BESS Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 12, 2019

G. Dawra, Ed.
LinkedIn
C. Filsfils
D. Dukes
P. Brissette
S. Sethuram
P. Camarilo
Cisco Systems
J. Leddy
Comcast
D. Voyer
D. Bernier
Bell Canada
D. Steinberg
Steinberg Consulting
R. Raszuk
Bloomberg LP
B. Decraene
Orange
S. Matsushima
SoftBank
S. Zhuang
Huawei Technologies
March 11, 2019

SRv6 BGP based Overlay services
draft-dawra-bess-srv6-services-00

Abstract

This draft defines procedures and messages for SRv6-based BGP services including L3VPN, EVPN and Internet services. It builds on RFC4364 "BGP/MPLS IP Virtual Private Networks (VPNs)" and RFC7432 "BGP MPLS-Based Ethernet VPN" and provides a migration path from MPLS-based VPNs to SRv6 based VPNs.

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 RFC8174 [RFC8174].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

SRv6 refers to Segment Routing instantiated on the IPv6 dataplane [I-D.filsfils-spring-srv6-network-programming][I-D.ietf-6man-segment-routing-header].

SRv6 based BGP services refers to the L3 and L2 overlay services with BGP as control plane and SRv6 as dataplane.

SRv6 SID refers to a SRv6 Segment Identifier as defined in [I-D.filsfils-spring-srv6-network-programming].

SRv6 Service SID refers to an SRv6 SID that MAY be associated with one of the service specific behavior on the advertising Provider Edge(PE) router, such as (but not limited to), in the case of L3VPN service, END.DT (Table lookup in a VRF) or END.DX (crossconnect to a nexthop) functions as defined in [I-D.filsfils-spring-srv6-network-programming].

To provide SRv6 service with best-effort connectivity, the egress PE signals an SRv6 Service SID with the BGP overlay service route. The ingress PE encapsulates the payload in an outer IPv6 header where the destination address is the SRv6 Service SID provided by the egress PE. The underlay between the PEs only need to support plain IPv6 forwarding [RFC2460].

To provide SRv6 service in conjunction with an underlay SLA from the ingress PE to the egress PE, the egress PE colors the overlay service route with a Color extended community[I-D.ietf-idr-segment-routing-te-policy]. The ingress PE encapsulates the payload packet in an outer IPv6 header with an SRH that contains the SR policy associated with the related SLA followed by the SRv6 Service SID associated with the route. The underlay nodes whose SRv6 SID's are part of the SRH must support SRv6 data plane.

BGP is used to advertise the reachability of prefixes of a particular service from an egress PE to ingress PE nodes.

This document describes how existing BGP messages between PEs may carry SRv6 Service SIDs as a means to interconnect PEs and form VPNs.

2. SRv6 Services TLVs

This document extends the BGP Prefix-SID attribute [I-D.ietf-idr-bgp-prefix-sid] to carry SRv6 SIDs and associated information.

The SRv6 Service TLVs are defined as two new TLVs of the BGP Prefix-SID Attribute to achieve signaling of SRv6 SIDs for L3 and L2 services.

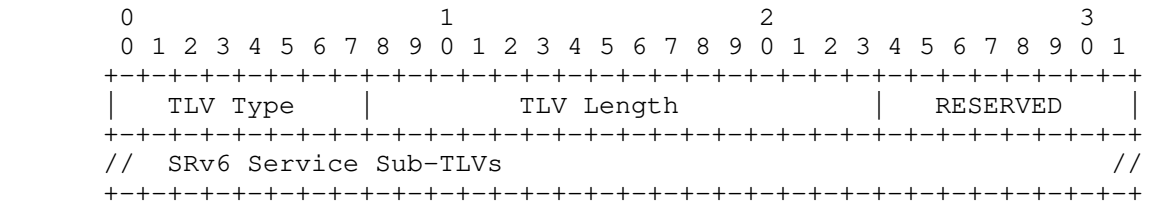
- SRv6 L3 Service TLV: This TLV encodes Service SID information for SRv6 based L3 services. It corresponds to the equivalent functionality provided by an MPLS Label when received with a Layer 3 service route. Some functions which may be encoded, but not limited to, are End.DX4, End.DT4, End.DX6, End.DT6, etc.

- SRv6 L2 Service TLV: This TLV encodes Service SID information for SRv6 based L2 services. It corresponds to the equivalent functionality provided by an MPLS Label1 for EVPN Route-Types as defined in[RFC7432]. Some functions which may be encoded, but not limited to, are End.DX2, End.DX2V, End.DT2U, End.DT2M etc.

BGP Prefix-SID Attribute [I-D.ietf-idr-bgp-prefix-sid] is referred to as BGP SID Attribute in the rest of the document.

When an egress PE is capable of SRv6 data-plane, it SHOULD signal one or more SRv6 Service SIDs enclosed in SRv6 Service TLV(s) within the BGP SID Attribute attached to MP-BGP NLRIs defined in [RFC4760][RFC4659][RFC5549][RFC7432][RFC4364].

