SPRING
Internet-Draft
Intended status: Standards Track
Expires: 11 August 2026

S. Agrawal, Ed.
C. Filsfils
Cisco Systems
D. Voyer
Bell Canada
G. Dawra
LinkedIn
Z. Li
Huawei Technologies
S. Hegde
Juniper Networks
7 February 2026

SRv6 and MPLS interworking
draft-ietf-spring-srv6-mpls-interworking-02

Abstract

This document describes interworking between SRv6 and MPLS domains to provide end to end path. Interworking problem is generalized into various interworking scenarios. These scenarios are stitched either by transport interworking or service interworking. New SRv6 SID endpoint behaviors are defined for the purpose. These new SRv6 SID behaviors and MPLS labels stitch end to end path across different data plane.

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.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

Copyright Notice

Table of Contents

1. Introduction

The incremental deployment of SRv6 into existing networks require SRv6 to interwork and co-exist with MPLS. This document introduces interworking scenarios and building blocks for solution to interconnect them.

This document assumes SR-MPLS-IPv4 for MPLS domain 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. Generally, a domain has a single data plane type applicable both for overlay and underlay.

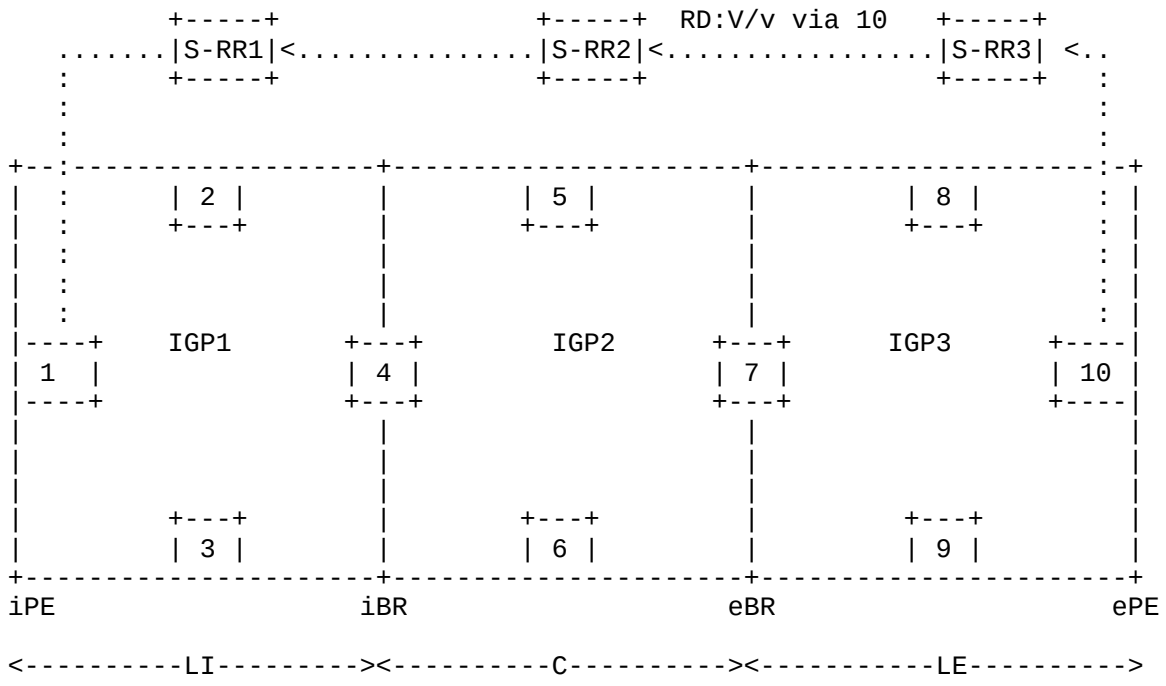

Figure 1: Reference multi-domain network topology

There are various SRv6 and MPLS 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. Transport IW

Provider Edge (PE) nodes deploy either MPLS-based [RFC4364] or SRv6 Service SID-based [RFC9252] BGP services (L3VPN, EVPN, GRT) through service Route Reflectors. Service endpoint (PE loopback address for MPLS or locator for SRv6) signaling through border nodes and corresponding forwarding state provide interworking over intermediate transport domains.

