Inter-Domain Routing
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2019

G. Dawra, Ed.
LinkedIn
C. Filsfils
D. Dukes
P. Brissette
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
October 22, 2018

BGP Signaling for SRv6 based Services.

draft-dawra-idr-srv6-vpn-05

Abstract

This draft defines procedures and messages for BGP SRv6-based L3VPN and EVPN. 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.

Status of This Memo

This Internet-Draft will expire on April 25, 2019.

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 PE, such as (but not limited to) in the case of L3VPN service, END.DT (crossconnect to a VRF) or END.DX (crossconnect to a nexthop) functions as defined in[I-D.filsfils-spring-srv6-network-programming].

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

To provide SRv6 Service service in conjunction with an underlay SLA from the ingress PE to the egress PE, the egress PE colors the overlay VPN route with a color extended community[I-D.ietf-idr-segment-routing-te-policy]. The ingress PE encapsulates the VPN 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 in a particular VPN from an egress Provider Edge (egress-PE) to ingress Provider Edge (ingress-PE) nodes.

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

2. SRv6 Services TLV

The SRv6 Service TLVs are defined as two new TLVs for BGP Prefix SID Attribute [I-D.ietf-idr-bgp-prefix-sid], to achieve signaling of SRv6 Service SID for L3 and L2 services.

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

When an egress-PE is capable of SRv6 data-plane, it SHOULD signal SRv6 Service SID TLV within the BGP SID Attribute attached to MP-BGP NLRI defined in [RFC4659][RFC5549][RFC7432]. [RFC4364]

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   TLV Type    |             Length            |   RESERVED    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   //  SRv6 Service Information (variable)                        //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This document defines the following two new TLVs for BGP SID Attribute.

- SRv6 L3 Service TLV. Type code 5 (to be assigned by IANA as described in section 8). This TLV encodes Service SID information for the SRv6 based L3 services. It corresponds to the equivalent functionality provided by an MPLS Label when received with a Layer 3 VPN route [RFC4364]. Some functions which MAY be encoded, but not limited to, are End.DX4, End.DT4, End.DX6, End.DT6, etc.

- SRv6 L2 Service TLV. Type code 6 (to be assigned by IANA as described in section 8). This TLV encodes Service SID information for the 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.

The "SRv6 Service Information" is encoded as an un-ordered list of sub-TLVs ("Type/Length/Value" blocks), as following:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Service       |                               |              //
 | information   |                               |              //
 | sub-TLV Type  |    sub-TLV Length             |   Value      //
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

This document defines a sub-TLV Type code to encode a single SRv6 SID value along with its properties as following:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |sub-TLV Type=1 |  sub-TLV Length               |   RESERVED1   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 //  SRv6 SID Value (16 bytes)                                   //
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | SID Flags     |  Endpoint Behavior            |   RESERVED2   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | SRv6 SID Optional Information                                 |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

- o Type is 1 (to be assigned by IANA as described in Section 8). As defined to be "SID information sub-TLV".

- o Length: 16 bit field. The total length of the value portion of the sub-TLV.

- o RESERVED1: 8 bit field. SHOULD be 0 on transmission and MUST be ignored on reception.

- o SRv6 SID Value: 128 bit field. Encodes an SRv6 SID as defined in [I-D.filsfils-spring-srv6-network-programming]

- o SID Flags: 8 bit field. Encodes SRv6 SID Flags. Value is opaque to BGP.

- o Endpoint Behavior : 16 bit field. Encodes Endpoint behavior. For SRv6 VPN services, this field is always set to (0xFFFF).

- o RESERVED2: 8 bit field. SHOULD be 0 on transmission and MUST be ignored on reception.

- o SRv6 SID Optional Information. Variable length. Encodes optional properties as described below.

SRv6 SID Optional information is encoded as a list of "SID optional information sub-TLV" blocks. Where each block is encoded as Type/Length/Value triplet.

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | SID Optional  |  sub-TLV Length               |   Value      //
    | information   |                               |              //
    | sub-TLV Type  |                               |              //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

No Type codes for SID Optional information sub-TLV are defined at this point.

3. BGP based L3 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-PE) receive these advertisements and may add the prefix to the RIB in an appropriate VRF.