The following depicts the SRv6 Service TLVs encoded in the BGP SID attribute:



- o TLV Type (1 octet): This field is assigned values from the IANA registry "BGP Prefix-SID TLV Types". It is set to [TBD1] (to be assigned by IANA) for SRv6 L3 Service TLV. It is set to [TBD2] (to be assigned by IANA) for SRv6 L2 Service TLV.

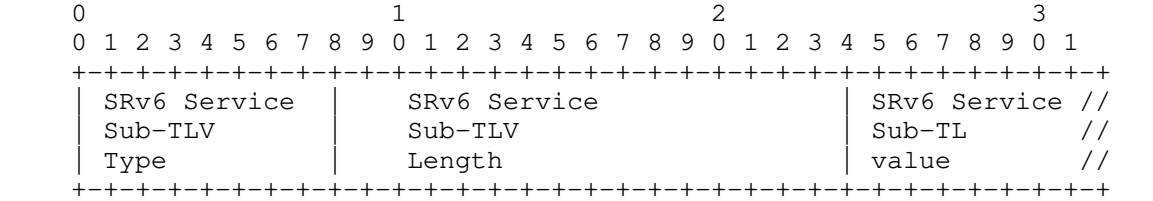
- o TLV Length (2 octets): Specifies the total length of the TLV Value.

- o RESERVED (1 octet): This field is reserved; it SHOULD be set to 0 by the sender and MUST be ignored by the receiver.

- o SRv6 Service Sub-TLVs (variable): This field contains SRv6 Service related information and is encoded as an unordered list of Sub-TLVs whose format is described below.

2.1. SRv6 Service Sub-TLVs

The format of a single SRv6 Service Sub-TLV is depicted below:



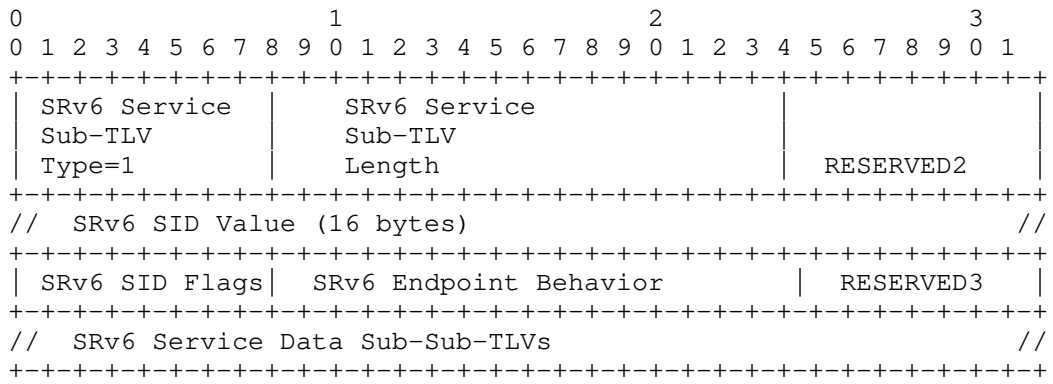
- o SRv6 Service Sub-TLV Type (1 octet): Identifies the type of SRv6 service information. It is assigned values from the IANA Registry "SRv6 Service Sub-TLV Types".

- o SRv6 Service Sub-TLV Length (2 octets): Specifies the total length of the Sub-TLV Value field.

- o SRv6 Service Sub-TLV Value (variable): Contains data specific to the Sub-TLV Type. In addition to fixed length data, this may also optionally contain other properties of the SRv6 Service encoded as a set of SRv6 Service Data Sub-sub-TLVs whose format is described in another sub-section below.

2.1.1. SRv6 SID Information Sub-TLV

SRv6 Service Sub-TLV Type 1 is assigned for SRv6 SID Information Sub-TLV. This Sub-TLV contains a single SRv6 SID along with its properties. Its encoding is depicted below:



- SRv6 Service Sub-TLV Type (1 octet): This field is set to 1 to represent SRv6 SID Informaton Sub-TLV.

- SRv6 Service Sub-TLV Length (2 octets): This field contains the total length of the Value field of the Sub-TLV.

- RESERVED2 (1 octet): SHOULD be set to 0 by the sender and MUST be ignored by the receiver.

- SRv6 SID Value (16 octets): Encodes an SRv6 SID as defined in [I-D.filsfils-spring-srv6-network-programming]

- SRv6 SID Flags (1 octet): Encodes SRv6 SID Flags - none are currently defined.