* SRv6 over MPLS (6oM)

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

  - L3/L2 BGP SRv6 service [RFC9252] extend between PEs. The ingress PE encapsulates the service traffic in an outer IPv6 header where the SRv6 Service SID is the last segment.

- 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 service ([RFC4364], [RFC7432]) extend between PEs. The ingress PE encapsulates the service traffic in an MPLS service label and tunnel 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.

2.2. Service IW

BGP L2/L3 service encapsulation interworking between SRv6 SID-based and MPLS-based PEs for service connectivity across domains of different data planes. BGP L2/L3 service route encapsulation type change and corresponding forwarding state at border node provide interworking between PEs.

* SRv6 to MPLS(6toM): The ingress PE encapsulates the service traffic in an outer IPv6 header where the destination address is the SRv6 Service SID[RFC9252]. Service traffic reaches egress PE with an MPLS encapsulation where bottom most label is a service label [RFC4364] that PE advertised with the service prefix.

* MPLS to SRv6 (Mto6): The ingress PE encapsulates the service traffic in an MPLS encapsulation where bottom most label is a service label. Service traffic reaches egress PE with IPv6 encapsulation where IPv6 destination address is a SRv6 service SID that PE advertised with the service prefix.

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 behaviors. These behaviors are executed on border node between the SRv6 and MPLS domain.

4.1. End.DTM

The "Endpoint with decapsulation and lookup in MPLS table" 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 MPLS table T
S04.    Submit the packet to the MPLS FIB lookup for
        transmission according to the lookup result.
S05. } Else {
S06.    Process as per RFC8986 section 4.1.1
S07. }
```

4.2. End.DTM46

The "Endpoint with decapsulation and lookup in MPLS table or Global IP table" behavior.

The End.DTM46 SID MUST be the last segment in a SR Policy, and a SID instance is associated with the MPLS table, the Global IPv4 FIB table and the Global IPv6 FIB table. This behavior is superset of End.DTM and the procedures defined in the document using End.DTM works with End.DTM46 as well.

When N receives a packet destined to S and S is a local End.DTM46 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.DTM46 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 table to MPLS table
S04.    Submit the packet to the MPLS FIB lookup for
            transmission according to the lookup result.
S05. } Else if (Upper-Layer header type == 4(IPv4) ) {
S06.    Remove the outer IPv6 header with all its extension headers
S07.    Set the packet’s associated FIB table to Global IPv4 table
S08.    Submit the packet to the IPv4 FIB lookup for
              transmission to the new destination
S09. } Else if (Upper-Layer header type == 41(IPv6) ) {
S10.    Remove the outer IPv6 header with all its extension headers
S11.    Set the packet’s associated FIB table to Global IPv6 table
S12     Submit the packet to the IPv6 FIB lookup for
              transmission to the new destination
S13  } Else {
S14.    Process as per RFC8986 section 4.1.1
S15. }
```

4.3. DXM behaviors

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

DXM behavior maps or translates SRv6 SID to MPLS label stack that operates similar to label cross-connect functionally at border node. DXM SID MUST be the last segment.

The behavior is associated to the FEC [RFC3031] and determine the expected Upper-layer header type in the IPv6 header. Different DXM variants are specified for various FECs (for example L3 IPv4, L3 IPv6, L2 service).

4.3.1. End.DXM4

SRv6 SID of End.DXM4 behavior is allocated for IPv4 layer 3 VPN/EVPN service label switch path such as signalled by BGP AFI=1 and SAFI=128 [RFC4364]. End.DXM4 behavior cross-connects virtual layer 3 IPv4 payload. Upper-Layer Header type is 4 as payload is IPv4 packet.

