Network Working Group S. Previdi,
Ed.
Internet-Draft C.
Filsfils
Intended status: Standards Track D. Jain,
Ed.
Expires: November 19, 2018 Cisco Systems,
Inc.
P.
Mattes
Microsoft
E.
Rosen
Juniper
Networks
S.
Lin
Google
May 18,
2018

Advertising Segment Routing Policies in BGP
draft-ietf-idr-segment-routing-te-policy-03

Abstract

This document defines a new BGP SAFI with a new NLRI in order to advertise a candidate path of a Segment Routing Policy (SR Policy). An SR Policy is a set of candidate paths consisting of one or more segment lists. The headend of an SR Policy may learn multiple candidate paths for an SR Policy. Candidate paths may be learned via a number of different mechanisms, e.g., CLI, NetConf, PCEP, or BGP. This document specifies the way in which BGP may be used to distribute candidate paths. New sub-TLVs for the Tunnel Encapsulation Attribute are defined.

Status of This Memo

This Internet-Draft will expire on November 19, 2018.

Copyright Notice

Table of Contents

1. Introduction

Segment Routing (SR) allows a headend node to steer a packet flow along any path. Intermediate per-flow states are eliminated thanks to source routing [I-D.ietf-spring-segment-routing].

The headend node is said to steer a flow into a Segment Routing Policy (SR Policy).

The header of a packet steered in an SR Policy is augmented with the ordered list of segments associated with that SR Policy.

[I-D.filsfils-spring-segment-routing-policy] details the concepts of SR Policy and steering into an SR Policy. These apply equally to the MPLS and SRv6 instantiations of segment routing.

As highlighted in section 2 of [I-D.filsfils-spring-segment-routing-policy]:

- o an SR policy may have multiple candidate paths learned via various mechanisms (CLI, NetConf, PCEP or BGP);
- o the SRTE process selects the best candidate path for a Policy;
- o the SRTE process binds a BSID to the selected path of the Policy;
- o the SRTE process installs the selected path and its BSID in the forwarding plane.

This document specifies the way to use BGP to distribute one or more of the candidate paths of an SR policy to the headend of that policy.

The SRTE process ([I-D.filsfils-spring-segment-routing-policy]) of the headend receives candidate paths from BGP, and possibly other sources as well, and the SRTE process then determines the selected path of the policy.

This document specifies a way of representing SR policies and their candidate paths in BGP UPDATE messages. BGP can then be used to propagate the SR policies and candidate paths. The usual BGP rules for BGP propagation and "bestpath selection" are used. At the headend of a specific policy, this will result in one or more candidate paths being installed into the "BGP table". These paths are then passed to the SRTE process. The SRTE process may compare them to candidate paths learned via other mechanisms, and will choose one or more paths to be installed in the data plane. BGP itself does not install SRTE candidate paths into the data plane.

This document defines a new BGP address family (SAFI). In UPDATE messages of that address family, the NLRI identifies an SR policy, and the attributes specify candidate paths of that policy.

While for simplicity we may write that BGP advertises an SR Policy, it has to be understood that BGP advertises a candidate path of an SR policy and that this SR Policy might have several other candidate paths provided via BGP (via an NLRI with a different distinguisher as defined in this document), PCEP, NETCONF or local policy configuration.

Typically, a controller defines the set of policies and advertise them to policy head-end routers (typically ingress routers). The policy advertisement uses BGP extensions defined in this document. The policy advertisement is, in most but not all of the cases, tailored for a specific policy head-end. In this case the advertisement may sent on a BGP session to that head-end and not propagated any further.

Alternatively, a router (i.e., a BGP egress router) advertises SR Policies representing paths to itself. In this case, it is possible to send the policy to each head-end over a BGP session to that head-end, without requiring any further propagation of the policy.

An SR Policy intended only for the receiver will, in most cases, not traverse any Route Reflector (RR, [RFC4456]).

In some situations, it is undesirable for a controller or BGP egress router to have a BGP session to each policy head-end. In these situations, BGP Route Reflectors may be used to propagate the advertisements, or it may be necessary for the advertisement to propagate through a sequence of one or more ASes. To make this possible, an attribute needs to be attached to the advertisement that enables a BGP speaker to determine whether it is intended to be a head-end for the advertised policy. This is done by attaching one or more Route Target Extended Communities to the advertisement ([RFC4360]).

The BGP extensions for the advertisement of SR Policies include following components:

- o A new Subsequent Address Family Identifier (SAFI) whose NLRI identifies an SR Policy.

- o A set of new TLVs to be inserted into the Tunnel Encapsulation Attribute (as defined in [I-D.ietf-idr-tunnel-encaps]) specifying candidate paths of the SR policy, as well as other information about the SR policy.

- o One or more IPv4 address format route-target extended community ([RFC4360]) attached to the SR Policy advertisement and that indicates the intended head-end of such SR Policy advertisement.

- o The Color Extended Community (as defined in [I-D.ietf-idr-tunnel-encaps]) and used in order to steer traffic into an SR Policy, as described in section 8.4 in [I-D.filsfils-spring-segment-routing-policy]. This document (Section 3) modifies the format of the Color Extended Community by using the two leftmost bits of the RESERVED field.

1.1. 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 RFC 2119 [RFC2119].

2. SR TE Policy Encoding

2.1. SR TE Policy SAFI and NLRI

A new SAFI is defined: the SR Policy SAFI, (codepoint 73 assigned by IANA (see Section 8) from the "Subsequent Address Family Identifiers (SAFI) Parameters" registry).

The SR Policy SAFI uses a new NLRI defined as follows:

```
+------------------+
|  NLRI Length     | 1 octet
+------------------+
|  Distinguisher   | 4 octets
+------------------+
|  Policy Color    | 4 octets
+------------------+
|  Endpoint        | 4 or 16 octets
+------------------+
```

where:

- NLRI Length: 1 octet of length expressed in bits as defined in [RFC4760].

- Distinguisher: 4-octet value uniquely identifying the policy in the context of <color, endpoint> tuple. The distinguisher has no semantic value and is solely used by the SR Policy originator to make unique (from an NLRI perspective) multiple occurrences of the same SR Policy.

