Workgroup: SPRING
Internet-Draft:
draft-agrawal-spring-srv6-mpls-interworking-14
Published: 4 March 2024
Intended Status: Standards Track
Expires: 5 September 2024
Authors: S. Agrawal, Ed. | C. Filsfils | D. Voyer
Cisco Systems | Cisco Systems | Bell Canada
G. Dawra | Z. Li | S. Hegde
LinkedIn | Huawei Technologies | Juniper Networks

SRv6 and MPLS interworking

Abstract

This document describes SRv6 and MPLS/SR-MPLS interworking and co-existence procedures.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The incremental deployment of SRv6 into existing networks require SRv6 to interwork and co-exist with SR-MPLS/MPLS. This document introduces interworking scenarios and building blocks for solutions to inter connect them.

This document assumes SR-MPLS-IPv4 for MPLS domains but the design equally works for SR-MPLS-IPv6, LDP-IPv4/IPv6 and RSVP-TE-MPLS label binding protocols.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Interworking(IW) scenarios

A multi-domain network (Figure 1) can be generalized as a central domain C with many leaf domains around it. Specifically, the document looks at a service flow from an ingress PE in an ingress leaf domain (LI), through the C domain and up to an egress PE of the egress leaf domain (LE). Each domain runs its own IGP instance. A domain has a single data plane type applicable both for its overlay and its underlay.

```
          +-----+                +-----+  RD:V/v via 10   +-----+
   .......|S-RR1|<...............|S-RR2|<.................|S-RR3| <..
   :      +-----+                +-----+                  +-----+    :
   :                                                                 :
   :                                                                 :
+--:-------------------+----------------------+----------------------:-+
|  :      | 2 |        |        | 5 |         |          | 8 |       : |
|  :      +---+        |        +---+         |          +---+       : |
|  :                   |                      |                      : |
|  :                   |                      |                      : |
|  :                   |                      |                      : |
|----+    IGP1       +---+        IGP2      +---+       IGP3      +----|
| 1  |               | 4 |                  | 7 |                 | 10 |
|----+               +---+                  +---+                 +----|
|                      |                      |                        |
|                      |                      |                        |
|                      |                      |                        |
|         +---+        |        +---+         |          +---+         |
|         | 3 |        |        | 6 |         |          | 9 |         |
+----------------------+----------------------+------------------------+
iPE                   iBR                    eBR                      ePE

<----------LI---------><----------C----------><------------LE---------->
```

Figure 1: Reference multi-domain network topology

There are various SRv6 and SR-MPLS-IPv4 interworking scenarios possible.

Below scenarios cover various cascading of SRv6 and MPLS networks, e.g., SR-MPLS-IPv4 <-> SRv6 <-> SR-MPLS-IPv4 <-> SRv6 <-> SR-MPLS-IPv4 etc, though not all combinations are described for brevity.

2.1. IW scenarios

2.1.1. Transport IW

Provider edge devices run MPLS based [RFC4364] or SRv6 Service SID based [RFC9252] BGP L3(e.g.VPN) or L2(e.g.EVPN) services through service Route Reflectors. Service endpoint signaling through borders routers and corresponding forwarding state provide interworking over intermediate transport domain.

* SRv6 over MPLS (6oM)

  - LI and LE domains are SRv6 data plane, C is MPLS data plane

  - L3/L2 BGP SRv6 services [RFC9252] extend between PEs. The ingress PE encapsulates the payload in an outer IPv6 header where the SRv6 Service SID is the last segment or destination address(DA).

  - Transport IW border nodes forward SRv6 encapsulated traffic destined to egress PE over MPLS C domain.

* MPLS over SRv6 (Mo6)

  - LI and LE domains are MPLS data plane, C is SRv6 data plane

  - L3/L2 BGP MPLS services [RFC4364], [RFC7432]. The ingress PE encapsulates the payload in an MPLS service label and sends it through MPLS LSP to egress PE.

  - Transport IW nodes forward encapsulated label stack to egress PE over SRv6 C domain.

Note: Easiest and most probable deployment is ships in the night i.e. supporting dual stack and IPv4 MPLS in each domain.

2.1.2. Service IW

L3/L2 service signaling discontinuity i.e. SRv6 service SID based PE interworks with BGP MPLS based PE for service connectivity. L3/L2 service BGP signaling and forwarding state provide interworking over intermediate domain.

* SRv6 to MPLS(6toM): The ingress PE encapsulates the payload in an outer IPv6 header where the destination address is the SRv6 Service SID[RFC9252]. Payload is delivered to egress PE with MPLS service label[RFC4364] that it advertised with service prefixes.