When N receives a packet destined to S and S is a local End.DXM4 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.DXM4 SID, N does:

```
S01. If (Upper-Layer Header type == 4(IPv4)) {
S02. Remove the outer IPv6 Header with all its extension headers
S03. Push the MPLS label stack associated with S
S04. and forward the resulting packet to next node in LSP path.
S05.  } Else {
S06.    Process as per RFC8986 section 4.1.1
S07. }
```

4.3.2. End.DXM6

SRv6 SID of End.DXM6 behavior is allocated for IPv6 layer 3 VPN/EVPN service label switch path such as signalled by BGP AFI=2 and SAFI=128 [RFC4364]. End.DXM6 behavior cross-connects virtual layer 3 IPv6 payload. Upper-Layer Header type is 41 as payload is IPv6 packet.

When N receives a packet destined to S and S is a local End.DXM6 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.DXM6 SID, N does:

```
S01. If (Upper-Layer Header type == 41(IPv6)) {
S02. Remove the outer IPv6 Header with all its extension headers
S03. Push the MPLS label stack associated with S
S04. and forward the resulting packet to next node in LSP path.
S05.  } Else {
S06.    Process as per RFC8986 section 4.1.1
S07. }
```

4.3.3. End.DXM2

SRv6 SID of End.DXM2 behavior is allocated for layer 2 virtual service label switch path such as specified in [RFC7432]. End.DXM2 behavior cross-connects virtual layer 2 payload. Upper-Layer Header type is 143 as payload is Ethernet.

When N receives a packet destined to S and S is a local End.DXM2 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.DXM2 SID, N does:

```
S01. If (Upper-Layer Header type == 143(ethernet)) {
S02. Remove the outer IPv6 Header with all its extension headers
S03. Push the MPLS label stack associated with S
S04. and forward the resulting packet to next node in LSP path.
S05.  } Else {
S06.    Process as per RFC8986 section 4.1.1
S07. }
```

5. SRv6 Policy Headend Behaviors

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

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

6. Interconnecting Binding SIDs

Binding Segment (BSID) is bound to an SR policy [RFC8402] and provides domain opacity. Opacity fits well for interworking because an SR-MPLS BSID label can be bound to an SRv6 Policy and an SRv6 BSID can be bound to an SR-MPLS Policy. Such BSIDs are called as interconnecting BSIDs and help to represent intermediate domain of different data plane type as a SID of ingress domain dataplane type in the headend policy. The IW SR-PCE solution Section 7.1.1 leverage these BSIDs as segments of SR policy on headend domain.

7. Interworking Procedures

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

Following is assumed of data plane support at various nodes:

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

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

* Nodes 4 and 7 are border nodes that support both the SRv6 and MPLS data plane.

A VPN route is advertised via service RRs (S-RR) from an egress PE(node 10) to 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, transport IW requires:

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

* For Mo6, tunnel MPLS encapsulated traffic destined to 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] that stitches end to end path 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 that advertise and propagate PE locator or Loopback address across domains. During propagation, domain border node set nexthop to self that result in allocation of label or SRv6 SID depending on dataplane type of the domain where prefix is propagated. These procedure can provide both best effort or intent aware end to end path.

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 it is 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. For interworking function, it programs an SR policy at border node 4 with segment list node 5 and 7 bounded to an End.BM BSID [RFC8986] (For example, SR-PCE creates an SR-MPLS policy (C1,7) at node 4 with segments <16005,16007>. This policy is bound to End.BM behavior with SRv6 BSID as B:4:BM-C1-7::). SR-PCE responds back to node 1 with SRv6 segments of requested SLA including End.BM at node 4 to traverse SR-MPLS-IPv4 C domain.

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

* Node 1 performs SRv6 operation 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 SRv6 End behavior on B:2:E:: SRv6 SID present in the DA

  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 node) performs End.BM behavior on B:4:BM-C1-7:: SRv6 SID present in the DA

  Packet leaving node 4 MPLS (16005,16007, IPv6 Explicit NULL)((A:1::, B:8:E::) (B:10:DT4::, B:8:E::, B:4:BM-C1-7-:: ; SL=1)).

