

BESS Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: January 4, 2020

G. Dawra, Ed.  
LinkedIn  
C. Filsfils  
P. Brissette  
S. Agrawal  
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  
J. Rabadan  
Nokia  
July 3, 2019

SRv6 BGP based Overlay services  
[draft-dawra-bess-srv6-services-01](#)

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".

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/>.

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 January 4, 2020.

Copyright Notice

Copyright (c) 2019 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- [1.](#) Introduction . . . . . [3](#)
- [2.](#) SRv6 Services TLVs . . . . . [4](#)
  - [2.1.](#) SRv6 Service Sub-TLVs . . . . . [5](#)
    - [2.1.1.](#) SRv6 SID Information Sub-TLV . . . . . [6](#)
    - [2.1.2.](#) SRv6 Service Data Sub-Sub-TLVs . . . . . [7](#)
- [3.](#) BGP based L3 service over SRv6 . . . . . [9](#)
  - [3.1.](#) IPv4 VPN Over SRv6 Core . . . . . [10](#)
  - [3.2.](#) IPv6 VPN Over SRv6 Core . . . . . [10](#)
  - [3.3.](#) Global IPv4 over SRv6 Core . . . . . [11](#)
  - [3.4.](#) Global IPv6 over SRv6 Core . . . . . [11](#)
- [4.](#) BGP based Ethernet VPN (EVPN) over SRv6 . . . . . [12](#)
  - [4.1.](#) Ethernet Auto-discovery route over SRv6 Core . . . . . [12](#)
    - [4.1.1.](#) Per-ES A-D route . . . . . [13](#)
    - [4.1.2.](#) Per-EVI A-D route . . . . . [13](#)



4.2.	MAC/IP Advertisement route over SRv6 Core . . . . .	14
4.3.	Inclusive Multicast Ethernet Tag Route over SRv6 Core . . . . .	16
4.4.	Ethernet Segment route over SRv6 Core . . . . .	17
4.5.	IP prefix route over SRv6 Core . . . . .	17
4.6.	EVPN multicast routes (Route Types 6, 7, 8) over SRv6 core . . . . .	18
5.	Encoding SRv6 SID information . . . . .	18
6.	Implementation Status . . . . .	19
7.	Error Handling . . . . .	20
8.	IANA Considerations . . . . .	21
8.1.	BGP Prefix-SID TLV Types registry . . . . .	21
8.2.	SRv6 Service Sub-TLV Types registry . . . . .	22
8.3.	SRv6 Service Data Sub-Sub-TLV Types registry . . . . .	22
9.	Security Considerations . . . . .	22
10.	Conclusions . . . . .	23
11.	References . . . . .	23
11.1.	Normative References . . . . .	23
11.2.	Informative References . . . . .	24
Appendix A.	Contributors . . . . .	26
Authors' Addresses	. . . . .	26

## 1. Introduction

SRv6 refers to Segment Routing instantiated on the IPv6 dataplane [I-D.ietf-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.ietf-spring-srv6-network-programming].

SRv6 Service SID refers to an SRv6 SID associated with one of the service specific behavior on the advertising Provider Edge(PE) router, such as (but not limited to), END.DT (Table lookup in a VRF) or END.DX (cross-connect to a nexthop) behaviors in the case of L3VPN service as defined in [I-D.ietf-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 [RFC8200].

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 segment list of 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.

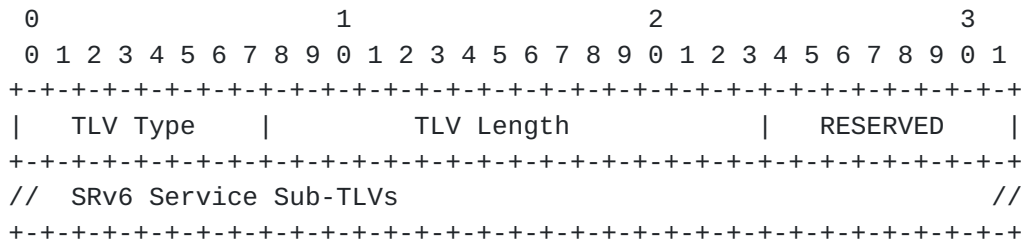
The SRv6 Service TLVs are defined as two new TLVs of the BGP Prefix-SID Attribute to achieve signaling of SRv6 SIDs for L3 and L2 services.