* MPLS to SRv6 (Mto6): The ingress PE encapsulates the payload in an MPLS service label. Payload is delivered to egress PE with IPv6 header with destination address as SRv6 service SID that it advertised with service prefixes.

3. Terminology

The following terms used within this document are defined in [RFC8402]: Segment Routing, SR-MPLS, SRv6, SR Domain, Segment ID (SID), SRv6 SID, Prefix-SID.

Domain: Without loss of the generality, domain is assumed to be instantiated by a single IGP instance or a network within IGP if there is clear separation of data plane.

Node k has a classic IPv6 loopback address Ak::1/128.

A SID at node k with locator block B and function F is represented by B:k:F::

A SID list is represented as <S1, S2, S3> where S1 is the first SID to visit, S2 is the second SID to visit and S3 is the last SID to visit along the SR path.

(SA,DA) (S3, S2, S1; SL) represents an IPv6 packet with:

IPv6 header with source address SA, destination addresses DA and SRH as next-header

SRH with SID list <S1, S2, S3> with SegmentsLeft = SL

Note the difference between the <> and () symbols: <S1, S2, S3> represents a SID list where S1 is the first SID and S3 is the last SID to traverse. (S3, S2, S1; SL) represents the same SID list but encoded in the SRH format where the rightmost SID in the SRH is the first SID and the leftmost SID in the SRH is the last SID. When referring to an SR policy in a high-level use-case, it is simpler to use the <S1, S2, S3> notation. When referring to an illustration of the detailed packet behavior, the (S3, S2, S1; SL) notation is more convenient.

4. SRv6 SID behavior

This document introduces a new SRv6 SID behavior. This behavior is executed on border routers between the SRv6 and MPLS domain.

4.1. End.DTM

The "Endpoint with decapsulation and MPLS table lookup" behavior.

The End.DTM SID MUST be the last segment in a SR Policy, and a SID instance is associated with an MPLS table.

When N receives a packet destined to S and S is a local End.DTM SID, N does:

```
S01. When an SRH is processed {
S02.   If (Segments Left != 0) {
S03.      Send an ICMP Parameter Problem to the Source Address,
          Code 0 (Erroneous header field encountered),
          Pointer set to the Segments Left field,
          interrupt packet processing and discard the packet.
S04.   }
S05.   Proceed to process the next header in the packet
S06. }
```

When processing the Upper-layer header of a packet matching a FIB entry locally instantiated as an End.DTM SID, N does:

```
S01. If (Upper-Layer Header type == 137(MPLS) ) {
S02.    Remove the outer IPv6 Header with all its extension headers
S03.    Set the packet's associated FIB table to T
S04.    Submit the packet to the MPLS FIB lookup for
        transmission according to the lookup result.
S05. } Else {
S06.    Process as per [ietf-spring-srv6-network-programming] section 4.
S07. }
```

4.2. End.DPM

The "Endpoint with decapsulation and MPLS label push" behavior.

The End.DPM SID MUST be the last segment and a SID instance is associated with label stack.

When N receives a packet destined to S and S is a local End.DPM SID, N does:

```
S01. When an SRH is processed {
S02.   If (Segments Left != 0) {
S03.      Send an ICMP Parameter Problem to the Source Address,
          Code 0 (Erroneous header field encountered),
          Pointer set to the Segments Left field,
          interrupt packet processing and discard the packet.
S04.   }
S05.   Proceed to process the next header in the packet
S06. }
```

When processing the Upper-layer header of a packet matching a FIB entry locally instantiated as an End.DPM SID, N does:

```
S01. Remove the outer IPv6 Header with all its extension headers
S02. Push the MPLS label stack associated with S
S03. Submit the packet to the MPLS engine for transmission
```

5. SRv6 Policy Headend Behaviors

5.1. H.Encaps.M: H.Encaps applied to MPLS label stack

The H.Encaps.M behavior encapsulates a received MPLS Label stack [RFC3032] packet in an IPv6 header with an SRH. Together MPLS label stack and its payload becomes the payload of the new IPv6 packet. The Next Header field of the SRH MUST be set to 137 [RFC4023].

