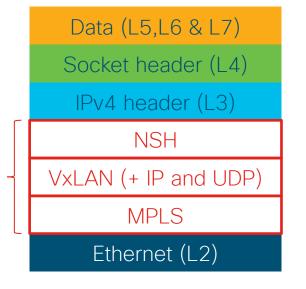
Socket header (L4)

IPv4 header (L3)

Ethernet (L2)

work-arounds

- $\rightarrow NAT$
- → MPLS Entropy Label, VxLAN UDP
- → MPLS VPN's, VxLAN
- → RSVP-TE, SR-TE MPLS
- → NSH



IPv4 limitations & work-arounds

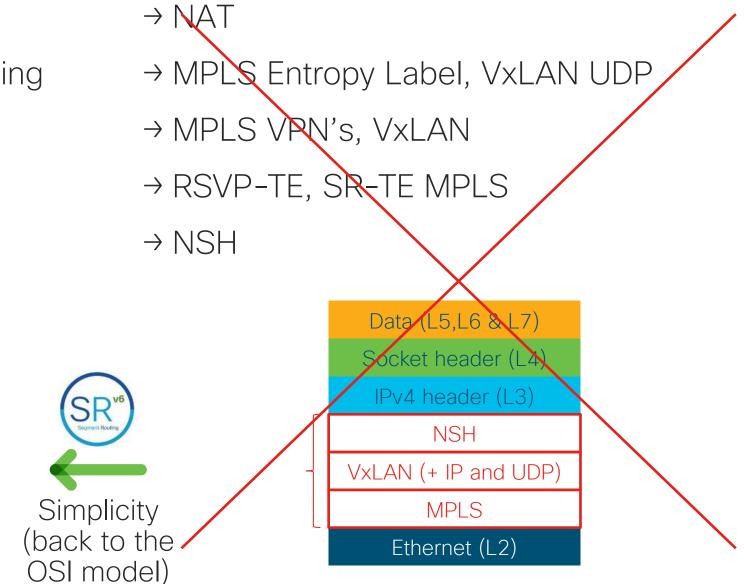
- × Limited address space
- × No engineered Load Balancing
- × No VPN
- × No Traffic Engineering
- × No Service Chaining

Data (L5,L6 & L7)

Socket header (L4)

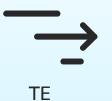
IPv4 header (L3)

Ethernet (L2)



SRv6 unleashes IPv6 potential

















VPN

NFV

Scalability A

Automation

Single protocol

SR for anything: Network as a Computer

Network instruction

Locator

Function

- 128-bit SRv6 SID
 - Locator: routed to the node performing the function
 - Function: any possible function either local to NPU or app in VM/Container
 - Flexible bit-length selection