- o 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 behaviors which MAY be encoded, but not limited to, are End.DX4, End.DT4, End.DX6, End.DT6, etc.
- o 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 behaviors which MAY be encoded, but not limited to, are End.DX2, End.DX2V, End.DT2U, End.DT2M etc.

When an egress PE is enabled for BGP Services over SRv6 data-plane, it MUST signal one or more SRv6 Service SIDs enclosed in SRv6 Service TLV(s) within the BGP Prefix-SID Attribute attached to MP-BGP NLRIs defined in [[RFC4760](#)][RFC4659][[RFC5549](#)][RFC7432][[RFC4364](#)] where applicable as described in [section 3](#) and 4.

The following depicts the SRv6 Service TLVs encoded in the BGP Prefix-SID Attribute:

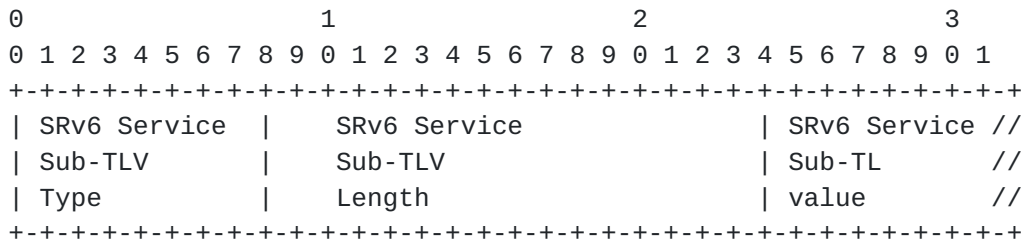




- o TLV Type (1 octet): This field is assigned values from the IANA registry "BGP Prefix-SID TLV Types". It is set to [TBD1] (to be assigned by IANA) for SRV6 L3 Service TLV. It is set to [TBD2] (to be assigned by IANA) for SRV6 L2 Service TLV.
- o TLV Length (2 octets): Specifies the total length of 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:



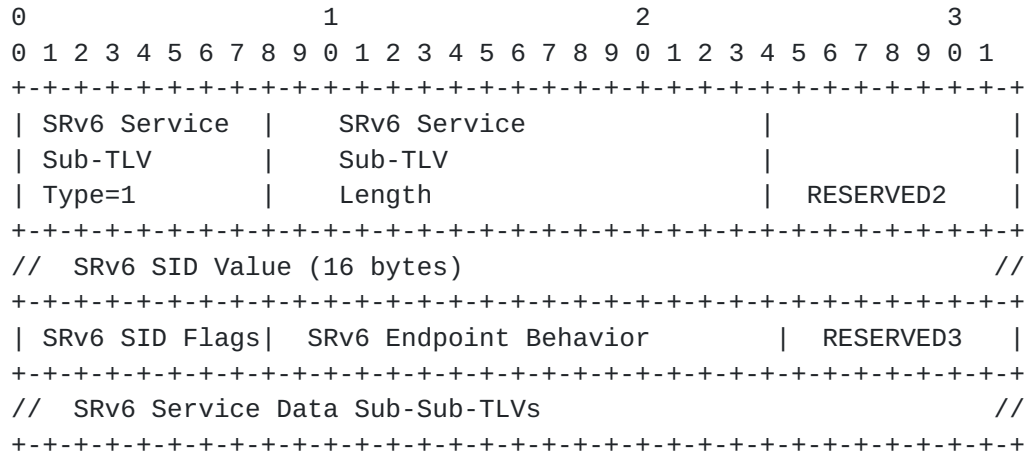
- o SRV6 Service Sub-TLV Type (1 octet): Identifies the type of SRV6 service information. It is assigned values from the IANA Registry "SRV6 Service Sub-TLV Types".
- o SRV6 Service Sub-TLV Length (2 octets): Specifies the total length of the Sub-TLV Value field.
- o SRV6 Service Sub-TLV Value (variable): 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:

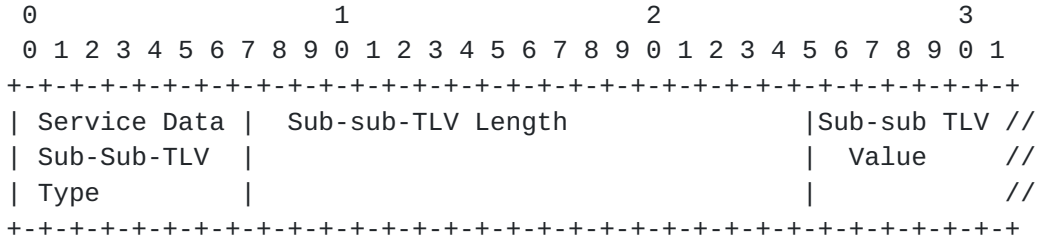


- o SRv6 Service Sub-TLV Type (1 octet): This field is set to 1 to represent SRv6 SID Information Sub-TLV.
- o SRv6 Service Sub-TLV Length (2 octets): This field contains the total length of the Value field of the Sub-TLV.
- o RESERVED2 (1 octet): SHOULD be set to 0 by the sender and MUST be ignored by the receiver.
- o SRv6 SID Value (16 octets): Encodes an SRv6 SID as defined in [\[I-D.ietf-spring-srv6-network-programming\]](#)
- o SRv6 SID Flags (1 octet): Encodes SRv6 SID Flags - none are currently defined.
- o SRv6 Endpoint Behavior (2 octets): Encodes SRv6 Endpoint behavior defined in [\[I-D.ietf-spring-srv6-network-programming\]](#). This field SHOULD be set to the value 0xFFFF indicating opaque behavior unless the router wants to signal the actual behavior.
- o RESERVED3 (1 octet): SHOULD be set to 0 by the sender and MUST be ignored by the receiver.
- o SRv6 Service Data Sub-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.



2.1.2. SRV6 Service Data Sub-Sub-TLVs

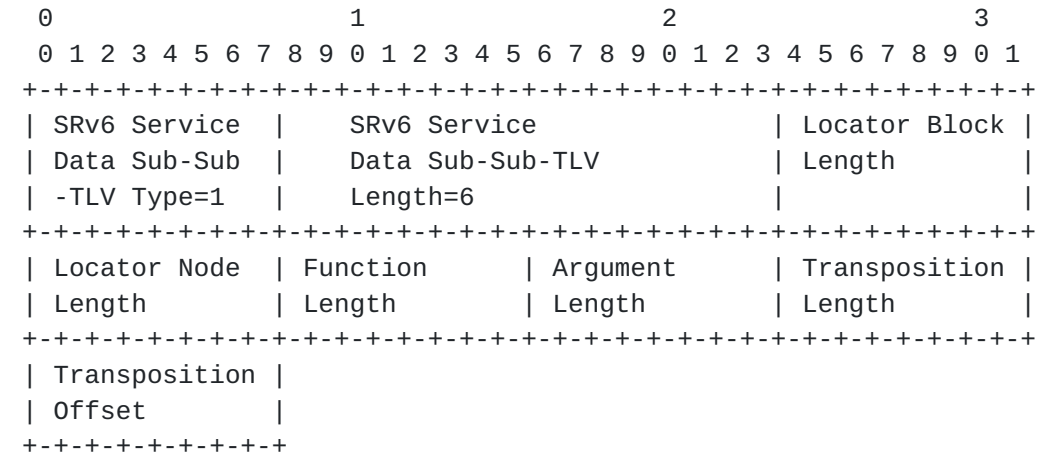
The format of the SRV6 Service Data Sub-Sub-TLV is depicted below:



- o 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".
- o SRV6 Service Data Sub-Sub-TLV Length (2 octets): Specifies the total length of the Sub-Sub-TLV Value field.
- o SRV6 Service Data Sub-Sub-TLV Value (variable): Contains data specific to the Sub-Sub-TLV Type.