5.2. H.Encaps.M.Red: H.Encaps.Red applied to MPLS label stack

The H.Encaps.M.Red behavior is an optimization of the H.Encaps.M behavior. H.Encaps.M.Red reduces the length of the SRH by excluding the first SID in the SRH of the pushed IPv6 header. The first SID is only placed in the Destination Address field of the pushed IPv6 header. The push of the SRH MAY be omitted when the SRv6 Policy only contains one segment and there is no need to use any flag, tag or TLV. In such case, the Next Header field of the IPv6 header MUST be set to 137 [RFC4023].

6. Interconnecting Binding SIDs

Binding Segment (BSID) is bound to SR policy [RFC8402]. Further an SR-MPLS label can be bound to an SRv6 Policy and an SRv6 SID can be bound to an SR-MPLS Policy. The IW SR-PCE solution Section 7.1.1 leverage these BSIDs as segments of SR policy on headend domain to represent intermediate domain of different dataplane type. In summary, an intermediate domain of different data plane type is represented by BSID of ingress domain data plane type in SID list.

7. Interworking Procedures

Figure 1 shows reference multi-domain network topology and Section 2 its description. The procedure in this section are illustrated using the topology.

Following is assumed for data plane support of various nodes:

* Nodes 2,3,5,6,8,9 are provider(P) routers which need to support single data plane type.

* 1 and 10 are PEs. They support single data plane type in overlay and underlay.

* Border routers 4 and 7 need to support both the SRv6 and SR-MPLS-IPv4 data plane.

A VPN route is advertised via service RRs (S-RR) between an egress PE(node 10) and an ingress PE (node 1).

For illustrations, the SRGB range starts from 16000 and prefix SID of a node is 16000 plus node number

7.1. Transport IW

As described in Section 2.1.1, transport IW requires:

* For 6oM, tunnel traffic destined to SRv6 Service SID of egress PE over MPLS C domain.

* For Mo6, tunnel MPLS label stack bound to IPv4 loopback address of egress PE over SRv6 C domain.

This draft enhances two well-known solutions to achieve above:

* An SR-PCE [RFC8664] multi-domain On Demand Next-hop (ODN) SR policy [RFC9256] stitching end to end across different data plane domains using interconnecting binding SIDs. These procedures can be used when overlay prefixes are signaled with a color extended community [RFC9012].

* BGP Inter-Domain routing procedures advertising PE locator or IPv4 Loopback address for best effort or intent aware end to end connectivity.

7.1.1. SR-PCE multi-domain On Demand Nexthop

This procedure provides a best-effort path as well as a path that satisfies the intent (e.g. low latency), across multiple domains. Service routes (VPN/EVPN) are received on ingress PE with color extended community from egress PE. A Color is a 32-bit numerical value that associates an SR Policy with an intent [RFC9256]. Ingress PE does not know how to compute the traffic engineered path through the multi-domain network to egress PE and requests SR-PCE for it. The SR-PCE is aware of interworking requirement at border nodes as its fed with BGP-LS topological information from each domain. It programs intermediate domain data plane specific policy on border nodes for the given intent and represents it in end to end path SID list on ingress PE leveraging Section 6.

Below sections describe 6oM and Mo6 IW with SR-PCE

7.1.1.1. 6oM

Service prefix (e.g. VPN or EVPN) is received on head-end (node 1) with color extended community (C1) from egress PE (node 10) with SRv6 service SID. The PCE computes (C1,10) path via node 2, 5 and 8. It programs an SR policy at border node 4 with segment list node 5 and 7 bounded to an End.BM BSID [RFC8986]. SR-PCE responds back to node 1 with SRv6 segments along required SLA including End.BM at node 4 to traverse SR-MPLS-IPv4 C domain.

For example, SR-PCE create SR-MPLS policy (C1,7) at node 4 with segments <16005,16007>. It is bound to End.BM behavior with SRv6 BSID as B:4:BM-C1-7::

The data plane operations for the above-mentioned interworking example are:

* Node 1 performs SRv6 function H.Encaps.Red with VPN service SID and SRv6 Policy (C1,10):

  Packet leaving node 1 IPv6 ((A:1::, B:2:E::) (B:10::DT4, B:8:E::, B:4:BM-C1-7:: ; SL=3))

* Node 2 performs End function

  Packet leaving node 2 IPv6 ((A:1::, B:4:BM-C1-7::) (B:10::DT4, B:8:E::, B:4:BM-C1-7:: ; SL=2))

* Node 4(border rout4er) performs End.BM function

  Packet leaving node 4 MPLS (16005,16007,2)((A:1::, B:8:E::) (B:10::DT4, B:8:E::, B:4:BM-C1-7-:: ; SL=1)).