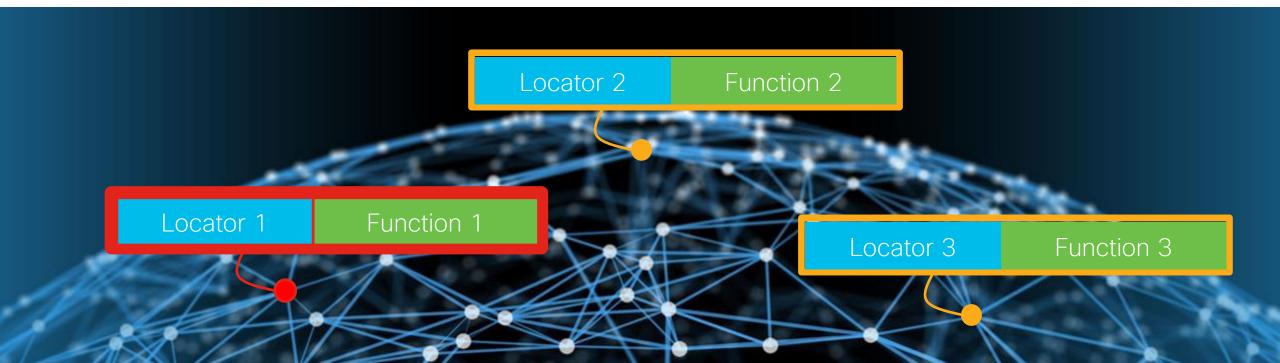
Network Program

Next Segment

Locator 1 Function 1

Locator 2 Function 2

Locator 3



Function 3

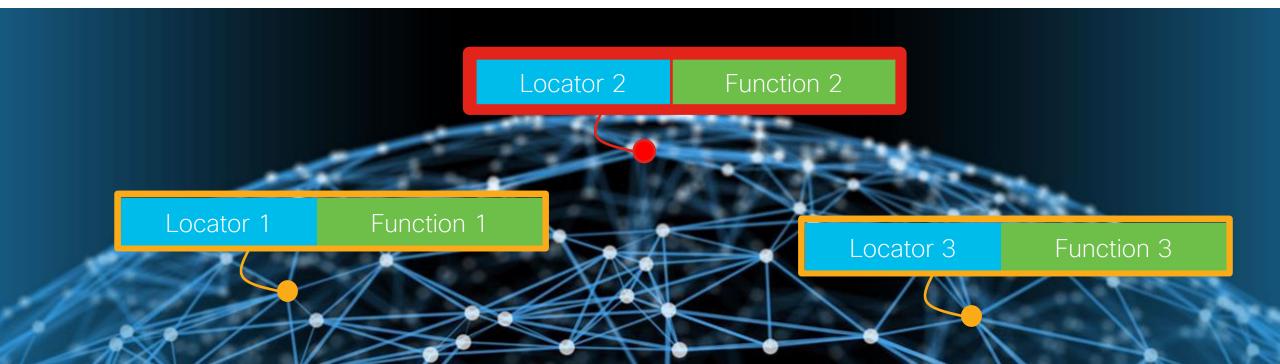
Network Program

Next Segment -

Locator 1 Function 1

Locator 2 Function 2

Locator 3 Function 3



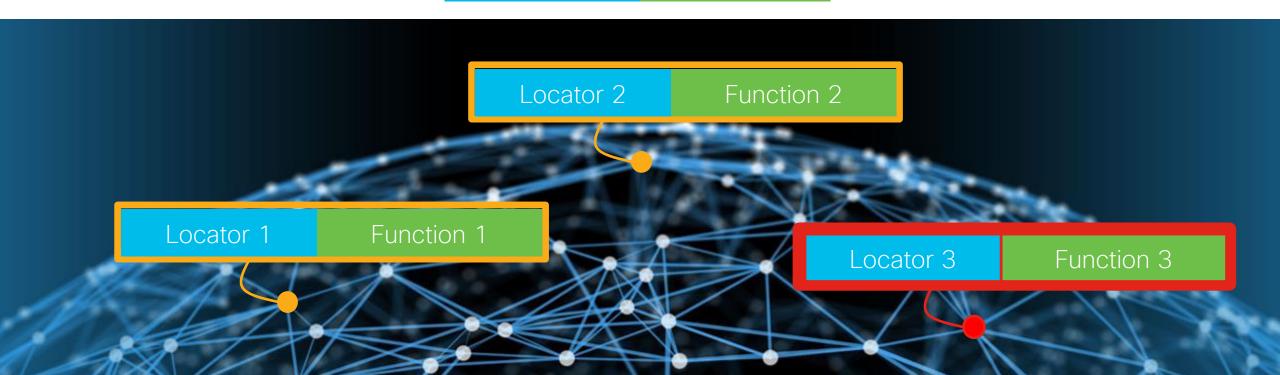
Network Program

Next Segment

Locator 1 Function 1

Locator 2 Function 2

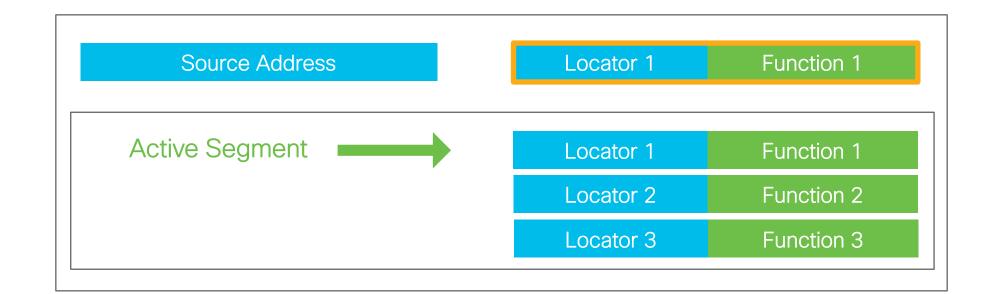
Locator 3 Function 3



Network Program in the Packet Header

IPv6 header

Segment Routing Header



IPv6 payload

TCP, UDP, QUIC

Network Program in the Packet Header

IPv6 header

Segment Routing Header Source Address

Locator 2

Function 2

Locator 1

Function 1

Locator 2

Function 2

Locator 3

Function 3

IPv6 payload

TCP, UDP, QUIC

Network Program in the Packet Header

IPv6 header

Segment Routing Header Source Address

Locator 3 Function 3

Locator 1 Function 1

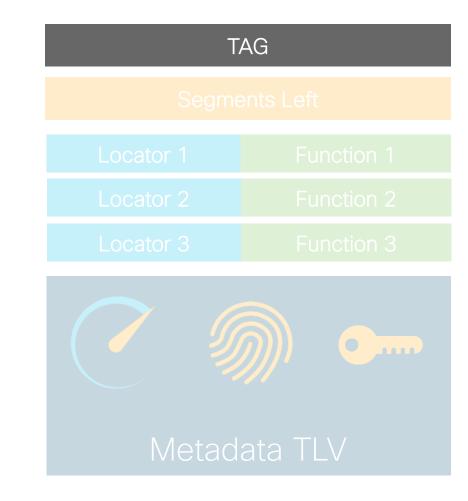
Locator 2 Function 2

Active Segment Locator 3 Function 3

IPv6 payload

TCP, UDP, QUIC

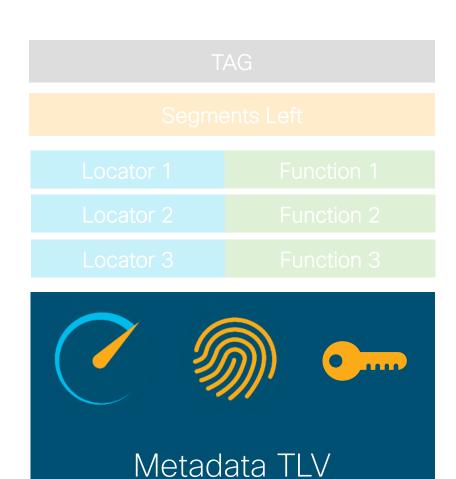
Group-Based Policy



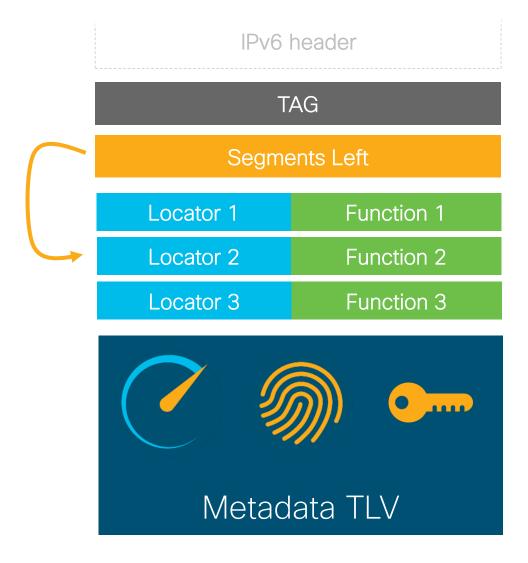
Argument shared between functions

"Global" Argument



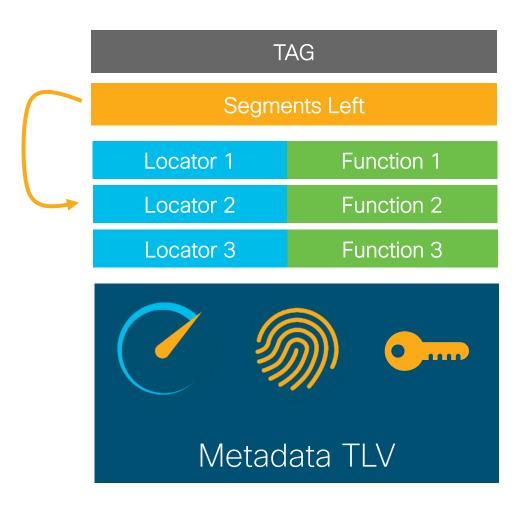


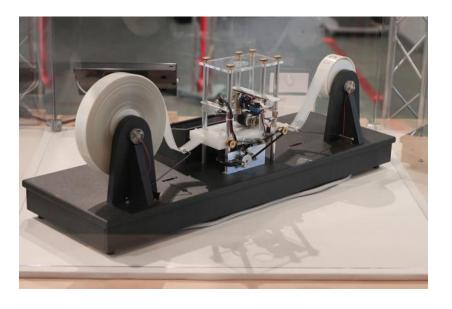
SRv6 Header



```
Hdr Ext Len | Routing Type | Segments Left
 Last Entry
                Flags
         Segment List[0] (128 bits IPv6 address)
                       . . .
         Segment List[n] (128 bits IPv6 address)
        Optional Type Length Value objects (variable)
```

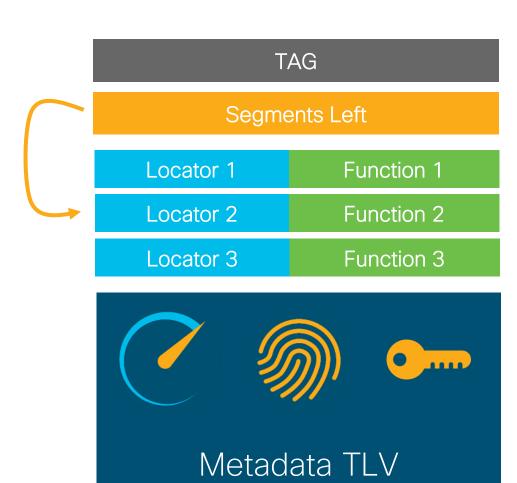
SRv6 for anything





Turing

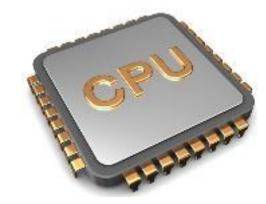
SRv6 for anything





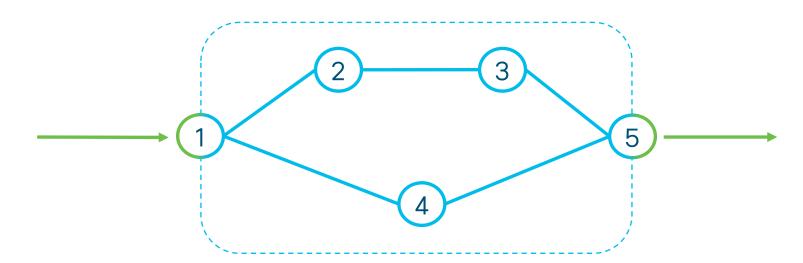
Optimized for HW processing e.g. Underlay & Tenant use-cases

Optimized for SW processing e.g. NFV, Container, Micro-Service



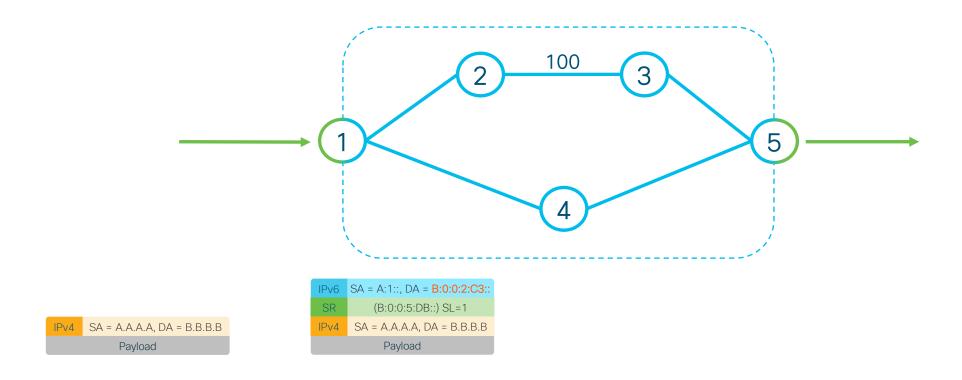
SRv6 Domain

IPv6 enabled provider infrastructure SR Domain



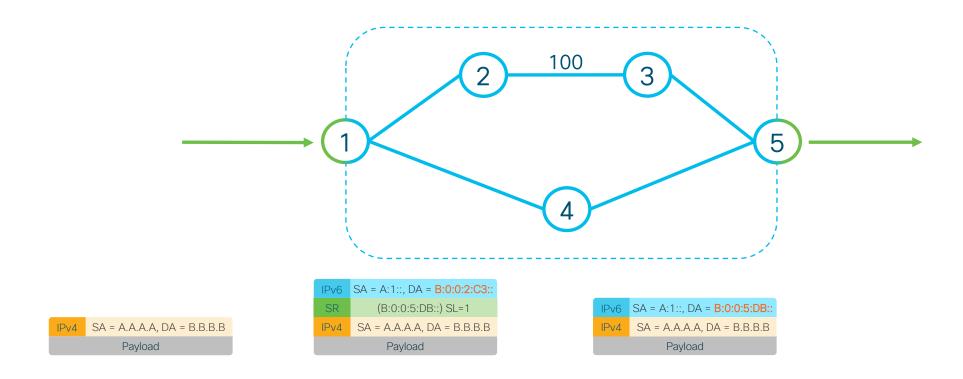
Encapsulation at the Domain ingress

- IPv4, IPv6 or L2 frame is encapsulated within the SR Domain
- Outer IPv6 header includes an SRH with the list of segments