- Policy Color: 4-octet value identifying (with the endpoint) the policy. The color is used to match the color of the destination prefixes to steer traffic into the SR Policy [I-D.filsfils-spring-segment-routing-policy].

- Endpoint: identifies the endpoint of a policy. The Endpoint may represent a single node or a set of nodes (e.g., an anycast address). The Endpoint is an IPv4 (4-octet) address or an IPv6 (16-octet) address according to the AFI of the NLRI.

The color and endpoint are used to automate the steering of BGP Payload prefixes on SR policy ([I-D.filsfils-spring-segment-routing-policy]).

The NLRI containing the SR Policy is carried in a BGP UPDATE message [RFC4271] using BGP multiprotocol extensions [RFC4760] with an AFI of 1 or 2 (IPv4 or IPv6) and with a SAFI of 73 (assigned by IANA from the "Subsequent Address Family Identifiers (SAFI) Parameters" registry).

An update message that carries the MP_REACH_NLRI or MP_UNREACH_NLRI attribute with the SR Policy SAFI MUST also carry the BGP mandatory attributes. In addition, the BGP update message MAY also contain any of the BGP optional attributes.

The next-hop network address field in SR Policy SAFI (73) updates may be either a 4 octet IPv4 address or a 16 octet IPv6 address,

independent of the SR Policy AFI. The length field of the next-hop address specifies the next-hop address family. If the next-hop length is 4, then the next-hop is an IPv4 address; if the next-hop length is 16, then it is a global IPv6 address; and if the next-hop length is 32, then it has a global IPv6 address followed by a link-local IPv6 address. The setting of the next-hop field and its attendant processing is governed by standard BGP procedures as described in section 3 in [RFC4760].

It is important to note that any BGP speaker receiving a BGP message with an SR Policy NLRI, will process it only if the NLRI is among the best paths as per the BGP best path selection algorithm. In other words, this document does not modify the BGP propagation or bestpath selection rules.

It has to be noted that if several candidate paths of the same SR Policy (endpoint, color) are signaled via BGP to a head-end, it is recommended that each NLRI use a different distinguisher. If BGP has installed into the BGP table two advertisements whose respective NLRIs have the same color and endpoint, but different distinguishers, both advertisements are passed to the SRTE process as different candidate paths. In addition, the originator information corresponding to the each candidate path, as described in section 2.4 ([I-D.filsfils-spring-segment-routing-policy]), is passed to the SRTE process.

2.2. SR TE Policy and Tunnel Encapsulation Attribute

The content of the SR Policy is encoded in the Tunnel Encapsulation Attribute originally defined in [I-D.ietf-idr-tunnel-encaps] using a new Tunnel-Type TLV (codepoint is 15, assigned by IANA (see Section 8) from the "BGP Tunnel Encapsulation Attribute Tunnel Types" registry).

The SR Policy Encoding structure is as follows:

```
   SR Policy SAFI NLRI: <Distinguisher, Policy-Color, Endpoint>
   Attributes:
      Tunnel Encaps Attribute (23)
         Tunnel Type: SR Policy
             Binding SID
             Preference
             Priority
             Policy Name
             Explicit NULL Label Policy (ENLP)
             Segment List
                 Weight
                 Segment
                 Segment
                 ...
             ...
```

where:

- SR Policy SAFI NLRI is defined in Section 2.1.

- Tunnel Encapsulation Attribute is defined in [I-D.ietf-idr-tunnel-encaps].

- Tunnel-Type is set to 15 (assigned by IANA from the "BGP Tunnel Encapsulation Attribute Tunnel Types" registry).

- Preference, Binding SID, Priority, Policy Name, ENLP, Segment-List, Weight and Segment sub-TLVs are defined in this document.

- Additional sub-TLVs may be defined in the future.

A Tunnel Encapsulation Attribute MUST NOT contain more than one TLV of type "SR Policy". If more than one TLV of type "SR Policy" appears, the update is considered malformed and the "treat-as-withdraw" strategy of [RFC7606] is applied.

Multiple occurrences of "Segment List" MAY be encoded within the same SR Policy.

Multiple occurrences of "Segment" MAY be encoded within the same Segment List.

2.3. Remote Endpoint and Color

The Remote Endpoint and Color sub-TLVs, as defined in [I-D.ietf-idr-tunnel-encaps], MAY also be present in the SR Policy encodings.

The Remote Endpoint and Color Sub-TLVs are not used for SR Policy encodings and therefore their value is irrelevant in the context of the SR Policy SAFI NLRI. If present, the Remote Endpoint sub-TLV and the Color sub-TLV MUST be ignored by the BGP speaker.

2.4. SR TE Policy Sub-TLVs

This section defines the SR Policy sub-TLVs.

Preference, Binding SID, Segment-List, Priority, Policy Name and Explicit NULL Label Policy sub-TLVs are assigned from the "BGP Tunnel Encapsulation Attribute Sub-TLVs" registry.

Weight and Segment sub-TLVs are assigned from a new registry defined in this document and called: "SR Policy List Sub-TLVs". See Section 8 for the details of the registry.

2.4.1. Preference Sub-TLV

The Preference sub-TLV does not have any effect on the BGP bestpath selection or propagation procedures. The contents of this sub-TLV are used by the SRTE process as described in section 2.9 in ([I-D.filsfils-spring-segment-routing-policy]).

The Preference sub-TLV is optional and it MUST NOT appear more than once in the SR Policy. If the Preference sub-TLV appears more than once, the update is considered malformed and the "treat-as-withdraw" strategy of [RFC7606] is applied.

The Preference sub-TLV has following format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Preference (4 octets)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- o Type: 12
- o Length: 6.
- o Flags: 1 octet of flags. None are defined at this stage. Flags SHOULD be set to zero on transmission and MUST be ignored on receipt.

o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

o Preference: a 4-octet value.

2.4.2. SR TE Binding SID Sub-TLV

The Binding SID sub-TLV is not used by BGP. The contents of this sub-TLV are used by the SRTE process as described in section 6 in ([I-D.filsfils-spring-segment-routing-policy]).