* Node 7 performs a native IPv6 lookup on due PHP behavior for 16007

  Packet leaving node 7 IPv6 ((A:1::, B:8:E::) (B:10::DT4, B:8:E::, B:4:BM-C1-7:: ; SL=1))

*Node 8 performs End(PSP) function

Packet leaving node 8 IPv6 ((A:1::, B:10::DT4))

*Node 10 performs End.DT function and lookups IP in VRF and send traffic to CE.

7.1.1.2. Mo6

Refer Section 2.1.1 for Mo6 scenario. MPLS Service prefix (e.g. VPN or EVPN) is received on head-end(node 1) with color extended community(C1) from egress PE(node 10). The PCE computes color-aware C1 path via node 2, 5 and 8. It programs a SRv6 policy bound to MPLS BSID at border node 4 with SRv6 segment list along required color-aware path with last segment of behavior End.DTM Section 4.1. SR-PCE responds back to node 1 with MPLS segment list including MPLS BSID of SRv6 policy at node 4 to traverse SRv6 core domain.

For example, SR-PCE create SRv6 policy (C1,7) at node 4 with segments <B:5:E::,B:7:DTM::>. It is bound to MPLS BSID 24407.

The data plan operations for the above-mentioned interworking example are:

1. Node 1 performs MPLS label stack encapsulation with VPN label and SR-MPLS Policy (C1,10):

   Packet leaving node 1 towards 2 (Note: PHP of node 2 prefix SID): MPLS packet (16004,24407,16008,16010,vpn_label)

2. Node 2 forwards traffic towards 4 (PHP of 16004)

   Packet leaving node 2 MPLS packet (24407,16008,16010,vpn_label)

3. Node 4 steers MPLS traffic into SRv6 policy bound to 24407

   Packet leaving node 4 IPv6(A:4::, B:5:E::) (B:7:DTM:: ; SL=1)NH=137) MPLS((16008,16010,vpn_label)

4. Node 7 receive IPv6 packet with DA=B:7:DTM::. It performs DTM behavior to remove IPv6 header and perform 16008 lookup in MPLS table.

   Packet leaves node 7 towards node 8(PHP of 16008) MPLS packet (16010,vpn_label)

5. Node 8 forwards traffic towards 10 (PHP of 16010)

   Packet leaving node 8 MPLS packet (vpn_label)

6. Node 10 performs vpn_label lookup and send traffic to CE.

7.1.2. BGP inter domain routing procedures

Procedures described below build upon BGP 3107 [I-D.ietf-mpls-seamless-mpls] and [RFC4798] to advertise transport reachability for PE IPv4 loopbacks or SRv6 locators across a multi-domain network. The procedures leverage existing SAFIs (for example, BGP-LU(1/4, 2/4)) and IPv6 (2/1)). Nexthop self on border routers provide independence of intra domain tunnel technology in different domains.

The sections below describe 6oM and Mo6 IW with BGP procedures for best effort paths to a locator or loopback prefix. The procedures are equally applicable to intent aware paths, i.e., locator assigned for a given intent, for instance from an IGP-FlexAlgo. They are also applicable to color-aware routes [I-D.ietf-idr-bgp-car] recursing over intent aware intra-domain paths.

7.1.2.1. 6oM

Refer Section 2.1.1 for 6oM scenario. SRv6 based L3/L2 BGP services are signaled with SRv6 Service SID allocated from egress PE locator prefix and with no BGP Color Extended community. Ingress PE learns the service routes and need to resolve SRv6 Service SID over egress PE locator or its summary. Below describes propagation of locators or its summary to create end to end underlay path.

* Egress border router learns LE domain PE locators through IGP. These are redistributed in BGP like any IPv6 global prefixes. Alternatively, locator is originated by egress PE in the BGP IPv6 unicast address family (AFI=2,SAFI=1) to border nodes.

* Egress border router advertise these locators in BGP IPv6 LU [AFI=2/SAFI=4] with locally allocated label or explicit NULL to ingress border router with IPv4 next hop. Next hop has SR MPLS IPv4 LSP paths built in C domain. Note: Egress border router may advertise summary prefix covering all PE locators in LE domain.

* Ingress border router advertise these remote locators in LI domain. Options to advertise are:

  - To ingress PE in BGP address family (AFI=2,SAFI=1) with SRv6 transport SID of local End behavior in Prefix-SID attribute TLV type 5 [RFC9252]. This option will result in additional SRv6 encapsulation at ingress PE.

