

SPRING Working Group
Internet-Draft
Intended status: Standards Track
Expires: 20 September 2024

M. Rajesh
R. Bonica
Juniper Networks
H. wang
Huawei Technologies
Shaofu. Peng
ZTE Corporation
19 March 2024

SRv6 inter-domain mapping SIDs
draft-salih-spring-srv6-inter-domain-sids-05

Abstract

This document describes three new SRv6 end-point behaviors, called END.REPLACE, END.REPLACEB6 and END.DB6. These behaviors are used in distributed inter-domain solutions and are normally executed on border routers. They also can be used to provide multiple intent-based paths across these domains.

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/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 20 September 2024.

Copyright Notice

Copyright (c) 2024 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components

extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Overview	2
2. Requirements Language	2
3. Usecases	3
3.1. usecase 1	3
3.2. usecase 2	3
4. SRv6 SID Behaviors	4
4.1. END.REPLACE	4
4.2. END.REPLACEB6	4
4.3. END.DB6	5
5. Interworking Procedures	6
5.1. Option C Transport Interworking	6
5.2. Option B service interworking	9
6. IANA Considerations	10
7. Security Considerations	10
8. Contributors	10
9. Acknowledgements	11
10. References	11
10.1. Normative References	11
10.2. Informative References	12
Authors' Addresses	12

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].

This document describes three new SRv6 end-point behaviors, called END.REPLACE, END.REPLACEB6 and END.DB6. These behaviors are used to build paths across SRv6 domains. They also facilitate end-to-end SRv6 intent-based path stitching.

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. Usecases

3.1. usecase 1

This use-case is mentioned in Section 4.1.1 of [I-D.hegde-spring-mpls-seamless-sr].

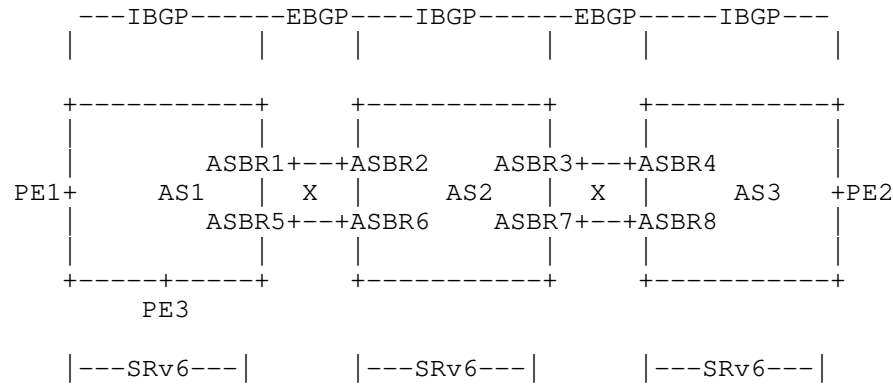


Figure 1: Multiple ASes connected with E-BGP

Figure 1 depicts three ASes (AS1, AS2 and AS3). All the three domains deploy SRv6. Inter-provider Option C[RFC4364] connectivity is maintained from PE1 to PE2.

3.2. usecase 2

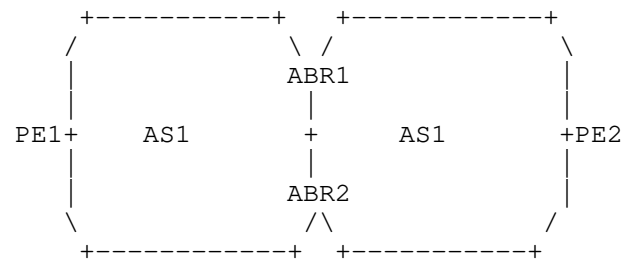


Figure 2: Singe AS with different IGP domains

The above diagram Figure 2 shows two different SRv6 IGP domains. Services are running between PE1 and PE2 in option B [RFC4364] style. The requirement here is to avoid service route lookup on ABR1 and ABR2 to provide option B style end to end connectivity

4. SRv6 SID Behaviors

4.1. END.REPLACE

The END.REPLACE behavior is applicable in the Multiple ASes Connected With E-BGP (Section 3.1) use-case.