The Binding SID sub-TLV is optional and it MUST NOT appear more than once in the SR Policy. If the Binding SID sub-TLV appears more than once, the update is considered malformed and the "treat-as-withdraw" strategy of [RFC7606] is applied.

The Binding SID sub-TLV has the following format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Binding SID (variable, optional)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 13

o Length: specifies the length of the value field not including
Type
   and Length fields. Can be 2 or 6 or 18.

o Flags: 1 octet of flags. Following flags are defined (to be assigned by IANA from the registry "SR Policy Binding SID Flags" defined in this document Section 8.5):

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|S|I|           |
+-+-+-+-+-+-+-+-+
```

where:

* S-Flag: This flag encodes the "Specified-BSID-only" behavior. It is used by SRTE process as described in section 6.2.3 in ([I-D.filsfils-spring-segment-routing-policy]).

      * I-Flag: This flag encodes the "Drop Upon Invalid" behavior.
It
         is used by SRTE process as described in section 8.2 in
         ([I-D.filsfils-spring-segment-routing-policy]).

      * Unused bits in the Flag octet SHOULD be set to zero upon
         transmission and MUST be ignored upon receipt.

   o RESERVED: 1 octet of reserved bits. SHOULD be unset on
      transmission and MUST be ignored on receipt.

   o Binding SID: if length is 2, then no Binding SID is present.

   o If length is 6 then the Binding SID contains a 4-octet SID.
Below
      format is used to encode the SID. TC, S, TTL(Total of 12bits)
are
      RESERVED and SHOULD be set to Zero and MUST be ignored.

```
       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
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |          Label                        | TC  |S|       TTL     |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

      If length is 18 then the Binding SID contains a 16-octet IPv6
SID.

2.4.3. Segment List Sub-TLV

The Segment List sub-TLV encodes a single explicit path towards the endpoint as described in section 5.1 in ([I-D.filsfils-spring-segment-routing-policy]). The Segment List sub-TLV includes the elements of the paths (i.e., segments) as well as an optional Weight sub-TLV.

The Segment List sub-TLV may exceed 255 bytes length due to large number of segments. Therefore a 2-octet length is required. According to [I-D.ietf-idr-tunnel-encaps], the first bit of the sub-TLV codepoint defines the size of the length field. Therefore, for the Segment List sub-TLV a code point of 128 (or higher) is used. See Section 8 for details of codepoints allocation.

The Segment List sub-TLV is optional and MAY appear multiple times
in
the SR Policy. The ordering of Segment List sub-TLVs, each sub-TLV encoding a Segment List, does not matter.

The Segment List sub-TLV contains zero or more Segment sub-TLVs and MAY contain a Weight sub-TLV.

The Segment List sub-TLV has the following format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |             Length            |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                           sub-TLVs                          //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- o Type: 128.
- o Length: the total length (not including the Type and Length fields) of the sub-TLVs encoded within the Segment List sub-TLV.
- o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.
- o sub-TLVs:
  - * An optional single Weight sub-TLV.
  - * Zero or more Segment sub-TLVs.

Validation of an explicit path encoded by the Segment List sub-TLV is completely within the scope of SRTE process as described in section 5 in ([I-D.filsfils-spring-segment-routing-policy]).

2.4.3.1. Weight Sub-TLV

The Weight sub-TLV specifies the weight associated to a given candidate path (i.e., a given segment list). The contents of this sub-TLV are used only by the SRTE process as described in section 2.11 in ([I-D.filsfils-spring-segment-routing-policy]).

The Weight sub-TLV is optional and it MUST NOT appear more than once inside the Segment List sub-TLV. If the Weight sub-TLV appears more than once, the update is considered malformed and the "treat-as-withdraw" strategy of [RFC7606] is applied.

The Weight sub-TLV has the following format:

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |     Type      |   Length      |     Flags     |   RESERVED    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                             Weight                            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Type: 9 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

Length: 6.

Flags: 1 octet of flags. None are defined at this stage. Flags SHOULD be set to zero on transmission and MUST be ignored on receipt.

RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

2.4.3.2. Segment Sub-TLV

The Segment sub-TLV describes a single segment in a segment list (i.e., a single element of the explicit path). Multiple Segment sub-TLVs constitute an explicit path of the SR Policy.

The Segment sub-TLV is optional and MAY appear multiple times in the Segment List sub-TLV.

The Segment sub-TLV does not have any effect on the BGP bestpath selection or propagation procedures. The contents of this sub-TLV are used only by the SRTE process as described in section 4 in ([I-D.filsfils-spring-segment-routing-policy]).

[I-D.filsfils-spring-segment-routing-policy] defines several types of Segments:

Type 1: SID only, in the form of MPLS Label
Type 2: SID only, in the form of IPv6 address
Type 3: IPv4 Node Address with optional SID
Type 4: IPv6 Node Address with optional SID for SR MPLS
Type 5: IPv4 Address + index with optional SID
Type 6: IPv4 Local and Remote addresses with optional SID
Type 7: IPv6 Address + index for local and remote pair with optional SID for SR MPLS
Type 8: IPv6 Local and Remote addresses with optional SID for SR MPLS
Type 9: IPv6 Node Address with optional SID for SRv6
Type 10: IPv6 Address + index for local and remote pair with optional SID for SRv6
Type 11: IPv6 Local and Remote addresses for SRv6

2.4.3.2.1. Type 1: SID only, in the form of MPLS Label

The Type-1 Segment Sub-TLV encodes a single SID in the form of an MPLS label. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Label                        | TC  |S|       TTL     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 1 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).
- Length is 6.
- Flags: 1 octet of flags as defined in Section 2.4.3.2.12.
- RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.
- Label: 20 bits of label value.
- TC: 3 bits of traffic class.
- S: 1 bit of bottom-of-stack.
- TTL: 1 octet of TTL.

The following applies to the Type-1 Segment sub-TLV:

o The S bit SHOULD be zero upon transmission, and MUST be ignored upon reception.

o If the originator wants the receiver to choose the TC value, it sets the TC field to zero.

o If the originator wants the receiver to choose the TTL value, it sets the TTL field to 255.

o If the originator wants to recommend a value for these fields, it puts those values in the TC and/or TTL fields.

o The receiver MAY override the originator's values for these fields. This would be determined by local policy at the receiver.
One possible policy would be to override the fields only if the fields have the default values specified above.