Egress-PEs which supports SRv6-VPN advertises a Service SID encoded within SRv6 Service TLV within BGP SID attribute, with the VPN routes. The Service SID thus signaled only has local significance at the egress-PE, where it is allocated or configured on a per-CE or per-VRF basis. In practice, the SID encodes 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 VPN identifier.

To support SRv6 based L3VPN overlay, a SID is advertised with BGP MPLS L3VPN route update[RFC4364]. SID is encoded in a SRv6 Service SID TLV within the optional transitive BGP SID attribute[I-D.ietf-idr-bgp-prefix-sid]. This attribute 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 to include in the SRH encapsulation. For L3VPN, only a single SRv6 Service SID MAY be necessary. A BGP speaker supporting an SRv6 underlay MAY distribute SID per route via the SRv6 Service TLV. If the BGP speaker supports MPLS based L3VPN simultaneously, it MAY also populate the Label values in L3VPN route

NLRI, and allow the BGP ingress device to decide which encapsulation to use. If the BGP speaker does not support MPLS based L3VPN services the MPLS Labels in L3VPN NLRI MUST be set to IMPLICIT-NULL.[RFC7432]

At an ingress-PE, BGP installs the advertised prefix in the correct RIB table, recursive 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 VPN route is colored with an extended color community C and signaled with Next-Hop N and the ingress PE has a valid SRv6 Policy (N, C) associated with SID list <S1,S2, S3> [I-D.filsfils-spring-segment-routing-policy] then the SR Policy is <S1, S2, S3, SRv6 Service SID>.

Multiple VPN routes MAY resolve recursively on 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 = Implicit-Null

SRv6 Service SID is encoded as part of the SRv6 Service SID TLV defined in Section 2. 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 enclosed as follows for an SRv6 Core:

- o AFI = 2
- 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 = IPv6-VPN routes
- o Label = Implicit-Null

SRv6 Service SID are encoded as part of the SRv6 Service SID TLV defined in Section 2. The function of the IPv6 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:

- 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 SID for Global IPv4 routes is encoded as part of the SRv6 Service SID defined in Section 2. 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.

3.4. 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 SID for Global IPv6 routes is encoded as part of the SRv6 Service SID defined in Section 2. 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.

Also, by utilizing the SRv6 Service SID TLV, as defined in Section 2, to encode the Global SID, 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. It defines 4 route-types which carry prefixes and MPLS Label attributes, the Labels each have specific use for MPLS encapsulation of EVPN traffic. The fifth route-type carrying MPLS label information (and thus encapsulation information) for EVPN is defined in[I-D.ietf-bess-evpn-prefix-advertisement]. The Route Types discussed below are:

o Ethernet Auto-discovery Route

o MAC/IP Advertisement Route

o Inclusive Multicast Ethernet Tag Route

o Ethernet Segment route

o IP prefix route

o Selective Multicast route

o IGMP join sync route

o IGMP leave sync route

To support SRv6 based EVPN overlays a SRv6 Service SID is advertised in route-type 1,2,3 and 5 above. The SRv6 Service SID (or list of those, when applicable) per route-type are advertised in SRv6 Service TLV, as described in section 2. Signaling of SRv6 Service SID 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 or SIDs to include in the SRH encapsulation. If the BGP speaker does not support MPLS based EVPN

services the MPLS Labels in 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 Type-1 route type 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 with the introduced ESI MPLS label extended community defined in [RFC7432].The extended community label is set to implicit-null. PEs may identify other PEs connected to the same Ethernet segment after the EVPN 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)                |
+---------------------------------------+
```

For a SRv6 only BGP speaker for an SRv6 Core:

o SRv6 Service SID TLV MAY be advertised with the route.

4.1.1. EVPN Route Type-1(Per ES AD)

Where:

o BGP next-hop: IPv6 address of an egress PE

o Ethernet Tag ID: all FFFF's

o MPLS Label: always set to zero value

o Extended Community: Per ES AD, ESI label extended community

BGP SID Attribute with SRv6 Service TLV MAY be advertised along with the route advertisement and the behavior of the SRv6 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. Prefix Type-1(Per EVI/ES AD)

Where:

- o BGP next-hop: IPv6 address of an egress PE

- o Ethernet Tag ID: non-zero for VLAN aware bridging, EVPN VPWS and FXC

- o MPLS Label: Implicit-Null

BGP SID Attribute with SRv6 Service TLV MAY be advertised along with the route advertisement and the behavior of the SRv6 Service SID 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(Type-2) with 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.

A MAC/IP Advertisement route type is encoded as follows for SRv6 Core:

```
                   +---------------------------------------+
                   |  RD (8 octets)                        |
                   +---------------------------------------+
                   |Ethernet Segment Identifier (10 octets)|
                   +---------------------------------------+
                   |  Ethernet Tag ID (4 octets)           |
                   +---------------------------------------+
                   |  MAC Address Length (1 octet)         |
                   +---------------------------------------+
                   |  MAC Address (6 octets)               |
                   +---------------------------------------+
                   |  IP Address Length (1 octet)          |
                   +---------------------------------------+
                   |  IP Address (0, 4, or 16 octets)      |
                   +---------------------------------------+
                   |  MPLS Label1 (3 octets)               |
                   +---------------------------------------+
                   |  MPLS Label2 (0 or 3 octets)          |
                   +---------------------------------------+
```

where:

- BGP next-hop: IPv6 address of an egress PE

- MPLS Label1: Implicit-null

- MPLS Label2: Implicit-null

BGP SID Attribute with SRv6 Service TLV MAY be advertised. The behavior of the SRv6 Service SID is entirely up to the originator of the advertisement. In practice, the behavior of the SRv6 SID is as follows:

- END.DX2, END.DT2U (Layer 2 portion of the route)

- END.DT6/4 or END.DX6/4 (Layer 3 portion of the route)

Described below are different types of Type-2 advertisements.

- MAC/IP Advertisement Route(Type-2) with MAC Only

  * BGP next-hop: IPv6 address of egress PE

  * MPLS Label1: Implicit-null

  * MPLS Label2: Implicit-null

  * SRv6 Service SID TLV within BGP SID Attribute MAY encode END.DX2 or END.DT2U behavior

- MAC/IP Advertisement Route(Type-2) with MAC+IP

  * BGP next-hop: IPv6 address of egress PE

  * MPLS Label1: Implicit-Null

  * MPLS Label2: Implicit-Null

  * SRv6 Service TLV within BGP SID Attribute MAY encode Layer2 END.DX2 or END.DT2U behavior and Layer3 END.DT6/4 or END.DX6/4 behavior

4.3. Inclusive Multicast Ethernet Tag Route with 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.

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

An Inclusive Multicast Ethernet Tag route type specific EVPN NLRI consists of the following [RFC7432] where:

- BGP next-hop: IPv6 address of egress PE

- SRv6 Service TLV MAY encode END.DX2/END.DT2M function.

- BGP Attribute: PMSI Tunnel Attribute[RFC6514] MAY contain MPLS implicit-null label and Tunnel Type would be similar to defined in EVPN Type-6 i.e. Ingress replication route.

The format of PMSI Tunnel Attribute attribute is encoded as follows for an 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: Implicit-Null
- o Tunnel Identifier: IP address of egress PE

SRv6 Service TLV may be encoded as part of BGP SID Attribute. The behavior of the SRv6 Service SID 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 function
- o The ESI Filtering argument(Arg.FE2) carried along with EVPN Route Type-1 (in SRv6 VPN SID), MAY be merged together with the applicable End.DT2M SID advertised by remote PE by doing a bitwise logical OR 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 with SRv6 Core

An Ethernet Segment route type specific EVPN NLRI consists of the following defined in [RFC7432]

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

where:

o BGP next-hop: IPv6 address of egress PE

As opposed to the previous route types, SRv6 Service TLV as part of BGP SID Attribute, is NOT advertised along with the route. The processing of that route has not changed; it remains as described in [RFC7432].

4.5. IP prefix router(Type-5) with 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 defined in[I-D.ietf-bess-evpn-prefix-advertisement]

An IP Prefix advertisement is encoded as follows for an 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: Implicit-Null

BGP SID Attribute with SRv6 Service TLV MAY be advertised. The behavior of the SRv6 Service SID is entirely up to the originator of the advertisement. In practice, the behavior of the SRv6 SID is an End.DT6/4 or End.DX6/4.

4.6. Multicast routes (EVPN Route Type-6, Type-7, Type-8)

These routes do not require any additional SRv6 Service TLV. As per 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 L3 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 IPv6 session) is a local configurable decision of the ingress-PE, and is outside the scope of this document.