The End.REPLACE SID cannot be the last segment in SRH or SR Policy.

Any SID instance of this behavior is associated with a set, J, of one or more L3 adjacencies of immediate BGP neighbors

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

```
S01. When an SRH is processed {
S02.   If (Segments Left == 0) {
S03.     Stop processing the SRH, and proceed to process the next
        header in the packet, whose type is identified by
        the Next Header field in the routing header. Procedure is as
        per Section 4.1.1 of [RFC8986].
S04.   }
S05.   If (IPv6 Hop Limit <= 1) {
S06.     Send an ICMP Time Exceeded message to the Source Address with Code 0
        (Hop limit exceeded in transit), interrupt packet processing, and di
scard packet
S07.   }
S08.   Decrement IPv6 Hop Limit by 1
S09.   Update IPv6 DA with new destination address(SID) mapped with END.REPLAC
E SID.
S10.   Submit the packet to the IPv6 module for transmission
        to the new destination via a member of J.
S11. }
```

4.2. END.REPLACEB6

The END.REPLACEB6 behavior is applicable in the Multiple ASes Connected With E-BGP (Section 3.1) use-case.

The End.REPLACEB6 SID cannot be the last segment in a SRH or SR Policy.

Node N is configured with an IPv6 address T (e.g., assigned to its loopback).

When Node N receives a packet destined to S and S is a locally instantiated End.REPLACEB6 SID, Node N executes the following procedure:

```

S01. When an SRH is processed {
S02.   If (Segments Left == 0) {
S03.     Stop processing the SRH, and proceed to process the next
        header in the packet, whose type is identified by
        the Next Header field in the routing header. Procedure is as
        per Section 4.1.1 of [RFC8986].
S04.   }
S05.   If (IPv6 Hop Limit <= 1) {
S06.     Send an ICMP Time Exceeded message to the Source Address with Code 0
        (Hop limit exceeded in transit), interrupt packet processing, and di
scard packet
S07.   }
S08.   Decrement IPv6 Hop Limit by 1
S09.   Update IPv6 DA with new destination address(SID) mapped with END.REPLAC
EB6.
S10.   Push an IPv6 header with an SRH.
S11.   Set outer IPv6 SA = T and outer IPv6 DA to the first SID in the segment
list
S12.   Set outer Payload Length, Traffic Class, Hop Limit, and Flow Label fiel
ds
S13.   Set the outer Next Header value
S14.   Submit the packet to the IPv6 module for transmission to the First SID.
S15. }
```

Note :

S10 - S13. Implemetation may choose to avoid outer encapsulation for flex-algo and best effort based SRv6 transport tunnels.

S12. The Payload Length, Traffic Class, Hop Limit, and Next Header fields are set as per [RFC2473]. The Flow Label is computed as per [RFC6437].

4.3. END.DB6

For the use-case mentioned under Section 3.2 END.DB6 SID is applicable.

The End.DB6 SID MUST be the last segment in SRH or SR Policy.

Node N is configured with an IPv6 address T (e.g., assigned to its loopback).

When Node N receives a packet destined to S and S is a locally instantiated End.DB6 SID, Node N executes the following procedure:


```

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.   If (Upper-Layer header type == 4(IPv4) OR Upper-Layer header type == 4
1(IPv6) OR
        Upper-Layer header type == 143(Ethernet)) {
S06.     Remove the outer IPv6 header with all its extension headers.
S07.     Push the new IPv6 header with the SRv6 SIDs associated with the END.D
B6 sid in an SRH.
S08.     Set outer IPv6 SA = T and outer IPv6 DA to the first SID in the segme
nt list.
S09.     Set outer Payload Length, Traffic Class, Hop Limit, and Flow Label fi
elds
S10.     Set the outer Next Header value
S11.     Submit the packet to the IPv6 module for transmission to First SID.
S12.   } else {
S13.     Process as per Section 4.1.1 of [RFC8986].
S14.   }
S15. }

```

Note :

S09. The Payload Length, Traffic Class, Hop Limit, and Next Header fields are set as per [RFC2473]. The Flow Label is computed as per [RFC6437].

