Workgroup: SPRING Working Group
Internet-Draft:
draft-bonica-spring-srv6-end-dtm-08
Published: 21 October 2022
Intended Status: Standards Track
Expires: 24 April 2023
Authors: S. Hegde, Juniper Networks
P. Kaneriya, Juniper Networks
R. Bonica, Juniper Networks
P. Shaofu, ZTE Corporation
G. Mirsky, ZTE Corporation
Z. Zhang, ZTE Corporation
B. Decraene, Orange

SR-MPLS / SRv6 Transport Interworking

Abstract

This document describes procedures for interworking between an SR-MPLS transit domain and an SRv6 transit domain. Each domain contains Provider Edge (PE) and Provider (P) routers. Area Border Routers (ABR) provide connectivity between domains.

The procedures described in this document require the ABR to carry a route to each PE router. However, they do not required the ABR to carry service (i.e., customer) routes. In that respect, these procedures resemble L3VPN Interprovider Option C.

Procedures described in this document support interworking for global IPv4 and IPv6 service prefixes. They do not support interworking for VPN services prefixes where the SR-MPLS domain uses MPLS service labels.

Status of This Memo

Copyright Notice

Table of Contents

1. Overview

Segment Routing (SR) [RFC8402] allows source nodes to steer packets through SR paths. It can be implemented over IPv6 [RFC8200] or MPLS [RFC3031]. When SR is implemented over IPv6, it is called SRv6 [RFC8986]. When SR is implemented over MPLS, it is called SR-MPLS [RFC8660].

This document describes procedures for interworking between an SR-MPLS transit domain and an SRv6 transit domain. Each domain contains Provider Edge (PE) and Provider (P) routers. Area Border Routers (ABR) provide connectivity between domains.

The procedures described in this document require the ABR to carry a route to each PE router. However, they do not required the ABR to carry service (i.e., customer) routes. In that respect, these procedures resemble L3VPN Interprovider Option C [RFC4364].

Procedures described in this document support interworking for global IPv4 and IPv6 service prefixes. They do not support interworking for VPN services prefixes where the SR-MPLS domain uses MPLS service labels.

2. 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.

3. Reference Topology

```
 ----------------------- SR Path 1 -------------------------->

 <---------------------- SR Path 2 --------------------------

  ------       ------       ------       ------       ------
 | PE    |    |  P    |    | ABR  |     |  P    |    | PE    |
 |Node 1| --- |Node 2| --- |Node 3| --- |Node 4| --- |Node 5|
  ------       ------       ------       ------       ------

  Seg. A       Seg. B       Seg. C       Seg. D       Seg. E

 <---------- SRv6 Domain ---------->
                          <--------- SR-MPLS Domain --------->
```

Figure 1: Interworking Between SR Domains

Figure 1 depicts interworking between an SR-MPLS domain and an SRv6 domain. The SRv6 domain contains PE Node 1 and P Node 2. The SR-MPLS domain contains P Node 4 and PE node 5. Both domains contain ABR Node 3.

Nodes 1 and 2 MUST support SRv6 but are NOT REQUIRED to support SR-MPLS. Nodes 4 and 5 MUST support SR-MPLS but are NOT required to support SRv6. Node 3 MUST support both SRv6 and SR-MPLS. It must also support interworking procedures.

Network operators configure a loopback interface on Nodes 1 through 5. These are called Loopback1 through Loopback5. They also configure 2 additional loopback interfaces on PE Node 5. These are called Loopback5.IPv4 and Loopback5.IPv6.

Each node instantiates an SR Segment (i.e., Segment A through Segment E). SR Path 1 begins on PE Node 1 and ends on PE Node 5. It visits Nodes 2, 3, 4, and 5, executing the instructions associated with Segments B, C, and D. SR Path 2 begins on PE Node 5 and ends on PE Node 1. It visits Nodes 4, 3, 2, and 1, executing the instructions associated with Segments D, C, B and A.