2.4.3.2.2. Type 2: SID only, in the form of IPv6 address

The Type-2 Segment Sub-TLV encodes a single SRv6 SID in the form of an IPv6 address. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                      SRv6 SID (16 octets)                    //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 2 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

o Length is 18.

o Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

o SRv6 SID: 16 octets of IPv6 address.

The IPv6 Segment Identifier (SRv6 SID) is defined in [I-D.ietf-6man-segment-routing-header].

2.4.3.2.3. Type 3: IPv4 Node Address with optional SID

The Type-3 Segment Sub-TLV encodes an IPv4 node address, SR Algorithm and an optional SID in the form of an MPLS label. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |  SR Algorithm |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 IPv4 Node Address (4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SID (optional, 4 octets)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 3 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

- Length is 6 or 10.

- Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

- SR Algorithm: 1 octet specifying SR Algorithm as described in section 3.1.1 in [I-D.ietf-spring-segment-routing], when A-Flag as defined in Section 2.4.3.2.12 is present. SR Algorithm is used by SRTE process as described in section 4 in ([I-D.filsfils-spring-segment-routing-policy]). When A-Flag is not encoded, this field SHOULD be unset on transmission and MUST be ignored on receipt.

- IPv4 Node Address: a 4 octet IPv4 address representing a node.

- SID: 4 octet MPLS label.

The following applies to the Type-3 Segment sub-TLV:

- The IPv4 Node Address MUST be present.

- The SID is optional and specifies a 4 octet MPLS SID containing label, TC, S and TTL as defined in Section 2.4.3.2.1.

- If length is 6, then only the IPv4 Node Address is present.

- If length is 10, then the IPv4 Node Address and the MPLS SID are present.

2.4.3.2.4. Type 4: IPv6 Node Address with optional SID for SR MPLS

The Type-4 Segment Sub-TLV encodes an IPv6 node address, SR
Algorithm
and an optional SID in the form of an MPLS label. The format is as
follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |  SR Algorithm |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                IPv6 Node Address (16 octets)                //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 SID (optional, 4 octets)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 4 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

- Length is 18 or 22.

- Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

- SR Algorithm: 1 octet specifying SR Algorithm as described in section 3.1.1 in [I-D.ietf-spring-segment-routing], when A-Flag
as
defined in Section 2.4.3.2.12 is present. SR Algorithm is used
by
SRTE process as described in section 4 in ([I-D.filsfils-spring-segment-routing-policy]). When A-Flag is not encoded, this field SHOULD be unset on transmission and MUST be ignored on receipt.

- IPv6 Node Address: a 16 octet IPv6 address representing a node.

- SID: 4 octet MPLS label.

The following applies to the Type-4 Segment sub-TLV:

- The IPv6 Node Address MUST be present.

- The SID is optional and specifies a 4 octet MPLS SID containing label, TC, S and TTL as defined in Section 2.4.3.2.1.

- If length is 18, then only the IPv6 Node Address is present.

- If length is 22, then the IPv6 Node Address and the MPLS SID are present.

2.4.3.2.5. Type 5: IPv4 Address + Local Interface ID with optional SID

The Type-5 Segment Sub-TLV encodes an IPv4 node address, a local interface Identifier (Local Interface ID) and an optional SID in the form of an MPLS label. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Local Interface ID (4 octets)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 IPv4 Node Address (4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SID (optional, 4 octets)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- o Type: 5 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).
- o Length is 10 or 14.
- o Flags: 1 octet of flags as defined in Section 2.4.3.2.12.
- o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.
- o Local Interface ID: 4 octets of interface index as defined in [I-D.ietf-pce-segment-routing].
- o IPv4 Node Address: a 4 octet IPv4 address representing a node.
- o SID: 4 octet MPLS label.

The following applies to the Type-5 Segment sub-TLV:

- o The IPv4 Node Address MUST be present.
- o The Local Interface ID MUST be present.
- o The SID is optional and specifies a 4 octet MPLS SID containing label, TC, S and TTL as defined in Section 2.4.3.2.1.
- o If length is 10, then the IPv4 Node Address and Local Interface ID are present.

o If length is 14, then the IPv4 Node Address, the Local Interface ID and the MPLS SID are present.

2.4.3.2.6. Type 6: IPv4 Local and Remote addresses with optional SID

The Type-6 Segment Sub-TLV encodes an adjacency local address, an adjacency remote address and an optional SID in the form of an MPLS label. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Local IPv4 Address (4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Remote IPv4 Address  (4 octets)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     SID (optional, 4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 6 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

o Length is 10 or 14.

o Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

o Local IPv4 Address: a 4 octet IPv4 address.

o Remote IPv4 Address: a 4 octet IPv4 address.

o SID: 4 octet MPLS label.

The following applies to the Type-6 Segment sub-TLV:

o The Local IPv4 Address MUST be present and represents an adjacency local address.

o The Remote IPv4 Address MUST be present and represents the remote end of the adjacency.

o The SID is optional and specifies a 4 octet MPLS SID containing label, TC, S and TTL as defined in Section 2.4.3.2.1.

o If length is 10, then only the IPv4 Local and Remote addresses are present.

o If length is 14, then the IPv4 Local address, IPv4 Remote address and the MPLS SID are present.

2.4.3.2.7. Type 7: IPv6 Address + Interface ID for local and remote pair with optional SID for SR MPLS

The Type-7 Segment Sub-TLV encodes an IPv6 Link Local adjacency with IPv6 local node address, a local interface identifier (Local Interface ID), IPv6 remote node address , a remote interface identifier (Remote Interface ID) and an optional SID in the form of an MPLS label. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Local Interface ID (4 octets)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                IPv6 Local Node Address (16 octets)          //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Remote Interface ID (4 octets)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                IPv6 Remote Node Address (16 octets)         //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SID (optional, 4 octets)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 7 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

o Length is 22, 26, 42 or 46.

o Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

o Local Interface ID: 4 octets of interface index as defined in [I-D.ietf-pce-segment-routing].

- o IPv6 Local Node Address: a 16 octet IPv6 address.

