SRv6/MPLS Option-BC Service Interworking

draft-zzhang-spring-service-interworking

Jeffrey Zhang, Shraddha Hegde, Krzysztof Szarkowicz
IETF116, Yokohama
Existing Interwork Solutions

• *draft-bonica-spring-srv6-end-dtm*
  • Transport interwork with BGP-LU

• *draft-agrawal-spring-srv6-mpls-interworking*
  • Transport interwork
    • 6oM/Mo6, star topology with core using one data plane and leaves using the other
  • Service Interwork
    • Single-plane PEs (SRv6 PEs don’t even do MPLSoIP/MPLSoUDP)
    • Dual-plane Interwork Nodes
      • Option-A style: Service header lookup in service instances on GW
      • Option-B style: allocate new service SIDs/labels when re-advertising received service routes
        • Per service SID/label FIB state required on the interwork nodes

• If SRv6 PEs can do MPLSoIP/MPLSoUDP, then Option-C works
  • Though this is really MPLS service all the way – not interwork
Option B Example

• SRv6 PE1 advertises 10k service prefixes with 1K Service SIDs
  • E.g., 1k VRF table “labels”, each embedded in the NLRI’s label field
  • All with the same Prefix SID attribute that includes:
    • A SID value
    • LOC/FUNCT/ARG and transposition offset/length
      • This allows the ingress PE to superimpose the “label” to the SID to get the Service SID

• IWN re-advertises to MPLS side with 1k locally allocated service labels
  • Creates 1k MPLS FIB entries to map the locally allocated service labels to individual SRv6 Service SIDs
  • Notice that this is for each SRv6 PE

• The goal is to reduce the 1k FIB entry (per PE) to 1
  • Similar in the other direction
Service Interwork Option-BC: SRv6 → MPLS

- When InterWork Node (IWN) re-advertises service routes from MPLS to SRv6 domain
  - Don’t change the NLRI (i.e. the service label)
  - Add a Prefix SID attribute
    - SRv6 SID Structure Sub-Sub-TLV’ transposition length/offset will direct the receiving SRv6 ingress PE to place the NLRI label into the lower part of FUNC bits
    - The higher FUNC bits indicate a new End.DBS behavior specific to the received BGP nexthop (e.g., the egress/advertising service PE)

- Ingress SRv6 PE
  - Send service traffic with the service SID resulting from superimposing NLRI label to the lower part of the FUNC bits in the SID received from the IWN

- End.DBS behavior on IWN:
  - “Decapsulation, Binding (to a particular MPLS PE), Shifting (part of FUNCT to label stack)”
  - Higher part of FUNC bits map to the DBS behavior for a particular MPLS PE
  - Lower part of FUNC bits become the service label being pushed first
Signaling of service prefix spfx2

<200, spfx2, IWN, PrefixSID>
Prefix SID Attr: SID value & structure

Re-advertise with added Prefix SID; high FUNC bits identify End.DBS for the received NH (PE2)

<PE2 SID, 200, payload>

Traffic for service prefix spfx2

SRv6 PE1  IWN  MPLS PE2

<IWN:hFUNC:200, payload>
IWN:hFUNC is the SID in the Prefix SID, with End.DBS behavior on IWN.
200 is from NLRI.
Note that hFUNC:200 together is the FUNCT

From IWN:hFUNC:, IWN finds base tunnel or Node SID for egress PE2, extract service label from lower FUNC and push to stack

<200, spfx2, PE2>
Service Interwork Option-BC: MPLS ➔ SRv6

• When re-advertising service route from SRv6 to MPLS domain:
  • Don’t change the NLRI (i.e. the service label)
  • Use a next hop address that maps to the SID in the Prefix SID attribute
    • If MPLS domain is IPv6, this can be the SID itself in Prefix SID attribute
    • If the NLRI label field is 24-bit then the extra 4-bit is appended to the SID

• For the above-mentioned next hop address, a transport/underlay route is advertised via BGP-LU with a distinct IW label
  • Ingress MPLS PE sends service traffic with the <IWN SID, IW label, service label in NLRI>

• IWN behavior
  • For an incoming IW label, find the corresponding SRv6 SID, superimpose the next (service) label in stack to the SID, and send traffic out with the resulting SID
Signaling of service prefix spfx1

Based on IW label 300, IWN finds corresponding SID advertised by PE1, superimpose the service label 100 (next in stack) and use the resulting Service SID as destination address of outer encapsulation.

Re-advertise with IWNH mapped to SID value in the received Prefix SID; Advertise IWNH with a IW label (e.g., 300) via BGP-LU

Traffic for service prefix spfx1
Pros and Cons

• Completely independent SRv6/MPLS domains
  • One side SRv6 and the other side can be MPLS – IPv4/IPv6, SR or not
  • Incremental transition (domain by domain)
  • Option-B advantage

• No per-service SID/label state on the interwork node
  • Option-C advantage

• Works with EVPN label-based multi-homing split-horizon
  • Details in draft

• Interwork node needs programmable or new ASIC
  • Extract lower FUNCT bits and push as label for SRv6→MPLS traffic
  • Pop next label and superimpose to SRv6 SID for MPLS→SRv6 traffic
Next Steps

• Comments and suggestions appreciated
• Figure out a plan to move forward