Network Working Group
Internet-Draft
Updates: 9012 (if approved)
Intended status: Standards Track
Expires: December 18, 2022

S. Previdi
Huawei Technologies
C. Filsfils
Cisco Systems
K. Talaulikar, Ed.
Arrcus Inc
P. Mattes
Microsoft
D. Jain
S. Lin
Google

June 16, 2022

Advertising Segment Routing (SR) Policies in BGP
draft-ietf-idr-segment-routing-te-policy-18

Abstract

This document introduces a BGP SAFI with two NLRIs to advertise a candidate path of a Segment Routing (SR) Policy. An SR Policy is a set of candidate paths, each 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~~ provided via several different mechanisms, e.g., CLI, ~~NetConf~~NETCONF, PCEP, or BGP. This document specifies how BGP may be used to distribute SR Policy candidate paths. It defines sub-TLVs for the Tunnel Encapsulation Attribute for signaling information about these candidate paths.

This documents updates RFC9012 with extensions to the Color Extended Community to support additional steering modes over SR Policy.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on December 18, 2022.

Copyright Notice

Commenté [BMI1]: Please expand acronyms at first use: https://www.rfc-editor.org/materials/abbrev.expansion.txt

Well-known ones can be left as such. See the guide for more details.

Commenté [BMI2]: I would use a wording that is aligned with the definition in RFC8402:

SR Policy: an ordered list of segments.

That definition is inherited in draft-ietf-spring-segment-routing-policy:

```
   A Segment Routing Policy (SR Policy)
[RFC8402] is an ordered list of
   segments (i.e., instructions) that
represent a source-routed policy.
```

draft-ietf-spring-segment-routing-policy-22#section-2.2 says:

An SR Policy is associated with one or more candidate paths.

Commenté [BMI3]: This seems to be redundant as the SR Policy is defined as a set of candidate paths.

a mis en forme : Soulignement

Commenté [BMI4]: As the SR policy already includes a set of paths.

Commenté [BMI5]: I would just delete this sentence.

a mis en forme : Surlignage

a mis en forme : Surlignage

Table of Contents

1. Introduction

Segment Routing (SR) [RFC8402] allows a headend node to steer a packet flow along ~~any~~ a specific path. Intermediate per-path states are eliminated thanks to source routing.

The headend node is said to steer a flow into an SR Policy [RFC8402].

The packets steered into an SR Policy carry an ordered list of segments associated with that SR Policy.

[I-D.ietf-spring-segment-routing-policy] further details the concepts of SR Policy and steering into an SR Policy. These apply equally to the SR-MPLS and Segment Routing for IPv6 (SRv6) data-plane instantiations of Segment Routing using ~~SR-MPLS and SRv6~~ Segment Identifiers (SIDs) as described in [RFC8402]. [RFC8660] describes the representation and processing of this ordered list of segments as an MPLS label stack for SR-MPLS. While [RFC8754] and [RFC8986] describe the same for SRv6 with the use of the Segment Routing Header (SRH).

Commenté [BMI6]: For the readability, I suggest to include a mention that basically says that an SR Policy is identified through the tuple <headend, color, endpoint>. This will prepare the readers to understand the NLRI format.

The SR Policy related functionality described in [I-D.ietf-spring-segment-routing-policy] can be conceptually viewed as being incorporated in an SR Policy Module (SRPM). Following is a reminder of the high-level functionality of SRPM:

- Learning multiple candidate paths for an SR Policy via various mechanisms (CLI, ~~NetConf~~NETCONF, PCEP, or BGP).
- Selection of the best candidate path for an SR Policy.
- Binding ~~B~~SID (BSID) to the selected candidate path of an SR Policy.
- Installation of the selected candidate path and its BSID in the forwarding plane.

Commenté [BMI7]: Is there always one selected path? Or multiple paths are possible? If so, I suggest to make this explicit.

This document specifies ~~the way~~how to use BGP to distribute one or more of the candidate paths of an SR Policy to the headend of that policy. The document describes the functionality provided by BGP and, as appropriate, provides references for the functionality which is outside the scope of BGP (i.e., resides within SRPM on the headend node).