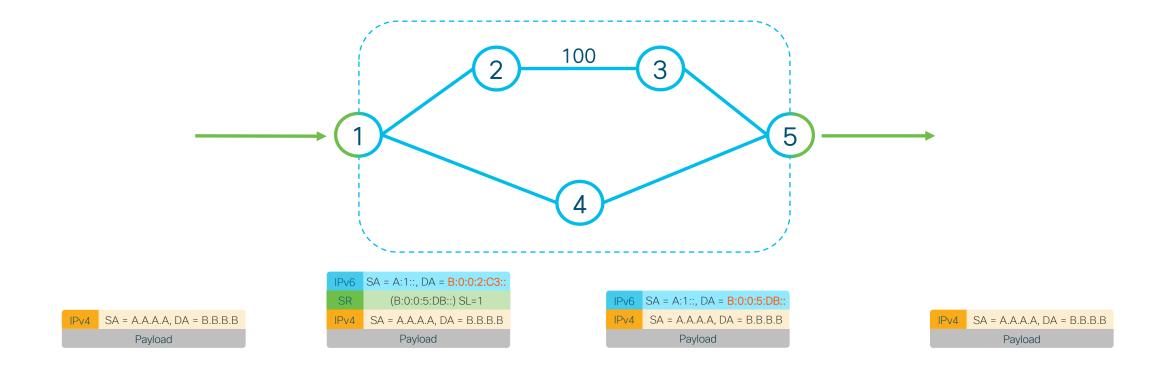
SRH of the outer IPv6 encapsulation

- Domain acts as a giant computer
- The network program in the outer SRH is executed



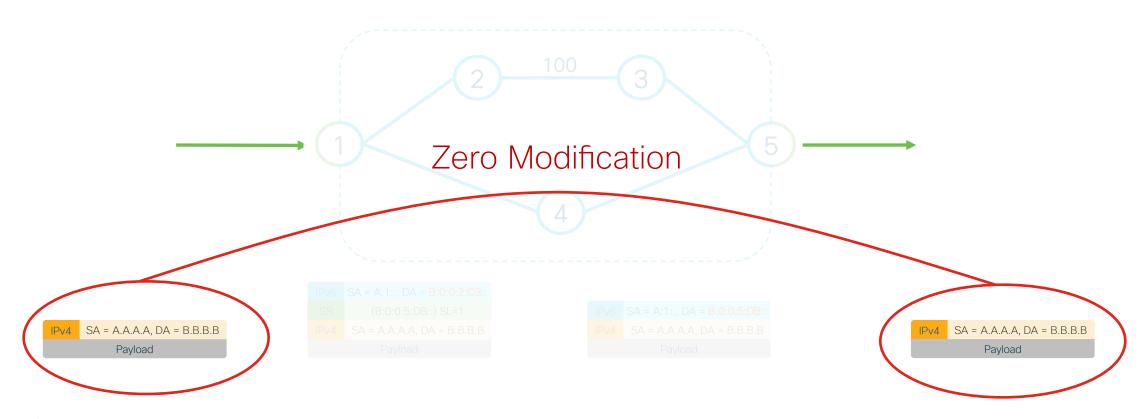
Decapsulation at Domain Egress

• Egress PE removes the outer IPv6 header as the packet leaves the SR domain



End-to-End Integrity

- End-to-end integrity principle is strictly guaranteed
 - Inner packet is unmodified
 - Same as SR-MPLS (MPLS stack is replaced by IPv6 outer header and SRH)



End and End.X SID behaviors

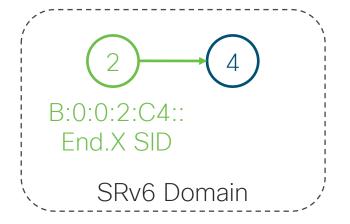
- End Default endpoint behavior
 - shortest-path to the SID's endpoint
 - endpoint updates DA with next SID
 - endpoint forwards according to updated DA

- End.X Endpoint with cross-connect
 - shortest-path to SID's endpoint
 - endpoint updates DA with next SID
 - endpoint forwards to interface associated with SID

Illustration convention:

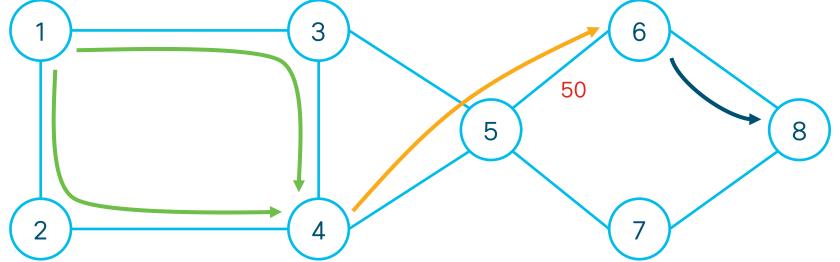
- IPv6 address of node k is A:<k>::
- SRv6 SID of node k is B:0:0:<k>:<function>::





Endpoint behaviors illustration

SR: 〈 B:0:0:4:1::, B:0:0:5:C6::, A:8:: 〉



Default metric 10

- B:0:0:4:1:: shortest path to node 4
- B:0:0:5:C6:: shortest path to node 5, then cross-connect towards 6
- A:8:: regular IPv6 address of node 8

IETF

SR Architecture

- RFC 8402 Proposed Standard
 - Defines SR-MPLS with MPLS dataplane and Label SID's
 - Defines SRv6 with SRH and SRv6 SID's

SR_V6

- RFC 8754 Proposed Standard
 - SRv6 DataPlane: SRH and SRv6 SID
- RFC 8986 Proposed Standard
 - Network Programming (END, END.X, END.DX/DT, H.Encaps)

- In the last IETF stages towards RFC Proposed Standard
 - Control Plane (ISIS, BGP-LS)
 - Policy
 - OAM
 - BGP

SR is IETF Proposed Standard

Architecture

- Segment Routing Architecture RFC 8402
- Source Packet Routing in Networking (SPRING) Problem Statement and Requirements RFC 7855
- Segment Routing with MPLS data plans DEC 9660
- IPv6 Segment Routing He
- Segment Routing over IPv

Protocol Exte

ISIS

- IS-IS Extensions for Segn
- Signaling MSD (Maximum
- Advertising L2 Bundle Me
- IS-IS Traffic Engineering (

BGP

- Segment Routing Prefix S
- BGP-LS Advertisement o Performance Metric Exter

Use-cases

- SR-MPLS over IP RFC 8663
- Resiliency Use Cases in SPRING Networks RFC 8355
- Use Cases for IPv6 Source Packet Routing in Networking (SPRING) RFC 8354

RFC 8670

t Routing RFC 8604

ing LSP co-existance RFC 8426

Cisco Leads Standards Bodies

Editor of

Editor of

96% IETF RFCs

Co-author of 100% IETF RFCs

77% IETF WG Drafts

Co-author of 84% IETF WG Drafts

RFC 8476 C 7471

OAM

- A Scalable and Topology-Aware MPLS Dataplane Monitoring System RFC 8403
- Label Switched Path (LSP) Ping/Trace for Segment Routing Networks Using MPLS Dataplane RFC 8287

Performance Measurement

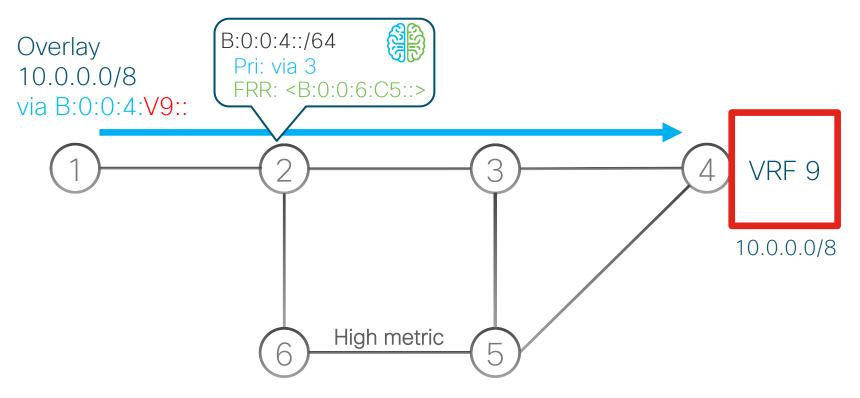
- Packet Loss and Delay Measurement for MPLS Networks RFC 6374
- UDP Return Path for Packet Loss and Delay Measurement for MPLS Networks RFC 7876

Up-to-date list on https://segment-routing.net/ietf

© 2021 Cisco and/or its affiliates. All rights reserved.

