Path Tracing in SRv6 networks draft-filsfils-spring-path-tracing-00

Clarence Filsfils - Cisco Systems (cf@cisco.com)
Ahmed Abdelsalam - Cisco Systems (ahabdels@cisco.com) - Presenter
Pablo Camarillo - Cisco Systems (pcamaril@cisco.com)
Mark Yufit - Broadcom (mark.yufit@broadcom.com)
Thomas Graf – Swisscom (thomas.graf@swisscom.com)
Yuanchao Su – Alibaba (yitai.syc@alibaba-inc.com)
Satoru Matsushima – SoftBank (satoru.matsushima@g.softbank.co.jp)

Agenda

- Goal
- Scope
- Terminologies
- Roles/Data Model
- Midpoint Compressed Data (MCD)
- PT Headers
- Ecosystem
- Next Steps

Goal

- Path Tracing (PT) provides a record of the packet path as a sequence of interface ids.
- In addition, it provides a record of end-to-end delay, per-hop delay, and load on each egress interface along the packet delivery path.
- It has been designed for linerate hardware implementation in the base pipeline.
  - Minimize header size (e.g., 36-Bytes HbH option allows recording end-to-end information for 14 transit nodes)
  - Minimize variability (no options; same editing by all nodes)

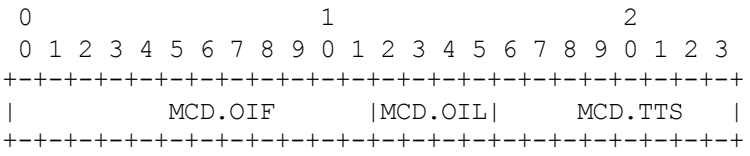
Scope

- This document defines the Path Tracing specification for the SRv6 dataplane.
  - Path Tracing is applicable to both SR-MPLS [RFC8660], as well as SRv6 [RFC8986].
  - The SR-MPLS dataplane will be detailed in a separate document.

Terminology

- PT: Path Tracing
- MCD: Midpoint Compressed Data. Information that every transit router adds to the packet for PT purposes.
- HbH-PT: IPv6 Hop-by-Hop [RFC8200] Path Tracing Option used for collecting PT data from Midpoints. It contains a stack of MCDs. Defined in this document.
- SRH PT-TLV: SRH TLV used to collect PT from source and Sink nodes. Defined in this document.

MCD – 3Bytes



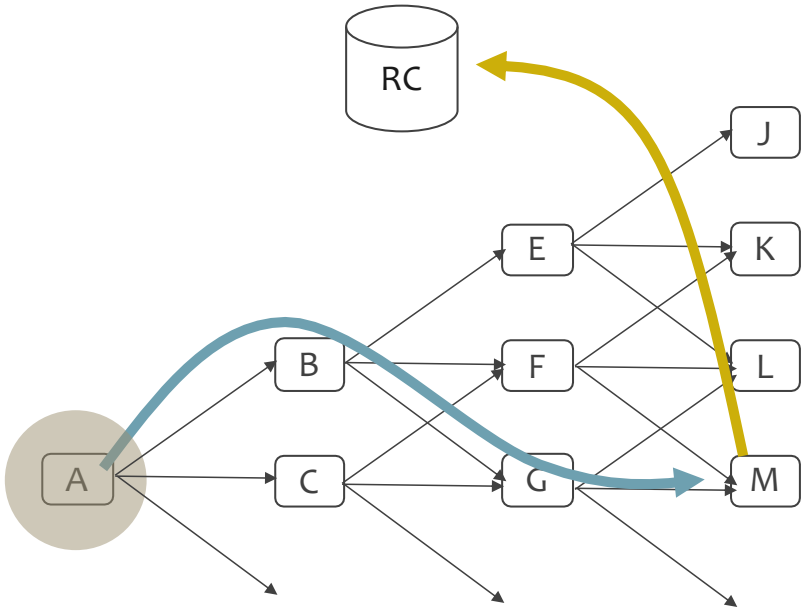
- MCD.OIF: Outgoing Interface ID
  - An 8-bit or 12-bit interface ID associated with the egress physical port of the router
  - Interface IDs are assigned by the operator and not globally unique across the network.
- MCD.OIL: Outgoing Interface Load
  - A 4-bit representation of the egress interface load (i.e., current throughout relative to the port bandwidth).
  - The load is represented using a 4-bit value in logarithmic scale.
- MCD.TTS: Truncated Timestamp
  - An 8-bit timestamp encoding the time at which the packet egress the router
  - The 8-bit TTS has various possible significance depending on the link type. This is known as Time Template, and it is configured by the operator (each egress port in the device is configured with one Time Template).