* Node 5 performs PHP pn 16007

* Node 7 performs native IPv6 lookup after IPv6 explicit NULL processing

  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) behavior B:8:E:: SRv6 SID present in the DA

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

* Node 10 performs End.DT4 behavior on B:10:DT4:: SRv6 SID present in the DA i.e. it looks up inner IP destination in the VRF corresponding to B:10:DT4:: SID and forwards traffic to CE accordingly.

7.1.1.2. Mo6

Refer Section 2.1 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) and vpn label. The SR-PCE computes color-aware C1 path via node 2, 5 and 8. For interworking function, it programs a SRv6 policy bound to MPLS BSID at border node 4 with SRv6 segment list along the required color-aware path with last segment of behavior End.DTM46 Section 4.2 (For example, SR-PCE create SRv6 policy (C1,7) at node 4 with segments <B:5:E::,B:7:DTM46::>. It is bound to MPLS BSID 24407). 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.

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: node 2’s prefix SID is not pushed due to PHP): 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 MPLS BSID label 24407

   Packet leaving node 4 IPv6(A:4::, B:5:E::) (B:7:DTM46:: ; SL=1)NH=137) MPLS((16008,16010,vpn_label)

4. Node 7 performs DTM46 behavior on B:7:DTM46:: SRv6 SID present in the DA i.e. remove IPv6 header and lookup label 16008 in MPLS table.

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

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

   Packet leaving node 8 MPLS packet (vpn_label)

6. Node 10 pops vpn_label, looks up inner IP destination in the VRF corresponding to the vpn_label and forwards to traffic to CE accordingly.

7.1.2. BGP inter domain routing procedures

Procedures described below build upon BGP Label Unicast (BGP-LU) [I-D.ietf-mpls-seamless-mpls] and [RFC4798] that advertise transport reachability of PE loopback address or SRv6 locator across a multi-domain network. The procedures leverage existing SAFIs (for example, BGP-LU(AFI=1 or 2 and SAFI=4) and IPv6 (AFI=2,SAFI=1)). Setting nexthop to self on border node 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 Flexible Algorithm. They are also applicable to intent-aware routes recursing over intent aware intra-domain paths.

7.1.2.1. 6oM

Refer Section 2.1 for 6oM scenario. SRv6 based L3/L2 BGP services are signaled with SRv6 Service SID allocated from egress PE locator prefix. Ingress PE learns the service routes and need to resolve SRv6 Service SID over egress PE locator or its summary. Below describes propagation of locator or its summary to create end to end underlay path.

* Egress border node learns LE domain PE locator through IGP and redistribute it in BGP. Alternatively, locator is advertised by egress PE in the BGP IPv6 Unicast (AFI value 2 and SAFI value 1) to border nodes.

* Egress border node uses [RFC4798] 6PE procedure that results in advertisement of locator as BGP-LU route with locally allocated label or IPv6 Explicit NULL and nexthop set to itself to ingress border node. SR-MPLS-IPv4 builds MPLS LSP path in C domain from ingress to egress border node. Note: Egress border router may advertise summary prefix covering all PE locators in LE domain.

* Ingress border node advertise route of remote locator or its summary in LI domain. Below are the options to advertise route:

  - BGP IPv6 SAFI (AFI=2 and SAFI=1) distribute the route with SRv6 transport SID of local End behavior in Prefix-SID attribute TLV type 5 [RFC9252]. This option results in additional SRv6 encapsulation at ingress PE.

  - BGP IPv6 SAFI (AFI=2 and SAFI=1) distribute the route to each of P and PE router through infrastructure route reflector. This option avoids additional SRv6 transport SID encapsulation at ingress PE and forwards traffic hop by hop in LI domain.

  - Leak remote locators or their summary in LI IGP (Typically on transport ABR only infrastructure prefixes are present in BGP. If that is not the case, proper filters need to be configured for such leaking into IGP).