SRv6 Deployed Use-Cases

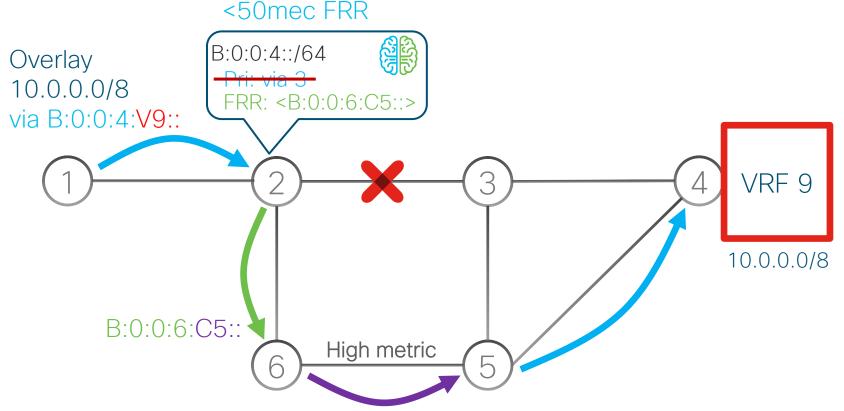
TI-LFA



- 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



- 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

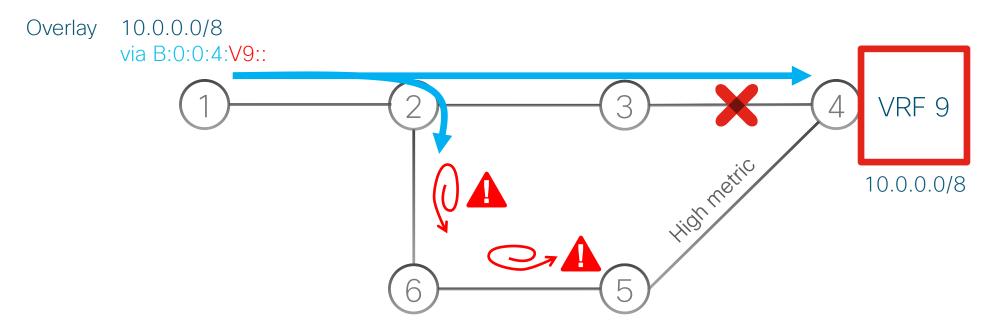
TILFA - Scaling

Do we need many SID's? No!



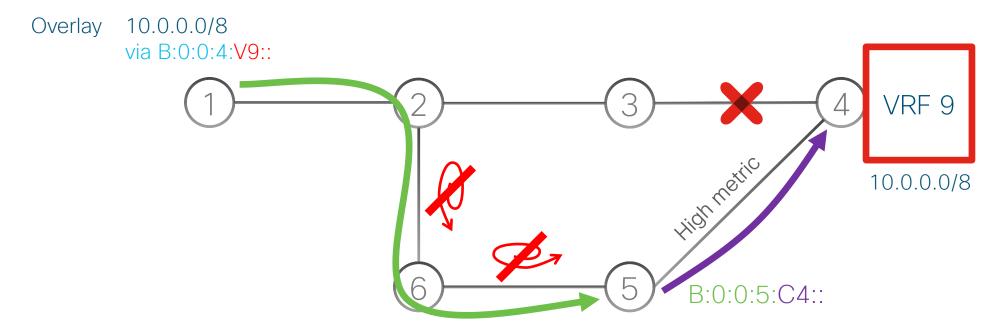
Note: in SRv6 using base format SIDs, >1 SID is rarely needed on the backup path due to the global reachability of the End.X SIDs (Adj-SIDs), essentially combining Node-SID + Adj-SID into a single SID

Microloops



- Microloops are a day-one IP drawback
 - Unsynchronized distributed convergence and IP hop-by-hop routing can cause transient packet loops after a topology change
- Microloops cause packet loss and out-of-order packets

SR Microloop Avoidance

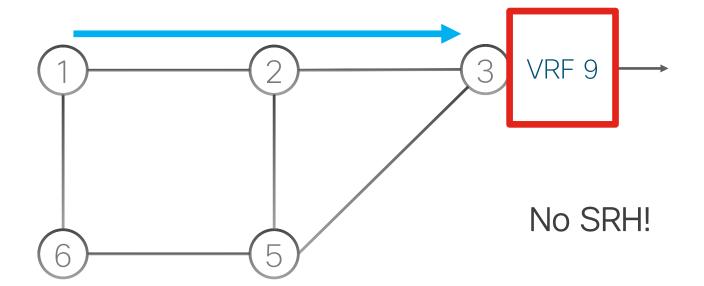


- SR Microloop Avoidance temporarily steers traffic on the loop-free postconvergence paths using SR Policies
- After the network has converged the SR Policies are deactivated

VPN over Best-Effort 5G Slice

Network Program: B:0:0:3:V9::

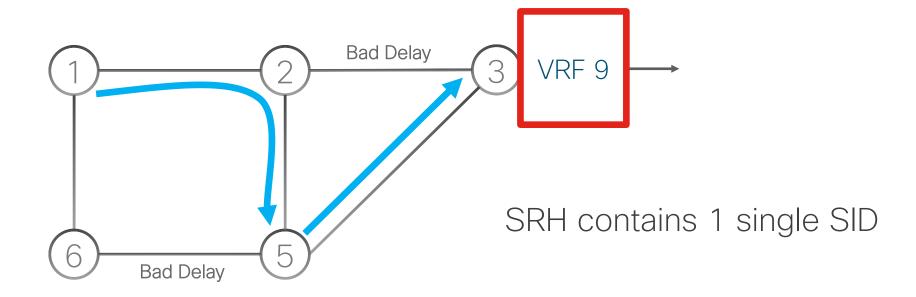
B::/40 locator block is associated with ISIS base algo (Low Cost, Best Effort)



VPN with Low-Delay 5G Slice - SR-TE Option

Network Program: B:0:0:2:C5:: then B:0:0:3:V9::

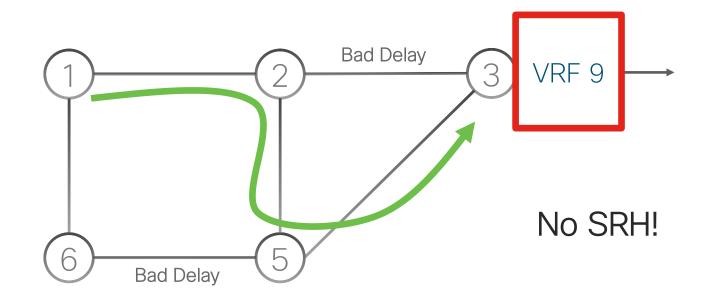
B::/40 locator block is associated with ISIS base algo (Low Cost)



VPN with Low-Delay 5G Slice - Flex-Algo Option

Network Program: D:0:0:3:V9::

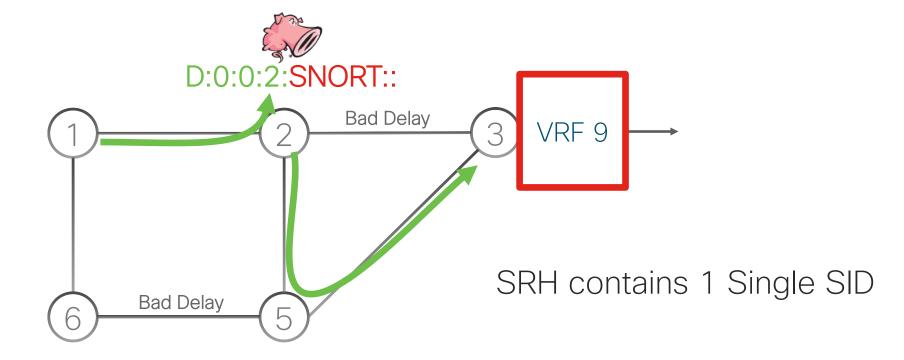
D::/40 locator block is associated with Low Delay Flex-Algo



Snort Firewall, VPN & Low-Delay Slice

Network Program: D:0:0:2:SNORT:: then D:0:0:3:V9::

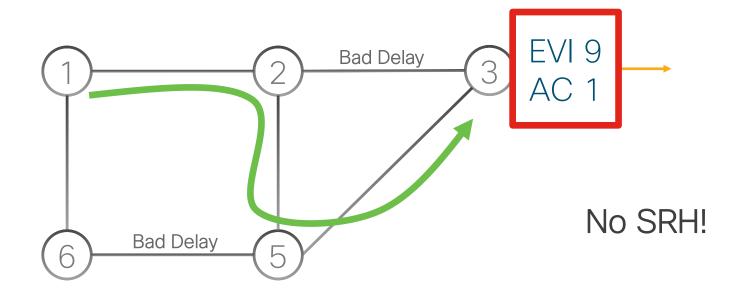
D::/40 locator block is associated with Low Delay Flex-Algo



EVPN VPWS Single-Home & Low-Delay 5G Slice

Network Program: D:0:0:3:X1::

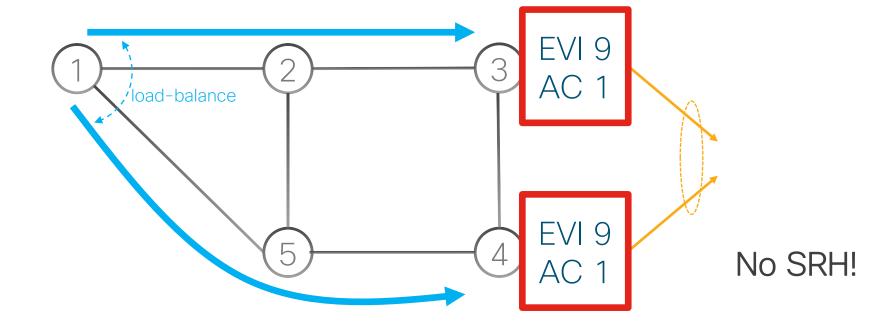
D::/40 locator block is associated with Low Delay Flex-Algo