4. Forwarding Plane

```
  ----------------------- SR Path 1 -------------------------->

    ------       ------       ------       ------       ------
   | PE   |     |  P   |     | ABR  |     |  P   |     | PE   |
   |Node 1| --- |Node 2| --- |Node 3| --- |Node 4| --- |Node 5|
    ------       ------       ------       ------       ------

     Seg. A       Seg. B       Seg. C       Seg. D       Seg. E

      IPv6:          IPv6:           SR-MPLS:         SR-MPLS:
       SA: Node 1     SA: Node 1      Seg. D           Exp. Null
       DA: Seg. B     DA: Seg. C      Exp. Null       Payload
      SRH:           SRH:            Payload
       SL:  1         SL: 0
       SID: Seg. C    SID: Seg. C
      Payload       Payload
```

Figure 2: Encapsulation: SRv6 To SR-MPLS

Figure 2 depicts the forwarding plane as a packet traverses SR Path 1, from Node 1 to Node 5. In this example, PE Node 1 receives an IPv4 packet.

PE Node 1 encapsulates the IPv4 packet in an SRv6 header. The SRv6 header contains an IPv6 header and a Segment Routing Header (SRH) [RFC8754]. The Destination Address in the IPv6 header is a Segment Identifier (SID) that represents Segment B. Segment B is an END instruction instantiated on P Node 2. The SRH contains a Segments Left field and one SID. The Segments Left field is equal to 1 and the SID represents Segment C, and END.DM (Section 4.1) instruction instantiated on ABR Node 3.

PE Node 1 forwards the packet to P Node 2. When P Node 2 receives the packet, it processed the END instruction. It decrements the Segments Left field in the SRH and copies the SID from the SRH to the Destination Address field of the IPv6 header. It then forwards the packet to ABR Node 3.

When ABR Node 3 receives the packet, it processes the END.DM instruction. It removes the SRv6 header and replaces it with an SR-MPLS label stack that contains two entries. The top entry represents a prefix SID instantiated on P Node 4. The bottom entry is an

Explicit Null instruction (i.e., MPLS label 0), instantiated on PE Node 5.

ABR Node 3 then forwards the packet to P Node 4. P Node 4 processes the prefix SID, removing the top entry from the SR-MPLS label stack and forwarding the packet to PE Node 5. PE Node 5 processes the Explicit Null instruction, removing the remaining SR-MPLS label stack entry and processing the payload.

```
    <----------------------- SR Path 2 -------------------------

      ------       ------       ------       ------       ------
     | PE   |     |  P   |     | ABR  |     |  P   |     | PE   |
     |Node 1| --- |Node 2| --- |Node 3| --- |Node 4| --- |Node 5|
      ------       ------       ------       ------       ------

       Seg. A       Seg. B       Seg. C       Seg. D       Seg. E

        IPv6:          IPv6:           SR-MPLS:          SR-MPLS:
         SA: Node 3     SA: Node 3      Seg. C            Seg. D
         DA: Seg. A      DA: Seg. B    Payload            Seg. C
        SRH:           SRH:                              Payload
         SL:  0         SL: 1
         SID: Seg. A    SID: Seg. A
        Payload       Payload
```

Figure 3: Encapsulation: SR-MPLS to IPv6

Figure 3 depicts the forwarding plane as a packet traverses SR Path 2, from Node 5 to Node 1. In this example, PE Node 5 receives an IPv4 packet.

PE Node 5 encapsulates the IPv4 packet in an SR-MPLS label stack that contains two entries. The top entry represents a prefix SID instantiated on P Node 4. The bottom entry is a binding SID instantiated on ABR Node 3.

PE Node 5 then forwards the packet to P Node 4. P Node 4 processes the prefix SID, removing the top entry from the SR-MPLS label stack and forwarding the packet to ABR Node 3. ABR Node 3 processes binding SID, removing the remaining SR-MPLS label stack entry and replacing it with an SRv6 header. The SRv6 header contains an IPv6 header and an SRH. The Destination Address in the IPv6 header is a Segment Identifier (SID) that represents Segment B. Segment B is an END instruction instantiated on P Node 2. The SRH contains a Segments Left field and one SID. The Segments Left field is equal to 1 and the SID represents Segment A, an END.DT46 instruction instantiated on PE Node 1. That instruction causes the packet to be forwarded using the main IP forwarding table, not a VPN forwarding table.