2.1.2.1. SRV6 SID Structure Sub-Sub-TLV

SRV6 Service Data Sub-Sub-TLV Type 1 is assigned for SRV6 SID structure Sub-Sub-TLV. SRV6 SID Structure Sub-Sub-TLV is used to advertise the lengths of each individual parts of the SRV6 SID as defined in [I-D.ietf-spring-srv6-network-programming]. It is carried as Sub-Sub-TLV in SRV6 SID Information Sub-TLV



- o SRV6 Service Data Sub-Sub-TLV Type (1 octet): This field is set to 1 to represent SRV6 SID Structure Sub-Sub-TLV.



- o SRv6 Service Data Sub-Sub-TLV Length (2 octets): This field contains the total length of 6 bytes.
- o Locator Block Length(1 octet): Contains length of SRv6 SID locator Block in bits.
- o Locator Node Length(1 octet): Contains length of SRv6 SID locator Node in bits.
- o Function Length(1 octet): Contains length of SRv6 SID Function in bits.
- o Arguments Length(1 octet): Contains length of SRv6 SID arguments in bits.
- o Transposition Length(1 octet): Size in bits for the part of SID that has been transposed (or shifted) into a label field
- o Transposition Offset(1 octet): The offset position in bits for the part of SID that has been transposed (or shifted) into a label field.

[Section 5](#) describes mechanisms for signaling of the SRv6 Service SID by transposing a variable part of the SRv6 SID value (function and/or the argument parts) and carrying them in existing label fields to achieve more efficient packing of those service prefix NLRIs in BGP update messages. The SRv6 SID Structure Sub-Sub-TLV MUST be included with the appropriate length fields when the SRv6 Service SID is signaled in split parts to enable the receiver to put together the SID accurately.

Transposition Offset indicates the bit position and Transposition Length indicates the number of bits that are being taken out of the SRv6 SID value and put into high order bits of label field. The bits that have been shifted out MUST be set to 0 in the SID value.

Transposition Length of 0 indicates nothing is transposed and that the entire SRv6 SID value is encoded in the SID Information sub-TLV. In this case, the Transposition Offset MUST be set to 0.

Since size of label field is 24 bits, only that many bits can be transposed from the SRv6 SID value into it.

The SRv6 SID Structure Sub-Sub-TLV is optional and MAY be included when the entire SRv6 Service SID value is encoded in the SID Information Sub-TLV.



Arguments MAY be generally applicable for SIDs of only specific behaviors (e.g. End.DT2M) and therefore the argument length MUST be set to 0 for SIDs where the argument is not applicable.

### 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 Prefix-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 MUST choose to perform IPv6 encapsulation and optionally insert an SRH when required; second, it indicates the value of the Service SID to be used in the 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 [[I-D.ietf-spring-srv6-network-programming](#)].

The SRv6 Service SID SHOULD 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.

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 community '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:

- o AFI = 1
- o SAFI = 128
- o Length of Next Hop Network Address = 16 (or 32)
- o Network Address of Next Hop = IPv6 address of the egress PE
- o NLRI = IPv4-VPN routes
- o Label = It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Function part of SRv6 SID when indicated as such by the SID Structure Sub-Sub-TLV.

SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The behavior of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the behavior SHOULD 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:

- o AFI = 2
- o SAFI = 128
- o Length of Next Hop Network Address = 24 (or 48)
- o Network Address of Next Hop = 8 octets of RD set to 0 followed by IPv6 address of the egress PE
- o NLRI = IPv6-VPN routes
- o Label = It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Function part of SRv6 SID when indicated as such by the SID Structure Sub-Sub-TLV.



SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The behavior of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the behavior SHOULD 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:

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

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

### **3.4. Global IPv6 over SRv6 Core**

The MP\_REACH\_NLRI is encoded with:

- o AFI = 2
- o SAFI = 1
- o Length of Next Hop Network Address = 16 (or 32)
- o Network Address of Next Hop = IPv6 address of Next Hop
- o NLRI = IPv6 routes

SRv6 Service SID is encoded as part of the SRv6 L3 Service TLV. The behavior of the SRv6 SID is entirely up to the originator of the advertisement. In practice, the behavior SHOULD be End.DX4 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 dataplane.



#### **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](#)].

- o Ethernet Auto-discovery Route (Route Type 1)
- o MAC/IP Advertisement Route (Route Type 2)
- o Inclusive Multicast Ethernet Tag Route (Route Type 3)
- o Ethernet Segment route (Route Type 4)
- o IP prefix route (Route Type 5)
- o Selective Multicast Ethernet Tag route (Route Type 6)
- o IGMP join sync route (Route Type 7)
- o 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 Prefix-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 MUST choose to perform IPv6 encapsulation and optionally insert an SRH when required; second, it indicates the value of the Service SID(s) to be used in the encapsulation.

##### **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 [\[RFC7432\]](#). 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:

```

+-----+
|  RD (8 octets)  |
+-----+
|Ethernet Segment Identifier (10 octets)|
+-----+
|  Ethernet Tag ID (4 octets)  |
+-----+
|  MPLS label (3 octets)  |
+-----+

```

#### **[4.1.1.](#) Per-ES A-D route**

- o BGP next-hop: IPv6 address of an egress PE
- o Ethernet Tag ID: set to MAX-ET per [\[RFC7432\] section 8.2.1](#)
- o MPLS Label: always set to zero per [\[RFC7432\] section 8.2.1](#)
- o ESI label extended community ESI label field: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Argument is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Argument part of SRV6 SID when indicated as such by the SID Structure Sub-Sub-TLV.

A Service SID enclosed in a SRV6 L2 Service TLV within the BGP Prefix-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. The Service SID is used to signal Arg.FE2 SID argument for applicable End.DT2M SIDs.

#### **[4.1.2.](#) Per-EVI A-D route**

- o BGP next-hop: IPv6 address of an egress PE
- o Ethernet Tag ID: non-zero for VLAN-aware bundling service, EVPN VPWS and FXC
- o MPLS Label: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for





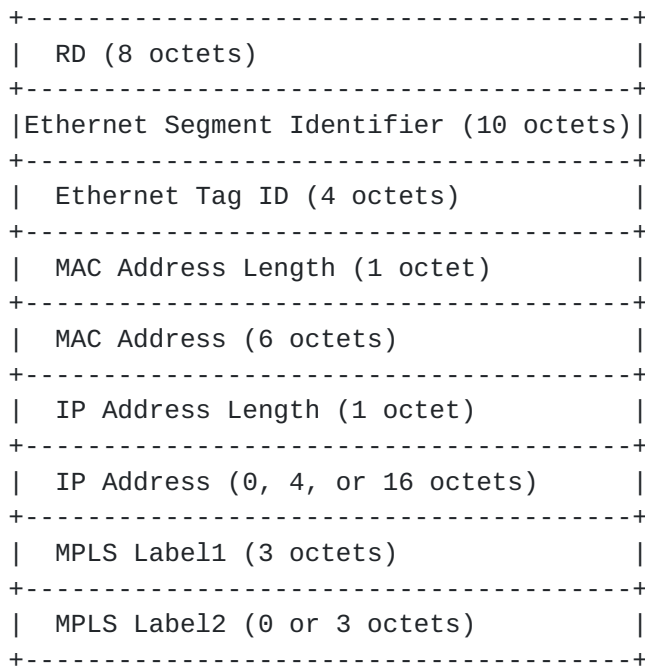
details). Otherwise it carries the Function part of SRv6 SID when indicated as such by the SID Structure Sub-Sub-TLV.