  - Alternatively, leak remote locators or their summary in LI IGP or propagate using BGP hop by hop routing model to each of P and PE routers through infrastructure route reflector. This option will avoid additional SRv6 transport SID encapsulation.

* Ingress PE learn remote locators or their summary with or without SRv6 SID encapsulation. When learnt with SRv6 SID, it builds the packet encapsulation that contains the SRv6 Service SID and SRv6 transport SID in the SID list and tunnels traffic to ingress border node in LI domain (P routers (node 2 and 3) does not need remote prefix). When learnt without SRv6 SID, the packet encapsulation contains the SRv6 Service SID and forwarded hop by hop based on remote locator IPv6 prefix lookup.

Control plane example:

1. Routing Protocol(RP) @10:

   * In ISIS advertise locator B:10::/48

   * BGP AFI=1,SAFI=128 originates a VPN route RD:V/v via B:10::1 and Prefix-SID attribute B:10:DT4::. This route is advertised to service RR.

2. RP @ 7:

   * ISIS redistribute B:10::/48 into BGP

   * BGP Originates B:10::/48 in AFI=2/SAFI=4 with next hop node 7 and label explicit null among border routers.

3. RP @ 4:

   * BGP learns B:10::/48 with next hop node 7 and outgoing label.

   * BGP advertise B:10::/48 in AFI=2/SAFI=1 with next hop B:4::1 and Prefix-SID attribute tlv type 5 carrying local End behavior function B:4:END:: to node 1

   * Alternatively, BGP redistributes remote locator or summary route in LI domain IGP.

4. RP @ 1:

   * BGP learns B:10::/48 via B:4::1 and Prefix-SID attribute TLV type 5 with SRv6 SID B:4:END::

   * Alternatively, B:10::/48 or summary route reachability is learned through ISIS

*BGP AFI=1, SAFI=128 learn service prefix RD:V/v, next hop B:10::1 and PrefixSID attribute TLV type 5 with SRv6 SID B:10:DT4

FIB state

```
@1: IPv4 VRF V/v => H.Encaps.red <B:4:END::, B:10:DT4::> with SRH, SRH.N
@4: IPv6 Table: B:4:END:: => Update DA with B:10:DT4::,set IPv6.NH=IPv4,
@4: IPv6 Table: B:10::/48 => push MPLS label 2 (Explicit NULL), push MPL
@7: MPLS label 2 => pop and lookup next IPv6 DA
@7: IPv6 Table B:10::/48 => forward via ISIS path to 10
@10: IPv6 Table B:10:DT4:: => pop the outer header and lookup the inner
```

7.1.2.2. Mo6

Refer Section 2.1.1 for Mo6 scenario. MPLS based L3/L2 BGP services are signaled with IPv4 next-hop of egress PE through Service RRs with no color extended community. Ingress PE need labelled reachability to remote PE IPv4 loopback address advertised as next hop with service routes.

BGP LU [RFC8277] advertise IPv4 PE loopbacks. Next hop self is performed on the border routers.

Following are options and protocol extensions to tunnel IPv4 PE loopback LSP through SRv6 C domain

7.1.2.2.1. Tunnel BGP LU LSP across SRv6 C domain

Intuitive solution for an MPLS-minded operator

* Existing BGP-LU label cross-connect on border routers for each PE IPv4 loopback address.

* The lookups at the ingress border router are based on BGP3107 label as usual

* Just the SR-MPLS IGP label to next hop is replaced by an IPv6 tunnel with DA = SRv6 SID associated with DTM behavior in C domain.

* Ingress border router forwarding perform 3107 label swap and H.Encaps.M with DA = SRv6 SID associated with DTM behavior

* Similar to MPLS-over-IP

Existing BGP LU updates between border routers need to signal SRv6 SID associated with DTM behavior. [I-D.agrawal-bess-bgp-srv6-mpls-interworking] proposes "SRv6 tunnel for label route" TLV of the BGP Prefix-SID Attribute to signal SRv6

SID to tunnel MPLS packet with label in NLRI at the top of its label stack through SRv6/IPv6 domain. Below describes the control plane and corresponding FIB state to achieve such tunneling:

Control plane example

1. Routing Protocol(RP) @10:

   * ISIS originates its IPv4 PE loopback with Node SID 16010

   * BGP AFI=1,SAFI=4 originate IPv4 loopback address with next hop node 10 and optionally label index=10 in Label-Index TLV of Prefix-SID attribute.

   * BGP AFI=1, SAFI=128 originates a VPN route RD:V/v next hop node 10. This route is advertised to service RR.