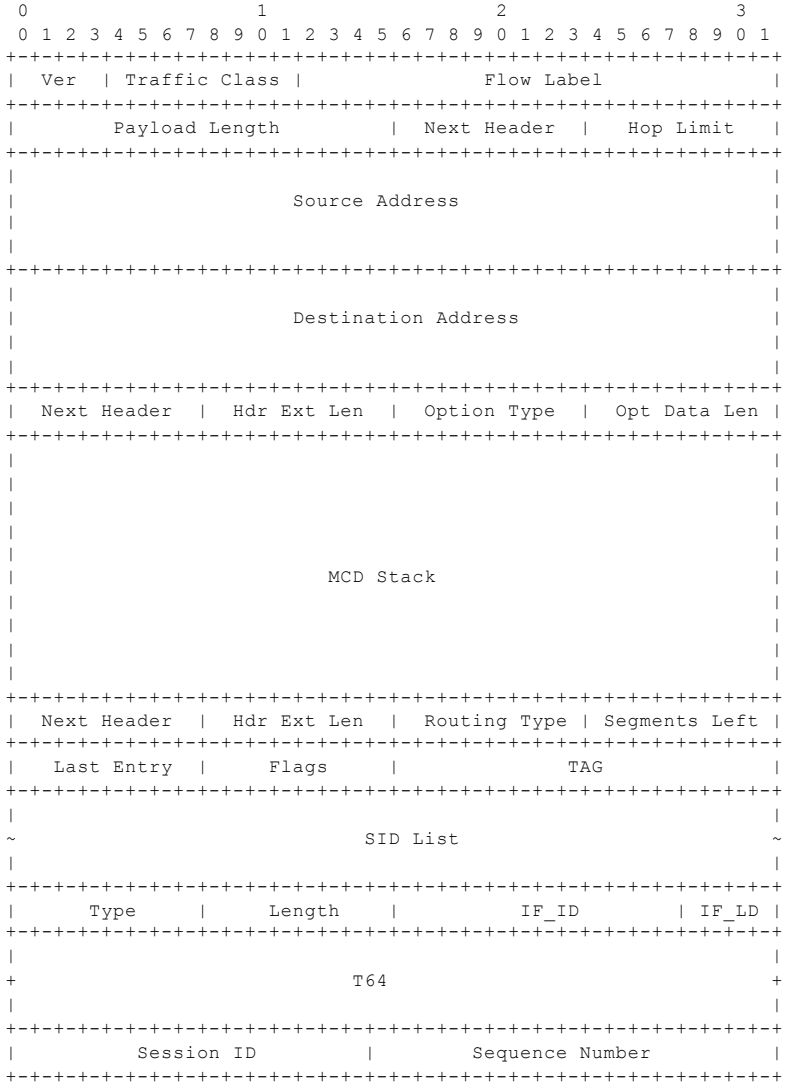
Roles

- PT Source: A node that starts a PT Probing Instance and generates PT probes.
- PT Midpoint: A transit node that performs plain IPv6 forwarding (or SR Endpoint processing) and in addition records its MCD in the HbH-PT.
- PT Sink: A node that receives PT probes sent from the SRC containing the MCDs recorded by every PT Midpoint along the path, and forwards them to a regional collector after recording its PT information.
- RC: Regional collector that receives PT probes, parses, and stores them in TimeSeries Database.

Data Model - Source

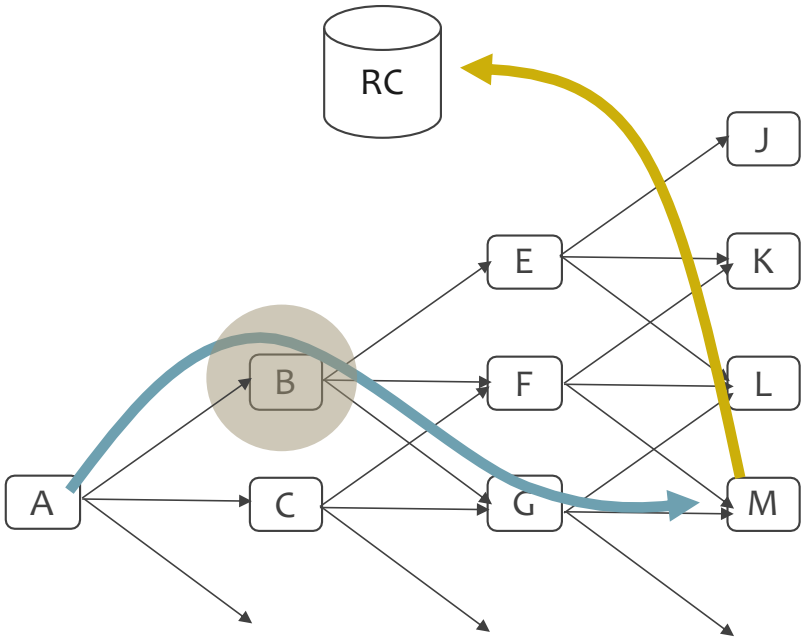
- Originates the probe
- Records transmission data