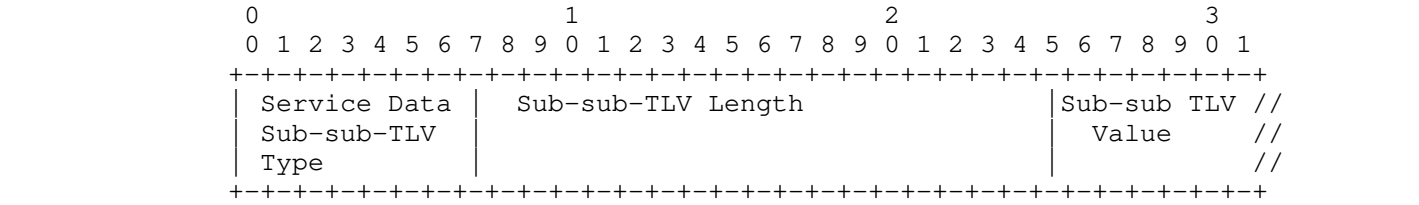
- SRv6 Endpoint Behavior (2 octets): Encodes SRv6 Endpoint behavior defined in [I-D.filsfils-spring-srv6-network-programming]. This field MUST be set to the Reserved value 0xFFFF.

- RESERVED3 (1 octet): SHOULD be set to 0 by the sender and MUST be ignored by the receiver.

- SRv6 Service Data Sub-TLV Value (variable): This field contains optional properties of the SRv6 SID. It is encoded as a set of SRv6 Service Data Sub-Sub-TLVs. None are applicable at this time.

2.1.2. SRv6 Service Data Sub-Sub-TLVs

The format of the SRv6 Service Data Sub-Sub-TLV is depicted below:



- SRv6 Service Data Sub-Sub-TLV Type (1 octet): Identifies the type of Sub-Sub-TLV. It is assigned values from the IANA Registry "SRv6 Service Data Sub-Sub-TLVs".

- SRv6 Service Data Sub-Sub-TLV Length (2 octets): Specifies the total length of the Sub-Sub-TLV Value field.

- SRv6 Service Data Sub-Sub-TLV Value (variable): Contains data specific to the Sub-Sub-TLV Type.

At this time, no Sub-Sub-TLV Types are defined.

3. BGP based L3 service over SRv6

BGP egress nodes (egress PEs) advertise a set of reachable prefixes. Standard BGP update propagation schemes[RFC4271], which may make use of route reflectors [RFC4456], are used to propagate these prefixes. BGP ingress nodes (ingress PEs) receive these advertisements and may add the prefix to the RIB in an appropriate VRF.

Egress PEs which supports SRv6 based L3 services advertises overlay service prefixes along with a Service SID enclosed in a SRv6 L3 Service TLV within the BGP SID attribute. This TLV serves two purposes - first, it indicates that the egress PE is reachable via an SRv6 underlay and the BGP ingress PE receiving this route MAY choose to encapsulate or insert an SRv6 SRH; second ,it indicates the value of the SID to include in the SRH encapsulation.

The Service SID thus signaled only has local significance at the egress PE, where it may be allocated or configured on a per-CE or per-VRF basis. In practice, the SID may encode a cross-connect to a specific Address Family table (END.DT) or next-hop/interface (END.DX) as defined in the SRv6 Network Programming Document [I-D.filsfils-spring-srv6-network-programming].

The SRv6 Service SID MAY be routable within the AS of the egress PE and serves the dual purpose of providing reachability between ingress PE and egress PE while also encoding the endpoint behavior.

If the BGP speaker supports MPLS based L3VPN services simultaneously, it MAY also populate a valid Label value in the service route NLRI encoding, and allow the BGP ingress PE to decide which encapsulation to use. If the BGP speaker does not support MPLS based L3VPN services the Label value in any service route NLRI encoding MUST be set to Implicit NULL [RFC3032].

At an ingress PE, BGP installs the received prefix in the correct RIB table, recursing via an SR Policy leveraging the received SRv6 Service SID.

Assuming best-effort connectivity to the egress PE, the SR policy has a path with a SID list made up of a single SID - the SRv6 Service SID received with the related BGP route update.

However, when the received route is colored with an extended color commmunity 'C' and Next-Hop 'N', and the ingress PE has a valid SRv6 Policy (C, N) associated with SID list <S1,S2, S3> [I-D.filsfils-spring-segment-routing-policy], then the effective SR Policy is <S1, S2, S3, SRv6-Service-SID>.

Multiple VPN routes MAY resolve recursively via the same SR Policy.

3.1. IPv4 VPN Over SRv6 Core

IPv4 VPN Over IPv6 Core is defined in [RFC5549]. The MP_REACH_NLRI is encoded as follows for an SRv6 Core:

- AFI = 1
- SAFI = 128
- Length of Next Hop Network Address = 16 (or 32)
- Network Address of Next Hop = IPv6 address of the egress PE
- NLRI = IPv4-VPN routes
- Label = Implicit NULL

SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The function of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the function may likely be End.DX4 or End.DT4.

3.2. IPv6 VPN Over SRv6 Core

IPv6 VPN over IPv6 Core is defined in [RFC4659]. The MP_REACH_NLRI is encoded as follows for an SRv6 Core:

- AFI = 2
- SAFI = 128
- Length of Next Hop Network Address = 24 (or 48)
- Network Address of Next Hop = 8 octets of RD set to 0 followed by IPv6 address of the egress PE
- NLRI = IPv6-VPN routes
- Label = Implicit NULL

SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The function of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the function may likely be End.DX6 or End.DT6.