EVPN VPWS MH All-Active & Best-Effort 5G Slice

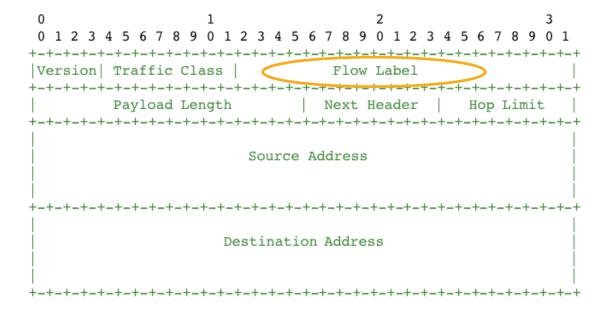
Network Program: B:0:0:3:X1:: or B:0:0:4:X1::

B::/40 locator block is associated with ISIS base algo (Low Cost)



Load-balancing

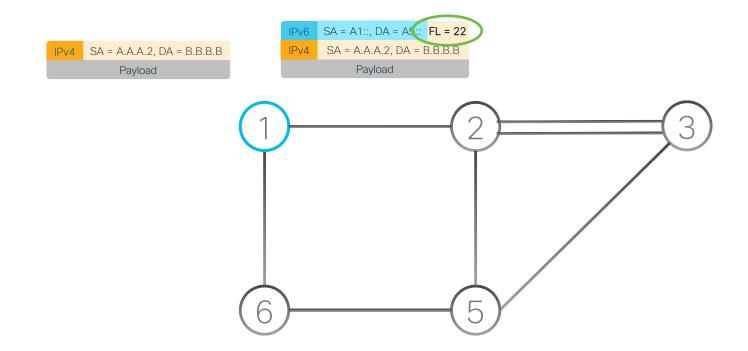
- 20-bit entropy
- No additional protocol
 - infamous mpls entropy label



Load-balancing

Action at the ingress of SRv6 domain

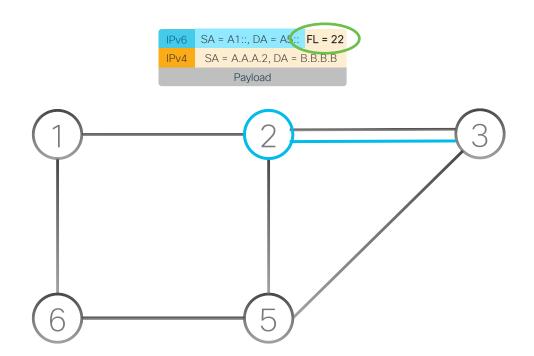
Flow Label is the result of the hash of the inner packet



Load-balancing

Action at a transit node

Outer Flow Label used for hashing



Seamless Incremental Deployment

- As soon as the network supports plain IPv6 forwarding
 - A new SRv6-VPN service only requires PE upgrade
 - TE objective can be achieved with a few well selected TE waypoints
 - FRR is deployed incrementally

On-Demand Nexthop / Automated Steering

- A service headend automatically instantiates an SR Policy to a BGP next-hop when required (on-demand)
- Color community is used as SLA indicator
- Reminder: an SR Policy is defined (color, endpoint)

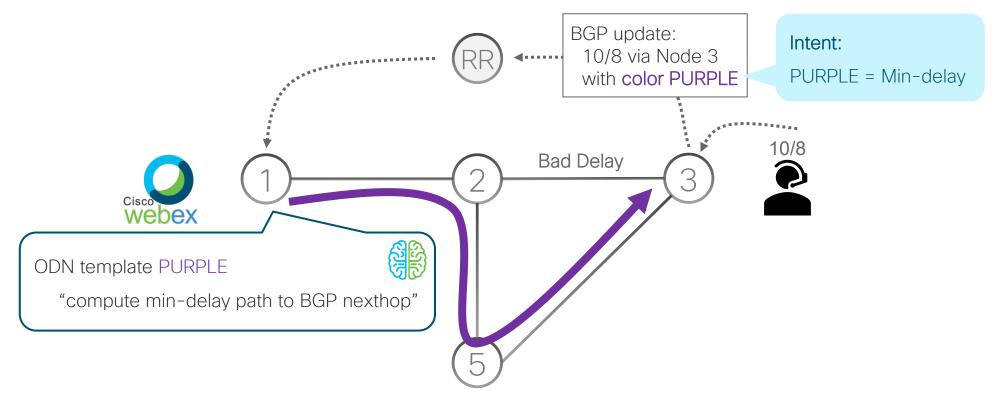
 BGP Color

 Community

 BGP

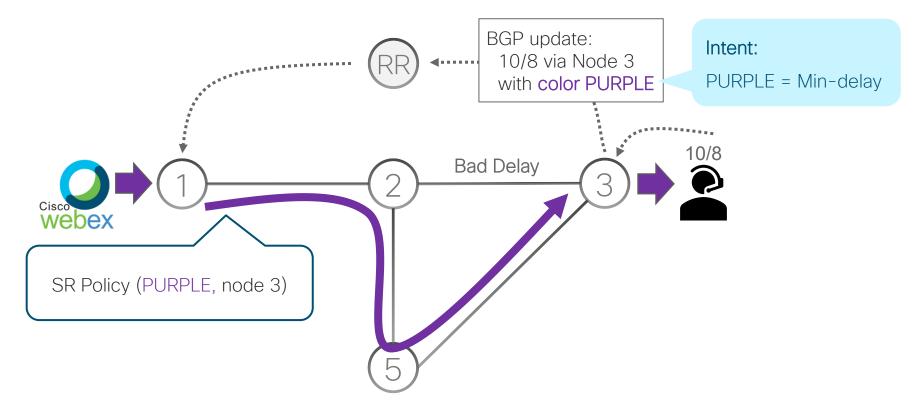
 Next-hop
- Automated Steering (AS) automatically steers the BGP traffic into this SR Policy, also based on nexthop and color

SR-TE On-Demand Nexthop (ODN)



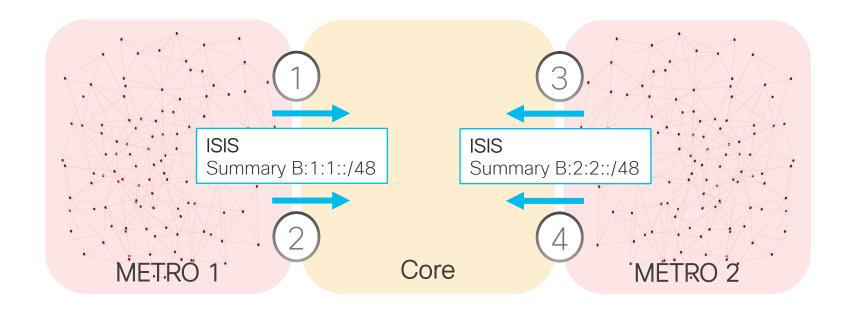
- Intent of service transport path is signaled by BGP (color community)
 - Best-effort reachability or reachability with SLA
- PE automatically computes or requests SR-PCE a path to the remote service endpoint and instantiates the SR Policy
 - No pre-configured mesh of SR Policies

SR-TE Automated Steering (AS)



- · Automatically steer services on the right SR Policy based on color of the service route
 - Provides per-destination SLA
 - No complex steering configuration
 - Data plane performant

Prefix Summarization



- Back to basic IP routing and summarization
- No BGP inter-AS Option A/B/C

SRv6 has excellent native Scale

- Many use-cases do not even use an SRH @
 - Any VPN (L3VPN, PW, eVPN)
 - Egress Peering Engineering
 - Low-Latency or Disjoint Slicing
 - Optimal Load-Balancing
- If SRH is needed, most cases will use 1 or 2 SID's
- Prefix Summarization gain
- Talk to the operators who deployed, they are happy to share experience

Negligible SRv6 SID block allocation - Iliad

As of the end of 2019, the SRv6 network consists of:

- o 1000 Cisco NCS 5500 routers.
- o 1800 Iliad's Nodeboxes.
- o The network services 4.5 million mobile subscribers (as of Q3 2019).
- The network is carrying 300 Gbps of commercial traffic at peak hours.
- o It is expected to grow to more than 4000 Nodeboxes in 2020. The SRv6 SIDs are allocated from a /40 sub-block of FC/8.