2. RP @ 7:

   * ISIS v6, advertise locator B:7::/48 in C domain

   * BGP learns node 10 IPv4 loopback address with outgoing label. It allocates local label (based on label index if present) and programs label swap to outgoing label and MPLS LSP to next hop.

   * BGP AFI=1, SAFI=4 advertise IPv4 loopback address of node 10 to node 4. NLRI label is set to local label and SRv6 SID B:7:DTM:: carried in SRv6 SID Information Sub-TLV of "SRv6 tunnel for label route" TLV in Prefix-Sid attribute. If received, label index=10 in Label-Index TLV of Prefix-SID attribute is also signaled.

3. RP @ 4:

   * ISIS v4 originates its IPv4 loopback with prefix SID 16004 in LI domain.

   * BGP learns node10 IPv4 loopback address from node 7 with outgoing label. It allocate local label (based on label index if present) and programs label swap and H.Encaps.M.red with IPv6 header destination address as SRv6 SID received in "SRv6 tunnel for label route" TLV of Prefix-Sid attribute i.e. B:7:DTM::.

   * BGP AFI=1, SAFI=4 advertise IPv4 Loopback address of node 10 to node 1. NLRI label is set to local label and do not signal "SRv6 tunnel for label route" TLV in Prefix-SID attribute.

4. RP @ 1:

* BGP learns IPv4 loopback address of node 10 from node 4 with outgoing label. It programs route to push outgoing label and MPLS LSP to next hop i.e. node 4

* BGP AFI=1, SAFI=128 learn service prefix RD:V/v, next hop IPv4 loopback address of node 10 and service label.

Forwarding state at different nodes:

```
@1: IPv4 VRF: V/v => out label=vpn_label, next hop=IPv4 address of node
@1: IPv4 table: IPv4 address of node 10 => out label=16010, next hop=nod
@1: IPv4 table: IPv4 address of node 4 => out label=16004, next hop=inte
@4: MPLS Table: 16010 => out label=16010, H.Encaps.M.red with DA=B:7:DTM
@4: IPv6 table: B:7::/48 => next hop=interface to reach 5
@7: SRv6 My SID table: B:7:DTM:: => decaps IPv6 header and lookup top la
@7: MPLS table: 16010 => out label=16010, next hop=interface to reach 8
@10: MPLS table: vpn label => pop label and lookup the inner IPv4 DA in
```

During transition when MPLS data plane is still enabled in C domain, an ABR that does not understand "SRv6 tunnel for label route" TLV in BGP Prefix-SID Attribute or based on operator configured local policy can continue MPLS encapsulation using label in NLRI and LSP to next hop.

7.1.2.2.2. Label and SRv6 SID translation per BGP LU route

For each PE IPv4 loopback address, existing BGP 3107 label cross-connect on area border router is replaced by label to SRv6 SID cross-connect or vice versa. In effect, it creates a translation between from 3107 label to SRv6 SID at ingress of SRv6 domain and SRv6 SID to 3107 label on egress.

* For each BGP LU route (IPv4 loopback address of PE) received from LE domain on egress border router, allocate SRv6 SID of DPM behavior bound to the PE address. Lookup of SRv6 SID result in decapsulation of IPv6 header and push of BGP LU outgoing label and MPLS LSP to next hop.

* Advertise BGP route to PE address with SRv6 SID to ingress border router.

* Ingress border router allocate local label and advertise to LI domain.

* The lookups at the ingress border router are based on BGP 3107 label as usual. Lookup results SRv6 SID of DPM behavior signaled

by egress border node. Decap BGP3107 label and perform H.Encaps.M with DA = SRv6 SID.

Section 2.2 of [I-D.agrawal-bess-bgp-srv6-mpls-interworking] describes how existing BGP advertisement can signal SRv6 SID associated with DPM behavior from egress to ingress border router.

7.2. Service IW

As described in Section 2.1.2 Service IW need BGP SRv6 based L2/L3 PE interworking with BGP MPLS based L2/L3 PE.

There are a number of different ways of handling this scenario as detailed below.

7.2.1. Gateway Interworking

Gateway is router which supports both BGP SRv6 based L2/L3 services and BGP MPLS based L2/L3 services for a service instance (e.g. L3 VRF, EVPN EVI). It terminates service encapsulation and perform L2/L3 destination lookup in service instance.

* A border router between SRv6 domain and SR-MPLS-IPv4 domain is suitable for Gateway role.