ABR Node 3 forwards the packet to P Node 2. When P Node 2 receives the packet, it processed the END instruction. It decrements the Segments Left field in the SRH and copies the SID from the SRH to the Destination Address field of the IPv6 header. It then forwards the packet to PE Node 1. PE Node 1 processes its END.DT46 instruction, removing the SRv6 header and processing the payload.

4.1. END.DM Processing

The End.DM SID MUST be the last segment in a SR Policy. Its arguments are associated with an SR-MPLS label stack.

When Node N receives a packet destined to S and S is a locally instantiated End.DM SID, Node N executes the following procedure:

```
S01. When an IPv6 Routing Header 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.DM SID, N executes the following procedure:

```
S01. Decapsulate the packet (i.e., remove the outer IPv6 Header and a
     its extension headers)
S02. Push the SR-MPLS label stack that is associated with the END.DM
     arguments. Set the MPLS Traffic Class and TTL values to reflect
     the Traffic Class and Hop count values received in the IPv6 head
S03. Submit the packet to the MPLS FIB lookup for transmission to the
     new destination
```

5. Control Plane

```
<------------------- Customer Routes (iBGP) -------------->

<-- PE and ABR Routes (iBGP) ->
                             <-- PE and ABR Routes(BGP-LU) ->
 ------       ------       ------       ------       ------
| PE   |     |  P   |     | ABR  |     |   P  |     | PE   |
|Node 1| --- |Node 2| --- |Node 3| --- |Node 4| --- |Node 5|
 ------       ------       ------       ------       ------