Less than 1 billionth of the FC/8 space - Negligible

Negligible SRv6 SID block allocation - SBB

- SBB currently has a /20 public IPv6 space from APNIC
- SBB SRv6 is supported by a /40 sub-block
- This is only 1 millionth of the current SBB allocation

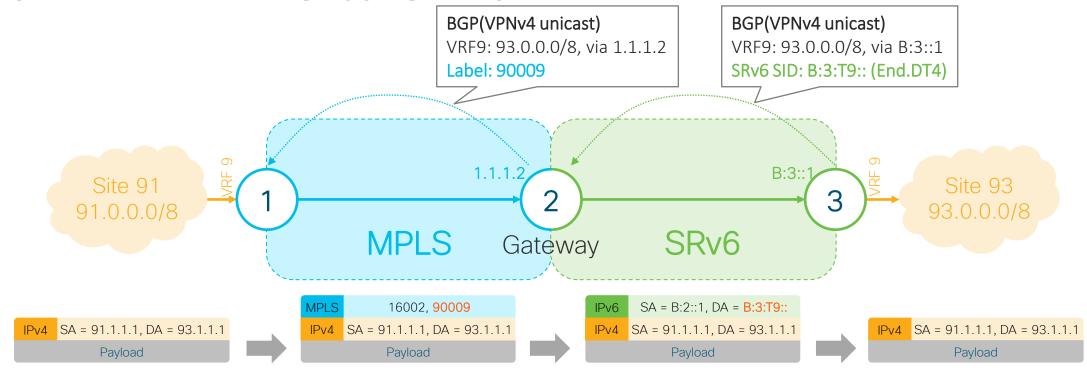
SRv6/MPLS L3 Service Interworking Gateway

Services - MPLS/SRv6 VPN Interworking Gateway

- L3VPN services
 - SRv6 VPNv4 unicast ↔ MPLS (SR or LDP) VPNv4 unicast

- L3 Global Routing Table (GRT) services
 - SRv6 4PE (global IPv4 unicast over IPv6)
 → MPLS (SR or LDP) global IPv4 unicast
 - SRv6 global IPv6 unicast ↔ MPLS (SR or LDP) 6PE (global IPv6 labeled unicast over IPv4)

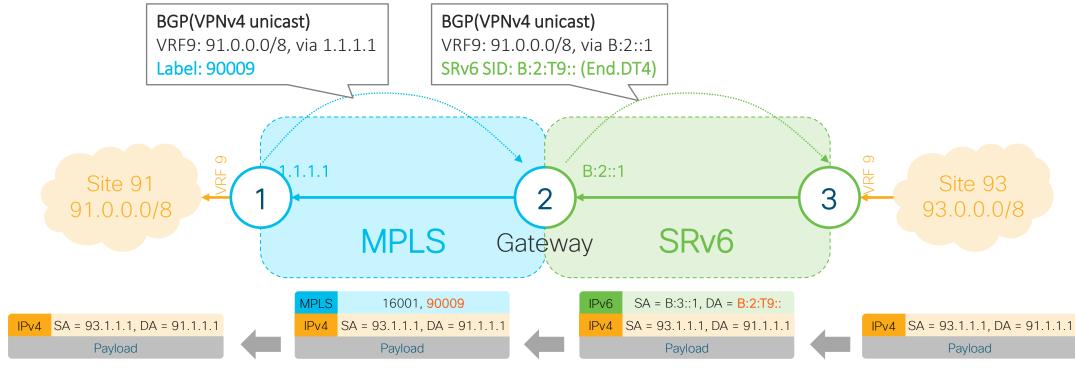
L3VPN - MPLS to SRv6



- Control plane: Gateway imports SRv6 service route 93.0.0.0/8
 and re-advertises in the MPLS domain with VPN label 90009
- Data plane: Gateway pops the VPN label 90009, looks up the destination in VRF 9 and pushes the appropriate SRv6 encapsulation

© 2021 Cisco and/or its affiliates. All rights reserved.

L3VPN - SRv6 to MPLS



- Control plane: Gateway imports MPLS service route 91.0.0.0/8 and re-advertises in the SRv6 domain with End.DT4 SID B:2:T9::
- Data plane: Gateway removes the outer IPv6 header, looks up the destination in VRF 9 and pushes the service and nexthop labels

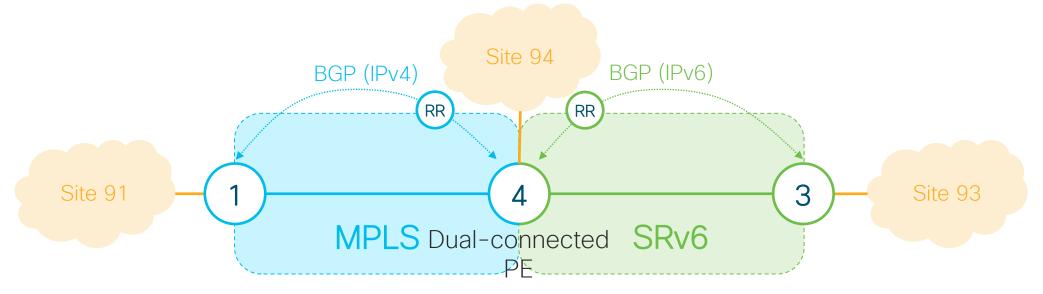
© 2021 Cisco and/or its affiliates. All rights reserved.

Services - Dual-connected PE

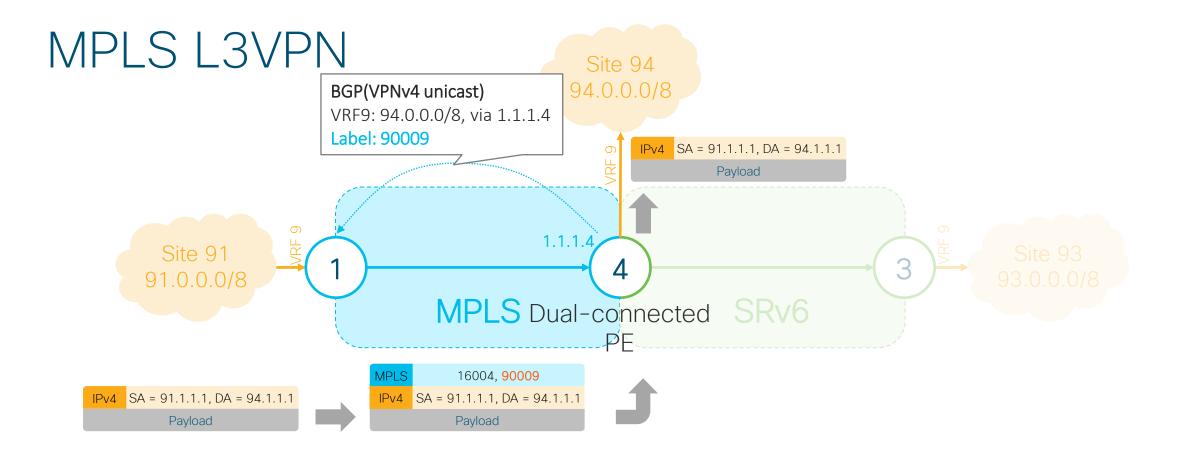
- L3VPN services
 - SRv6 VPNv4 unicast + MPLS (SR or LDP) VPNv4 unicast
 - SRv6 VPNv6 unicast + MPLS (SR or LDP) 6vPE (VPNv6 unicast over IPv4)

- L3 Global Routing Table (GRT) services
 - SRv6 4PE (global IPv4 unicast over IPv6) + MPLS (SR or LDP) global IPv4 unicast
 - SRv6 global IPv6 unicast + MPLS (SR or LDP) 6PE (global IPv6 labeled unicast over IPv4)

SRv6/MPLS Dual-Connected PE

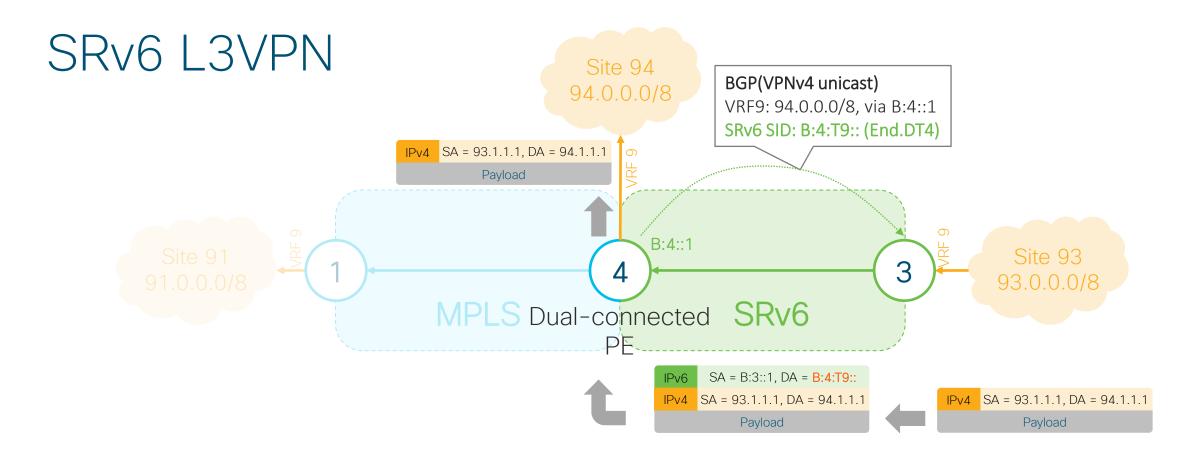


- The Dual-Connected PE enables customers to directly reach its services from both SRv6 and MPLS remote PEs without passing through a Gateway
- The Dual-Connected PE advertises its service routes to both MPLS and SRv6 remote PEs
 - MPLS service using IPv4 BGP session
 - SRv6 service using IPv6 BGP session
- The Dual-Connected PE directly (without GW) reaches service routes of remote SRv6 and MPLS PEs
 - A service route of a remote Dual-Connected PE can be reached via SRv6 XOR MPLS (apply BGP policy to select desired path)



- Dual-Connected PE4 advertises VPN route 94.0.0.0/8 in IPv4 BGP session with VPN label 90009
- Remote MPLS PE1 sends VPN packets to Site 94 using MPLS encapsulation and VPN label
- PE1 can use SR MPLS or LDP transport to reach PE4

© 2021 Cisco and/or its affiliates. All rights reserved.