3.3. Global IPv4 over SRv6 Core

IPv4 over IPv6 Core is defined in [RFC5549]. The MP_REACH_NLRI is encoded with:

- AFI = 1
- SAFI = 1
- Length of Next Hop Network Address = 16 (or 32)
- Network Address of Next Hop = IPv6 address of Next Hop
- NLRI = IPv4 routes

SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The function of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the function may likely be End.DX4/6 or End.DT4.

3.4. Global IPv6 over SRv6 Core

The MP_REACH_NLRI is encoded with:

- AFI = 2

- SAFI = 1
- Length of Next Hop Network Address = 16 (or 32)
- Network Address of Next Hop = IPv6 address of Next Hop
- NLRI = IPv6 routes

SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The function of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the function may likely be End.DX4/6 or End.DT6.

Also, by utilizing the SRv6 L3 Service TLV to encode the Global SID, a BGP free core is possible by encapsulating all BGP traffic from edge to edge over SRv6.

4. BGP based Ethernet VPN (EVPN) over SRv6

Ethernet VPN(EVPN), as defined in [RFC7432] provides an extendable method of building an EVPN overlay. It primarily focuses on MPLS based EVPNs but calls out the extensibility to IP based EVPN overlays. [RFC7432] defines 4 Route Types which carry prefixes and MPLS Label fields; the Label fields have specific use for MPLS encapsulation of EVPN traffic. Route Type 5 carrying MPLS label information (and thus encapsulation information) for EVPN is defined in [I-D.ietf-bess-evpn-prefix-advertisement]. Route Types 6, 7 and 8 are defined in [I-D.ietf-bess-evpn-igmp-mld-proxy].

- Ethernet Auto-discovery Route (Route Type 1)
- MAC/IP Advertisement Route (Route Type 2)
- Inclusive Multicast Ethernet Tag Route (Route Type 3)
- Ethernet Segment route (Route Type 4)
- IP prefix route (Route Type 5)
- Selective Multicast Ethernet Tag route (Route Type 6)
- IGMP join sync route (Route Type 7)
- IGMP leave sync route (Route Type 8)

To support SRv6 based EVPN overlays, one or more SRv6 Service SIDs are advertised with Route Type 1,2,3 and 5. The SRv6 Service SID(s) per Route Type are advertised in SRv6 L3/L2 Service TLVs within the

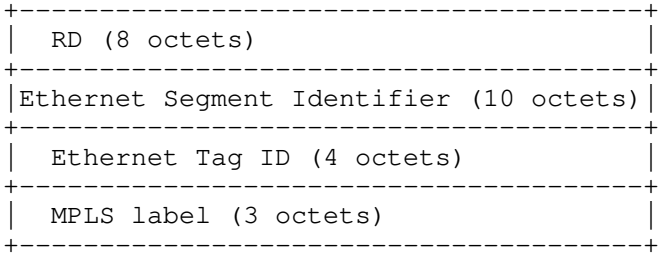
BGP SID attribute. Signaling of SRv6 Service SID(s) serves two purposes - first, it indicates that the BGP egress device is reachable via an SRv6 underlay and the BGP ingress device receiving this route MAY choose to encapsulate or insert an SRv6 SRH; second, it indicates the value of the SID(s) to include in the SRH encapsulation. If the BGP egress device does not support MPLS based EVPN services, the MPLS Label fiellds withing EVPN Route Types MUST be set to Implicit NULL.

4.1. Ethernet Auto-discovery route over SRv6 Core

Ethernet Auto-Discovery (A-D) routes are Route Type 1 defined in [RFC7432]and may be used to achieve split horizon filtering, fast convergence and aliasing. EVPN Route Type 1 is also used in EVPN-VPWS as well as in EVPN flexible cross-connect; mainly used to advertise point-to-point services ID.

Multi-homed PEs MAY advertise an Ethernet Auto-Discovery route per Ethernet segment along with the ESI Label extended community defined in [RFC7432]. The extended community label MUST be set to Implicit NULL. PEs may identify other PEs connected to the same Ethernet segment after the EVPN Route Type 4 ES route exchange. All the multi-homed and remote PEs that are part of same EVI may import the Auto-Discovery route.

EVPN Route Type 1 is encoded as follows for SRv6 Core:



4.1.1. Per-ES A-D route

- BGP next-hop: IPv6 address of an egress PE
- Ethernet Tag ID: set to 0xFFFF
- MPLS Label: always set to zero value
- Extended Community: Per ES AD, ESI label extended community

A Service SID enclosed in a SRv6 L2 Service TLV within the BGP SID attribute is advertised along with the A-D route. The behavior of the Service SID thus signaled is entirely up to the originator of the advertisement. This is typically used to signal Arg.FE2 SID argument for applicable End.DT2M SIDs.