<---------- SRv6 Domain ---------->
                         <--------- SR-MPLS Domain --------->
```

Figure 4: BGP NLRI Exchange

In the Figure 4, PE Node 1 and PE Node 5 exchange customer Network Layer Reachability Information (NLRI) [RFC4271] using either a direct BGP session or a route reflector [RFC4456]. All customer routes exchanged between PE Node 1 and PE Node 5 belong to the general routing instance. They cannot belong to a VPN.

PE Node 1 exchanges loopback routes with ABR Node 3, using either a direct BGP session or a route reflector. Likewise, ABR Node 3 exchanges loopback with PE Node 5, using either a direct BGP session or a route reflector.

PE Node 1 and ABR Node 3 bind SIDs to the loopback routes that they exchange, as described in [I-D.ietf-bess-srv6-services]. PE Node 5 and ABR Node 3 bind labels to the loopback routes that they exchange, as described in [RFC8277].

Both domains use an IGP to distribute link state information and establish connectivity within the domain.

5.1. Signaling SR Paths That Originate In The SRv6 Domain

PE Node 5 advertises an IPv4 customer route to PE Node 1 using BGP as follows:

* IPv4 Prefix

* Next-hop: Loopback5.IPv4

This causes PE Node 1 to resolve the customer route through its route to Loopback5.IPv4. The following paragraphs describe how PE Node 1 acquires a route to Loopback5.IPv4.

PE Node 5 advertises Loopback5.IPv4 to ABR Node 3 using BGP Labeled Unicast (BGP-LU) as follows:

* Prefix: Loopback5.IPv4

* Next-hop: Loopback5

* Color Community: Color to distinguish between paths between ABR Node 3 and PE Node 5

* MPLS Label: Explicit Null (0)

Now, ABR Node 3 resolves its route to Loopback5.IPv4 through its IGP route to Loopback5. Therefore, when forwarding traffic bound for Loopback5.IPv4, it imposes:

* An SR-MPLS label stack associated with the IGP route to Loopback5

* An additional Explicit Null label

ABR Node 3 advertises Loopback5.IPv4 to PE Node 1 using BGP as follows:

* Prefix: Loopback5.IPv4

* Next-hop: Loopback3

* Color Community: Color to distinguish between paths between ABR Node 3 and PE Node 1

* SID: SID C (i.e., an END.DM SID instantiated on ABR Node 3)

Now, PE Node 1 resolves its route to Loopback5.IPv4 through its IGP route to Loopback3. Therefore, when forwarding traffic bound for Loopback5.IPv4, it imposes an SRv6 header that includes the following SIDs:

* SIDS associated with the IGP route to Loopback3

* SID C

5.2. Signaling SR Paths That Originate In The SR-MPLS Domain

PE Node 1 advertises an IPv4 customer route to PE Node 5 using BGP as follows:

* IPv4 Prefix

* Next-hop: Loopback1

This causes PE Node 5 to resolve the customer route through its route to Loopback1. The following paragraphs describe how PE Node 5 acquires a route to Loopback1.

PE Node 1 advertises Loopback1 to ABR Node 3 using BGP as follows:

* Prefix: Loopback1
* Next-hop: Loopback1
* Color Community: Color to distinguish between paths between ABR Node 3 and PE Node 1
* SID: SID A (i.e., An END.DT46 SID instantiation on PE Node 1. This instruction causes a packet to be forwarded using the main IP forwarding table, not a VPN forwarding table.)

Now, ABR Node 3 resolves its route to Loopback1 through its IGP route to Loopback1. Therefore, when forwarding traffic bound for Loopback1, it imposes an SRv6 header that includes:

* SIDS associated with the IGP route to Loopback1
* SID A

ABR Node 3 advertises Loopback1 to PE Node 5 using BGP-LU as follows:

* Prefix: Loopback1
* Next-hop: Loopback3
* Color Community: Color to distinguish between paths between ABR Node 3 and PE Node 5
* MPLS Label: A binding label that represents the SRv6 path between ABR Node 3 and PE Node 5

Now, PE Node 5 resolves its route to Loopback1 through its IGP route to Loopback3. Therefore, when forwarding traffic bound for Loopback1, it imposes:

* An SR-MPLS label stack associated with the IGP route to ABR3
* An additional label representing a binding SID. The binding SID maps to the SRv6 path between ABR Node 3 and PE Node 5

6. IANA Considerations

The authors will request an early allocation from the "SRv6 Endpoint Behaviors" sub-registry of the "Segment Routing Parameters" registry.

7. Security Considerations

Because SR inter-working requires co-operation between inter-working domains, this document introduces no security consideration beyond those addressed in [RFC8402], [RFC8754] and [RFC8986].

8. Acknowledgements

Thanks to Melchior Aelmans, Takuya Miyasaka and Jeff Tantsura for their comments.

9. References

9.1. Normative References

[I-D.ietf-bess-srv6-services]
Dawra, G., Talaulikar, K., Raszuk, R., Decraene, B., Zhuang, S., and J. Rabadan, "SRv6 BGP based Overlay Services", Work in Progress, Internet-Draft, draft-ietf-bess-srv6-services-15, 22 March 2022, <https://www.ietf.org/archive/id/draft-ietf-bess-srv6-services-15.txt>.

[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>.

[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>.

[RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, DOI 10.17487/RFC4456, April 2006, <https://www.rfc-editor.org/info/rfc4456>.

[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>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[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>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/info/rfc8660>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

[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>.

9.2. Informative References

[I-D.hegde-spring-mpls-seamless-sr] Hegde, S., Bowers, C., Xu, X., Gulko, A., Bogdanov, A., Uttaro, J., Jalil, L., Khaddam, M., Alston, A., and M. Luis Contreras, "Seamless SR Problem Statement", Work in Progress, Internet-Draft, draft-hegde-spring-mpls-seamless-sr-07, 8 July 2022, <https://www.ietf.org/archive/id/draft-hegde-spring-mpls-seamless-sr-07.txt>.

[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>.

Authors' Addresses

Shraddha Hegde
Juniper Networks
Embassy Business Park
Bangalore 560093
KA
India

Email: shraddha@juniper.net

Parag Kaneriya
Juniper Networks
Elnath-Exora Business Park Survey
Bangalore 560103
Karnataka
India

Email: pkaneria@juniper.net

Ron Bonica
Juniper Networks
Herndon, Virginia 20171
United States of America

Email: rbonica@juniper.net

Peng Shaofu
ZTE Corporation

Email: peng.shaofu@zte.com.cn

Greg Mirsky
ZTE Corporation
United States of America

Email: gregimirsky@gmail.com

Zheng Zhang
ZTE Corporation

Email: zhang.zheng@zte.com.cn

Bruno Decraene
Orange
France

Email: bruno.decraene@orange.com