This document specifies a way of representing SR Policy candidate paths in BGP UPDATE messages. BGP can then be used to propagate the SR Policy candidate paths to the headend nodes in ~~the~~ a network. The usual BGP rules for BGP propagation and best-path 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 SRPM. The SRPM 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 SR Policy candidate paths into the data plane.

a mis en forme : Surlignage

Commenté [BMI8]: Need to be consistent with the second bullet.

a mis en forme : Surlignage

Commenté [BMI9]: Do we need to include hints how the selection is made. Adding a pointer to where this is discussed would be useful.

Commenté [BMI10]: What does that mean?

This document introduces a BGP subsequent address family (SAFI) for IPv4 and IPv6 address families. In UPDATE messages of those AFIs/SAFI~~s~~, the NLRI identifies an SR Policy Candidate Path while the attributes encode the segment lists and other details of that SR Policy Candidate Path.

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.

Commenté [BMI11]: Cited the section where this is defined.

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

Commenté [BMI12]: Which controller?

Commenté [BMI13]: For consistency with RFC8402

Commenté [BMI14]: Elaborate further on the reasoning.

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~~headend 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]).

Commenté [BMI15]: Isn't this deployment-specific? Not sure I would maintain "most" but simply provide a description of use in the presence of RRs.

In some situations, it is undesirable for a controller or BGP egress router to have a BGP session to each policy ~~head-end~~headend. 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~~headend for the advertised policy. This is done by attaching one or more Route Target Extended Communities to the advertisement ~~(~~[RFC4360]~~)~~.

Commenté [BMI16]: Does these ASes belong to the same "single trusted domain"?

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

Commenté [BMI17]: I guess you mean those specified in this draft. Please consider adding pointers to locate where each of these extensions is defined.

o A Subsequent Address Family Identifier (SAFI) whose NLRIs identifies an SR Policy candidate path.

o A Tunnel Type identifier for SR Policy, and a set of sub-TLVs to be inserted into the Tunnel Encapsulation Attribute (as defined in [RFC9012]) specifying segment lists of the SR Policy candidate path, 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~~headend of such an SR Policy advertisement.

The Color Extended Community (as defined in [RFC9012]) is used to steer traffic into an SR Policy, as described in ~~section~~ Section 8.8 of [I-D.ietf-spring-segment-routing-policy]. ~~This document (~~Section 3~~)~~ updates [RFC9012] with modifications to 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", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. SR Policy Encoding

2.1. SR Policy SAFI and NLRI

A SAFI is introduced ~~by~~ in this document: the SR Policy SAFI with codepoint 73. The AFI used MUST be IPv4(1) or IPv6(2).

The SR Policy SAFI uses the NLRI format 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]. ~~When AFI = 1 value~~ It MUST be set 96 when AFI = 1 and ~~when AFI = 2 value MUST be~~ 192 when AFI = 2.

- Distinguisher: 4-octet value uniquely identifying the policy in the context of <color, endpoint> ~~tuple~~pair. The distinguisher has no semantic value and is solely used by the SR Policy originator to make unique (from an NLRI perspective) both for multiple candidate paths of the same SR Policy as well as candidate paths of different SR Policies (i.e., with different segment lists) with the same Color and Endpoint but meant for different ~~head-end~~headends.

- 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 as specified in [I-D.ietf-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

Commenté [BMI18]: Actually, the format is not updated. You are just associating a meaning with some of these (reserved) flags.

Commenté [BMI19]: I don't see such field out there. Do you mean the "flags" bits?

Commenté [BMI20]: Consider adding pointers where terms used in the document are defined.

Commenté [BMI21]: The current registry points to [draft-previdi-idr-segment-routing-te-policy]. Need to update that entry.

Commenté [BMI22]: Please add the explicit section where this is defined.

(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 as described in [I-D.ietf-spring-segment-routing-policy].