4.1.2. Per-EVI A-D route

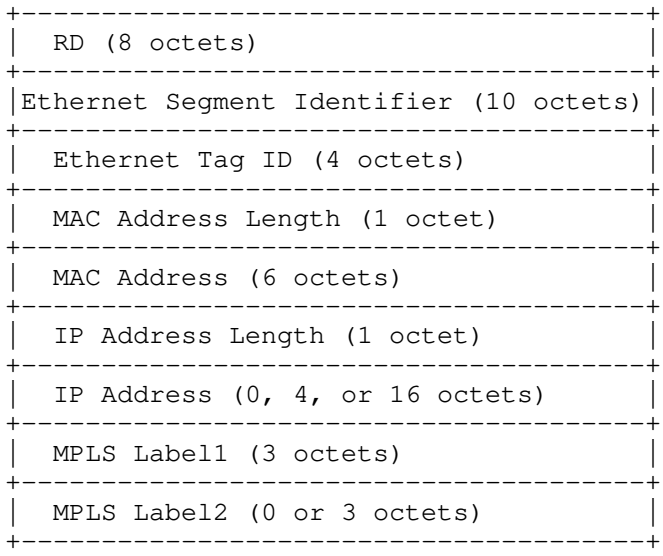
- BGP next-hop: IPv6 address of an egress PE
- Ethernet Tag ID: non-zero for VLAN aware bridging, EVPN VPWS and FXC
- MPLS Label: Implicit NULL

A Service SID enclosed in a SRv6 L2 Service TLV within the BGP SID attribute is advertised along with the A-D route. The behavior of the Service SID thus signaled is entirely up to the originator of the advertisement. In practice, the behavior would likely be END.DX2, END.DX2V or END.DT2U.

4.2. MAC/IP Advertisement route over SRv6 Core

EVPN Route Type 2 is used to advertise unicast traffic MAC+IP address reachability through MP-BGP to all other PEs in a given EVPN instance.

EVPN Route Type 2 is encoded as follows for SRv6 Core:



- BGP next-hop: IPv6 address of an egress PE
- MPLS Label1: Implicit NULL
- MPLS Label2: Implicit NULL

Service SIDs enclosed in SRv6 L2 Service TLV and optionally in SRv6 L3 Service TLV within the BGP SID attribute is advertised along with the MAC/IP Advertisement route.

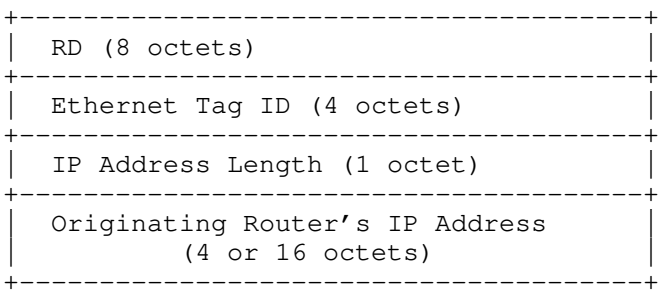
Described below are different types of Route Type 2 advertisements.

- MAC/IP Advertisement route with MAC Only
  - BGP next-hop: IPv6 address of egress PE
  - MPLS Label1: Implicit NULL
  - MPLS Label2: Implicit NULL
- A Service SID enclosed in a SRv6 L2 Service TLV within the BGP SID attribute is advertised along with the route. The behavior of the Service SID thus signaled is entirely up to the originator of the advertisement. In practice, the behavior would likely be END.DX2 or END.DT2U.
- MAC/IP Advertisement route with MAC+IP
  - BGP next-hop: IPv6 address of egress PE
  - MPLS Label1: Implicit NULL
  - MPLS Label2: Implicit NULL
- An L2 Service SID enclosed in a SRv6 L2 Service TLV within the BGP SID attribute is advertised along with the route. In addition, an L3 Service SID enclosed in a SRv6 L3 Service TLV within the BGP SID attribute MAY also be advertised along with the route. The behavior of the Service SID(s) thus signaled is entirely up to the originator of the advertisement. In practice, the behavior would likely be END.DX2 or END.DT2U for the L2 Service SID, and END.DT6/4 or END.DX6/4 for the L3 Service SID.

4.3. Inclusive Multicast Ethernet Tag Route over SRv6 Core

EVPN Route Type 3 is used to advertise multicast traffic reachability information through MP-BGP to all other PEs in a given EVPN instance.

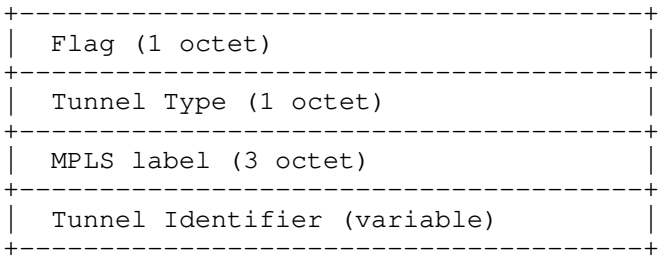
EVPN Route Type 3 is encoded as follows for SRv6 core:



- BGP next-hop: IPv6 address of egress PE

PMSI Tunnel Attribute [RFC6514] MAY contain MPLS Implicit NULL label and Tunnel Type would be similar to that defined in EVPN Route Type 6 i.e. Ingress replication route.

