SRv6

Network Programming

Francois Clad

Acknowledgements

- Clarence Filsfils
- Pablo Camarillo

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

Segment Routing

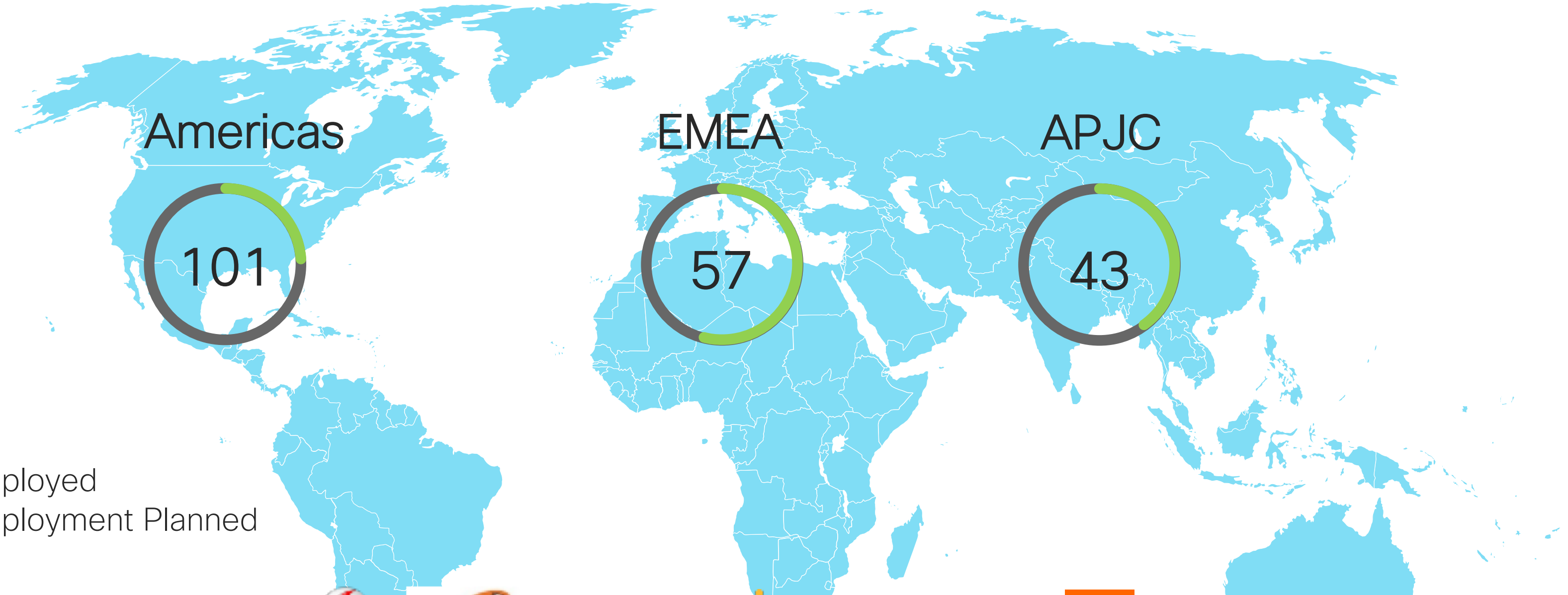
MPLS

- leverage the mature MPLS HW with only SW upgrade
- 1 segment = 1 label
- a segment list = a label stack

IPv6

- leverages ~~RFC2460~~ RFC8200 provision for source routing extension header
- 1 segment = 1 address
- a segment list = an address list in the SRH

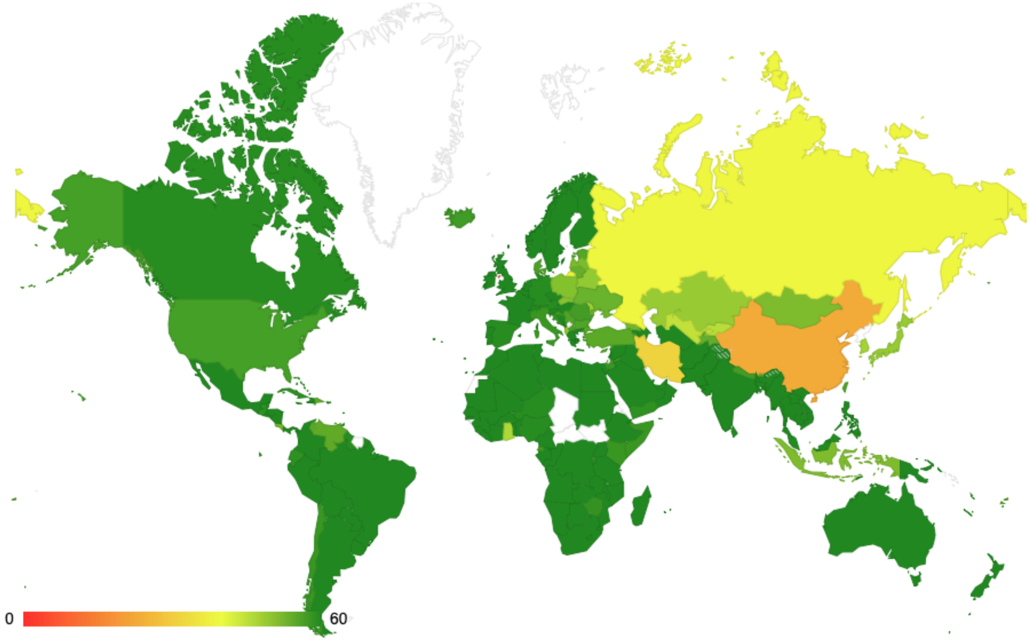
SR-MPLS: de-facto IPv4 solution



- Deployed
- Deployment Planned

Let's focus on SRv6: SR for IPv6

IPv6 adoption is a reality



% Web pages available over IPv6

Sources: 6lab.cisco.com – Web content
Cisco VNI Global IP Traffic Forecast, 2017-2022