The NLRI containing ~~the~~ an SR Policy candidate path is carried in a BGP UPDATE message [RFC4271] using BGP multi-protocol extensions [RFC4760] with an AFI of 1 or 2 (IPv4 or IPv6) and with a SAFI of 73.

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. Concretely~~-~~, i~~I~~f 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 unicast IPv6 address; 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~~ 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 leverages the existing BGP propagation and best-path selection rules. Details of the procedures are described in Section 4.

~~It has to be noted that i~~If several candidate paths of the same SR Policy (endpoint, color) are signaled via BGP to a ~~head-end~~headend, it is RECOMMENDED that each NLRI uses 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 SRPM as different candidate paths along with their respective originator information (i.e., ASN and BGP Router-ID) as described in ~~section~~ Section 2.4 of [I-D.ietf-spring-segment-routing-policy]. The ASN would be the ASN of the origin and the BGP Router-ID is determined in the following order:

- From the Route Origin Community [RFC4360] if present and carrying an IP Address

- As the BGP Originator ID [RFC4456] if present

- As the BGP Router-ID of the peer from which the update was received as a last resort.

2.2. SR Policy and Tunnel Encapsulation Attribute

Commenté [BMI23]: Not sure to get what is meant here.

Commenté [BMI24]: Please add the exact section

Commenté [BMI25]: If they are mandatory, why MUST is needed here?

Commenté [BMI26]: That is the definition of optional attributes. I would delete or reword.

Commenté [BMI27]: Just out of curiosity, I wonder whether ULAs are allowed?
What is the behavior to follow if this is not a GUA?

Commenté [BMI28]: Which procedure?

Commenté [BMI29]: What are the implications if the reco is not followed?

The content of the SR Policy Candidate Path is encoded in the Tunnel Encapsulation Attribute defined in [RFC9012] using a Tunnel-Type called SR Policy Type with codepoint 15.

The use of SR Policy
Tunnel-type is applicable only for the AFI/SAFI pairs of (1/73, 2/73).

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 (15)
             Binding SID
             SRv6 Binding SID
             Preference
             Priority
             Policy Name
             Policy Candidate Path 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 [RFC9012].

- Tunnel-Type is set to 15.

- Preference, Binding SID, SRv6 Binding SID, Priority, Policy Name, Policy Candidate Path Name, ENLP, Segment-List, Weight, and Segment sub-TLVs are defined in ~~this document~~Section 2.4.

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

A Tunnel Encapsulation Attribute MUST NOT contain more than one TLV of type "SR Policy".

Commenté [BMI30]: Add an explicit mention about error handling if this is not the case.

2.3. Remote Endpoint and Color

The Tunnel Egress Endpoint and Color sub-TLVs, as defined in [RFC9012], may also be present in the SR Policy encodings.

a mis en forme : Surlignage

The Tunnel Egress Endpoint and Color Sub-TLVs of the Tunnel Encapsulation Attribute 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 Tunnel Egress Endpoint sub-TLV and the Color sub-TLV MUST be ignored by the BGP speaker and MUST NOT~~not~~ be removed from the Tunnel Encapsulation Attribute during propagation.

Commenté [BMI31]: The rationale should be explicited.

2.4. SR Policy Sub-TLVs

This section specifies the sub-TLVs defined for encoding the information about the SR Policy Candidate Path:~~.~~

Preference, Binding SID, SRv6 Binding SID, Segment-List, Priority, Policy Name, Policy Candidate Path Name, and Explicit NULL Label Policy ~~are the~~ sub-TLVs~~ introduced for the BGP Tunnel Encapsulation Attribute [RFC9012] being defined in this section~~.

Weight and Segment are sub-TLVs of the Segment-List sub-TLV mentioned above.

None of the sub-TLVs defined in the following sub-sections have any effect on the BGP best-path selection or propagation procedures. These sub-TLVs are not used by BGP path selection process and are instead passed on to SRPM as SR Policy Candidate Path information for further processing described in [I-D.ietf-spring-segment-routing-policy].