The format of PMSI Tunnel Attribute attribute is encoded as follows for SRv6 Core:



- Flag: zero value defined per [RFC7432]

- Tunnel Type: defined per [RFC6514]

- MPLS label: Implicit NULL

- Tunnel Identifier: IP address of egress PE

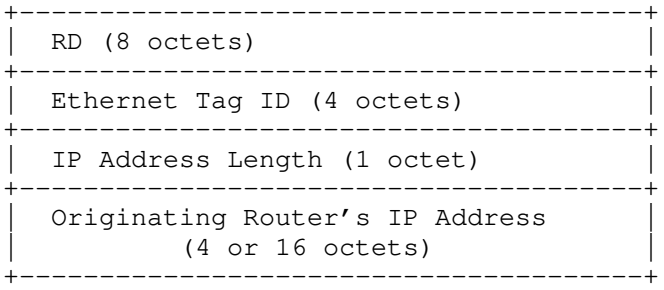
A Service SID enclosed in a SRv6 L2 Service TLV within the BGP SID attribute is advertised along with the route. The behavior of the Service SID thus signaled, is entirely up to the originator of the advertisement. In practice, the behavior of the SRv6 SID is as follows:

- END.DX2 or END.DT2M function

- The ESI Filtering argument (Arg.FE2) of the Service SID carried along with EVPN Route Type 1 route MAY be merged together with the applicable End.DT2M SID of Type 3 route advertised by remote PE by doing a bitwise logical-OR operation to create a single SID on the ingress PE for Split-horizon and other filtering mechanisms. Details of filtering mechanisms are described in [RFC7432].

4.4. Ethernet Segment route over SRv6 Core

An Ethernet Segment route i.e. EVPN Route Type 4 is encoded as follows for SRv6 core:



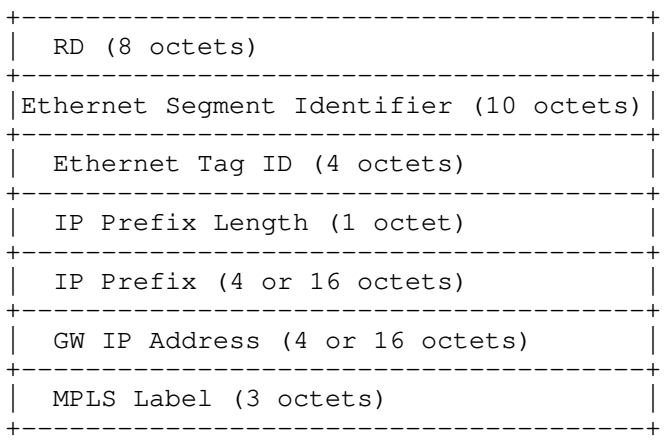
- BGP next-hop: IPv6 address of egress PE

SRv6 Service TLVs within BGP SID attribute are not advertised along with this route. The processing of the route has not changed - it remains as described in [RFC7432].

4.5. IP prefix route over SRv6 Core

EVPN Route Type 5 is used to advertise IP address reachability through MP-BGP to all other PEs in a given EVPN instance. IP address may include host IP prefix or any specific subnet.

EVPN Route Type 5 is encoded as follows for SRv6 core:



- o BGP next-hop: IPv6 address of egress PE
- o MPLS Label: Implicit NULL

SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The function of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the function may likely be End.DT6/4 or End.DX6/4.

4.6. EVPN multicast routes (Route Types 6, 7, 8) over SRv6 core

These routes do not require the advertisement of SRv6 Service TLVs along with them. Similar to EVPN Route Type 4, the BGP Nexthop is equal to the IPv6 address of egress PE. More details may be added in future revisions of this document.

5. Migration from MPLS based Segment Routing to SRv6 Segment Routing

Migration from IPv4 to IPv6 is independent of SRv6 BGP endpoints, and the selection of which route to use (received via the IPv4 or the IPv6 session) is a local configurable decision of the ingress PE, and is outside the scope of this document.

Migration from MPLS based underlay to an SRv6 underlay with BGP speakers is achieved with a few simple rules at each BGP speaker.

At Egress PE:
If BGP offers an SRv6 service, then:
BGP allocates an SRv6 Service SID for the L3 service
and includes it in the BGP SRv6 L3 Service TLV while advertising the
overlay prefixes.
If BGP offers an MPLS service, then:
BGP allocates an MPLS Label for the L3 service and
encode it as part of the NLRI as normal for MPLS based address-families;
else, the MPLS label value for the L3 service is set to Implicit NULL.

