



## SRv6 Record-Speed Standardization and Deployment

Pablo Camarillo pcamaril@cisco.com

DKNOG12 - Copenhagen

### Thank you

- Lead operators
- EcoSystem Partners
- Academic Partners
- IETF Partners
- Cisco SR Team



### segment-routing.net

- SRv6 uSID: training
- SRv6 Stateless Slicing: <u>demo</u>
- SRv6 Ultra-Scale SR Policy: 26 uSID push at linerate: <u>demo</u>
- BGP PIC Edge with SRv6 Summarization: ISIS UPA: <u>demo</u>
- Path Tracing: <u>demo</u>

Deployment Status

### Record-Speed Deployment

- 3 years of commercial deployment (2019-2022)
- ~100M SRv6 subscribers
- ~100 deployments, with ~14 public reports
- Across markets (Web, SP, Enterprise) and geographies (Asia, EU, US)



### Alibaba - Full Stack SRv6 Deployment



- "Predictable network services to every single application, based on the full stack SRv6 innovations across endpoints(container, host, P4 gateway), network devices and controller/network service center" YuanChao Su, Alibaba
- SID ⇔ Service
- Massive IPv6 space (HyperScale)
- Seamless Deployment
  - Many use-cases do not require SRH

#### **Co-Development**

- ✓ SRv6 Service Anchors on C8000
- 2. The uSID POC in C8000

### Full Stack SRv6 Towards a 'Predictable Network'



### Commitment to SR Lead-Operators



## Standard-Based Technology



Vendor eco-system



Open-Source eco-system

### Mature Standardization

- Proposed Standard
  - RFC 8402 SR Architecture
  - RFC 8754 SRv6 DataPlane
  - RFC 8986 SRv6 Network Programming
  - RFC xxxx
    SRv6 ISIS Extension
  - RFC xxxx
    SRv6 BGP Extension
  - RFC xxxx SRv6 OAM
  - RFC xxxx
    SR Policy

Much faster standardization than usual

Sign of the SRv6 Industry Endorsement

### 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, ExaBGP, GoBGP, GoBMP, Contiv-VPP, SONiC, SAI, SERA

### Many Successful Interops

- 2020/04: EANTC: SRv6 interop between Cisco, Huawei, Juniper, Arrcus, Ixia (link)
  - Classic IPv6 nodes as SRv6 transit nodes
  - SRv6-L3VPN for IPv4 and IPv6 services
  - SRv6 TI-LFA FRR link protection with SRH insert
  - SRv6 EVPN for E-Line and EVPN L3VPN services
  - SRv6 TE SR Policy
- 2021/02: NetOne Systems (<u>link</u>)
  - Cisco XR, Cisco NX, Juniper

#### • 2021/09: EANTC: SRv6 interop between Cisco, Huawei, Juniper, Nokia, Spirent (link)

- SRv6-Based Global IPv4 and IPv6 services
- SRv6-L3VPN for IPv4 and IPv6 services
- SRv6 TI-LFA FRR local SRLG protection with SRH insert
- SRv6 EVPN for E-Line and EVPN L3VPN services
- IGP Flex-Algo using TWAMP-measured link delays



- Stateless: DA, PARTITION-ID and DSCP are independent fields in the packet header
- Scalable: Routing, Partitioning & QoS are orthogonal spaces
- Seamless Deployment: PARTITION-ID is seamlessly hashed by legacy devices
- IETF Terminology: Network Resource Partitioning (draft-filsfils-spring-srv6-stateless-slice-id)





- Stateless: DA, PARTITION-ID and DSCP are independent fields in the packet header
- Scalable: Routing, Partitioning & QoS are orthogonal spaces
- Seamless Deployment: PARTITION-ID is seamlessly hashed by legacy devices
- IETF Terminology: Network Resource Partitioning (draft-filsfils-spring-srv6-stateless-slice-id)





- Stateless: DA, PARTITION-ID and DSCP are independent fields in the packet header
- Scalable: Routing, Partitioning & QoS are orthogonal spaces
- Seamless Deployment: PARTITION-ID is seamlessly hashed by legacy devices
- IETF Terminology: Network Resource Partitioning (draft-filsfils-spring-srv6-stateless-slice-id)





- Stateless: DA, PARTITION-ID and DSCP are independent fields in the packet header
- Scalable: Routing, Partitioning & QoS are orthogonal spaces
- Seamless Deployment: PARTITION-ID is seamlessly hashed by legacy devices
- IETF Terminology: Network Resource Partitioning (draft-filsfils-spring-srv6-stateless-slice-id)



## SRv6 uSIDs

### SRv6 uSID Terminology

- Industry:
  - SRv6 Micro Segment
  - SRv6 uSID
  - Briefly: uSID
- IETF: NEXT-C-SID
  - Briefly: Next
  - IETF document: draft-ietf-spring-srv6-srh-compression-01
  - Training: <u>link</u>

### Rakuten – SRv6 uSID in Deployment



- 5G End-to-End Network Slicing based on SRv6 uSID Flex-Algo
  - SRv6 uSIDs are allocated from the ULA address range
  - SRv6 uSID ISIS Flex-Algo: Low-Cost vs Low-Delay
  - SRv6 uSID BGP services
  - SRv6 uSID TILFA
  - Cisco NCS5500 and NCS-540 series
- Innovation in partnership with Cisco
  - BGP PIC Core and Edge with SRv6 Summarization: ISIS UPA (demo)
  - SR BW counters for deterministic and scalable capacity planning and BW guarantee

### Bell Canada - SRv6 uSID Deployment

- Bell promptly switched from SR-MPLS to SRv6 uSID
- Continued Simplification (remove MPLS dataplane)
- Better Routing Scale: Summarization
- Better HW Scale: linerate 26 uSID push for end-to-end SR Policy