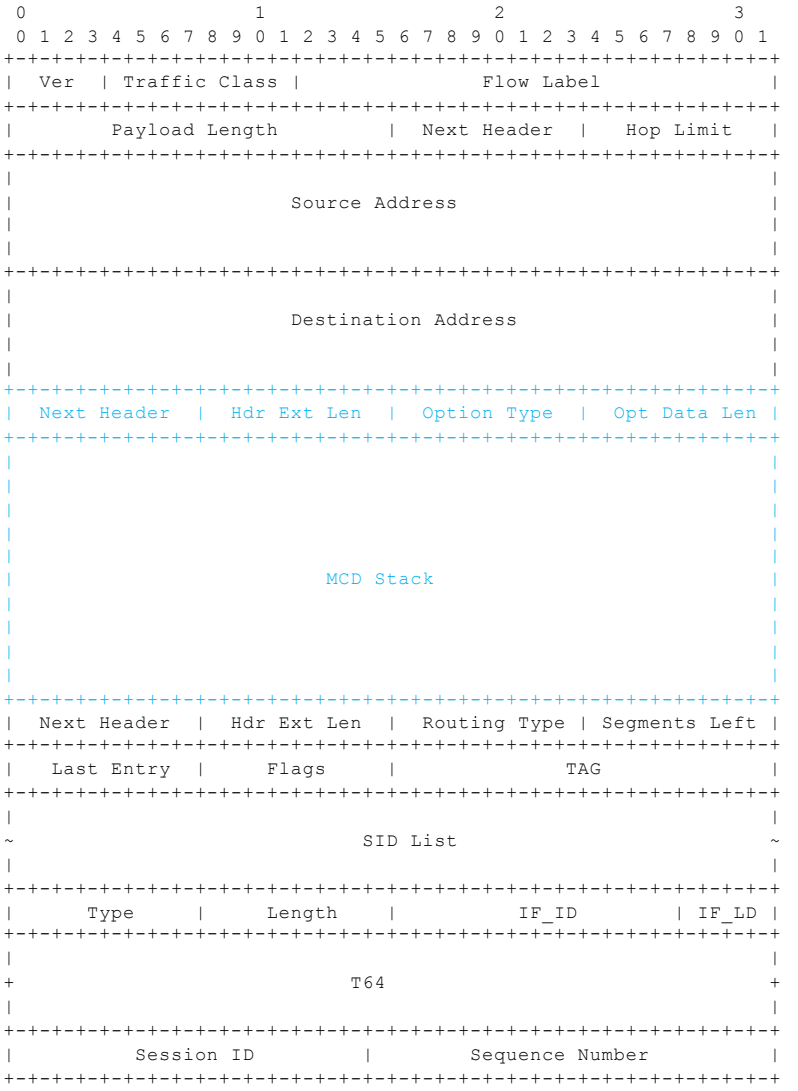




Data Model - Midpoint

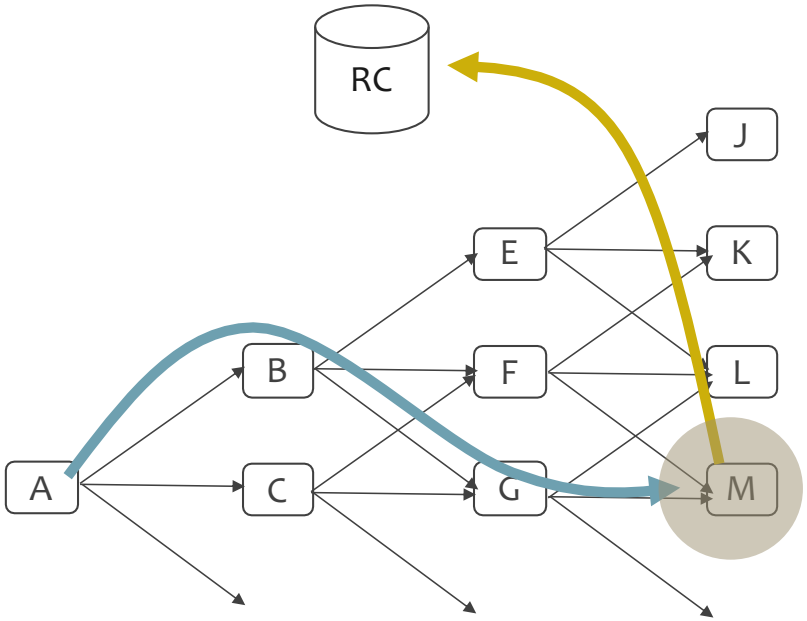
- Records MCD in HbH-PT MCD Stack

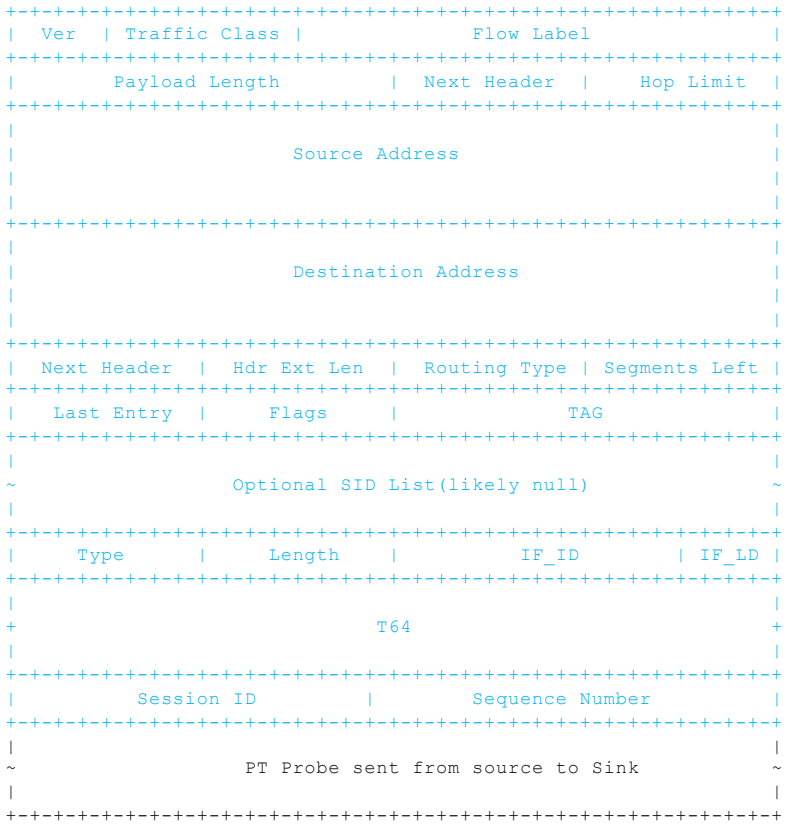