* Ingress PE learns remote locator or summary with or without SRv6 transport SID. When learnt with SRv6 transport SID, it builds the packet encapsulation that contains the SRv6 Service SID and SRv6 transport SID in the SID list. SRv6 transport SID tunnels traffic to ingress border node in LI domain (P routers like 2 and 3 does

not need egress PE locator reachability). When learnt without SRv6 transport SID, the packet encapsulation contains the SRv6 Service SID as DA and forwarded hop by hop based on remote locator IPv6 prefix lookup in LI domain.

Example to advertise node 10 locator to node 1 with SRv6 SID encapsulation:

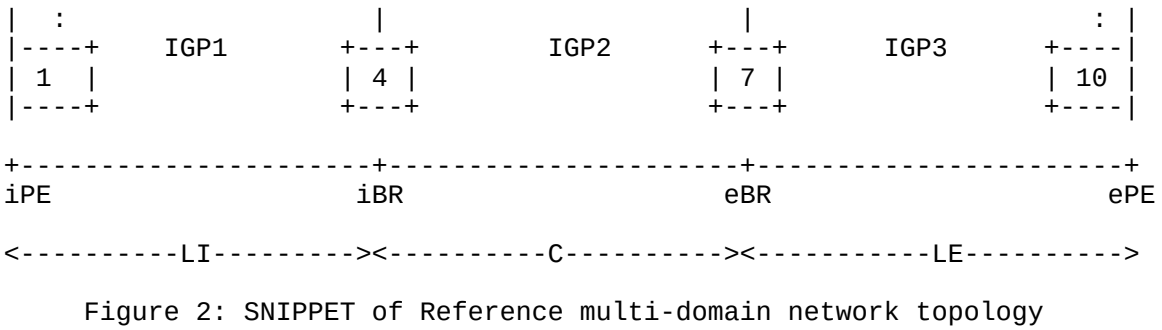

Figure 2: SNIPPET of Reference multi-domain network topology

1. Routing @ node 10:

   * SIS advertises locator B:B:10::/48

   * BGP (AFI=1,SAFI=128) originates a VPN route RD:V/v via B:B:10::1 with SRv6 Service SID B:B:10:DT4::. This route is advertised to service RR.

2. Routing @ node 7:

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

   * BGP advertise B:B:10::/48 in (AFI=2,SAFI=4) among border routers with nexthop as node 7 and IPv6 Explicit Null label.

3. Routing @ node 4:

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

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

4. Routing @ node 1:

* BGP learns B:B:10::/48 with nexthop node 4 and outgoing SRv6 SID B:B:4:END:: in Prefix-SID attribute TLV type 5

* BGP learns service prefix RD:V/v, with SRv6 service SID B:B:10:DT4::

* ISIS learns locator prefix B:B:4::/48 for node 4's SID reachability

Forwarding state at different nodes:

```
@1: IPv4 VRF V/v => H.Encaps.red <B:B:4:END::, B:B:10:DT4::>
                                        with SRH (SL=1 and NH=IPv4)
@1: IPv6 Table: B:B:4::/48 => forward via ISIS path to node 4
@4: IPv6 Table: B:B:4:END:: => PSP Processing (Update DA with B:B:10:DT4::,
                                        set IPv6.NH=IPv4, pop the SRH)
@4: IPv6 Table: B:B:10::/48 => push MPLS label stack {16007, 2 (IPv6 Explicit NULL)}
@7: MPLS label 2 => pop and lookup inner IPv6 DA
@7: IPv6 Table B:B:10::/48 => forward via ISIS path to node 10
@10: IPv6 Table B:B:10:DT4:: => DT4 SID processing (pop the outer header
                            and lookup the inner IPv4 DA in the VRF
```

7.1.2.2. Mo6

Refer Section 2.1 for Mo6 scenario. MPLS-based L3/L2 BGP service is signaled with IPv4 nexthop of egress PE through Service RRs. Ingress PE needs MPLS reachability for egress PE's IPv4 loopback address in the LE domain to transport services.