5. Interworking Procedures

Here we will describe the control plane and data plane procedures by taking examples.

Node n has a classic IPv6 loopback address An::<1/128. One of the SID at node n with locator block B and function F is represented by B:n:F::sid_num.

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.

5.1. Option C Transport Interworking

Here we will discuss the use-case mentioned under Section 3.1

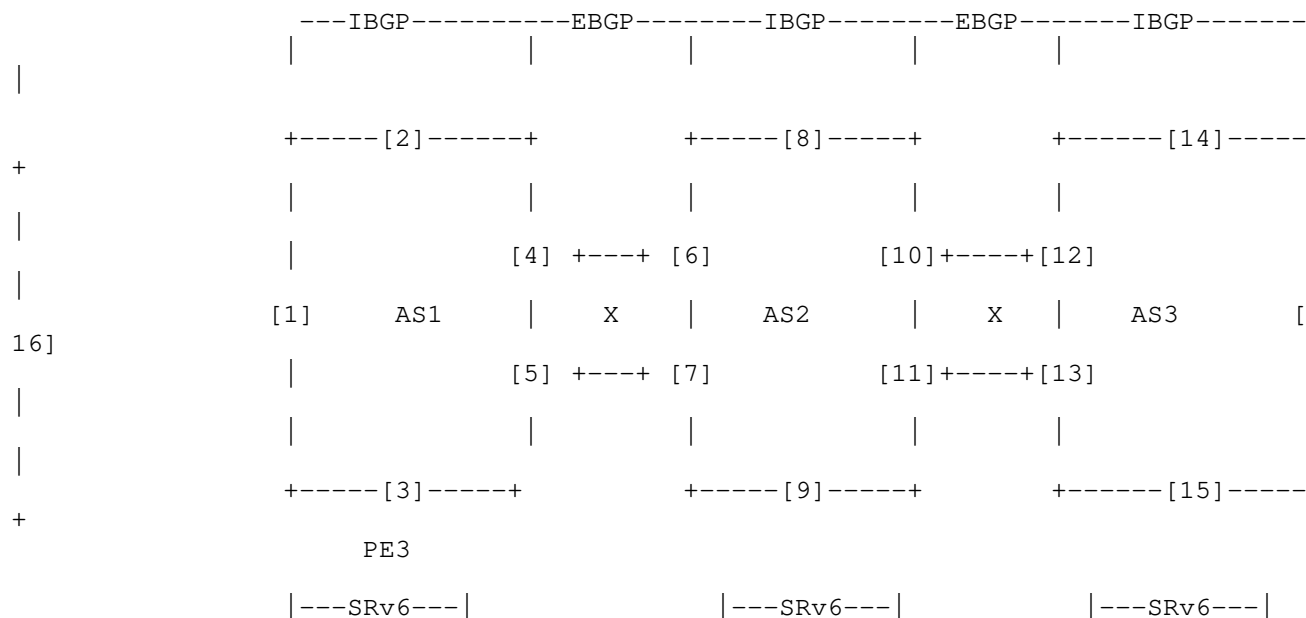


Figure 3: Option C Style Interworking

Node [1] acts as ingress PE and Node [16] acts as egress PE.

Nodes [2], [3], [8], [9], [14] and [15] are P routers.

Nodes [4], [5], [6], [7], [10], [11], [12] and [13] are ASBR routers.

A VPN route is advertised via service RRs between an egress PE (node 16) and an ingress PE (node 1). The example below shows IBGP-CT connection between border routers in each domain and single hop EBGP-CT for inter-domain connections. However the forwarding procedure for the sids remains the same irrespective of the the various inter-domain protocol extensions used to advertise the sids. AS1, AS2 and AS3 has SRTE policy for the required intent paths.

Control plane example:

For simplicity only one path is tracked.