A Service SID enclosed in a SRv6 L2 Service TLV within the BGP Prefix-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 SHOULD 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:



- o BGP next-hop: IPv6 address of an egress PE
- o MPLS Label1: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Function part of SRv6 SID when indicated as such by the SID Structure Sub-Sub-TLV.
- o MPLS Label2: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for



details). Otherwise it carries the Function part of SRV6 SID when indicated as such by the SID Structure Sub-Sub-TLV.

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.

- o MAC/IP Advertisement route with MAC Only
  - \* BGP next-hop: IPv6 address of egress PE
  - \* MPLS Label1: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Function part of SRV6 SID when indicated as such by the SID Structure Sub-Sub-TLV.
- o A Service SID enclosed in a SRV6 L2 Service TLV within the BGP Prefix-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 SHOULD be END.DX2 or END.DT2U.
- o MAC/IP Advertisement route with MAC+IP
  - \* BGP next-hop: IPv6 address of egress PE
  - \* MPLS Label1: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Function part of SRV6 SID when indicated as such by the SID Structure Sub-Sub-TLV.
  - \* MPLS Label2: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Function part of SRV6 SID when indicated as such by the SID Structure Sub-Sub-TLV.
- o An L2 Service SID enclosed in a SRV6 L2 Service TLV within the BGP Prefix-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 SHOULD 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:

```

+-----+
|  RD (8 octets)                |
+-----+
|  Ethernet Tag ID (4 octets)   |
+-----+
|  IP Address Length (1 octet)  |
+-----+
|  Originating Router's IP Address |
|                               |
|          (4 or 16 octets)      |
+-----+

```

- o 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 is encoded as follows for SRv6 Core:

```

+-----+
|  Flag (1 octet)                |
+-----+
|  Tunnel Type (1 octet)         |
+-----+
|  MPLS label (3 octet)         |
+-----+
|  Tunnel Identifier (variable)  |
+-----+

```

- o Flag: zero value defined per [[RFC7432](#)]
- o Tunnel Type: defined per [[RFC6514](#)]
- o MPLS label: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for



details). Otherwise it carries the Function part of SRv6 SID when indicated as such by the SID Structure Sub-Sub-TLV.

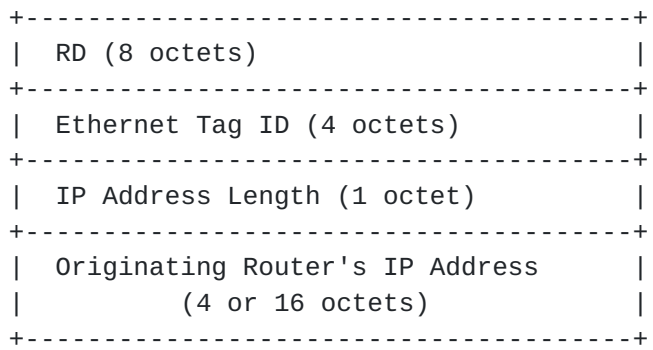
- o Tunnel Identifier: IP address of egress PE

A Service SID enclosed in a SRv6 L2 Service TLV within the BGP Prefix-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:

- o END.DX2 or END.DT2M behavior
- o The ESI Filtering argument (Arg.FE2) of the Service SID carried along with EVPN Route Type 1 route SHOULD 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:



- o 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:

```

+-----+
|  RD (8 octets)  |
+-----+
|Ethernet Segment Identifier (10 octets)|
+-----+
|  Ethernet Tag ID (4 octets)  |
+-----+
|  IP Prefix Length (1 octet)  |
+-----+
|  IP Prefix (4 or 16 octets)  |
+-----+
|  GW IP Address (4 or 16 octets)  |
+-----+
|  MPLS Label (3 octets)  |
+-----+

```

- o BGP next-hop: IPv6 address of egress PE
- o MPLS Label: It is set to Implicit NULL when the SID Structure Sub-Sub-TLV is not present or when it is present and indicates that the Function is encoded in the SID value (refer [Section 5](#) for details). Otherwise it carries the Function part of SRv6 SID when indicated as such by the SID Structure Sub-Sub-TLV.