Data Model - Sink

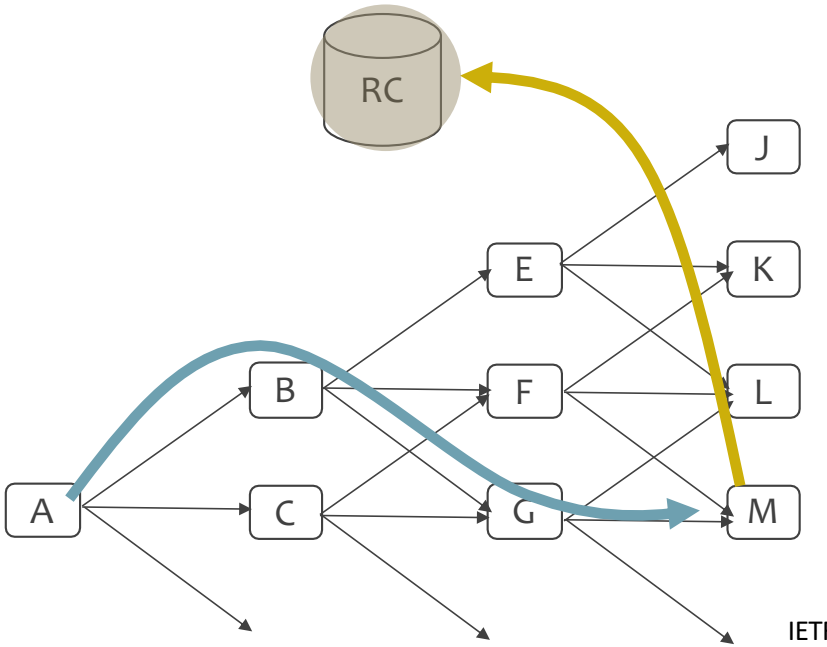
- SRv6 End.B6.TEF
  - Add Outer IPv6 + SRH + SRH PT-TLV
- Records PT data in the SRH PT-TLV