At Ingress PE:
Selection of either MPLS encapsulation or SRv6 encapsulation is defined
by local BGP policy.

If BGP supports SRv6 based services and receives overlay routes with
BGP SID attribute containing SRv6 L3 Service TLV(s) encoding
SRv6 Service SID(s), then:
BGP programs the destination prefix in RIB recursive via
the related SR Policy.
If BGP supports MPLS service, and the MPLS Label value is not
Implicit NULL, then:
the MPLS label is used as the overlay service label and inserted with the
prefix into RIB via the BGP Nexthop.

6. Implementation Status

The SRv6 Service is available for SRv6 on various Cisco hardware and other software platforms. An end-to-end integration of SRv6 L3VPN, SRv6 Traffic-Engineering and Service Chaining. All of that with data-plane interoperability across <http://www.segment-routing.net> different implementations:

- Three Cisco Hardware-forwarding platforms: ASR 1K, ASR 9k and NCS 5500
- Two Cisco network operating systems: IOS XE and IOS XR
- Huawei Hardware-forwarding platforms: ATN, CX, ME, NE5000E, NE9000, NG-OLT
- Huawei network operating systems: VRPv8
- Barefoot Networks Tofino on OCP Wedge-100BF
- Linux Kernel officially upstreamed in 4.10
- Fd.io

7. Error Handling

In case of any errors encountered while processing SRv6 Service TLVs, the details of the error SHOULD be logged for further analysis.

If multiple instances of SRv6 L3 Service TLV is encountered, all but the first instance MUST be ignored.

If multiple instances of SRv6 L2 Service TLV is encountered, all but the first instance MUST be ignored.

An SRv6 Service TLV is considered malformed in the following cases:

- the TLV Length is less than 1
- the TLV Length is inconsistent with the length of BGP SID attribute
- atleast one of the constituent Sub-TLVs is malformed

An SRv6 Service Sub-TLV is considered malformed in the following cases:

- the Sub-TLV Length is inconsistent with the length of the enclosing SRv6 Service TLV

An SRv6 SID Information Sub-TLV is considered malformed in the following cases:

  * the Sub-TLV Length is less than 21
  * the Sub-TLV Length is inconsistent with the length of the enclosing SRv6 Service TLV
  * atleast one of the constituent Sub-Sub-TLVs is malformed

An SRv6 Service Data Sub-sub-TLV is considered malformed in the following cases:

- the Sub-Sub-TLV Length is inconsistent with the length of the enclosing SRv6 service Sub-TLV

Any TLV or Sub-TLV or Sub-Sub-TLV is not considered malformed because its Type is unrecognized.

Any TLV or Sub-TLV or Sub-Sub-TLV is not considered malformed because of failing any semantic validation of its Value field.

The BGP SID attribute is considered malformed if it contains atleast one constituent SRv6 Service TLV that is malformed. In such cases, the attribute MUST be discarded [RFC7606]and not propagated further. Note that if a path whose BGP SID attribute is discarded in this manner is selected as the best path to be installed in the RIB, traffic forwarding for the corresponding prefix may be affected. Implementations MAY choose to make such paths less preferable or even ineligible during the selection of best path for the corresponding prefix.

A BGP speaker receiving a path containing BGP SID attribute with one or more SRv6 Service TLVs observes the following rules when advertising the received path to other peers:

- if the nexthop is unchanged during advertisement, the SRv6 Service TLVs, including any unrecognized Types of Sub-TLV and Sub-Sub-TLV, SHOULD be propagated further. In addition, all Reserved fields in the TLV or Sub-TLV or Sub-Sub-TLV MUST be propagated unchanged.

- if the nexthop is changed during advertisement, any unrecognized Sub-TLVs and Sub-Sub-TLVs MUST NOT be propagated.

- if the nexthop is changed during advertisement, the TLVs, Sub-TLVs and Sub-Sub-TLVs SHOULD be re-originated if appropriate, and not merely propagated unchanged. The interpretation of the meaning of re-origination versus propagation is a matter of local implementation.

A received VPN NLRI [RFC4364][RFC4659][RFC7432]that has neither a valid MPLS label nor a valid SRv6 Service TLV MUST be considered unreachable i.e. apply the -treat as withdraw- action specified in [RFC7606].

8. IANA Considerations

8.1. BGP Prefix-SID TLV Types registry

This document defines two new TLV Types of the BGP Prefx-SID attribute. IANA is requested to assign Type values in the registry "BGP Prefix-SID TLV Types" as follows:

Value	Type	Reference
[TBD1]	SRv6 L3 Service TLV	<this document>
[TBD2]	SRv6 L2 Service TLV	<this document>

IANA is also requested to reserve the following Type value. This was used in some implementations of previous versions of this draft.