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 behavior may SHOULD be End.DT4/6 or End.DX4/6.

#### **[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.](#) Encoding SRv6 SID information**

The SRv6 Service SID(s) for a BGP Service Prefix are carried in the SRv6 Services TLVs of the BGP Prefix-SID Attribute.

For certain types of BGP Services like L3VPN where a per-VRF SID allocation is used (i.e. End.DT4 or End.DT6 behaviors), the same SID is shared across multiple NLRIs thus providing efficient packing. However, for certain other types of BGP Services like EVPN VPWS where a per-PW SID allocation is required (i.e. End.DX2 behavior), each



NLRI would have its own unique SID there by resulting in inefficient packing.

To achieve efficient packing, this document allows flexibility in the advertisement of the SRV6 Service SID either as a whole in the SRV6 Services TLVs or the encoding of only the common parts of the SRV6 SID (e.g. Locator parts) in the SRV6 Services TLVs and encoding the variable (e.g. Function and Argument parts) in the existing label fields specific to that service encoding. The SRV6 SID Structure Sub-Sub-TLV describes the sizes of the parts of the SRV6 SID. It also indicate offset of variable part and its length in SRV6 SID value.

As an example, for the EVPN VPWS service prefix described in [section 4.1.2](#), the function part of the SRV6 SID is encoded in the MPLS Label field of the NLRI and the SID value in the SRV6 Services TLV carries only the locator parts with the SRV6 SID Structure Sub-Sub-TLV included. The SRV6 SID Structure sub-sub-TLV defines the lengths of locator block, locator node and function parts (arguments are not applicable for the End.DX2 behavior). Transposition Offset indicates the bit position and Transposition Length indicates the number of bits that are being taken out of the SID and put into label field.

In yet another example, for the EVPN Per-ES A-D route described in [section 4.1.1](#), only the argument of the SID needs to be signaled. This argument part of the SRV6 SID MAY be Transposed in the ESI Label field of the ESI Label Extended Community and the SID value in the SRV6 Services TLV is set to 0 with the SRV6 SID Structure Sub-Sub-TLV included. The SRV6 SID Structure sub-sub-TLV defines the lengths of locator block, locator node, function and argument parts. The offset and length of argument part SID value moved to label field is set in Transposition offset and length of SID structure TLV. The receiving router is then able to put together the entire SRV6 Service SID (e.g. for the End.DT2M behavior) placing the label value received in the ESI Label field of the Per-ES A-D route into the correct transposition offset and length in the SRV6 SID with the End.DT2M behavior received for a EVPN Route Type 3 value.

## 6. Implementation Status

The [[I-D.matsushima-spring-srv6-deployment-status](#)] describes the current deployment and implementation status of SRV6 which also includes the BGP services over SRV6 as specified in this document.



## 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:

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

An SRV6 Service Sub-TLV is considered malformed in the following cases:

- o 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:

- o 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 Prefix-SID attribute is considered malformed if it contains at least 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 Prefix-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.

SRv6 SID value in SRv6 Service Sub-TLV is invalid when SID Structure Sub-Sub-TLV is present and transposition length is greater than 24. Path pointing to such Prefix-SID Attribute should be ineligible during the selection of best path for the corresponding prefix.