- o Remote Interface ID: 4 octets of interface index as defined in [I-D.ietf-pce-segment-routing].

- o IPv6 Remote Node Address: a 16 octet IPv6 address.

- o SID: 4 octet MPLS label.

The following applies to the Type-7 Segment sub-TLV:

- o The Local Interface ID and IPv6 Local Node Address MUST be present.

- o The Remote Interface ID and Remote Node Address pair is optional. If Remote Interface ID is present, the Remote Node Address MUST be present as well. Similarly, if Remote Node Address is present, the Remote Interface ID MUST be present as well.

- o The SID is optional and specifies a 4 octet MPLS SID containing label, TC, S and TTL as defined in Section 2.4.3.2.1.

- o If length is 22, then the Local Interface ID and the Local IPv6 Address are present.

- o If length is 26, then the Local Interface ID, Local IPv6 Address and the MPLS SID are present.

- o If length is 42, then the Local Interface ID, Local IPv6 Node Address, Remote Interface ID, and the Remote IPv6 Node Address are present.

- o If length is 46, then the Local Interface ID, Local IPv6 Node Address, Remote Interface ID, Remote IPv6 Node Address and the MPLS SID are present.

2.4.3.2.8. Type 8: IPv6 Local and Remote addresses with optional SID for SR MPLS

The Type-8 Segment Sub-TLV encodes an adjacency local address, an adjacency remote address and an optional SID in the form of an MPLS label. The format is as follows:

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |     Type      |   Length      |     Flags     |   RESERVED    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //               Local IPv6 Address (16 octets)                //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    //               Remote IPv6 Address  (16 octets)              //
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     SID (optional, 4 octets)                  |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 8 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).
- Length is 34 or 38.
- Flags: 1 octet of flags as defined in Section 2.4.3.2.12.
- RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.
- Local IPv6 Address: a 16 octet IPv6 address.
- Remote IPv6 Address: a 16 octet IPv6 address.
- SID: 4 octet MPLS label.

The following applies to the Type-8 Segment sub-TLV:

- The Local IPv6 Address MUST be present and represents an adjacency local address.
- The Remote IPv6 Address MUST be present and represents the remote end of the adjacency.
- The SID is optional and specifies a 4 octet MPLS SID containing label, TC, S and TTL as defined in Section 2.4.3.2.1.
- If length is 34, then only the IPv6 Local and Remote addresses are present.
- If length is 38, then IPv6 Local and Remote addresses and the MPLS SID are present.

2.4.3.2.9. Type 9: IPv6 Node Address with optional SRv6 SID

The Type-9 Segment Sub-TLV encodes an IPv6 node address, SR Algorithm and an optional SID in the form of an IPv6 address. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |    Length     |     Flags     | SR Algorithm  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                 IPv6 Node Address (16 octets)               //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                     SID (optional, 16 octets)               //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- o Type: 10 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

- o Length is 18 or 34.

- o Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

- o SR Algorithm: 1 octet specifying SR Algorithm as described in section 3.1.1 in [I-D.ietf-spring-segment-routing], when A-Flag as defined in Section 2.4.3.2.12 is present. SR Algorithm is used by SRTE process as described in section 4 in ([I-D.filsfils-spring-segment-routing-policy]). When A-Flag is not encoded, this field SHOULD be unset on transmission and MUST be ignored on receipt.

- o IPv6 Node Address: a 16 octet IPv6 address.

- o SID: 16 octet IPv6 address.

The following applies to the Type-9 Segment sub-TLV:

- o The IPv6 Node Address MUST be present.

- o The SID is optional and specifies a SRv6 SID in the form of 16 octet IPv6 address.

- o If length is 18, then only the IPv6 Node Address is present.

- o If length is 34, then the IPv6 Node Address and the SRv6 SID are present.

2.4.3.2.10. Type 10: IPv6 Address + Interface ID for local and remote pair for SRv6 with optional SID

The Type-10 Segment Sub-TLV encodes an IPv6 Link Local adjacency with local node address, a local interface identifier (Local Interface ID), remote IPv6 node address , a remote interface identifier (Remote Interface ID) and an optional SID in the form of an IPv6 address. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Local Interface ID (4 octets)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                IPv6 Local Node Address (16 octets)          //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Remote Interface ID (4 octets)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                IPv6 Remote Node Address (16 octets)         //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                SID (optional, 16 octets)                    //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- o Type: 11 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

- o Length is 22, 38, 42 or 58.

- o Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

- o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

- o Local Interface ID: 4 octets of interface index as defined in [I-D.ietf-pce-segment-routing].

- o IPv6 Local Node Address: a 16 octet IPv6 address.

- o Remote Interface ID: 4 octets of interface index as defined in [I-D.ietf-pce-segment-routing].

- o IPv6 Remote Node Address: a 16 octet IPv6 address.

- o SID: 16 octet IPv6 address.

The following applies to the Type-10 Segment sub-TLV:

- o The Local Interface ID and the Local IPv6 Node Addresses MUST be present.

- o The Remote Interface ID and Remote Node Address pair is optional. If Remote Interface ID is present, the Remote Node Address MUST be present as well. Similarly, if Remote Node Address is present, the Remote Interface ID MUST be present as well.

- o The SID is optional and specifies a SRv6 SID in the form of 16 octet IPv6 address.

- o If length is 22, then the Local Interface ID, Local IPv6 Node Address, are present.

- o If length is 38, then the Local Interface ID, Local IPv6 Node Address and the SRv6 SID are present.

- o If length is 42, then the Local Interface ID, Local IPv6 Node Address, Remote Interface ID, and the Remote IPv6 Node Address are present.

- o If length is 58, then the Local Interface ID, Local IPv6 Node Address, Remote Interface ID, Remote IPv6 Node Address and the SRv6 SID are present.

2.4.3.2.11. Type 11: IPv6 Local and Remote addresses for SRv6 with optional SID

The Type-11 Segment Sub-TLV encodes an adjacency local address, an adjacency remote address and an optional SID in the form of IPv6 address. The format is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length      |     Flags     |   RESERVED    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//               Local IPv6 Address (16 octets)                //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//               Remote IPv6 Address  (16 octets)              //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                SID (optional, 16 octets)                    //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 12 (to be assigned by IANA from the registry "SR Policy List Sub-TLVs" defined in this document).

o Length is 34 or 50.

o Flags: 1 octet of flags as defined in Section 2.4.3.2.12.

o RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

o Local IPv6 Address: a 16 octet IPv6 address.

o Remote IPv6 Address: a 16 octet IPv6 address.

o SID: 16 octet IPv6 address.