* Transport reachability to SRv6 PE and gateway locators in SRv6 domain or MPLS LSP to PE/gateway IPv4 Loopbacks can be exchanged in IGP or through mechanism detailed in Section 2.1.1.

* Gateway exchange BGP L2/L3 service prefix with SRv6 based Service PEs via set of service RRs. This session will learn/advertise L3/L2 service prefixes with SRv6 service SID in prefix SID attribute [RFC9252].

* Gateway exchange BGP L2/L3 service prefix with MPLS based Service PEs via set of distinct service RRs. This session will learn/advertise L3/L2 service prefixes with service labels [RFC4364] [RFC7432].

* L2/L3 prefix received from a domain is locally installed in service instance and re advertised to other domain with modified service encapsulation information.

* Prefix learned with SRv6 service SID from SRv6 PE is installed in service instance with instruction to perform H.Encaps. It is advertised to MPLS service PE with service label. When gateway receives traffic with service label from MPLS service PE, it perform destination lookup in service instance. Lookup result in instruction to perform H.Encaps with DA being SRv6 Service SID learnt with prefix from SRv6 PE.

*Prefix learned with MPLS service label from MPLS service PE is installed in service instance with instruction to perform service label encapsulation and send to MPLS LSP to nexthop. It is advertised to SRv6 service PE with SRv6 service SID of behavior (e.g. DT4/DT6/DT2U) [RFC8986]. When gateway receives traffic with SRv6 Service SID as DA of IPv6 header from SRv6 service PE, it perform destination lookup in service instance after decaps of IPv6 header. Lookup result in instruction to push service label and send it to nexthop.

```
+-------------------+                            +-------------
|   ....|S-RR|....  |                            |  ....|S-RR|.
|   :   +----+   :  |                            |  :   +----+
|   :            :  |                            |  :
|----+          +--------------------------------------+
|PE1 |          |             IW border node           |
|----+          | VPN Label<->L2/L3 lookup<->SRv6 SID  |
|               |                                      |
|               +--------------------------------------+
|      MPLS         |                            |       SRv6
+-------------------+                            +-------------
<------MPLS VPN----->                            <------SRv6 VP
```

Figure 2: Gateway IW

Couple of border routers can act as gateway for redundancy. It can scale horizontally by distributing service instance among them.

7.2.2. Translation between Service labels and SRv6 service SIDs

This is similar to inter-as option B procedures described in [RFC4364] just that service label cross-connect on border router is replaced with service label to SRv6 service SID or vice verse translation on IW node.

*IW node does not need service instance like VRF or EVI.

*IW node exchange BGP L2/L3 service prefix with SRv6 based Service PEs via set of service RRs. This BGP session will learn/advertise L3/L2 service prefixes with SRv6 service SID in prefix SID attribute [RFC9252].

*IW node exchange BGP L2/L3 service prefix with MPLS based service PEs via set of distinct service RRs. This BGP session will learn/advertise L3/L2 service prefixes with service labels [RFC4364] [RFC7432].

* IW node allocates SRv6 SID of behavior End.DPM that result in pushing service label and MPLS label stack to service nexthop for BGP L2/L3 service learnt from MPLS PE. It advertises the service to SRv6 domain.

* IW node allocates service label that results in H.Encaps with IPv6 header DA set to SRv6 SID signaled in BGP L2/L3 service learnt from SRv6 PE. Advertises the service to MPLS domain with allocated service label.

```
+-------------------+                             +-------------
|   ....|S-RR|....  |                             |  ....|S-RR|.
|   :   +----+   :  |                             |  :   +-----+
|   :            :  |                             |  :
|----+           +-------------------------------------+
|PE1 |           |             IW border node          |
|----+           |          VPN Label -> SRv6 SID      |
|                | VPN label, LSP PE1 <- SRv6 SID      |
|                +-------------------------------------+
|      MPLS         |                             |       SRv6
+-------------------+                             +-------------
<------MPLS VPN----->                             <------SRv6 VP
```

Figure 3: Service translation

Certain L2 service specific information (eg. control word) translation is out of the scope. It will be covered in separate document.

8. Migration and co-existence

In addition, the draft also addresses migration and coexistence of the SRv6 and SR-MPLS-IPv4. Co-existence means a network that supports both SRv6 and MPLS in a given domain. This may be a transient state when brownfield SR-MPLS-IPv4 network upgrades to SRv6 (migration) or permanent state when some devices are not capable of SRv6 but supports native IPv6 and SR-MPLS-IPv4.