The use of SR Policy Sub-TLVs is applicable only for the AFI/SAFI pairs of (1/73, 2/73). Future documents may extend their applicability to other AFI/SAFI.

2.4.1. Preference Sub-TLV

The Preference sub-TLV is used to carry the preference of ~~the~~ an SR Policy candidate path. The contents of this sub-TLV are used by the SRPM as described in ~~section~~ Section 2.7 ~~in~~of [I-D.ietf-spring-segment-routing-policy].

The Preference sub-TLV is optional. This sub-TLV ~~and it~~ MUST NOT appear more than once in the SR Policy encoding.

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:

- Type: 12
- Length: 6 octets.
- Flags: 1 octet of flags. None are defined ~~at this stage~~in this document. Flags ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.
- RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

Commenté [BMI32]: Already mentioned.

Commenté [BMI33]: Which section are you referring to?

Commenté [BMI34]: This is redundant with

```
The use of SR Policy
   Tunnel-type is applicable only for the
AFI/SAFI pairs of (1/73,
   2/73).
```

Commenté [BMI35]: What is the error handling if this is not followed.

Commenté [BMI36]: Do you need a registry for maintaining future assignments?

Commenté [BMI37]: Can these be assigned in the future? If so, I would use "Unassigned" as per RFC8126.

FWIW, 8126 says the following:

```
      Unassigned:  Not currently
assigned, and available for assignment
          via documented procedures.
While it's generally clear that
          any values that are not
registered are unassigned and
          available for assignment, it
is sometimes useful to
          explicitly specify that
situation.  Note that this is
          distinctly different from
"Reserved".


      Reserved:  Not assigned and not
available for assignment.
          Reserved values are held for
special uses, such as to extend
          the namespace when it becomes
exhausted.  "Reserved" is also
          sometimes used to designate
values that had been assigned
          but are no longer in use,
keeping them set aside as long as
          other unassigned values are
available.  Note that this is
          distinctly different from
"Unassigned".
```

o Preference: a 4-octet value. A higher value indicates higher preference and the default preference value is 100.

Commenté [BMI38]: Provide more description.

2.4.2. Binding SID Sub-TLV

The Binding SID sub-TLV is used to signal the binding SID related information of the SR Policy candidate path. The contents of this sub-TLV are used by the SRPM as described in ~~section~~ Section 6 in [I-D.ietf-spring-segment-routing-policy].

The Binding SID sub-TLV is optional and ~~it~~ MUST NOT appear more than once in the SR Policy encoding.

When the Binding SID sub-TLV is used to signal an SRv6 SID, the choice of its SRv6 Endpoint Behavior [RFC8986] to be instantiated is left to the headend node. It is RECOMMENDED that the SRv6 Binding SID sub-TLV defined in Section 2.4.3, that enables the specification of the SRv6 Endpoint Behavior, be used for signaling of an SRv6 Binding SID for an SR Policy candidate path.

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~~ “Binding SID + 2” not including Type and Length fields. Can be 2 or 6 or 18.

o Flags: 1 octet of flags. The following initial flags are defined in the registry "SR Policy Binding SID Flags" as described in Section 6.6:

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

Commenté [BMI39]: Add a mention that multiple flags can be set simultaneously unless this is precluded in future flag assignment.

where:

- S-Flag: This flag encodes the "Specified-BSID-only" behavior. It is used by SRPM as described in ~~section~~ Section 6.2.3 in [I-D.ietf-spring-segment-routing-policy].

- I-Flag: This flag encodes the "Drop Upon Invalid" behavior. It is used by SRPM as described in S~~s~~ection 8.2 in [I-D.ietf-spring-segment-routing-policy].

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

- o RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

Commenté [BMI40]: Idem as above

- o Binding SID: if the length is 2, then no Binding SID is present. If the length is 6 then the Binding SID is encoded in 4 octets using the format below. Traffic Class (TC), S, and TTL (Total of 12 bits) are RESERVED and ~~SHOULD~~ MUST be set to zero and MUST be ignored.