Kal

- Seamless Deployment (6 uSID's in DA without SRH)
- End-to-End IP Unified Dataplane from socket to Internet Peering
  - SRv6-TE Policy: topological and service uSID's
- Service Programming
- Reduce network service costs by up to 90% footprint by 75% power consumption by as much as 66%

### SRv6 uSID - Feature Parity with SR-MPLS

- TILFA & uLoop Avoidance
- Flex-Algo Low-Cost/Low-Delay with VPN Automated Steering
  - Performance Monitoring: Link Latency (for Low-Delay slice)
- SRv6-TE Policy: topological and service uSID's
- L3VPN (IPv4 and IPv6), IPv4 Internet, IPv6 Internet, PW, BD
- VPN GW to interconnect with legacy VPN
- Seamless Inter-Domain with Summarization
  - No need for complex BGP3107 mechanisms

### SRv6 uSID - Rich Eco-System

- Cisco, NoviFlow, Arrcus, Nokia, Ciena
- Merchant: Silicon One, Broadcom, Marvell, Barefoot

Linux

......

**CISCO** 

NOKIA

ΔREF

KS | an Intel company

MARVELL

na

• Open Source: Linux, FD.io, P4, eBPF, Cillium, SAI

### SRv6 uSID – Pure IP Routing

• Classless Routing (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

### IP and uSID – Pure IP Forwarding

- Longest-Prefix-Match (LPM)
- Benefits
  - Straightforward Implementation
  - Support for different uSID sizes in the same uSID list
  - Capability to process multiple uSIDs with one single LPM
- Other technologies cannot do this
  - An MPLS lookup processes one single label at a time
  - All MPLS labels have a fixed 20-bit size
  - GSID (CSID-REPLACE) has the same limitations as MPLS

### Illustration – LPM flexibility

• 1 single LPM to resolve across different scales of services at PE2



### Perfect SRv6 Integration

- uSID reuses SRH (RFC8754) without any change
- uSID strictly applies the SRv6 Network Programming (RFC8986)
- uSID can be bound to any Network Programming Instruction

### SRv6 uSID offers the best SRv6 Compression



### SRv6 uSID offers the best SRv6 Compression



### SRv6 uSID offers the best SRv6 Compression



### uSID is 59% more efficient than GSID (CSID-Replace)

62/39 = 1.59

### Better HW Scale



	SRv6 uSID	MPLS
Linerate steering into SR Policy of N SID's (NCS-5700, J2)	N=26	N=~12
Consumed counters associated to a remote ISIS node	1	4
Consumed dataplane entries associated to remote ISIS node	1	4



### Better Routing Scale



	SRv6 uSID	MPLS
Unique Nodes in the SR domain	15M-240M	0.8M
Unique Services per node	512k	200k
ISIS Summarization	Yes	No
BGP3107 complexity tax to scale ISIS Host Routes	No	Yes

### Seamless Deployment

- Many use-cases leverage classic IPinIP
- The outer DA already holds 6 uSIDs !! <u>D. Cai Alibaba podcast SRv6 uSID</u>
- SDN coherence
  - The SDN controller packs the uSID policy in the outer DA
  - The host/access node only sees an opaque DA
- VPN GW to interconnect with legacy VPN

# Path Tracing

### The exact path from A to M is not known



### The exact path from A to M is not known

- 7 possible "valid" ECMP path
  - ABFM, ABGM, ACFM, ACGM, ACHM, ADGM, ADHM ...



### The exact path from A to M is not known

- 7 possible "valid" ECMP path
  - ABFM, ABGM, ACFM, ACGM, ACHM, ADGM, ADHM ...
- The path may be invalid
  - Routing or FIB corruption @ B
- Timestamp at each hop
- Interface Load at each hop



### Stamping Trajectory in PT Header



### Stamping Trajectory in PT Header



### Stamping Trajectory in PT Header



### The PT idea

- Stamping in the Packet Header
- Implemented in the most basic HW pipeline
  - Linerate for any packet
- Ultra-MTU-efficient: only 3 bytes per hop!
  - 12-bit Interface, 8-bit Timestamp, 4-bit Load
- For IPv6, with or without SRH
  - MPLS solution also designed
- Interwork with legacy node

### Dataplane Encapsulation

- Minimize NPU parsing
- Minimize # of Read/Write
- Minimize depth of Read/Write
- Maximize Read/Write at fixed positions
- Avoid Header Insert/Resize
- Minimize MTU

### Minimize HW complexity by leveraging SDN analytics

- Analytics
  - translates the list of collected IDs into a path
  - deduces the timing and load history at each hop
  - Highlights hotspots
- Consistency check: FIB (PT) vs RIB (PCE)
- Feedback loop to applications
  - Trigger a change of path (SR, MTCP)
  - Trigger a change of rate



### Product, Deployment & EcoSystem

- Cisco Shipping in CY22
  - PT Demo and Training (<u>link</u>)
- Strong Operator Interest
- Rich Eco-System

• Rich Open-Source



......

CISCO



• At IETF: draft-filsfils-spring-path-tracing

# Conclusion





-LDP
-RSVP-TE
-BGP 3107
MPLS
-UDP/VxLAN
NSH



LDP
-RSVP-TE
-BGP 3107
MPLS
-UDP/VxLAN
-NSH

#### Furthermore with more scale







#### Furthermore with more scale and functionality





### Stay up-to-date



#### amzn.com/B01I58LSUO



#### amazon.com/dp/B07N13RDM9

SRv6 Part III Coming by Sept 2022



twitter.com/SegmentRouting



segment-routing.net



facebook.com/SegmentRouting/





#### ask-segment-routing@cisco.com

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