- Dual-Connected PE4 advertises VPN route 94.0.0.0/8 in IPv6 BGP session with End.DT4 SID B:4:T9::
- Remote SRv6 PE3 sends VPN packets to Site 94 using IPv6 packets with DA = SID B:4:T9::

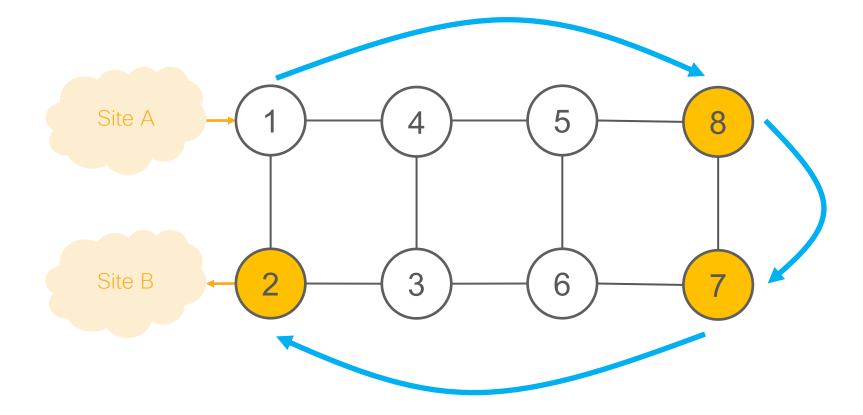
© 2021 Cisco and/or its affiliates. All rights reserved.

Micro-Program

Intuitive SRv6 Network Program

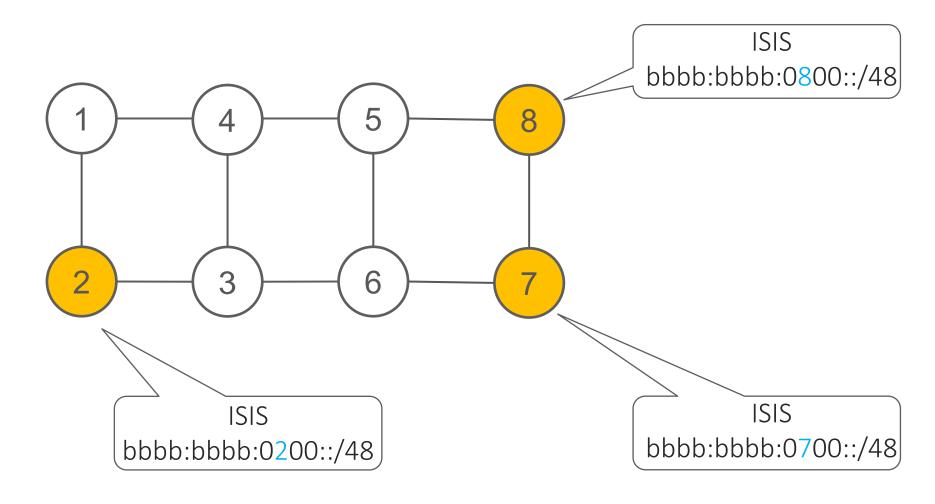
- Program
 - list of instructions contained in DA/SRH
- Instruction
 - SRv6 SID
- Micro-program
 - SRv6 SID (called container) that contains a list of micro-instructions
- Micro-Instruction
 - SRv6 uSID, can represent any behavior: TE, VPN, Service

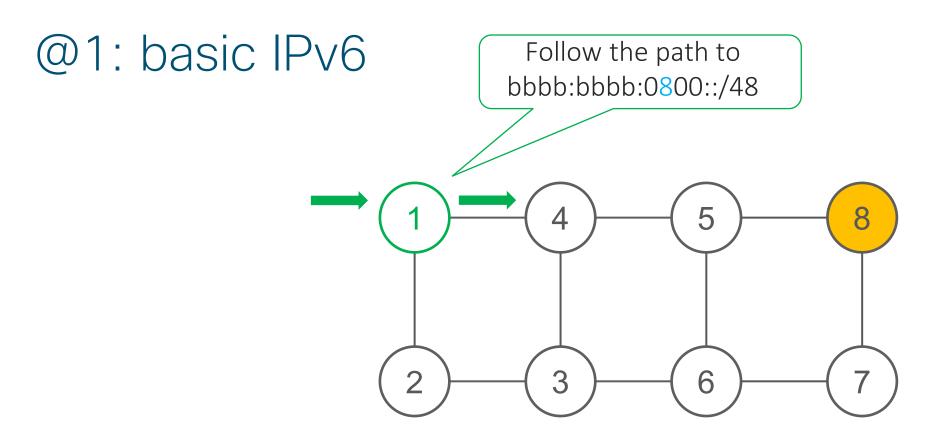
Illustration: go to 8 then 7 then 2 and decaps

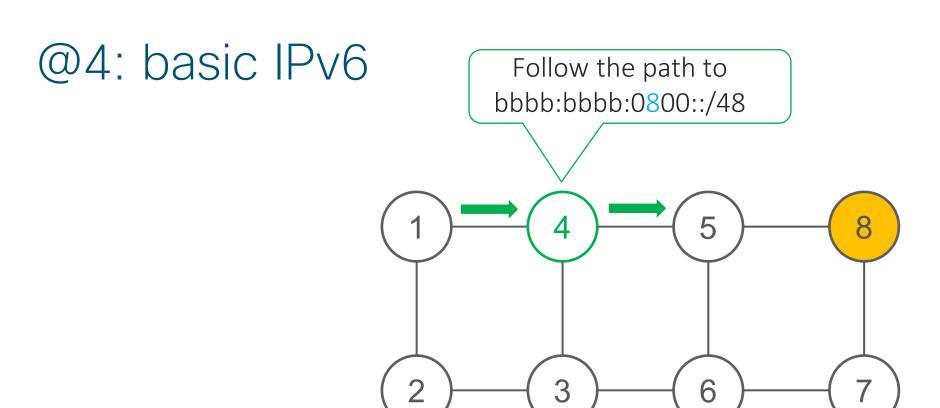


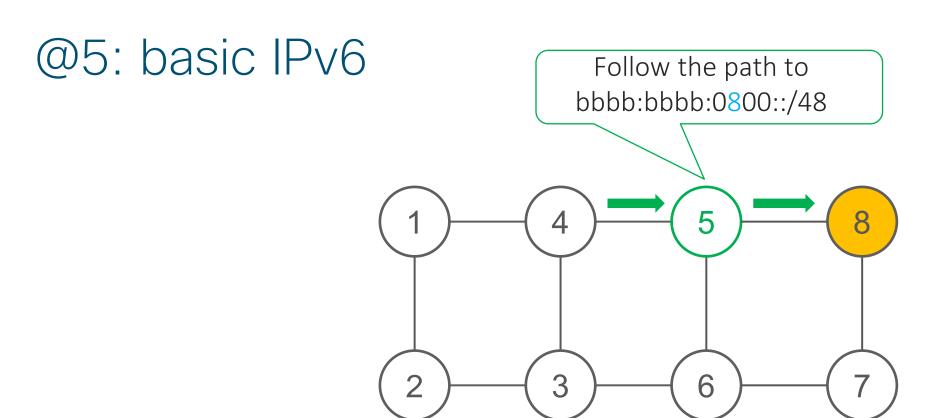
One single micro-program in the DA is enough

Basic IP Routing: no new extension

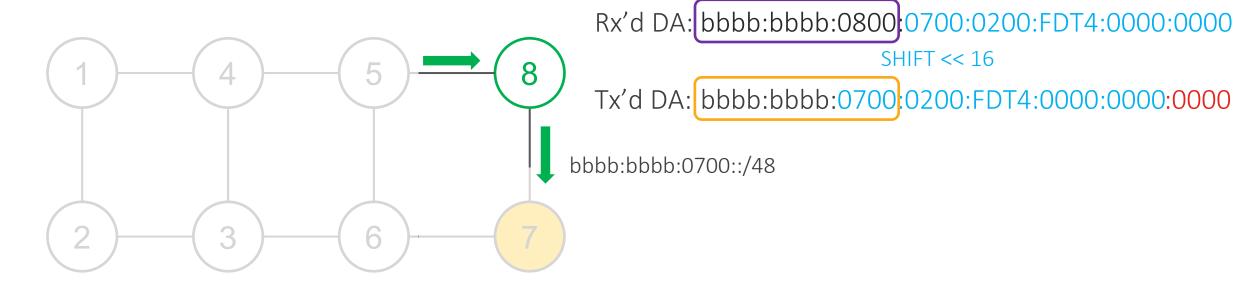








@8: Shift and Forward



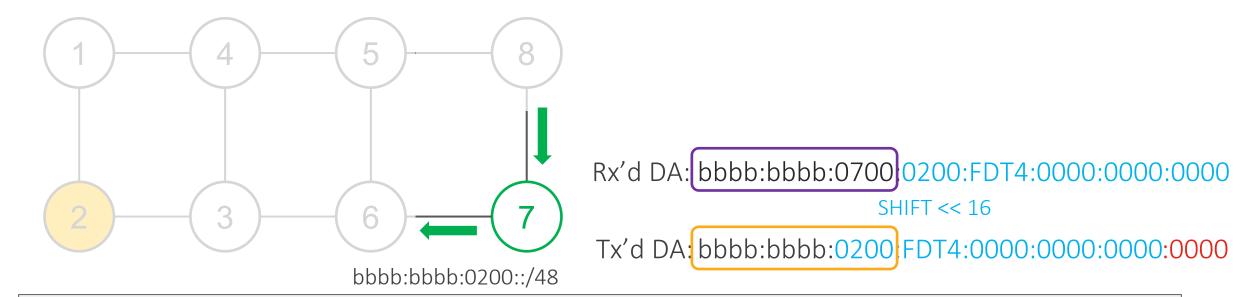
```
FIB Longest-Match bbbb:bbbb:0800::/48 → SRv6 Instruction:

Shift micro-program by one micro-instruction

Set last micro-instruction to "End-of-Container"

Lookup the updated DA and forward
```