Commenté [BMI41]: It is weird to define a semantic but then ask to ignore it.

```
  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     |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Commenté [BMI42]: Add a description. See e.g., Section 3.6 of RFC9012

If the length is 18 then the Binding SID contains a 16-octet SRv6 SID.

2.4.3. SRv6 Binding SID Sub-TLV

The SRv6 Binding SID sub-TLV is used to signal the SRv6 Binding SID related information of ~~the~~ an SR Policy candidate path. It enables the specification of the SRv6 Endpoint Behavior [RFC8986] to be instantiated on the headend node. The contents of this sub-TLV are used by the SRPM as described in ~~section~~ Section 6 in [I-D.ietf-spring-segment-routing-policy].

The SRv6 Binding SID sub-TLV is optional. More than one SRv6 Binding SID sub-TLVs MAY be signaled in the same SR Policy encoding to indicate one or more SRv6 SIDs, each with potentially different SRv6 Endpoint Behaviors to be instantiated.

Commenté [BMI43]: Does the ordering have an importance? If so, please reflect that in the tex.

a mis en forme : Surlignage

The SRv6 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    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 SRv6 Binding SID (16 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//     SRv6 Endpoint Behavior and SID Structure (optional)     //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- o Type: TBD1

Commenté [BMI44]: To ease updating the document when a value is assigned

I guess the temporary assigned value is 20.

- o Length is variable. It MUST be set to the length of “SRv6 Endpoint Behavior and SID Structure” + 18 octets.

- o Flags: 1 octet of flags. The following flags are defined in the registry "SR Policy Binding SID Flags" as described in Section 6.7:

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

Commenté [BMI45]: Add a mention that multiple flags can be set simultaneously unless this is precluded in future flag assignment.

where:

- S-Flag: This flag encodes the "Specified-BSID-only" behavior. It is used by SRPM as described in S~~s~~ection 6.2.3 in [I-D.ietf-spring-segment-routing-policy].

- I-Flag: This flag encodes the "Drop Upon Invalid" behavior. It is used by SRPM as described in S~~s~~ection 8.2 in [I-D.ietf-spring-segment-routing-policy].

- B-Flag: This flag, when set, indicates the presence of the SRv6 Endpoint Behavior and SID Structure encoding specified in Section 2.4.4.2.13.

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

o RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

o SRv6 Binding SID: Contains a 16-octet SRv6 SID.

o SRv6 Endpoint Behavior and SID Structure: Optional, as defined in Section 2.4.4.2.13.

2.4.4. Segment List Sub-TLV

The Segment List sub-TLV encodes a single explicit path towards the endpoint as described in S~~s~~ection 5.1 in [I-D.ietf-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 in length due to a large number of segments. ~~Therefore a~~A 2-octet length is thus required.
According to Section 2 of [RFC9012], Sub-TLV Type indicates the ~~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 in the range from 128 to 255 ~~of 128 or higher~~ is used.

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

Commenté [BMI46]: Is it legal to include a list with no segment/weight?

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~~ MUST be set to zero on transmission and MUST be ignored on receipt.

- o sub-TLVs currently defined:

  * 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 beyond the scope of BGP and performed by the SRPM as described in ~~section~~ Section 5 in [I-D.ietf-spring-segment-routing-policy].

2.4.4.1. Weight Sub-TLV

The Weight sub-TLV specifies the weight associated with a given segment list. The contents of this sub-TLV are used only by the SRPM as described in ~~section~~ Section 2.11 in [I-D.ietf-spring-segment-routing-policy].

The Weight sub-TLV is optional and it MUST NOT appear more than once inside the Segment List sub-TLV.

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:

- o Type: 9.

- o Length: 6

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

- RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

- Weight: 4 octets value.

2.4.4.2. Segment Sub-TLVs

A Segment sub-TLV describes a single segment in a segment list (i.e., a single element of the explicit path). One or more Segment sub-TLVs constitute an explicit path of the SR Policy candidate path. The contents of these sub-TLVs are used only by the SRPM as described in ~~section~~ Section 4 in [I-D.ietf-spring-segment-routing-policy].

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

Commenté [BMI47]: Any note about ordering?

Section 4 of [I-D.ietf-spring-segment-routing-policy] defines several Segment Types:

```
Type  A: SR-MPLS Label
Type  B: SRv6 SID
Type  C: IPv4 Prefix with optional SR Algorithm
Type  D: IPv6 Global Prefix with optional SR Algorithm for SR-MPLS
Type  E: IPv4 Prefix with Local Interface ID
Type  F: IPv4 Addresses for link endpoints as Local, Remote pair
Type  G: IPv6 Prefix and Interface ID for link endpoints as Local,
         Remote pair for SR-MPLS
Type  H: IPv6 Addresses for link endpoints as Local, Remote pair
         for SR-MPLS
Type  I: IPv6 Global Prefix with optional SR Algorithm for SRv6
Type  J: IPv6 Prefix and Interface ID for link endpoints as Local,
         Remote pair for SRv6
Type  K: IPv6 Addresses for link endpoints as Local, Remote pair
         for SRv6
```

The following sub-sections specify the sub-TLVs used for encoding each of these Segment Types.

2.4.4.2.1. Segment Type A

The Type A Segment Sub-TLV encodes a single SR-MPLS SID. 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.

- Length is 6 octets.

- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

- RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero 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.

Commenté [BMI48]: Please include a description of the fields.

Commenté [BMI49]: This is not a description.

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

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

Commenté [BMI50]: What is the meaning of setting this bit?

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

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

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

- 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.

2.4.4.2.2. Segment Type B

The Type B Segment Sub-TLV encodes a single SRv6 SID. 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)                  //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//     SRv6 Endpoint Behavior and SID Structure (optional)      //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 13.

- Length is 26 when the SRv6 Endpoint Behavior and SID Structure is present else it is 18.

- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

- RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

- SRv6 SID: 16 octets of IPv6 address.

- SRv6 Endpoint Behavior and SID Structure: Optional, as defined in Section 2.4.4.2.13.

The TLV 2 defined for the advertisement of Segment Type B in the earlier versions of this document has been deprecated to avoid backward compatibility issues.

Commenté [BMI51]: Are we sure we still need this?

2.4.4.2.3. Segment Type C

The Type C Segment Sub-TLV encodes an IPv4 node address, SR Algorithm and an optional SR-MPLS SID. 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)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SR-MPLS SID (optional, 4 octets)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 3.

- Length is 10 when the SR-MPLS SID is present else it is 6.

- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

- SR Algorithm: 1 octet specifying SR Algorithm as described in ~~section~~ Section 3.1.1 in [RFC8402] when A-Flag as defined in Section 2.4.4.2.12 is present. SR Algorithm is used by SRPM as described in S~~s~~ection 4 in [I-D.ietf-spring-segment-routing-policy]. When A-Flag is not encoded, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

Commenté [BMI52]: MUST?

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

- SR-MPLS SID: optional, 4 octet field containing label, TC, S and TTL as defined in Section 2.4.4.2.1.

2.4.4.2.4. Segment Type D

The Type D Segment Sub-TLV encodes an IPv6 node address, SR Algorithm and an optional SR-MPLS SID. 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)                //
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 SR-MPLS SID (optional, 4 octets)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 4
- Length is 22 when the SR-MPLS SID is present else it is 18.
- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.
- SR Algorithm: 1 octet specifying SR Algorithm as described in section 3.1.1 in [RFC8402] when A-Flag as defined in Section 2.4.4.2.12 is present. SR Algorithm is used by SRPM as described in section 4 in [I-D.ietf-spring-segment-routing-policy]. When A-Flag is not encoded, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.
- IPv6 Node Address: ~~a 16-octet~~an IPv6 address representing a node.
- SR-MPLS SID: optional, 4 octet field containing label, TC, S and TTL as defined in Section 2.4.4.2.1.

Commenté [BMI53]: MUST?

2.4.4.2.5. Segment Type E

The Type E Segment Sub-TLV encodes an IPv4 node address, a local interface Identifier (Local Interface ID), and an optional SR-MPLS SID. 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)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 SR-MPLS SID (optional, 4 octets)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 5.
- Length is 14 when the SR-MPLS SID is present else it is 10.
- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.
- RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.