These procedures would be detailed in a future revision

9. Availability

* Failure within domain are taken care by existing FRR mechanisms [I-D.ietf-rtgwg-segment-routing-ti-lfa].

* Procedures listed in [RFC9256] provides protection in SR-PCE multi-domain On Demand Nexthop (ODN) SR policy based approach.

*Convergence on failure of border routers can be achieved by well known methods for BGP inter domain routing approach:

- -BGP Add Path provide diverse path visibility
- -BGP backup path pre-programming
- -Sub-second convergence on border router failure notified by local IGP.

10. IANA Considerations

10.1. SRv6 Endpoint Behaviors

This document introduces a new SRv6 Endpoint behaviors "End.DTM" and "End.DPM". IANA is requested to assign identifier value in the "SRv6 Endpoint Behaviors" sub-registry under "Segment Routing Parameters" registry.

Value	Hex	Endpoint behavior	Reference
TBD	TBD	End.DTM	<this document>
TBD	TBD	End.DPM	<this document>

11. Security Considerations

12. Contributors

Zafar Ali
Cisco Systems
Email: zali@cisco.com

Srihari Sangli
Juniper Networks
Email: ssangli@juniper.net

13. Acknowledgements

The authors would like to acknowledge Kamran Raza, Dhananjaya Rao, Stephane Litkowski, Pablo Camarillo, Ketan Talaulikar, Jorge Rabadan for their comments and review.

14. References

14.1. Normative References

[I-D.agrawal-bess-bgp-srv6-mpls-interworking]

Agrawal, S., Rao, D., Ali, Z., Filsfils, C., Voyer, D., and Z. Li, "BGP extensions for SRv6 and MPLS interworking", Work in Progress, Internet-Draft, draft-agrawal-bess-bgp-srv6-mpls-interworking-00, 24 October 2022, <https://datatracker.ietf.org/doc/html/draft-agrawal-bess-bgp-srv6-mpls-interworking-00>.

[I-D.ietf-idr-bgp-car] Rao, D., Agrawal, S., and Co-authors, "BGP Color-Aware Routing (CAR)", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-car-06, 2 February 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-car-06>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, DOI 10.17487/RFC3032, January 2001, <https://www.rfc-editor.org/info/rfc3032>.

[RFC4023] Worster, T., Rekhter, Y., and E. Rosen, Ed., "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", RFC 4023, DOI 10.17487/RFC4023, March 2005, <https://www.rfc-editor.org/info/rfc4023>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[RFC4798] De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur, "Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)", RFC 4798, DOI 10.17487/RFC4798, February 2007, <https://www.rfc-editor.org/info/rfc4798>.

[RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8277] Rosen, E., "Using BGP to Bind MPLS Labels to Address Prefixes", RFC 8277, DOI 10.17487/RFC8277, October 2017, <https://www.rfc-editor.org/info/rfc8277>.

[RFC8402]
Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Hendrickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/info/rfc8986>.

[RFC9252] Dawra, G., Ed., Talaulikar, K., Ed., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "BGP Overlay Services Based on Segment Routing over IPv6 (SRv6)", RFC 9252, DOI 10.17487/RFC9252, July 2022, <https://www.rfc-editor.org/info/rfc9252>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

14.2. Informative References

[I-D.ietf-mpls-seamless-mpls]
Leymann, N., Decraene, B., Filsfils, C., Konstantynowicz, M., and D. Steinberg, "Seamless MPLS Architecture", Work in Progress, Internet-Draft, draft-ietf-mpls-seamless-mpls-07, 28 June 2014, <https://datatracker.ietf.org/doc/html/draft-ietf-mpls-seamless-mpls-07>.

[I-D.ietf-rtgwg-segment-routing-ti-lfa]
Bashandy, A., Litkowski, S., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", Work in Progress, Internet-Draft, draft-ietf-rtgwg-segment-routing-ti-lfa-13, 16 January 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-rtgwg-segment-routing-ti-lfa-13>.

[RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <https://www.rfc-editor.org/info/rfc9012>.

Authors' Addresses

Swadesh Agrawal (editor)
Cisco Systems

Email: swaagraw@cisco.com

Clarence Filsfils
Cisco Systems

Email: cfilsfil@cisco.com

Daniel Voyer
Bell Canada
Canada

Email: daniel.voyer@bell.ca

Gaurav dawra
LinkedIn
United States of America

Email: gdawra.ietf@gmail.com

Zhenbin Li
Huawei Technologies
China

Email: lizhenbin@huawei.com

Shraddha Hegde
Juniper Networks

Email: shraddha@juniper.net