@7: Shift and Forward



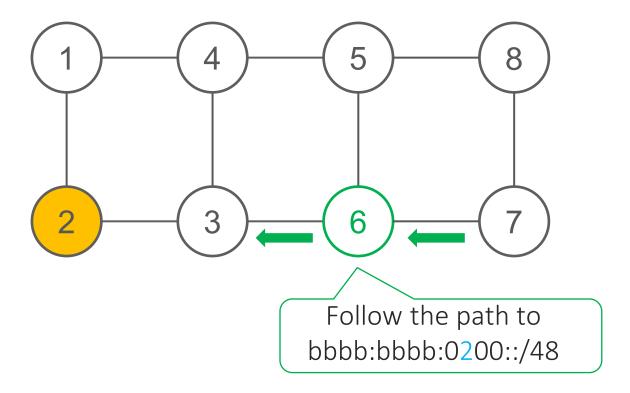
FIB Longest-Match bbbb:bbbb:0700::/48 → SRv6 Instruction:

Shift micro-program by one micro-instruction

Set last micro-instruction to "End-of-Container"

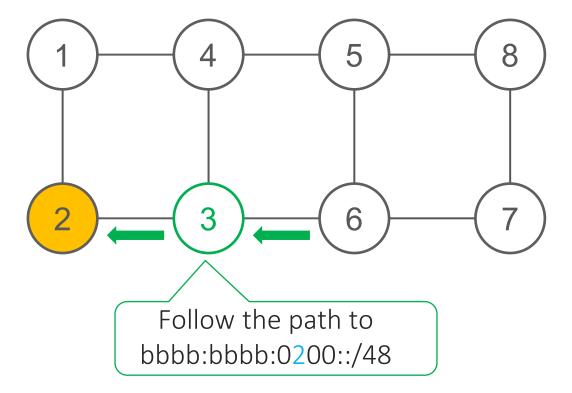
Lookup the updated DA and forward

@6: basic IPv6



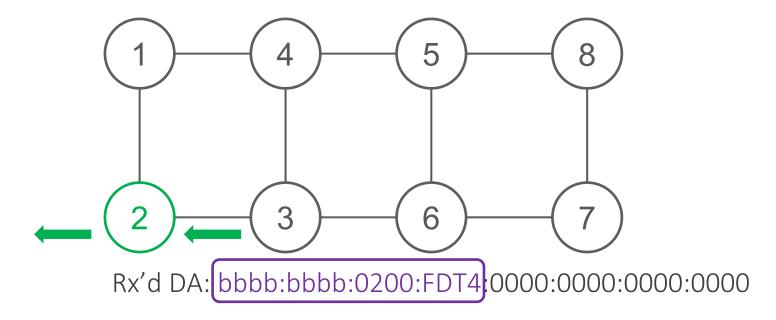
DA = bbbb:bbbb:0200:FDT4:0000:0000:0000:0000

@3: basic IPv6



DA = bbbb:bbbb:0200:FDT4:0000:0000:0000:0000

@2: SRv6 End.DT4 behavior



```
FIB Longest-Match bbbb:bbbb:0200:FDT4::/64 → SRv6 Instruction:

Decapsulate and forward inner IPv4 packet to Site B
```

Compliant with SRv6, Net Prog and IPv6

100% SRv6 and Net Prog compliant

- ✓ Just another SID, just another pseudocode
- ✓ Any SID in SRH or DA can be a uSID container
- ✓ uSIDs can be combined with any other SID

IPv6 compliant

- ✓ Leverage classic IP longest-match lookup
- ✓ Leverage classic IP-in-IP
- ✓ Use any IPv6 block available to the operator

uSID Benefits

Data Plane

- ✓ Best MTU efficiency (6 uSIDs without SRH)
- ✓ Hyper-Scalable SR-TE (18 uSIDs with 40 bytes overhead)
- ✓ Hardware-friendly (linerate on merchant silicon)

Control Plane

- ✓ Scalable number of globally unique uSIDs per domain
- ✓ No new protocol extensions

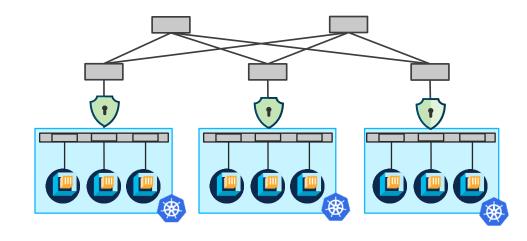
IP Power

- ✓ IP summarization and longest match is POWERFUL
- ✓ FIB efficiency 2 to 3 times gain vs MPLS
- ✓ Optimal IPv6 load-balancing (flow label)

SRv6 to the host

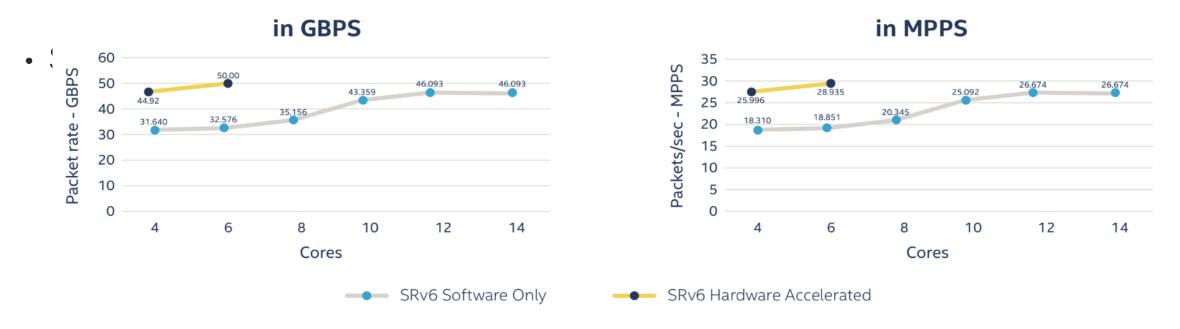
Containers networking

- Kubernetes (K8s) is the defacto container orchestration platform
 - Cisco (CCP) and others provides on-premise K8s solutions
 - Supported by most cloud providers (Alibaba, Azure, AWS, Google)
- SRv6-based K8s Fabric
 - Automated overlay
 - Scalable group-based network policies.
 - Scalable service programming.
 - Supported in Contiv (k8s network plugins)



SmartNICs

- Increase data path performances while saving server CPU's for to run VNFs and applications workload.
- SRv6 (decap and proxy) is supported in Intel FPGA N3000 SmartNIC
- SRv6 encapsulation/insertion on their roadmap¹



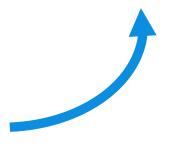
Conclusion

Simplicity Always Prevails



RSVP-TE
Inter-AS Option A/B/C
MPLS
UDP/VxLAN
NSH

Furthermore with more scaleand functionality

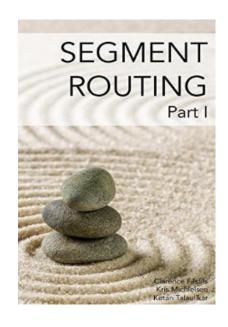




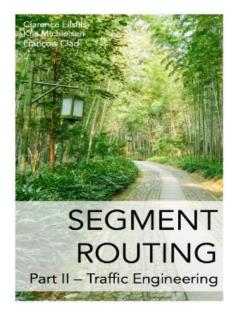
At record speed

- 9+ large-scale commercial deployments
 - Softbank, Iliad, China Telecom, LINE corporation,
 China Unicom, CERNET2, China Bank, MTN Uganda, NOIA Network, ...
- 25+ HW linerate implementations
 - Cisco Systems, Huawei, Juniper
 - Arrcus, Broadcom, Barefoot, Intel, Marvell, Mellanox, Kaloom
 - Spirent, Ixia
 - Multiple Interop Reports
- 11+ open-source platforms/ Applications
 - Linux, FD.io VPP, P4, iptables, nftables, snort, SERA, ExaBGP, GoBGP, GoBMP, Contiv-VPP, ...

Stay up-to-date



amzn.com/B01I58LSUO



amazon.com/dp/B07N13RDM9

SRv6 Part III Coming CY 21





facebook.com/SegmentRouting/



segment-routing.net



linkedin.com/groups/8266623

· I | I · I | I · I CISCO