Migration from IPv6 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 service
      Then BGP allocates an SRv6 Service SID for the VPN service
      and adds the BGP SRv6 Service SID TLV while advertising VPN prefixes.
  If BGP offers an MPLS VPN service
      Then BGP allocates an MPLS Label for the VPN service and
      use it in NLRI as normal for MPLS L3 VPNs.
  else MPLS label for VPN service is set to IMPLICIT-NULL.

At Ingress-PE
  *Selection of which encapsulation below (SRv6 Service or MPLS-VPN) is
   defined by local BGP policy
  If BGP supports SRv6 Service service, and
  receives a BGP SID Attribute with an SRv6 Service TLV encoding a SRv6 Service
SID
      Then BGP programs the destination prefix in RIB recursive via
      the related SR Policy.
  If BGP supports MPLS VPN service, and
  the MPLS Label is not Implicit-Null
      Then the MPLS label is used as a VPN 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 different implementations [1]:

- o Three Cisco Hardware-forwarding platforms: ASR 1K, ASR 9k and NCS 5500

- o Huawei network operating system

- o Two Cisco network operating systems: IOS XE and IOS XR

- Barefoot Networks Tofino on OCP Wedge-100BF

- Linux Kernel officially upstreamed in 4.10

- Fd.io

7. Error Handling of BGP SRv6 SID Updates

If the SRv6 Service TLV within the received BGP SID Attribute is malformed, consider the entire BGP SID Attribute as malformed, discard it and not propagate it further to other peers i.e. use the -attribute discard- action specified in [RFC7606] an error MAY be logged for further analysis.

The SRv6 Service TLV is not considered to be malformed in the following cases. The rest of the BGP SID Attribute MUST be processed normally. An error MAY be logged for further analysis.

- The Service Information sub-TLV Type is unrecognized: all unrecognized sub-TLV Types must be stored locally and propagated further to other peers. It is a matter of local implementation whether to use locally any recognized SID Types that may be present in the TLV along with the unrecognized Types.

In addition, the following rules apply for processing NLRIs received with BGP SID Attribute containing SRv6 Service TLV:

- If the TLV is advertised by a CE peer, the receiving PE may discard it before advertising the route to its PE peers.

- If the received NLRI has neither a valid SRv6 Service SID nor a valid MPLS label as specified in [RFC4659][RFC5549][RFC7432] , the NLRI MUST be considered unreachable i.e. apply the -treat as withdraw- action specified in [RFC7606].

8. IANA Considerations

This document defines a new TLV, SRv6 Service TLV, within BGP SID attribute. This document defines the following new TLV Types of BGP SID attribute:

- Type 5: SRv6 Layer3 Service

- Type 6: SRv6 Layer2 Service

and are assigned to SRv6 Layer3 Service TLV and SRv6 Layer2 Service TLV defined in this document.

Further, this document defines a new sub-TLV; namely Service information sub-TLV, within SRv6 Service TLV, as described in Section 2. A new registry "BGP SRv6 Service Information sub-TLV Types" is required and a new Type code point with value 1, is requested in this registry, to denote "SID information sub-TLV".

Further, this document defines new optional sub-TLVs, namely "SID optional information sub-TLV" within Service information sub-TLV, as described in Section 2. New registry for this purpose is required.

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-05 (work in progress), July 2018.

[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-14 (work in progress), June 2018.

[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-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-04 (work in progress), July 2018.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., Litkowski, S., Decraene, B., and J. Tantsura, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-19 (work in progress), July 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.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

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

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <https://www.rfc-editor.org/info/rfc4364>.

[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>.

[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>.

11.3. URIs

[1] http://www.segment-routing.net

Appendix A. Acknowledgements

The authors would like to thank Shyam Sethuram for comments and discussion of TLV processing and validation.

Appendix B. 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

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