Global IPv6 traffic **grew 226%** in 2017

Globally IPv6 traffic **will grow 18-fold** from 2017 to 2022

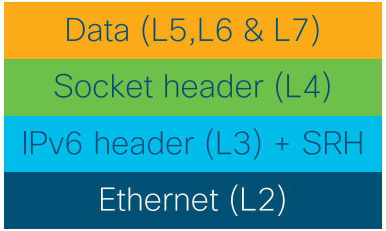
IPv6 **will be 38%** of total Internet traffic in 2022

IPv4 limitations & work-arounds

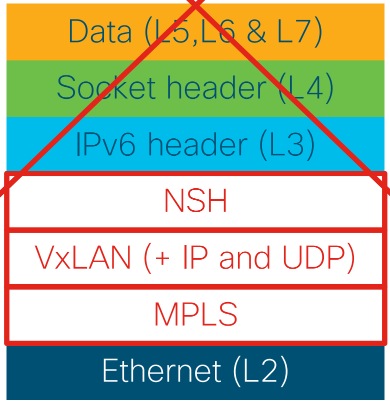
- × Limited address space → NAT
- × No engineered Load Balancing → MPLS Entropy Label, VxLAN UDP
- × No VPN → MPLS VPN’s, VxLAN
- × No Traffic Engineering → RSVP-TE, SR-TE MPLS
- × No Service Chaining → NSH

SRv6 Solution

✓128-bit address space	→ NAT
✓IPv6 flow label	→ MPLS Entropy Label, VxLAN UDP
✓SRv6 VPN	→ MPLS VPN's, VxLAN
✓SRv6 Traffic Engineering	→ RSVP-TE, SR-TE MPLS
✓SRv6 Service Chaining	→ NSH



Simplicity (back to the OSI model)



SRv6 fundamentals

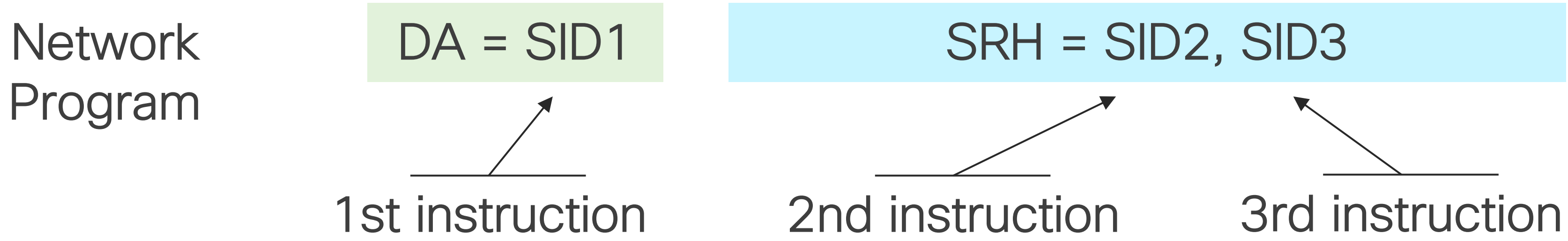
Network instruction

- 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 instruction

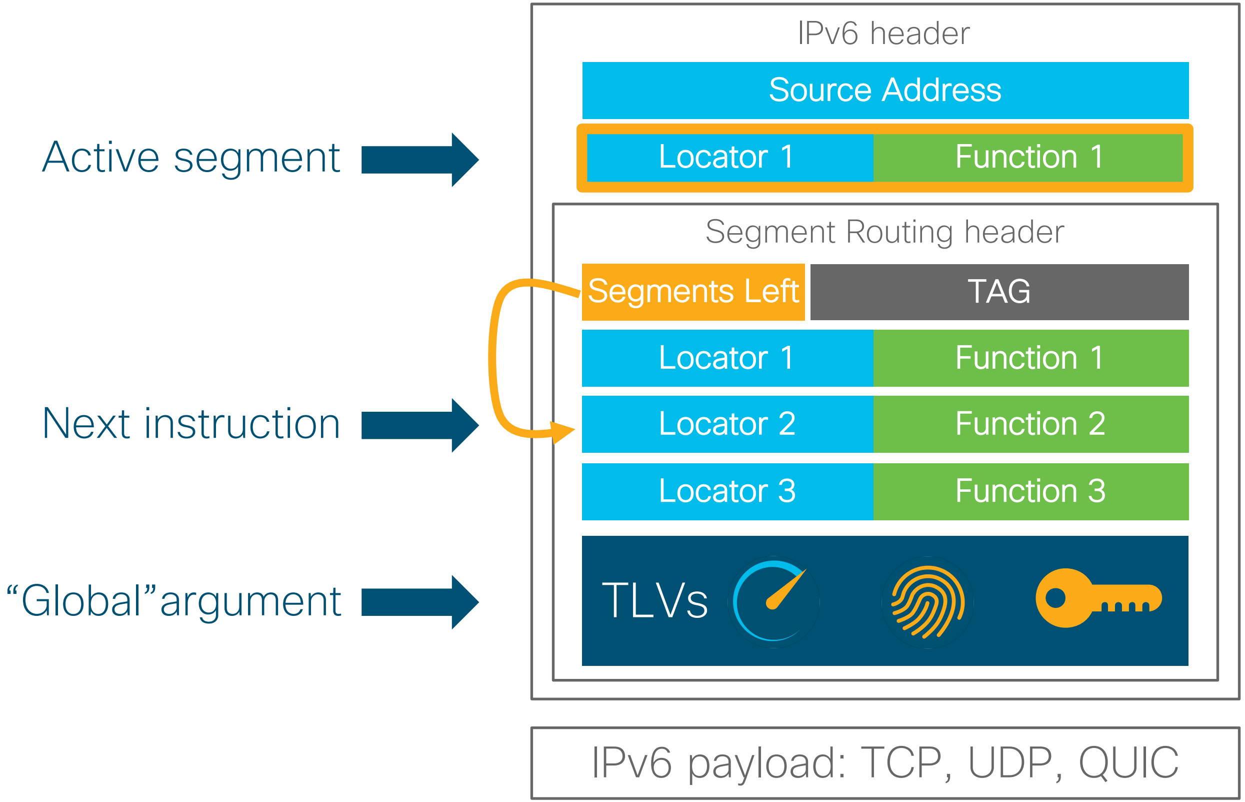
- 128-bit SRv6 SID
  - Locator: routed to the node performing the function
  - Function: any possible function
    either local to NPU or app in VM/Container
  - Arguments: optional argument bits to be used only by that SID
  - Flexible bit-length selection