The following applies to the Type-11 Segment sub-TLV:

o The Local IPv6 Node Address MUST be present.

o The Remote IPv6 Node Address MUST be present.

o The SID is optional and specifies a SRv6 SID in the form of 16 octet IPv6 address.

o If length is 34, then the Local IPv6 Node Address and the Remote IPv6 Node Address are present.

o If length is 50, then the Local IPv6 Node Address, the Remote IPv6 Node Address and the SRv6 SID are present.

2.4.3.2.12. Segment Flags

The Segment Types described above MAY contain following flags in the "Flags" field (codes to be assigned by IANA from the registry "SR Policy Segment Flags" defined in this document Section 8.6):

```
 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+
|V|A|           |
+-+-+-+-+-+-+-+-+
```

where:

V-Flag: This flag encodes the "Segment Validation" behavior. It is used by SRTE process as described in section 5 in ([I-D.filsfils-spring-segment-routing-policy]).

A-Flag: This flag indicates the presence of SR Algorithm id in the "SR Algorithm" field applicable to various Segment Types. SR Algorithm is used by SRTE process as described in section 4 in ([I-D.filsfils-spring-segment-routing-policy]).

Unused bits in the Flag octet SHOULD be set to zero upon transmission and MUST be ignored upon receipt.

The following applies to the Segment Flags:

- V-Flag is applicable to all Segment Types.

- A-Flag is applicable to Segment Types 3, 4 and 9. If A-Flag appears with any other Segment Type, it MUST be ignored.

2.4.4. Explicit NULL Label Policy Sub-TLV

In order to steer an unlabeled IP packet into an SR policy, it is necessary to create a label stack for that packet, and to push one or more labels onto that stack.

The Explicit NULL Label Policy sub-TLV is used to indicate whether an Explicit NULL Label [RFC3032] must be pushed on an unlabeled IP packet before any other labels.

If an Explicit NULL Label Policy Sub-TLV is not present, the decision of whether to push an Explicit NULL label on a given packet is a matter of local policy.

The contents of this sub-TLV are used by the SRTE process as described in section 4.1 in [I-D.filsfils-spring-segment-routing-policy].

```
 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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     Type      |    Length     |     Flags     |   RESERVED    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     ENLP      |
 +-+-+-+-+-+-+-+-+
```

Where:

Type: TBD1 (to be assigned by IANA from the registry "BGP Tunnel Encapsulation Attribute sub-TLVs" defined in this document Section 8.3).

Length: 3.

Flags: 1 octet of flags. None are defined at this stage. Flags SHOULD be set to zero on transmission and MUST be ignored on receipt.

RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

ENLP(Explicit NULL Label Policy): Indicates whether Explicit NULL labels are to be pushed on unlabeled IP packets that are being steered into a given SR policy. This field has one of the following 4 values:

1: Push an IPv4 Explicit NULL label on an unlabeled IPv4 packet, but do not push an IPv6 Explicit NULL label on an unlabeled IPv6 packet.

2: Push an IPv6 Explicit NULL label on an unlabeled IPv6 packet, but do not push an IPv4 Explicit NULL label on an unlabeled IPv4 packet.

3: Push an IPv4 Explicit NULL label on an unlabeled IPv4 packet, and push an IPv6 Explicit NULL label on an unlabeled IPv6 packet.

4: Do not push an Explicit NULL label.

The policy signaled in this Sub-TLV MAY be overridden by local policy.

2.4.5. Policy Priority Sub-TLV

An operator MAY set the Policy Priority sub-TLV to indicate the order in which the SR policies are re-computed upon topological change.

The Priority sub-TLV does not have any effect on the BGP bestpath selection or propagation procedures. The contents of this sub-TLV are used by the SRTE process as described in section 2.11 in ([I-D.filsfils-spring-segment-routing-policy]).

The Priority sub-TLV is optional and it MUST not appear more than once in the SR Policy TLV. If the Priority sub-TLV appears more than once, the update is considered malformed and the "treat-as-withdraw" strategy of [RFC7606] is applied.

The Priority sub-TLV has following format:

```
 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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     Type      |    Length     |   Priority    |   RESERVED    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: TBD2 (to be assigned by IANA from the registry "BGP Tunnel Encapsulation Attribute sub-TLVs" defined in this document Section 8.3).

Length: 2.

Priority: a 1-octet value.

RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

2.4.6. Policy Name Sub-TLV

An operator MAY set the Policy Name sub-TLV to attach a symbolic name to the SR Policy candidate path.

Usage of Policy Name sub-TLV is described in section 2 in ([I-D.filsfils-spring-segment-routing-policy]).

The Policy Name sub-TLV may exceed 255 bytes length due to long policy name. Therefore a 2-octet length is required. According to [I-D.ietf-idr-tunnel-encaps], the first bit of the sub-TLV codepoint defines the size of the length field. Therefore, for the Policy Name sub-TLV a code point of 128 (or higher) is used. See Section 8 for details of codepoints allocation.

The Policy Name sub-TLV is optional and it MUST not appear more than once in the SR Policy TLV. If the Policy Name sub-TLV appears more than once, the update is considered malformed and the "treat-as-withdraw" strategy of [RFC7606] is applied.

The Policy Name sub-TLV has following format:

```
 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
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |     Type      |    Length                     |   RESERVED    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 //                          Policy Name                         //
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: TBD3 (to be assigned by IANA from the registry "BGP Tunnel Encapsulation Attribute sub-TLVs" defined in this document Section 8.3).

Length: Variable.

RESERVED: 1 octet of reserved bits. SHOULD be unset on transmission and MUST be ignored on receipt.

Policy Name: Symbolic name for the policy. It SHOULD be a string of printable ASCII characters, without a NULL terminator.