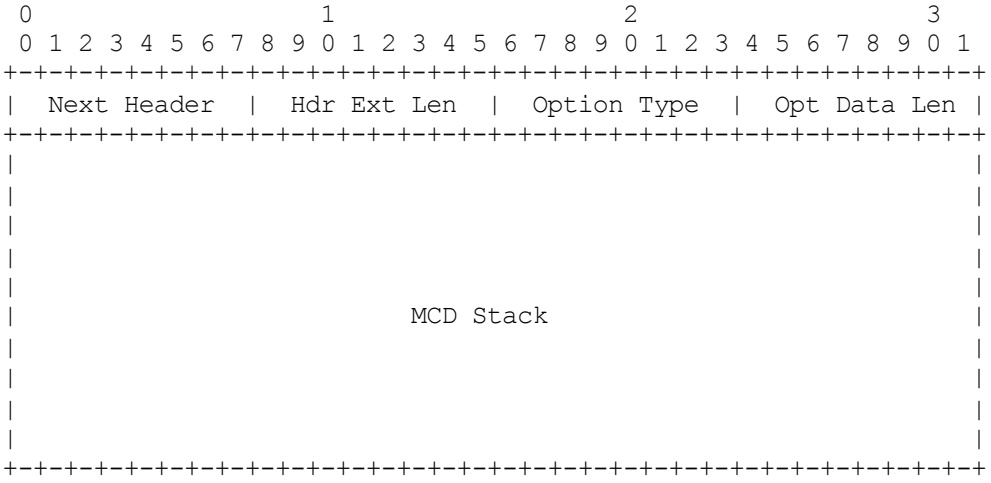
Data Model - RC

- Regional Collector
- Receives PT probes, parses, and stores PT data in TimeSeries Database.



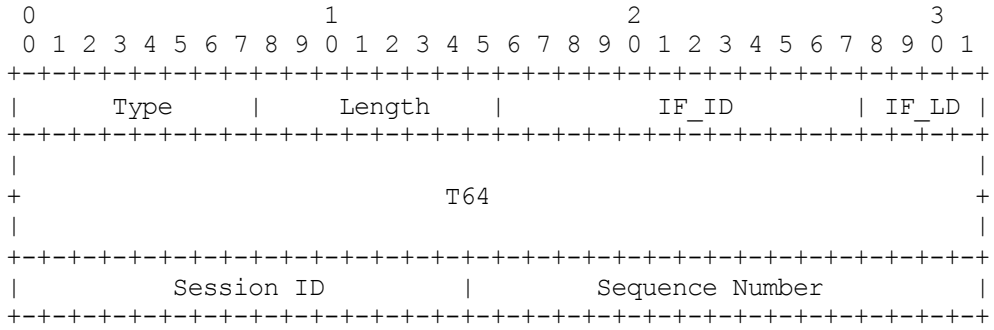
PT Header – HbH PT

- Option Type: TBA
  - The 3 high-order bits of the option must be set to 001
  - 00: Skip HbH for nodes that don't support the HbH-PT Option
  - 1: update HbH-PT for nodes that support the HbH-PT Option
- Opt Data Len: the length of the MCD stack in bytes. Recommended length 36B.
- MCD Stack: used to collects MCDs from Midpoints



PT Header – SRH PT-TLV

- Type: TBA
- Length: 14
- IF_ID: 12-bit Interface ID
- IF_LD: 4-bit Interface Load
- T64: 64-bit PTP Timestamp
- Session ID: Session identifier set by SRC node generating the probes.
  - Used to co-relate probes of the same session.
  - Value of zero means unset.
- Sequence Number: the sequence number of the probe set by SRC node generating the probes.
  - Value of zero means unset.



Security

- Leverages the SR Domain [RFC8754]
  - Any border router drops any external packet destined towards an internal interface with a PT HBH Option
  - Any border router drops any internal packet destined towards an external interface with a PT HBH Option
- The PT HBH option must be processed in the datapath (i.e. at linerate in NPU; Not CPU; Not co-processors). Therefore it cannot be used as an attack vector to the CPU.

Ecosystem

- Experimental and interoperable implementations available on:
  - Cisco 8802 (based Cisco Silicon One Q200)
  - Cisco ASR9904 with Lightspeed linecard
  - Cisco NCS5508 (based on Broadcom Jericho2 platform)
  - Cisco Nexus N3K-C3464C (based on Barefoot Tofino)
  - Marvel Prestera Falcon
- Open-Source implementations available:
  - FD.io VPP, Linux, Wireshark, Tcpdump, BMv2 P4 implementation

Next Steps

- Please review and provide feedback

Thank you