Network Programming



- A network program is a list of instructions (128-bit SRv6 SID)
- An instruction can be bound to any behavior
  - TE/FRR: END, END.X
  - VPN: END.DX, END.DT

Network Program in the Packet Header

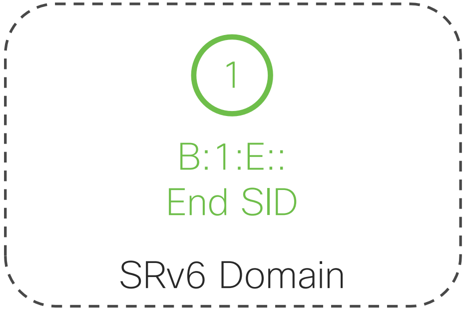


End and End.X SID behaviors

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

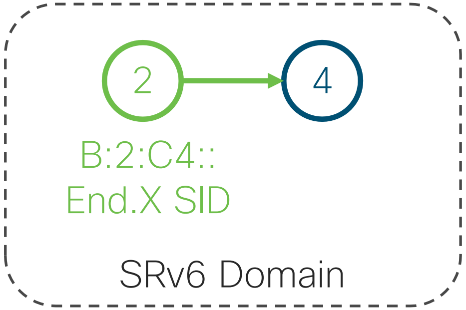
- End – Default endpoint behavior
  - shortest-path to the SID's endpoint
  - endpoint updates DA with next SID
  - endpoint forwards according to updated DA

Illustration convention B:<k>:E::



- 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 B:<k>:C<j>::, *where j identifies the remote node*

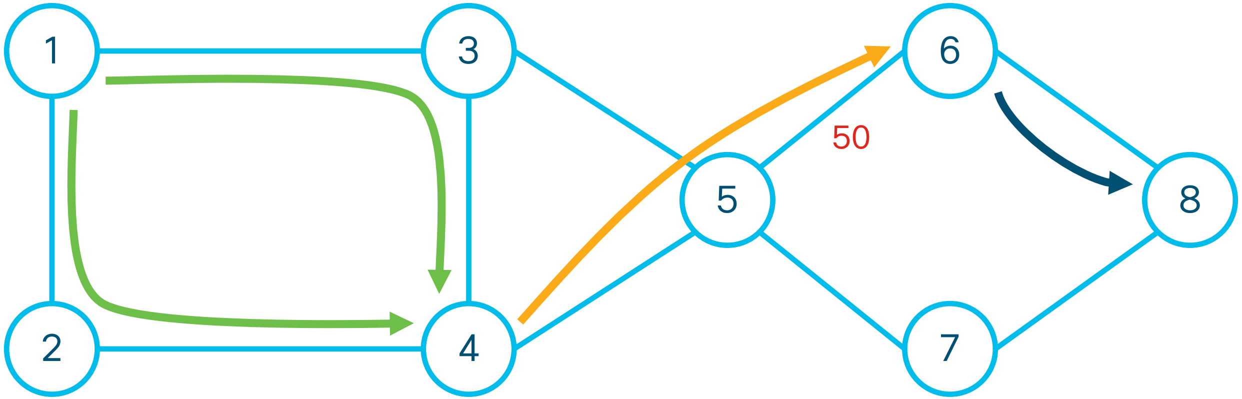


Endpoint behaviors illustration

Illustration convention:

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

SR: ⟨ B:4:E::, B:5:C6::, A:8:: ⟩

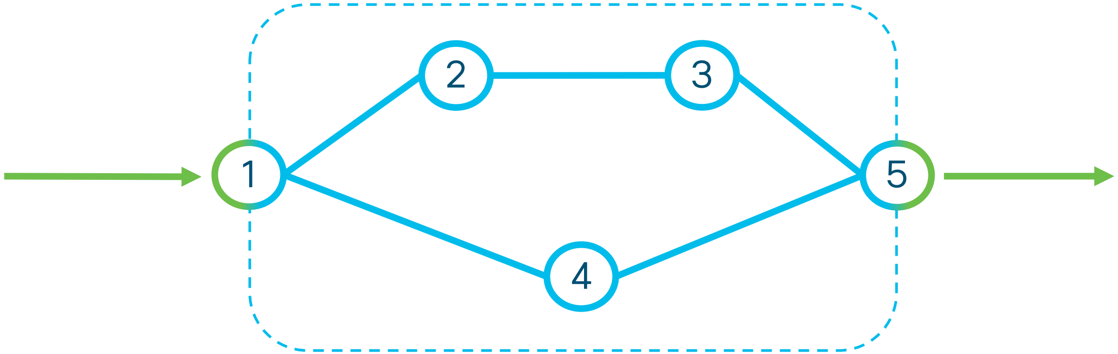


Default metric 10

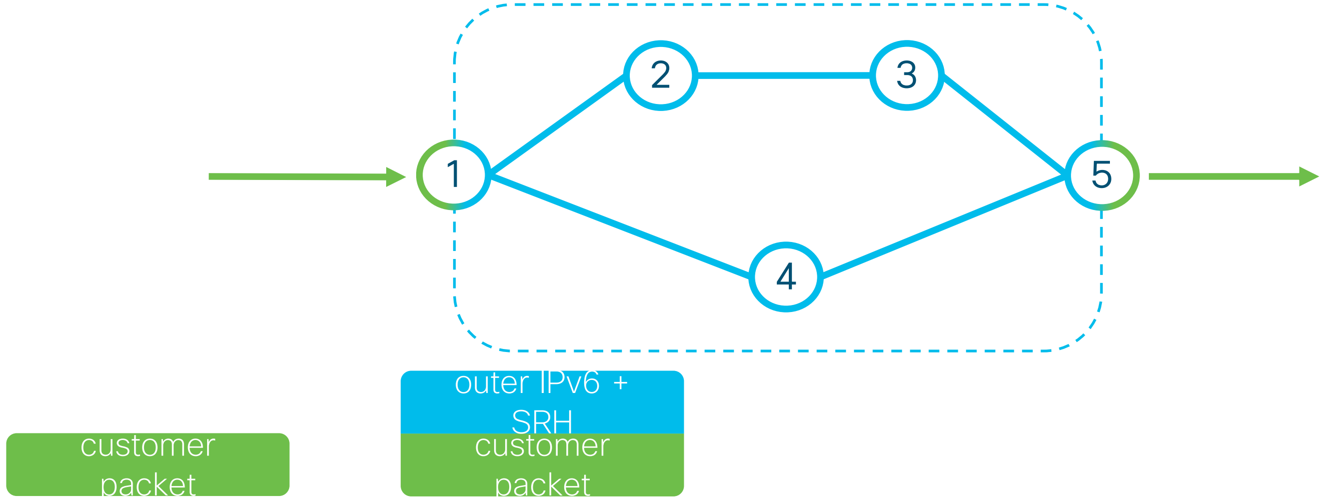
- B:4:E:: shortest path to node 4
- B:5:C6:: shortest path to node 5, then cross-connect towards 6
- A:8:: regular IPv6 address of node 8

SRv6 Domain

IPv6 enabled provider infrastructure
SR Domain

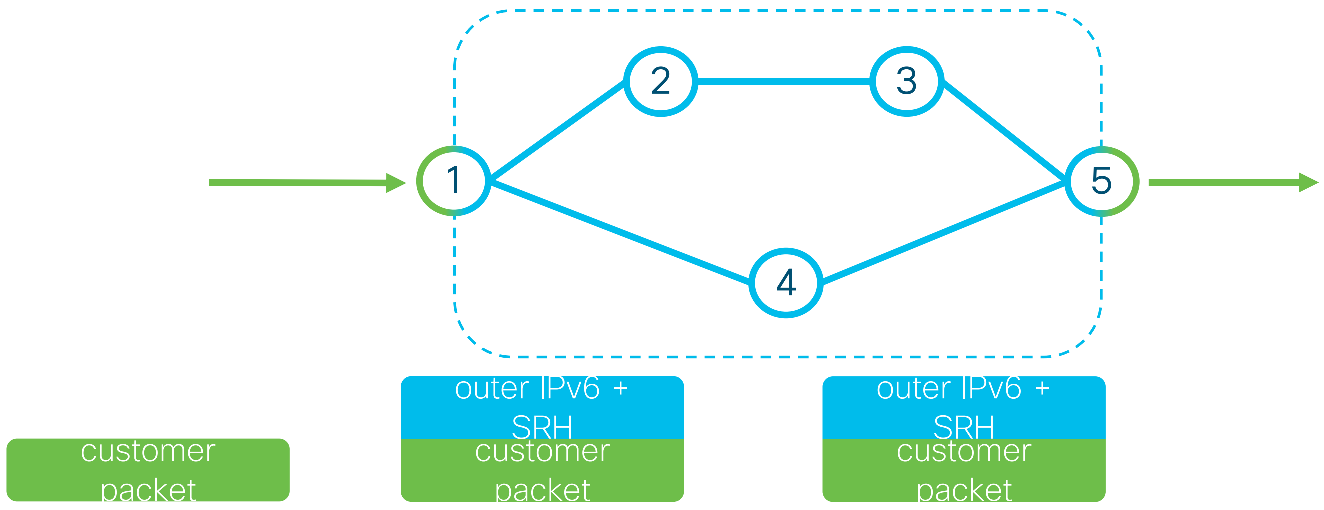


Encapsulation at the Domain ingress



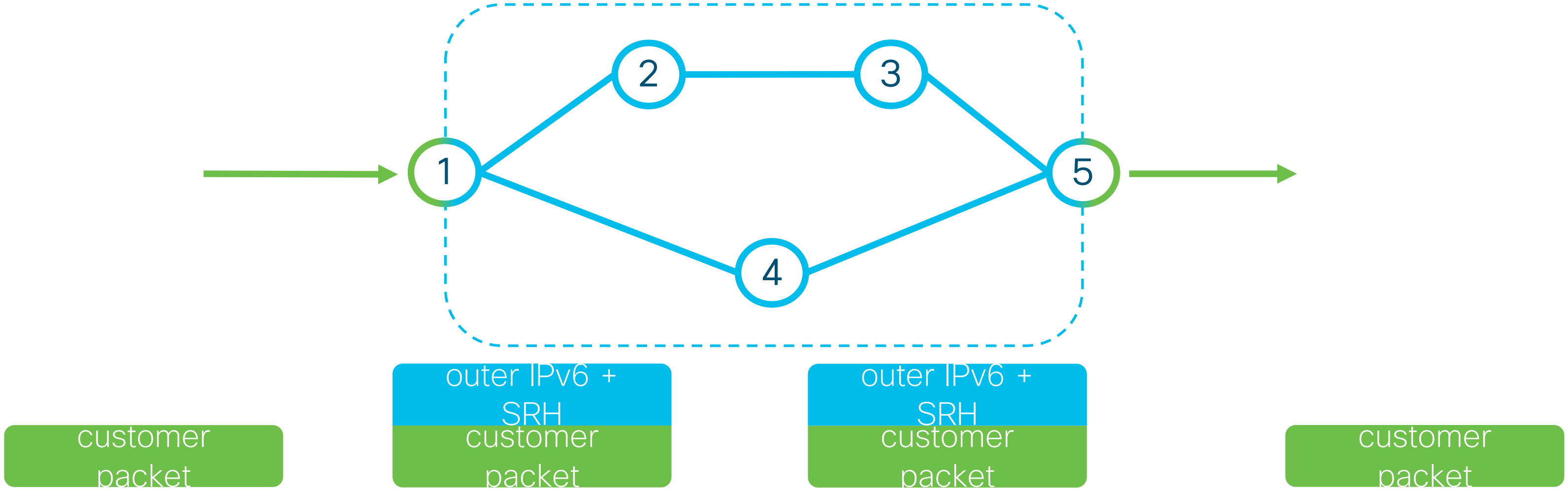
- IPv4, IPv6 or L2 frame is encapsulated within the SR Domain
- Outer IPv6 header includes an SRH with the list of segments

