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

**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
Let’s focus on SRv6: SR for IPv6
IPv6 adoption is a reality

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

% Web pages available over IPv6

Sources: 6lab.cisco.com - Web content
Cisco VNI Global IP Traffic Forecast, 2017-2022
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
- IPv6 flow label
- SRv6 VPN
- SRv6 Traffic Engineering
- SRv6 Service Chaining

→ NAT
→ MPLS Entropy Label, VxLAN UDP
→ MPLS VPN’s, VxLAN
→ RSVP-TE, SR-TE MPLS
→ NSH

Simplicity (back to the OSI model)
SRv6 fundamentals
• 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
• 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
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

IPv6 header
- Source Address
- Locator 1 Function 1
- Locator 2 Function 2
- Locator 3 Function 3

Segment Routing header
- Segments Left
- TAG
- Locator 1 Function 1
- Locator 2 Function 2
- Locator 3 Function 3

TLVs
- "Global" argument
- Next instruction
- Active segment

IPv6 payload: TCP, UDP, QUIC

© 2020 Cisco and/or its affiliates. All rights reserved.
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

*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

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

SR: (B:4:E::, B:5:C6::, A:8::)

• 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

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

Default metric 10
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
• 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

```
Site A
A.0.0.0/8

Overlay
B.0.0.0.0/8
via B:2:CB::

Site B
B.0.0.0.0/8
```
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, tcpcut, iptables, nftables, snort, ExaBGP, Contiv-VPP
SRv6 Ecosystem

Network Equipment Manufacturers

- Cisco
- Huawei

Merchant Silicon

- Barefoot Networks
- Broadcom
- Marvell

Open-Source Applications

- Pyroute2
- SERA

Open-Source Networking Stacks

- P4
- Kubernetes
- Docker

Smart NIC

- Intel
- Mellanox

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

Furthermore with more scale and functionality
Stay up-to-date

amzn.com/B01I58LSUO
amazon.com/dp/B07N13RDM9

twitter.com/SegmentRouting
facebook.com/SegmentRouting/
segment-routing.net
linkedin.com/groups/8266623