3. Extended Color Community

The Color Extended Community as defined in [I-D.ietf-idr-tunnel-encaps] is used to steer traffic into a policy.

When the Color Extended Community is used for the purpose of steering the traffic into an SRTE policy, the RESERVED field (as defined in [I-D.ietf-idr-tunnel-encaps] is changed as follows:

```
                     1
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|C O|        RESERVED           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where CO bits are defined as the "Color-Only" bits. [I-D.filsfils-spring-segment-routing-policy]defines the influence of these bits on the automated steering of BGP Payload traffic onto SRTE policies.

4. SR Policy Operations

As described in this document, the consumer of a SR Policy NLRI is not the BGP process. The BGP process is in charge of the origination and propagation of the SR Policy NLRI but its installation and use is outside the scope of BGP ([I-D.filsfils-spring-segment-routing-policy]).

4.1. Configuration and Advertisement of SR TE Policies

Typically, but not limited to, an SR Policy is configured into a controller.

Multiple SR Policy NLRIs may be present with the same <color, endpoint> tuple but with different content when these SR policies are intended to different head-ends.

The distinguisher of each SR Policy NLRI prevents undesired BGP route selection among these SR Policy NLRIs and allow their propagation across route reflectors [RFC4456].

Moreover, one or more route-target SHOULD be attached to the advertisement, where each route-target identifies one or more intended head-ends for the advertised SR policy.

If no route-target is attached to the SR Policy NLRI, then it is assumed that the originator sends the SR Policy update directly (e.g., through a BGP session) to the intended receiver. In such case, the NO_ADVERTISE community MUST be attached to the SR Policy update.

4.2. Reception of an SR Policy NLRI

On reception of an SR Policy NLRI, a BGP speaker MUST determine if it's first acceptable, then it determines if it is usable.

4.2.1. Acceptance of an SR Policy NLRI

When a BGP speaker receives an SR Policy NLRI from a neighbor it has to determine if it's acceptable. The following applies:

- o The SR Policy NLRI MUST include a distinguisher, color and endpoint field which implies that the length of the NLRI MUST be either 12 or 24 octets (depending on the address family of the endpoint).

- o The SR Policy update MUST have either the NO_ADVERTISE community or at least one route-target extended community in IPv4-address format. If a router supporting this document receives an SR policy update with no route-target extended communities and no NO_ADVERTISE community, the update MUST NOT be sent to the SRTE process. Furthermore, it SHOULD be considered to be malformed, and the "treat-as-withdraw" strategy of [RFC7606] is applied.

- o The Tunnel Encapsulation Attribute MUST be attached to the BGP Update and MUST have a Tunnel Type TLV set to SR Policy (codepoint is 15, assigned by IANA (see Section 8) from the "BGP Tunnel Encapsulation Attribute Tunnel Types" registry).

A router that receives an SR Policy update that is not valid according to these criteria MUST treat the update as malformed. The

route MUST NOT be passed to the SRTE process, and the "treat-as-withdraw" strategy of [RFC7606] is applied.

A unacceptable SR Policy update that has a valid NLRI portion with invalid attribute portion MUST be considered as a withdraw of the SR Policy.

4.2.2. Usable SR Policy NLRI

If one or more route-targets are present, then at least one route-target MUST match one of the BGP Identifiers of the receiver in order for the update to be considered usable. The BGP Identifier is defined in [RFC4271] as a 4 octet IPv4 address. Therefore the route-target extended community MUST be of the same format.

If one or more route-targets are present and no one matches any of the local BGP Identifiers, then, while the SR Policy NLRI is acceptable, it is not usable. It has to be noted that if the receiver has been explicitly configured to do so, it MAY propagate the SR Policy NLRI to its neighbors as defined in Section 4.2.4.

Usable SR Policy NLRIs are sent to the Segment Routing Traffic Engineering (SRTE) process. The description of the SRTE process is outside the scope of this document and it's described in [I-D.filsfils-spring-segment-routing-policy].

4.2.3. Passing a usable SR Policy NLRI to the SRTE Process

Once BGP has determined that the SR Policy NLRI is usable, BGP passes the path to the SRTE process described in ([I-D.filsfils-spring-segment-routing-policy]). Note that, along with the path details, BGP also passes the originator information for breaking ties in the path-selection process as described in section 2.4 in [I-D.filsfils-spring-segment-routing-policy].

The SRTE process applies the rules defined in section 2 [I-D.filsfils-spring-segment-routing-policy] to determine whether a path is valid and to select the best path among the valid paths.

4.2.4. Propagation of an SR Policy

By default, a BGP node receiving an SR Policy NLRI MUST NOT propagate it to any EBGP neighbor.

However, a node MAY be explicitly configured to advertise a received SR Policy NLRI to neighbors according to normal BGP rules (i.e., EBGP propagation by an ASBR or iBGP propagation by a Route-Reflector).

SR Policy NLRIs that have been determined acceptable and valid can be propagated, even the ones that are not usable.

Only SR Policy NLRIs that do not have the NO_ADVERTISE community attached to them can be propagated.

4.3. Flowspec and SR Policies

The SR Policy can be carried in context of a Flowspec NLRI ([RFC5575]). In this case, when the redirect to IP next-hop is specified as in [I-D.ietf-idr-flowspec-redirect-ip], the tunnel to the next-hop is specified by the segment list in the Segment List sub-TLVs. The Segment List (e.g., label stack or IPv6 segment list) is imposed to flows matching the criteria in the Flowspec route to steer them towards the next-hop as specified in the SR Policy SAFI NLRI.

5. Contributors

Arjun Sreekantiah
Cisco Systems
US

Email: asreekan@cisco.com

Acee Lindem
Cisco Systems
US

Email: acee@cisco.com

Siva Sivabalan
Cisco Systems
US

Email: msiva@cisco.com

Imtiyaz Mohammad
Arista Networks
India

Email: imtiyaz@arista.com

Gaurav Dawra
Cisco Systems
US

Email: gdawra.ietf@gmail.com