SRH of the outer IPv6 encapsulation



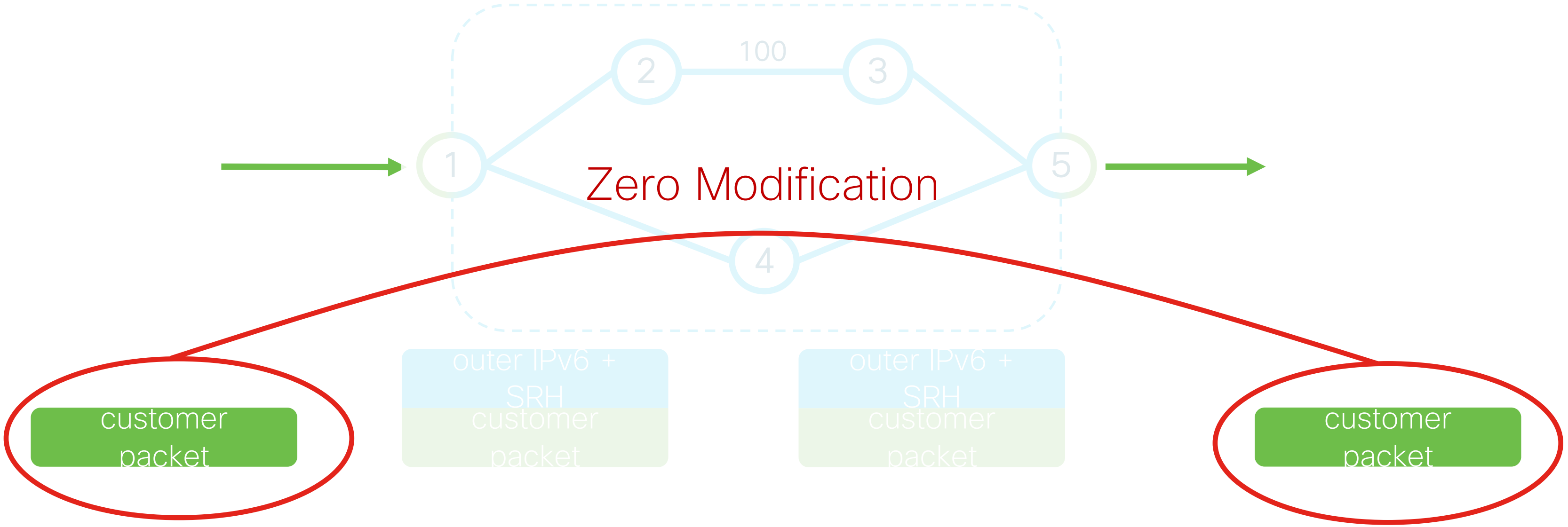
- Domain acts as a giant computer
- The network program in the outer SRH is executed

Decapsulation at Domain Egress



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

End-to-End Integrity

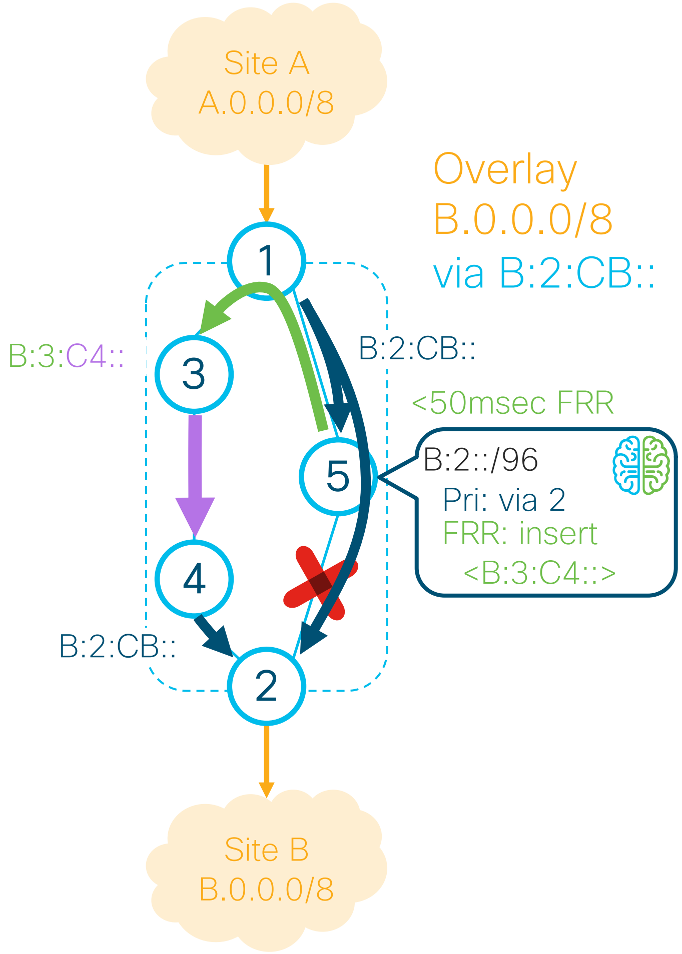


- End-to-end integrity principle is strictly guaranteed
  - Inner packet is unmodified
  - Same as SR-MPLS (MPLS stack is replaced by IPv6 outer header and SRH)