Egress PE originate route to its loopback address in BGP-LU [RFC8277]in LE domain. Egress border node sets nexthop to itself and signals MPLS-over-IP to ingress border node that results in tunneling of BGP-LU LSP across SRv6 C domain.

* There are existing BGP-LU label cross-connect on border nodes for each PE loopback address.

* The lookup at the ingress border node are based on BGP-LU label as usual

* Just the SR-MPLS IGP label to next hop is replaced by an SRv6 encapsulation with DA = SRv6 SID associated with DTM46 behavior of egress border node in C domain.

* Ingress border node forwarding performs RFC3107 label swap, followed by H.Encaps.M operation setting DA = SRv6 SID associated with DTM46 behavior.

Existing BGP-LU updates between border nodes need to signal SRv6 SID associated with DTM46 behavior. BGP procedures to signal SRv6 SID is described in [I-D.sa-idr-bgp-srv6-mpls-transport-iw] and outside the scope of this document. Below illustrate the example control plane and corresponding FIB state to achieve such tunneling:

Control plane example

1. Routing @10:

    * SR ISIS originates IPv4 PE loopback with SR Node SID 16010

    * BGP AFI=1,SAFI=4 originates 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 via next hop node 10. This route is advertised to service RR.

2. Routing @ 7:

    * ISIS IPv6 advertise its locator B:B:7::/48 in C domain

    * BGP learns node 10 IPv4 loopback address with label. It allocates local label (based on label-index, if present) and programs BGP-LU label swap to received label and deliver it to next hop.

    * BGP AFI=1,SAFI=4 advertises loopback address of node 10 to node 4. NLRI label is set to local label and SRv6 SID B:B:7:DTM46:: is carried in SRv6 SID Information Sub-TLV of "SRv6 Transport" TLV in Prefix-Sid attribute. If received, Label-Index TLV of Prefix-SID attribute is also signaled.

3. Routing @ 4:

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

    * BGP learns node 10 IPv4 loopback address from node 7 with outgoing label. It allocates local label (based on label-index, if present) and programs label swap entry to be SRv6 tunnled to BGP nexthop by performing H.Encaps.M.red operation where IPv6 header is set to SRv6 SID B:B:7:DTM46:: (received in "SRv6 transport TLV" of Prefix-Sid attribute).

* BGP AFI=1,SAFI=4 advertise loopback address of node 10 to node 1 with locally allocated label and nexthop to self. This results in removal of SRv6 Transport TLV in Prefix-SID attribute.

4. Routing @ 1:

* BGP learns IPv4 loopback address of node 10 from node 4 with outgoing label. It programs route to push outgoing label delivered to nexthop node 4.

* BGP AFI=1,SAFI=128 learns service prefix RD:V/v with service label via node 10 loopback address.

Forwarding state at different nodes:

```
@1: IPv4 VRF: V/v => out label=vpn_label, next hop=IPv4 loopback of node 10
@1: IPv4 table: IPv4 loopback of node 10 => out label=16010, next hop=node4
@1: IPv4 table: IPv4 loopback of node 4 => out label=16004, next hop=interface to node 2
@4: MPLS Table: 16010 (BGP-LU) => out label=16010, H.Encaps.M.red with DA=B:B:7:DTM46::
@4: IPv6 table: B:B:7::/48 => next hop=interface to node 5
@7: SRv6 My SID table: B:B:7:DTM46:: => decaps IPv6 header and lookup top label.
@7: MPLS table: 16010 (SR ISIS)=> out label=16010, next hop=interface to node 8
@10: MPLS table: vpn label => pop label and lookup the inner IPv4 DA in the VRF
```

During migration, when MPLS data plane is still enabled in C domain, a SRv6 capable ABR an select relevant encapsulation and legacy ABR can continue MPLS encapsulation using label in NLRI.