Value	Type	Reference
4	Reserved	<this document>

8.2. SRv6 Service Sub-TLV Types registry

IANA is requested to create and maintain a new registry called "SRv6 Service Sub-TLV Types". The allocation policy for this registry is:

0 : Reserved
1-127 : IETF Review
128-254 : First Come First Served
255 : Reserved

The following Sub-TLV Types are defined in this document:

Value	Type	Reference
1	SRv6 SID Information Sub-TLV	<this document>

9. Security Considerations

This document introduces no new security considerations beyond those already specified in [RFC4271] and [RFC8277].

10. Conclusions

This document proposes extensions to the BGP to allow advertising certain attributes and functionalities related to SRv6.

11. References

11.1. Normative References

[I-D.filsfils-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., Hegde, S., daniel.voyer@bell.ca, d., Lin, S., bogdanov@google.com, b., Krol, P., Horneffer, M., Steinberg, D., Decraene, B., Litkowski, S., Mattes, P., Ali, Z., Talaulikar, K., Liste, J., Clad, F., and K. Raza, "Segment Routing Policy Architecture", draft-filsfils-spring-segment-routing-policy-06 (work in progress), May 2018.

[I-D.filsfils-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J., daniel.voyer@bell.ca, d., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-filsfils-spring-srv6-network-programming-07 (work in progress), February 2019.

[I-D.ietf-6man-segment-routing-header]
Filsfils, C., Previdi, S., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-16 (work in progress), February 2019.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <https://www.rfc-editor.org/info/rfc2460>.

[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>.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, DOI 10.17487/RFC6514, February 2012, <https://www.rfc-editor.org/info/rfc6514>.

[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>.

[RFC7606] Chen, E., Ed., Scudder, J., Ed., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", RFC 7606, DOI 10.17487/RFC7606, August 2015, <https://www.rfc-editor.org/info/rfc7606>.

[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>.

11.2. Informative References

[I-D.ietf-bess-evpn-igmp-mld-proxy]
Sajassi, A., Thoria, S., Patel, K., Yeung, D., Drake, J., and W. Lin, "IGMP and MLD Proxy for EVPN", draft-ietf-bess-evpn-igmp-mld-proxy-02 (work in progress), June 2018.

[I-D.ietf-bess-evpn-prefix-advertisement]
Rabadan, J., Henderickx, W., Drake, J., Lin, W., and A. Sajassi, "IP Prefix Advertisement in EVPN", draft-ietf-bess-evpn-prefix-advertisement-11 (work in progress), May 2018.

[I-D.ietf-idr-bgp-prefix-sid]
Previdi, S., Filsfils, C., Lindem, A., Sreekantiah, A., and H. Gredler, "Segment Routing Prefix SID extensions for BGP", draft-ietf-idr-bgp-prefix-sid-27 (work in progress), June 2018.

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Jain, D., Mattes, P., Rosen, E., and S. Lin, "Advertising Segment Routing Policies in BGP", draft-ietf-idr-segment-routing-te-policy-05 (work in progress), November 2018.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-22 (work in progress), December 2018.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-15 (work in progress), January 2018.

[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>.

[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>.

[RFC4659] De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC 4659, DOI 10.17487/RFC4659, September 2006, <https://www.rfc-editor.org/info/rfc4659>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, DOI 10.17487/RFC4760, January 2007, <https://www.rfc-editor.org/info/rfc4760>.

[RFC5549] Le Faucheur, F. and E. Rosen, "Advertising IPv4 Network Layer Reachability Information with an IPv6 Next Hop", RFC 5549, DOI 10.17487/RFC5549, May 2009, <https://www.rfc-editor.org/info/rfc5549>.

[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>.

Appendix A. Contributors

Bart Peirens
Proximus
Belgium

Email: bart.peirens@proximus.com

Authors' Addresses

Gaurav Dawra (editor)
LinkedIn
USA

Email: gdawra.ietf@gmail.com

Clarence Filsfils
Cisco Systems
Belgium

Email: cfilsfil@cisco.com

Darren Dukes
Cisco Systems
Canada

Email: ddukes@cisco.com

Patrice Brissette
Cisco Systems
Canada

Email: pbrisset@cisco.com

Shyam Sethuram
Cisco Systems
USA

Email: shsethur@cisco.com

Pablo Camarilo
Cisco Systems
Spain

Email: pcamaril@cisco.com

Jonn Leddy
Comcast
USA

Email: john_leddy@cable.comcast.com

Daniel Voyer
Bell Canada
Canada

Email: daniel.voyer@bell.ca

Daniel Bernier
Bell Canada
Canada

Email: daniel.bernier@bell.ca

Dirk Steinberg
Steinberg Consulting
Germany

Email: dws@steinberg.net

Robert Raszuk
Bloomberg LP
USA

Email: robert@raszuk.net

Bruno Decraene
Orange
France

Email: bruno.decraene@orange.com

Satoru Matsushima
SoftBank
1-9-1,Higashi-Shimbashi,Minato-Ku
Japan 105-7322

Email: satoru.matsushima@g.softbank.co.jp

Shunwan Zhuang
Huawei Technologies
China

Email: zhuangshunwan@huawei.com