SRv6 Deployed Use-Cases

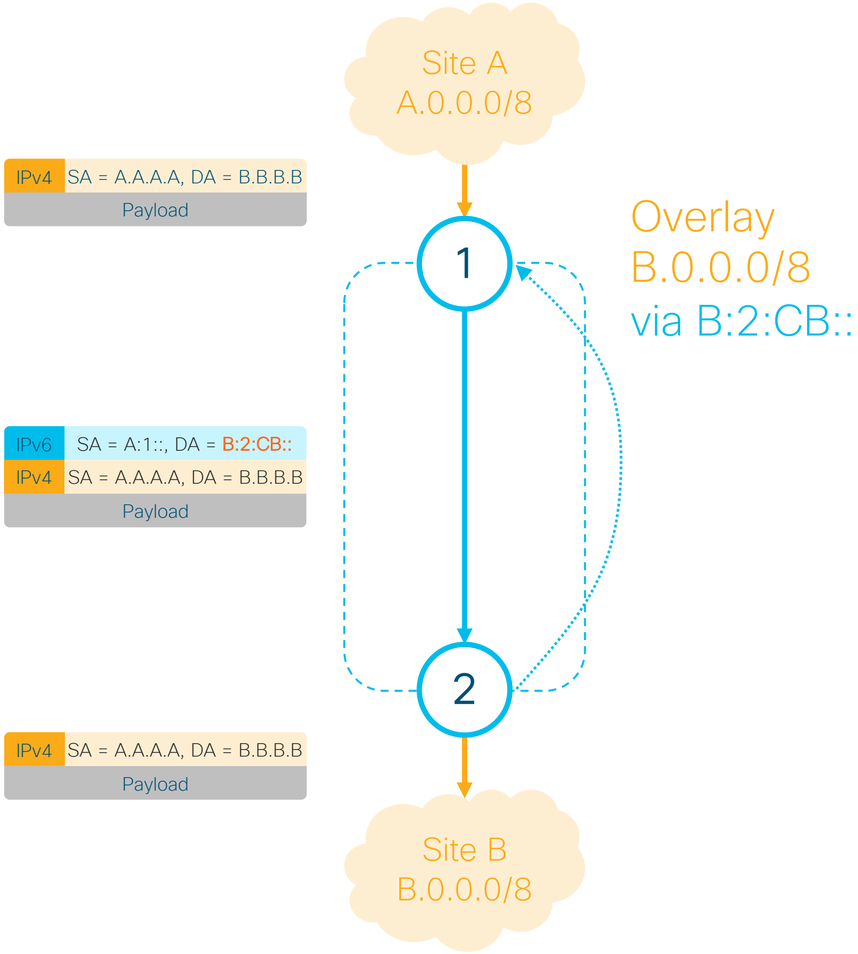
Fast reroute – TILFA

- 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 = postconvergence)
- Optimum backup path
  - leverages the post-convergence path, planned to carry the traffic
  - avoid any intermediate flap via alternate path
- Incremental deployment
- Distributed and Automated Intelligence



VPN over best-effort path

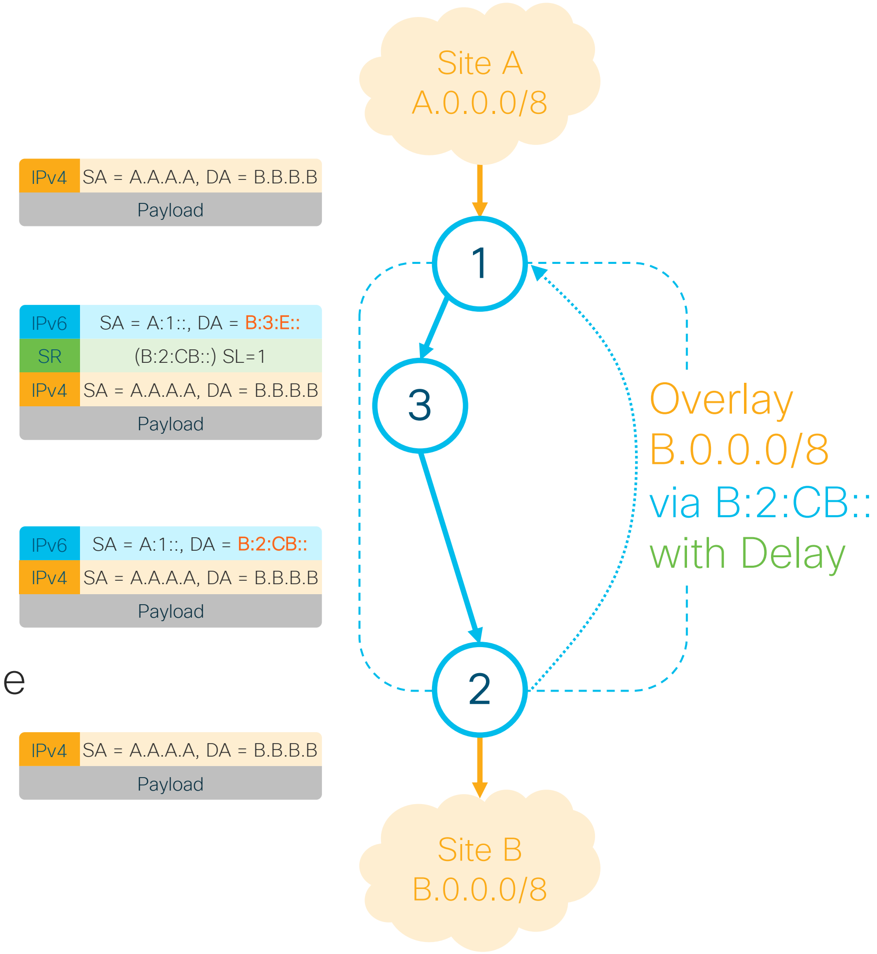
- Automated
  - No tunnel to configure
- Simple
  - Protocol elimination
  - Leverage existing control plane
    - >No new SAFI
    - >Lightweight extension to BGP Prefix-SID attribute



