Segment Routing Introduction

Igor Milojevic
Director Transport Development
Telekom Srbija
Overview

• Control Plane: Distribute labels through IGP: ISIS or OSPF
• Labels are segments
• At ingess express path as a list of segments – source routing concept
• Data Plane: ANY (MPLS or IPv6)
• Use existing MPLS hardware and MPLS dataplane
• Congruent with MPLS based services
Segment Routing with IS-IS Routing Protocol

draft-previdi-filsfils-isis-segment-routing-00

Abstract

Segment Routing (SR) enables any node to select any path (explicit or derived from IGPs SPT computations) for each of its traffic classes. The path does not depend on a hop-by-hop signaling technique (neither LDP nor RSVP). It only depends on a set of "segments" that are advertised by the IS-IS routing protocol. These segments act as topological sub-paths that can be combined together to form the desired path.
SR brings...

• Simple – less protocols (no LDP, no RSVP, no T-LDP)

• Scalable – less states in the network, less labels kept in the router, less tunnels to be signaled...

• FRR in place for any topology

• Responsive network - no waiting for signaling of new paths, virtualization
SR Technology

• Forwarding established by IGP
• MPLS or IPv6 dataplane
• Segments are labels
• Standard Push, Swap, Pop actions
• Two types of Segments: Node and Adjacency
• More general: A Segment identifies the prefix
  → A Node Segment is for the “loopback” prefix
Segments

- IGP automatically builds and maintains segments
  - Node: a shortest-path to the related node
  - Adjacency: one-hop through the related adjacency
  - Easy for implementing in ISIS – extensions with sub-TLVs for encoding Segment IDs (SIDs)
Node Segment as a Global Label

- Operator allocates a label block for Segment Routing
  - [64, 5000] is the SR Block
- Operator allocates a label from the SR block to each node
  - Z is given SID 65
  - Z advertises 65, remote nodes install 65 for FEC Z in MPLS dataplane

A packet injected anywhere with top label 65 will reach Z
Node Segment Encoding in ISIS

- Node-SID Sub-TLV present in TLV-135, TLV-235, TLV-236 or TLV-237
- Node-SID sub-TLV:
  - Type, Length, Flags, Node-SID
  - 32 bit SID
  - P flag (PHP flag)
  - E flag (External flag)
  - L flag (Level flag)
- Propagated through levels
Adj Segment as a Local Label

- Node automatically allocates a local label for each adjacency
  - label taken outside the SR block
  - B advertises 9003 for BC adjacency
  - Only B installs the adjacency segment in its MPLS dataplane

A packet entering B with top label 9003 is forced through datalink BC
Adjacency Segment Encoding in ISIS

- Adj-SID Sub-TLV present in TLV-22, TLV-222, TLV-23 or TLV-223

- Adj-SID sub-TLV:
  - Type, Length, Flags, Adj-SID
  - 32 bit SID
  - B flag (Bundle flag)
  - F flag (Forwarding Adjacency flag):
    - if set then the “ERO” sub-TLV must be added expressing the explicit path as a list of SIDs

- NOT propagated through levels
Source Routing

- Path is expressed at the source
- Any path can be expressed as a list of segments
- Nodal segment expresses an ECMP on SPT
- Implicit ECMP behaviour “inherited” from IGP
Routing Scenarios

• Non-ECMP
  – Express path with only Adj segments or
  – An intermediate node can advertise an Adj-SID for an
    FA expressing a desired SR explicit path

• Bundle Adj-SID
  – An additional common Adj-SID for multiple
    adjacencies
  – Used for load balancing on parallel links
FRR

- LFA FRR support
- Any topology
- Directed LFA concept
  - Overcomes the scenario with no PQ node in RLFA
  - Applicable when Q is connected to P
  - Repair tunnel: Node segment to P + Adj segment to Q node
  - Simpler than with T-LDP
  - Improves RLFA when having same metrics in the topology
FRR

Backbone

Node segment to P node

Adj segment to Q node

100
Scaling

• No LDP or RSVP
• No T-LDP
• No LDP/IGP synchronization
  → Less operational effort
• Label database on a router keeps much less labels comparing to LDP deployment
• A core router keeps \( \sim N^2 \) states in RSVP deployment and \( \sim N \) states with SR

\( N = \) number of edge nodes
Use cases – Disjoint planes

A sends traffic with [65]
Classic ecmp “a la IP”

A sends traffic with [11, 65]
Packet gets attracted in blue plane and then uses classic ecmp “a la IP”
Use cases – contd.

• CoS based TE: Define per-flow CoS based policies
  - Based on BW, latency requirements
  - Encode a list of segments to fulfill CoS demands

• OAM
  - Segments make it flexible for OAM tools
  - e.g. path and link probing scenario
Centralized Optimization

- find a path meeting the SLA requirement
- encode it at a list of nodal and adj segments
SDN and Central Optimization

- The architecture leverages BGP-LS, PCEP, Netflow...
- Standard stateful PCE + SR = no signalling delay → faster and more responsive
IPv6 Dataplane - SR IPv6 Address

<table>
<thead>
<tr>
<th>SRB-64bit-Block</th>
<th>Active SID</th>
<th>Flags</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0xCAFFE0123456789A</td>
<td>32bits</td>
<td>8 bits</td>
<td>24bits</td>
</tr>
</tbody>
</table>

- For SR IPv6 dataplane a /64 block is dedicated
  - SR block SRB (e.g. 0xCAFFE0123456789A/64)
- All SR domain nodes are configured with the SR block and thus know what to do with it
- The SID is encoded in the IPv6 destination address and SR nodes switch on SIDs
- Backward compatible
  - Non-SR capable nodes on the path switch the traffic classically
SR Forwarding

<table>
<thead>
<tr>
<th>SRB-64bit-Block</th>
<th>Active SID</th>
<th>Flags</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>0xCAFFE0123456789A</td>
<td>32bits</td>
<td>8 bits</td>
<td>24bits</td>
</tr>
</tbody>
</table>

- Use “SR Extension Header” NH=43, Type=T with segment list and active pointer
- For the incoming packet, the SR node checks whether the DA is in the SR block
- If yes, it forwards based on the 32-bit SID
  - It uses the 24-bit entropy to feed ECMP selection
    - ECMP efficiency is preserved despite the outer tunnel
Tunnel Solution for IPv6 Dataplane

- A packet that requires Source Routing is tunneled from I to E
Outer header + SR Extension Header

IPv6 outer header from IS to ES
SR domain

{v6, TC', FL', PL', NH=43, HL', 0xCAFFE0123456789A-SID1, 0xCAFFE0123456789A-SID1} {t=T, NH=6, ptr, SID1, SID2, SID3} {v6, TC, FL, PL, NH, HL, S, D, payload}
Outer header + SR Extension Header

- Intermediate nodes switch on outer header only
  - A, B, D, G do not process the SR extension header
  - A, B, D, G could be SR-non-capable as well
SR Extension Header processing

- Ingress node pushes the outer and SR headers and set the outer DA address to the first SR hop: SID-C
- Egress node pops the outer and SR headers
Intermediate segment switching routers update:
- the pointer in the SR extension header
- the outer DA address to the next segment
Conclusion – We expect from SR...

• Congruent with MPLS properties and functions and services
• Simple

• Network preserves only segments – application state is not in the network – it is in the cloud and can be manipulated in the cloud!
• Network highly programmable, SDN friendly
Thank You!