If domain border node is a pure transport node without any services, either End.DTM46 or End.DTM can be advertised and it is upto the implementation to choose. If domain border node does have global table IPv4 and IPv6 (Section 7.1.2.3), then it MUST advertise End.DTM46. END.DTM46 is a superset of END.DTM.

7.1.2.3. Global table services over BGP-LU transport

Procedures as defined in Section 7.1.2.2 work for global table services ([RFC4271], [RFC2545]) over BGP-LU transport when service endpoint is beyond border node (example node 10). In scenarios where service endpoint is border node, additional SRv6 decapsulation behavior is required that performs service traffic IP destination lookup in global IPv4 or IPv6 table. In such deployment, border node advertise its existing IPv4 PE loopback address in BGP-LU as per Section 7.1.2.2 where SRv6 SID is associated with End.DTM46 behavior instead of END.DTM.

7.2. Service IW

As described in Section 2.2 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

In Gateway Interworking role, node supports both BGP SRv6 based L2/L3 service and BGP MPLS based L2/L3 service for a given service instance (e.g. L3 VRF, EVPN EVI). In dataplane, it terminates service encapsulation of ingress PE and perform L2 MAC route or L3 IP destination lookup in service instance. Lookup provide egress PE service encapsulation that is used to send packet to egress PE. This is similar to inter-as option A that is implemented within a single node instead of implementing on two back2back ABRs/ASBRs nodes connected with VRF interface.

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

* 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 exchanges 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.Red. It is advertised to MPLS service PE with service label. When gateway receives traffic with service label from MPLS service PE, it performs MAC/destination IP lookup in service instance. The lookup results 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 towards the 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 performs inner destination lookup in service instance after decaps of IPv6 header. The Lookup result in instruction to push service label and send it over MPLS LSP towards the nexthop.

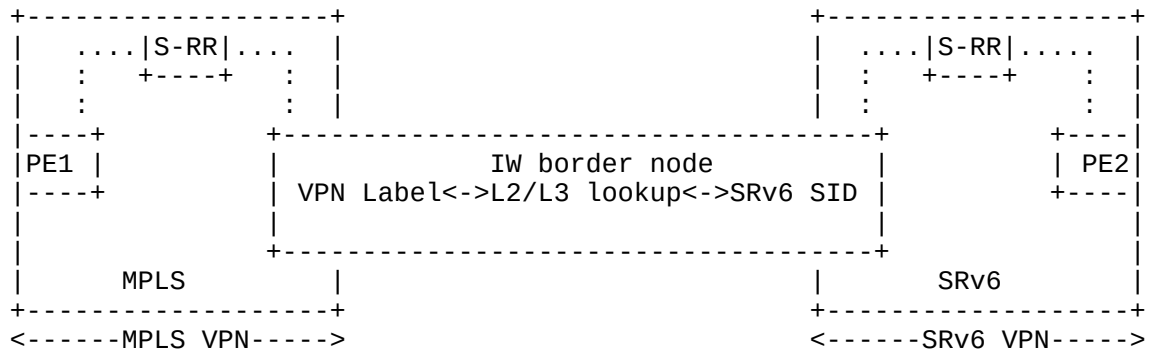

Figure 3: Gateway IW

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 approach is similar to inter-AS option B procedures described in [RFC4364], except that service label cross-connect on border node is replaced with service label to SRv6 service SID (or vice versa) translation on the IW node.

* IW node does not require service instance such as VRF or EVI.

* IW node exchanges BGP L2/L3 service prefix with SRv6 based Service PEs through a set of service RRs. This BGP session will learn and advertise L3/L2 service prefixes with SRv6 SIDs in the prefix SID attribute [RFC9252].

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

* IW node that sets nexthop to self, allocates SRv6 SID of DXM behavior variant based on service route AFI and SAFI i.e. End.DXM4 for IPv4 service prefix, End.DXM6 for IPv6 service prefix and End.DXM2 for layer 2 prefix learned from the MPLS PE. The FIB entry lookup on SRv6 local service SID provides service route outgoing service label and BGP nexthop (MPLS PE). The IW node then advertises the service route in the SRv6 domain with locally allocated SRv6 SID and its corresponding behavior. During packet forwarding, when an IPv6 packet arrives with the DA matching the locally allocated SRv6 SID, the node decapsulates the IPv6 header, pushes the outgoing service label, and delivers the packet to the BGP next-hop."