VPN over low-delay slice

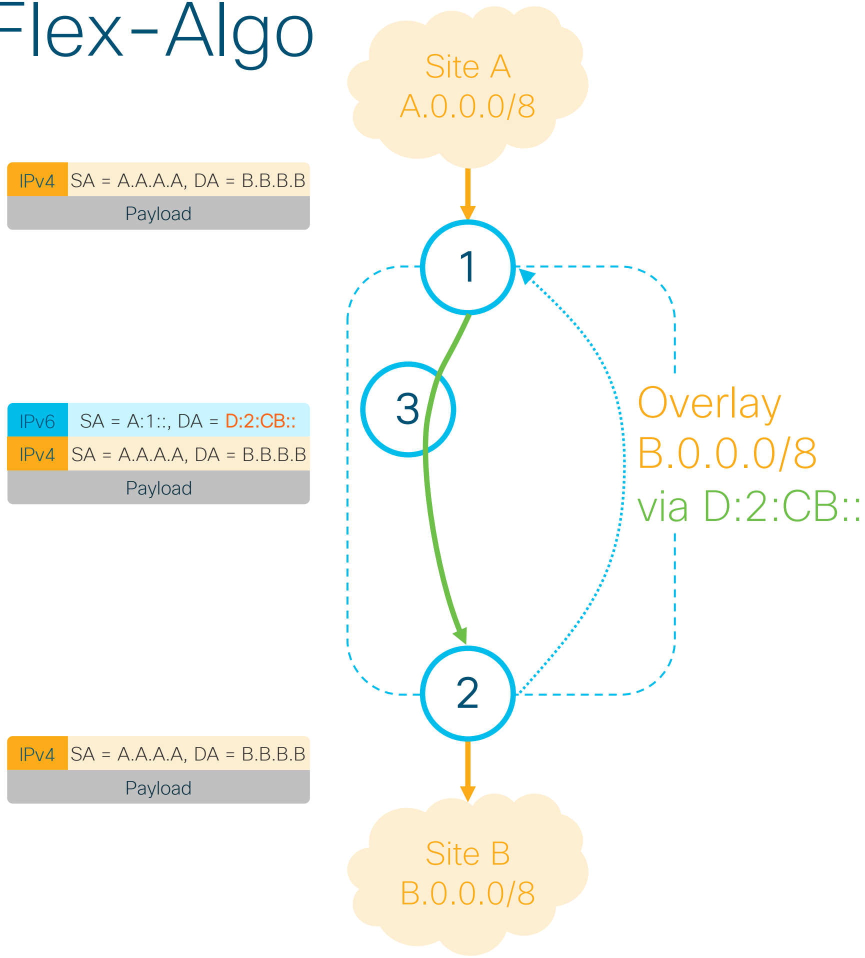
- Active SID is the IPv6 DA
- Any remaining SID stored in the SRH
- SL indicates how many SIDs remain

- Data plane optimizations:
  - First SID may be omitted in the SRH
  - Penultimate SR endpoint may remove the SRH



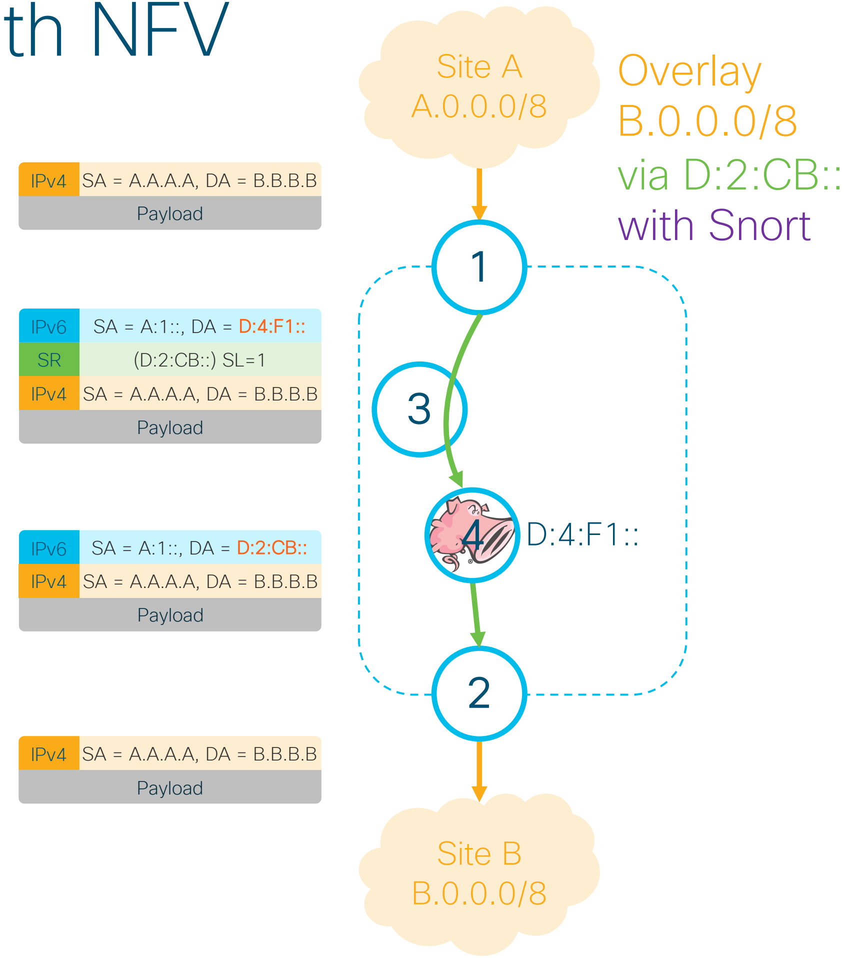
VPN over low-delay slice – Flex-Algo

- SR IGP Flexible Algorithms (Flex-Algo)
- Prefix segment bound to a custom shortest-path algorithm (e.g. low-delay)
- Fully distributed shortest path calculation
- Flex-Algo support and definitions advertised in the IGP