For a route if the next hop is one hop away then while advertising use END.REPLACE SID. For a route if the next hop is multi hop away then while advertising use END.REPLACEB6 SID. For single hop neighbor case, no encaps required as it is just replace and forward on specific link while in multihop case one encaps will be required.

Routing Protocol(RP) @16:
* In ISIS advertise locator B:16::/48 and an END SID B:16::END::1.
* BGP AFI=1,SAFI=128 originates a VPN route RD:V/v via A:16::1 and Prefix-SID attribute B:16:DT4::1.
This route is advertised to service RR with color extended community red.
* BGP originates prefix A:16::1 with color red to ASBR [12] with SRv6 SID B:16:END::1 since its the egress node.
RP @12:
* BGP receives the route A:16::1 over the ibgp session and readvertises with nexthop self to ASBR [10].
it advertises the SRv6 SID B:12:End.B6.Encaps::1 in the protocol extensions. As the prefix A:16::1 advertisement was received with End SID, this node allocates a End.B6.Encaps sid.
RP @10:
* BGP receives the route A:16::1 over the ebgp session and readvertises with nexthop self to ASBR [6].
it advertises the SRv6 SID B:10:REPLACE::1 in the protocol extensions. As the advertisement was received on a single hop e-bgp session this node allocates a REPLACE sid.
RP @6:
* BGP receives the route A:16::1 over the ibgp session and readvertises with nexthop self to ASBR [4].
it advertises the SRv6 SID B:6:REPLACEB6::1 in the protocol extensions. As the advertisement was received on a multihop i-bgp session this node allocates a REPLACEB6 sid.
RP @4:
* BGP receives the route A:16::1 over the ebgp session and readvertises with nexthop self to PE [1].
it advertises the SRv6 SID B:4:REPLACE::1 in the protocol extensions. As the advertisement was received on a single hop e-bgp session this node allocates a REPLACE sid.
RP @1:
* BGP receives the route A:16::1 with color red over the ibgp session.
* BGP AFI=1, SAFI=128 learn service prefix RD:V/v, next hop A:16::1 and PrefixSID attribute TLV type 5 with SRv6 SID B:16:DT4

FIB State:

```

    @1: IPv4 VRF V/v => H.Encaps.red <B:2:END::1, B:4:REPLACE::1, B:16:DT4::1
> with SRH, SRH NextHeader=IPv4 where the first
    sid B:2:END::1 belongs to the SR-policy in AS1.
    @2: IPv6 Table: B:2:END::1 => Update DA with B:4:REPLACE::1, decrement SL
and forward towards the ASBR [4].
    @4: IPv6 Table: B:4:REPLACE::1 => Update DA with B:6:REPLACEB6::1 and for
ward on the interface/interfaces identified by the
    ebgp neighbor; the SL remains at 1.
    @6: IPv6 Table: B:6:REPLACEB6::1 => Update DA with B:10:REPLACE::1 AND do
a fresh H.Encaps.red <B:8:END::1, B:10:END::1>
    with SRH where the new SRH SIDs belong to SR policy in AS2.
    @8: IPv6 Table: B:8:END::1 => Update outer IPv6 packet DA with B:10:END::
1 and forward towards ASBR [10]
    @10: IPv6 table: B:10:END::1 => Decap Outer IPv6 header and lookup next I
Pv6 DA B:10:REPLACE::1 => Update DA to B:12:End.B6.Encaps::1
    and forward on the interface/interfaces identified by the ebgp neighb
our. SL remains at 1.
    @12: IPv6 Table B:12:End.B6.Encaps::1 => Update DA with Next Segment in S
RH(B:16:DT4::1)
    and do a fresh H.Encaps.red <B:15:END::1, B:16:END::1> with SRH, whe
re the new SIDs belong to the SR policy in AS3.
    @15: IPv6 Table B:15:END::1 => Update outer IPv6 packet DA with B:16:END:
:1 and forward towards [16].
    @16: IPv6 Table B:16:END::1 => Decap the outer header and lookup the inne
r DA which results in B:16:DT4::1 lookup. DT4 lookup
    results in Decap and inner IPv4 packet DA lookup in the correspondin
g VRF.