6. Acknowledgments

The authors of this document would like to thank Shyam Sethuram, John Scudder, Przemyslaw Krol, Alex Bogdanov, Nandan Saha and Ketan Talaulikar for their comments and review of this document.

7. Implementation Status

Note to RFC Editor: Please remove this section prior to publication, as well as the reference to RFC 7942.

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

Several early implementations exist and will be reported in detail in a forthcoming version of this document. For purposes of early interoperability testing, when no FCFS code point was available, implementations have made use of the following values:

- o Preference sub-TLV: 12
- o Binding SID sub-TLV: 13
- o Segment List sub-TLV: 128

When IANA-assigned values are available, implementations will be updated to use them.

8. IANA Considerations

This document defines new Sub-TLVs in following existing registries:

- Subsequent Address Family Identifiers (SAFI) Parameters
- BGP Tunnel Encapsulation Attribute Tunnel Types
- BGP Tunnel Encapsulation Attribute sub-TLVs

This document also defines following new registries:

- SR Policy List Sub-TLVs
- SR Policy Binding SID Flags
- SR Policy Segment Flags

8.1. Existing Registry: Subsequent Address Family Identifiers (SAFI) Parameters

This document defines a new SAFI in the registry "Subsequent Address Family Identifiers (SAFI) Parameters" that has been assigned by IANA:

Codepoint	Description	Reference
73	SR Policy SAFI	This document

8.2. Existing Registry: BGP Tunnel Encapsulation Attribute Tunnel Types

This document defines a new Tunnel-Type in the registry "BGP Tunnel Encapsulation Attribute Tunnel Types" that has been assigned by IANA:

Codepoint	Description	Reference
15	SR Policy Type	This document

8.3. Existing Registry: BGP Tunnel Encapsulation Attribute sub-TLVs

This document defines new sub-TLVs in the registry "BGP Tunnel Encapsulation Attribute sub-TLVs" to be assigned by IANA:

Codepoint	Description	Reference
12	Preference sub-TLV	This document
13	Binding SID sub-TLV	This document
128	Segment List sub-TLV	This document
TBD1	ENLP sub-TLV	This document
TBD2	Priority sub-TLV	This document
TBD3	Policy Name sub-TLV	This document

8.4. New Registry: SR Policy List Sub-TLVs

This document defines a new registry called "SR Policy List Sub-TLVs". The allocation policy of this registry is "First Come First Served (FCFS)" according to [RFC8126].

Following Sub-TLV codepoints are defined:

Value	Description	Reference
1	MPLS SID sub-TLV	This document
2	SRv6 SID sub-TLV	This document
3	IPv4 Node and SID sub-TLV	This document
4	IPv6 Node and SID for SR-MPLS sub-TLV	This document
5	IPv4 Node, index and SID sub-TLV	This document
6	IPv4 Local/Remote addresses and SID sub-TLV	This document
7	IPv6 Node, index for remote and local pair and SID for SR-MPLS sub-TLV	This document
8	IPv6 Local/Remote addresses and SID sub-TLV	This document
9	Weight sub-TLV	This document
10	IPv6 Node and SID for SRv6 sub-TLV	This document
11	IPv6 Node, index for remote and local pair and SID for SRv6 sub-TLV	This document
12	IPv6 Local/Remote addresses and SID for SRv6 sub-TLV	This document

8.5. New Registry: SR Policy Binding SID Flags

This document defines a new registry called "SR Policy Binding SID Flags". The allocation policy of this registry is "First Come First Served (FCFS)" according to [RFC8126].

Following Flags are defined:

Bit	Description	Reference
0	Drop Upon Invalid Flag (I-Flag)	This document
1	Specified-BSID-Only Flag (S-Flag)	This document
2-7	Unassigned	

8.6. New Registry: SR Policy Segment Flags

This document defines a new registry called "SR Policy Segment Flags". The allocation policy of this registry is "First Come First Served (FCFS)" according to [RFC8126].

Following Flags are defined:

Bit	Description	Reference
0	Segment Validation Flag (V-Flag)	This document
1	SR Algorithm Flag (A-Flag)	This document
2-7	Unassigned	

9. Security Considerations

TBD.

10. References

10.1. Normative References

[I-D.filsfils-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., Raza, K., Liste, J., Clad, F., Talaulikar, K., Ali, Z., Hegde, S., daniel.voyer@bell.ca, d., Lin, S., bogdanov@google.com, b., Krol, P., Horneffer, M., Steinberg, D., Decraene, B., Litkowski, S., and P. Mattes, "Segment Routing Policy for Traffic Engineering", draft-filsfils-spring-segment-routing-policy-05 (work in progress), February 2018.

[I-D.ietf-idr-tunnel-encaps]
Rosen, E., Patel, K., and G. Velde, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-09 (work in progress), February 2018.

[I-D.ietf-pce-segment-routing]
Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W.,
and J. Hardwick, "PCEP Extensions for Segment Routing", draft-ietf-pce-segment-routing-11 (work in progress), November 2017.

[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>.

[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>.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, DOI 10.17487/RFC4360, February 2006, <https://www.rfc-editor.org/info/rfc4360>.

[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>.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <https://www.rfc-editor.org/info/rfc5575>.

[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>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

10.2. Informational References

[I-D.ietf-6man-segment-routing-header]
Previdi, S., Filsfils, C., Leddy, J., Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-12 (work in progress), April 2018.

[I-D.ietf-idr-flowspec-redirect-ip]
Uttaro, J., Haas, J., Texier, M., Andy, A., Ray, S., Simpson, A., and W. Henderickx, "BGP Flow-Spec Redirect to IP Action", draft-ietf-idr-flowspec-redirect-ip-02 (work in progress), February 2015.

[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>.

[RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <https://www.rfc-editor.org/info/rfc7942>.

Authors' Addresses

Stefano Previdi (editor)
Cisco Systems, Inc.
IT

Email: stefano@previdi.net

Clarence Filsfils
Cisco Systems, Inc.
Brussels
BE

Email: cfilsfil@cisco.com

Dhanendra Jain (editor)
Cisco Systems, Inc.
San Jose
USA

Email: dhjain@cisco.com

Paul Mattes
Microsoft
One Microsoft Way
Redmond, WA 98052
USA

Email: pamattes@microsoft.com

Eric Rosen
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
US

Email: erosen@juniper.net

Steven Lin
Google

Email: stevenlin@google.com