VPN over low-delay path with NFV

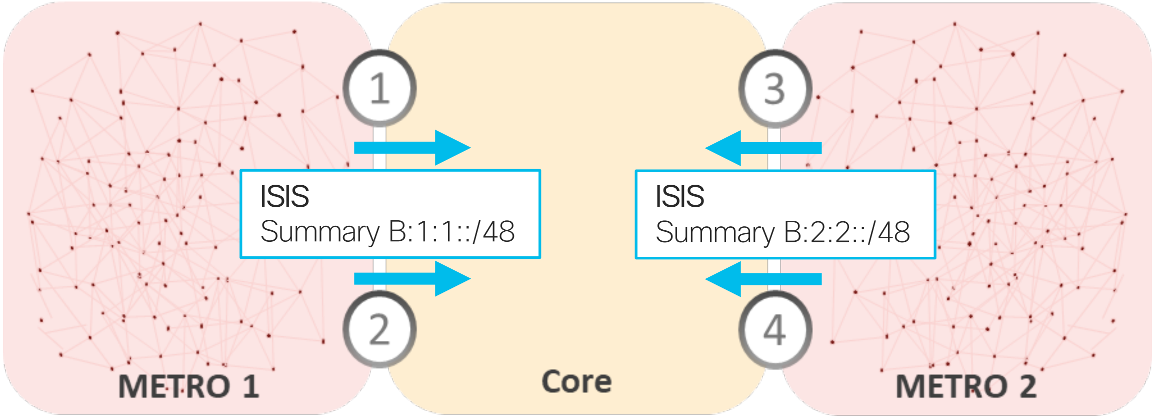
- SID bound to a Network Function
  - Just another type of segment
  - Stateless in the fabric
  - Seamless integration with VPN and TE
- NF can leverage the SRH
  - Implement branching operation
  - Read / write metadata
- Open-source SR-aware NFs
  - Snort, iptables, nftables
  - Leverage native SRv6 support in Linux kernel



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

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

SRv6 Eco-System

Lead Operators and Academia

At record speed

- In 2019: 8 large-scale commercial deployments
  - Softbank, Iliad, China Telecom, LINE corporation, China Unicom, CERNET2, China Bank and Uganda MTN.
- 18 HW linerate implementations
  - Cisco Systems, Huawei
  - Broadcom, Barefoot, Intel, Marvell, Mellanox
  - Multiple Interop Reports
- 11 open-source platforms/ Applications
  - Linux, FD.io VPP, P4, Wireshark, tcpdump, iptables, nftables, snort, ExaBGP, Contiv-VPP

SRv6 Ecosystem

Network Equipment Manufacturers

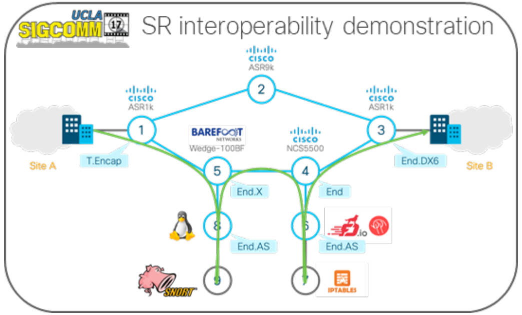
- CISCO
- HUAWEI

Merchant Silicon

- BAREFOOT NETWORKS
- BROADCOM
- MARVELL

Open-Source Applications

- Pyroute2
- SERA IPTABLES
- SNORT
- TCPDUMP
- Contiv
- WIRESHARK
- exa



Open-Source Networking Stacks

- fd.io

Smart NIC

- intel
- Mellanox TECHNOLOGIES

NFV Partners

- TREND MICRO
- ENEA Qosmos Division

Custom SRv6 behaviors with eBPF (End.BPF)

- Associates local SRv6 SID with user-defined eBPF program
  - Leverage Extended Berkeley Packet Filter (eBPF) functionality of the Linux kernel
  - User-defined C function inserted into the networking pipeline at run-time
  - No kernel compilation required
  - Guaranteed stability
- Provides helper functions to
  - Apply basic SRv6 behaviors (End, End.X,…)
  - Steer traffic into an SR policy
  - Add, modify or delete TLVs
- Available in Linux kernel 4.18 (August 2018)

SRv6 is a Proposed Standard

- RFC 8402 - Architecture
  - Defines SRv6 with SRH and SRv6 SID’s
- RFC 8754 (AUTH48) – SR Extension Header (SRH)
  - Defines the SRv6 dataplane encapsulation
- Last-Call status
  - Net Pgm
  - ISIS
  - OAM

Conclusion

Simplicity Always Prevails

~~LDP~~

~~RSVP-TE~~

~~Inter-AS Option~~

~~A/B/C~~

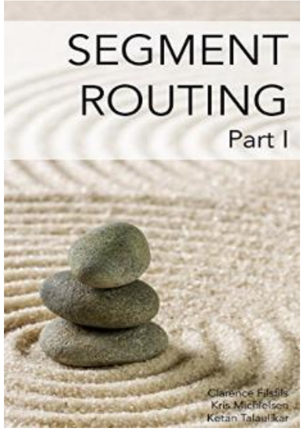
~~MPLS~~

~~UDP/VxLAN~~

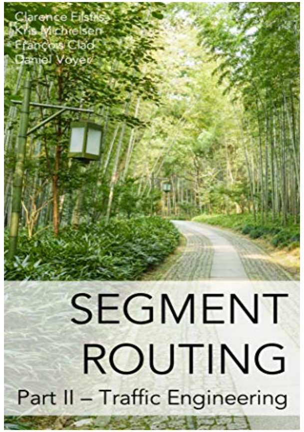
~~NSH~~

Furthermore with more scale and functionality

Stay up-to-date



amzn.com/B01I58LSUO



amazon.com/dp/B07N13RDM9

twitter.com/SegmentRouting

segment-routing.net

facebook.com/SegmentRouting/

linkedin.com/groups/8266623 cisco