* IW node that sets nexthop to self, allocates local label for service route learnt from SRv6 PE attached with SRv6 SID in Prefix-SID attribute. FIB label entry lookup results in H.Encaps with SRv6 SID learned from SRv6 PE. IW node then advertises the service route to MPLS domain with locally allocated label. During packet forwarding, when an MPLS packet arrives with locally allocated label, the node pops the service label and performs H.Encaps.Red operation, setting DA as remote SRv6 SID and Upper-layer header based on behavior of SRv6 SID.

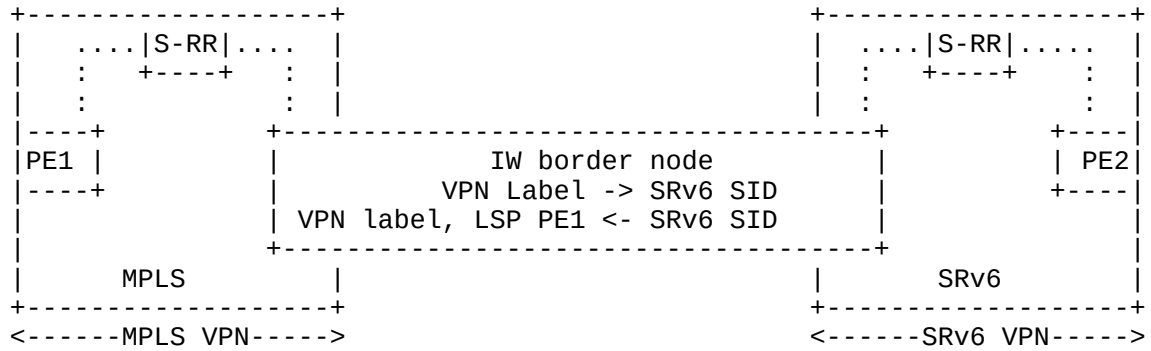

Figure 4: Service translation

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

8. Migration and co-existence

In addition to interworking, this 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", "End.DTM46", "End.DXM4", "End.DXM6" and "End.DXM2". IANA is requested to assign identifier value in the "SRv6 Endpoint Behaviors" sub-registry under "Segment Routing Parameters" registry.

Value	Hex	Endpoint behavior	Reference
73	0x0049	End.DTM	<this document>
TBD	TBD	End.DTM46	<this document>
TBD	TBD	End.DXM4	<this document>
TBD	TBD	End.DXM6	<this document>
TBD	TBD	End.DXM2	<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, Bruno Decraene for their comments and review.

14. References

14.1. Normative References

[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>.

[RFC2545] Marques, P. and F. Dupont, "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", RFC 2545, DOI 10.17487/RFC2545, March 1999, <https://www.rfc-editor.org/info/rfc2545>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <https://www.rfc-editor.org/info/rfc3031>.

[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>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[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., Henderickx, 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-21, 12 February 2025, <https://datatracker.ietf.org/doc/html/draft-ietf-rtgwg-segment-routing-ti-lfa-21>.

[I-D.sa-idr-bgp-srv6-mpls-transport-iw]
Agrawal, S., Filsfils, C., Rao, D., Dong, J., and R. Manur, "BGP extensions for SRv6/MPLS Transport Interworking", Work in Progress, Internet-Draft, draft-sa-idr-bgp-srv6-mpls-transport-iw-02, 6 February 2026, <https://datatracker.ietf.org/doc/html/draft-sa-idr-bgp-srv6-mpls-transport-iw-02>.

[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: robinli314@163.com

Shraddha Hegde
Juniper Networks
Email: shraddha@juniper.net