```

Note: At [1] we have optimized the solution by single Encapsulation with a SRH header. This can be supported by Most of the ASICs. Here we can even use two encapsulation, this mechanism will also work.

5.2. Option B service interworking

Here we will discuss the use-case mentioned under Section 3.2

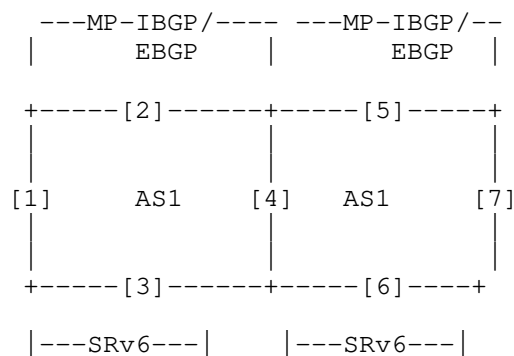


Figure 4: Option B style Service Interworking

Nodes [1] and [7] are PE routers. Node [4] is an option B style configured ABR/RR.

Control Plane example:

Routing Protocol (RP) @7:
* BGP AFI=1,SAFI=128 originates a VPN route RD:V/v via A:7::1 and Prefix-SID attribute B:7:DT4::1. This route is advertised to service RR [4].

RP @4:
* BGP receives the route over MP-IBGP/MP-EBGP session and readvertises with next hop self to PE [1].
it advertises the SRv6 SID B:4:DB6::1 in the Prefix-SID attribute TLV along with it. For all prefixes having SRv6 service SID B:7:DT4::1; the same DB6 SID B:4:DB6::1 will be reused. if a different service sid B:7:DT4::2 comes then a different DB6 SID B:4:DB6::2 will be allocated. This ensures that if the egress allocates per CE sid; the translation at border also ensure per CE sid.

RP @1:
* BGP AFI=1, SAFI=128 learn service prefix RD:V/v, next hop A:4::1 and PrefixSID attribute TLV type 5 with SRv6 SID B:4:DB6::1

FIB State:

@1: IPv4 VRF V/v => H.Encaps.red <B:4:DB6::1> with SRH, SRH NextHeader=IPv4 where the first sid belongs to the SR-policy in AS1
@4: IPv6 Table: B:4:DB6::1 => Decapsulate the incoming IPv6 header and H.Encaps <B:7:DT4::1>
@7: IPv6 Table: B:7:DT4::1 => Decapsulate the header and lookup the inner IPv4 packet DA in the VRF

6. IANA Considerations

This document requires no IANA action.

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. Contributors

Salih K A
Juniper Networks
Email: salih@juniper.net

Shraddha Hegde
Juniper Networks
Email: shraddha@juniper.net

Jie Dong
Huawei Technologies
Email: jie.dong@huawei.com

Swamy SRK
Juniper Networks
Email: swamys@juniper.net

G. Sri Karthik Goud
Juniper Networks
Email: gkarthik@juniper.net

9. Acknowledgements

Thanks to Ram Santhanakrishnan, Srihari Sangli, Rajendra Prasad Bollam and Kiran Kushalad for their valuable comments.

10. References

10.1. 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>>.
- [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>>.
- [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>>.

- [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>>.
- [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>>.

10.2. 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 L. M. Contreras, "Seamless SR Problem Statement", Work in Progress, Internet-Draft, draft-hegde-spring-mpls-seamless-sr-07, 8 July 2022, <<https://datatracker.ietf.org/doc/html/draft-hegde-spring-mpls-seamless-sr-07>>.
- [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

Rajesh
Juniper Networks
Embassy Business Park
Bangalore 560093
KA
India
Email: mrajesh@juniper.net

Ron Bonica
Juniper Networks
Herndon, Virginia 20171
United States of America
Email: rbonica@juniper.net

Haibo Wang
Huawei Technologies
Huawei Campus, No. 156 Beiqing Road
Beijing
100095
China
Email: rainsword.wang@huawei.com

Peng Shaofu
ZTE Corporation
China
Email: peng.shaofu@zte.com.cn