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

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

## **8. IANA Considerations**

### **8.1. BGP Prefix-SID TLV Types registry**

This document defines two new TLV Types of the BGP Prefix-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>

**8.3. SRv6 Service Data Sub-Sub-TLV Types registry**

IANA is requested to create and maintain a new registry called "SRv6 Service Data Sub-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-Sub-TLV Types are defined in this document:

Value	Type	Reference
1	SRv6 SID Structure Sub-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.ietf-6man-segment-routing-header]  
Filsfils, C., Dukes, D., Previdi, S., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", [draft-ietf-6man-segment-routing-header-21](#) (work in progress), June 2019.
- [I-D.ietf-spring-srv6-network-programming]  
Filsfils, C., Camarillo, P., Leddy, J., daniel.voyer@bell.ca, d., Matsushima, S., and Z. Li, "SRv6 Network Programming", [draft-ietf-spring-srv6-network-programming-01](#) (work in progress), July 2019.
- [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>>.
- [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>>.
- [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>>.

## **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-03](#) (work in progress), June 2019.
- [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-06](#) (work in progress), May 2019.



- [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-25](#) (work in progress), May 2019.
- [I-D.matsushima-spring-srv6-deployment-status]  
Matsushima, S., Filsfils, C., Ali, Z., and Z. Li, "SRV6 Implementation and Deployment Status", [draft-matsushima-spring-srv6-deployment-status-01](#) (work in progress), May 2019.
- [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>>.





[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>>.

## **Appendix A. Contributors**

Ali Sajassi  
Cisco

Email: [sajassi@cisco.com](mailto:sajassi@cisco.com)

Bart Peirens  
Proximus  
Belgium

Email: [bart.peirens@proximus.com](mailto:bart.peirens@proximus.com)

Darren Dukes  
Cisco

Email: [ddukes@cisco.com](mailto:ddukes@cisco.com)

Pablo Camarilo  
Cisco

Email: [pcamaril@cisco.com](mailto:pcamaril@cisco.com)

Shyam Sethuram  
Cisco

Email: [shsethur@cisco.com](mailto:shsethur@cisco.com)

Zafar Ali  
Cisco

Email: [zali@cisco.com](mailto:zali@cisco.com)

Ketan Talaulikar  
Cisco

Email: [ketant@cisco.com](mailto:ketant@cisco.com)

Authors' Addresses



Gaurav Dawra (editor)  
LinkedIn  
USA

Email: [gdawra.ietf@gmail.com](mailto:gdawra.ietf@gmail.com)

Clarence Filsfils  
Cisco Systems  
Belgium

Email: [cfilsfil@cisco.com](mailto:cfilsfil@cisco.com)

Patrice Brissette  
Cisco Systems  
Canada

Email: [pbrisset@cisco.com](mailto:pbrisset@cisco.com)

Swadesh Agrawal  
Cisco Systems  
USA

Email: [swaagraw@cisco.com](mailto:swaagraw@cisco.com)

John Leddy  
Comcast  
USA

Email: [john\\_leddy@cable.comcast.com](mailto:john_leddy@cable.comcast.com)

Daniel Voyer  
Bell Canada  
Canada

Email: [daniel.voyer@bell.ca](mailto:daniel.voyer@bell.ca)

Daniel Bernier  
Bell Canada  
Canada

Email: [daniel.bernier@bell.ca](mailto:daniel.bernier@bell.ca)



Dirk Steinberg  
Steinberg Consulting  
Germany

Email: [dws@steinberg.net](mailto:dws@steinberg.net)

Robert Raszuk  
Bloomberg LP  
USA

Email: [robert@raszuk.net](mailto:robert@raszuk.net)

Bruno Decraene  
Orange  
France

Email: [bruno.decraene@orange.com](mailto:bruno.decraene@orange.com)

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

Email: [satoru.matsushima@g.softbank.co.jp](mailto:satoru.matsushima@g.softbank.co.jp)

Shunwan Zhuang  
Huawei Technologies  
China

Email: [zhuangshunwan@huawei.com](mailto:zhuangshunwan@huawei.com)

Jorge Rabadan  
Nokia  
USA

Email: [jorge.rabadan@nokia.com](mailto:jorge.rabadan@nokia.com)