- Local Interface ID: 4 octets of interface index as defined in [RFC8664].
- IPv4 Node Address: ~~a 4-octet~~an IPv4 address representing a node.

o SR-MPLS SID: optional, 4 octet field containing label, TC, S and TTL as defined in Section 2.4.4.2.1.

2.4.4.2.6.   Segment Type F

The Type F Segment Sub-TLV encodes an adjacency local address, an adjacency remote address, and an optional SR-MPLS SID. 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)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SR-MPLS SID (optional, 4 octets)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 6.

o Length is 14 when the SR-MPLS SID is present else it is 10.

o Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

o RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

o Local IPv4 Address: an ~~4-octet~~ IPv4 address.

o Remote IPv4 Address: an ~~4-octet~~ IPv4 address.

Commenté [BMI54]: Update the description to be meaningful.

o SR-MPLS SID: optional, 4 octet field containing label, TC, S and TTL as defined in Section 2.4.4.2.1.

2.4.4.2.7.   Segment Type G

The Type G 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 SR-MPLS SID. 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)         //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SR-MPLS SID (optional, 4 octets)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 7

- Length is 46 when the SR-MPLS SID is present else it is 42.

- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

- RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

- Local Interface ID: 4 octets of interface index as defined in [RFC8664].

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

- Remote Interface ID: 4 octets of interface index as defined in [RFC8664]. The value MAY be set to zero when the local node address and interface identifiers are sufficient to describe the link.

- IPv6 Remote Node Address: a 16-octet IPv6 address. The value MAY be set to zero when the local node address and interface identifiers are sufficient to describe the link.

- SR-MPLS SID: optional, 4 octet field containing label, TC, S, and TTL as defined in Section 2.4.4.2.1.

Commenté [BMI55]: Update the description

2.4.4.2.8. Segment Type H

The Type H Segment Sub-TLV encodes an adjacency local address, an adjacency remote address, and an optional SR-MPLS SID. 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)              //
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                SR-MPLS SID (optional, 4 octets)               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 8

- Length is 38 when the SR-MPLS SID is present else it is 34.

- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

- RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

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

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

Commenté [BMI56]: Update the description

- SR-MPLS SID: optional, 4 octet field containing label, TC, S, and TTL as defined in Section 2.4.4.2.1.

2.4.4.2.9. Segment Type I

The Type I Segment Sub-TLV encodes an IPv6 node address, SR Algorithm, and an optional SRv6 SID. 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)               //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                    SRv6 SID (optional, 16 octets)           //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//     SRv6 Endpoint Behavior and SID Structure (optional)     //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

- Type: 14

- Length is variable.

- Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

- SR Algorithm: 1 octet specifying SR Algorithm as described in section 3.1.1 in [RFC8402] when A-Flag as defined in Section 2.4.4.2.12 is present. SR Algorithm is used by SRPM as described in section 4 in [I-D.ietf-spring-segment-routing-policy]. When A-Flag is not encoded, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

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

- SRv6 SID: optional, a 16-octet IPv6 address.

Commenté [BMI57]: Update the description

- SRv6 Endpoint Behavior and SID Structure: Optional, as defined in Section 2.4.4.2.13.

The TLV 10 defined for the advertisement of Segment Type I in the earlier versions of this document has been deprecated to avoid backward compatibility issues.

Commenté [BMI58]: Do we need to include this in the final RFC? I would delete this sentence.

2.4.4.2.10. Segment Type J

The Type J 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 SRv6 SID. 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  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Local Interface ID (4 octets)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                IPv6 Local Node Address (16 octets)          //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Remote Interface ID (4 octets)                |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                IPv6 Remote Node Address (16 octets)         //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                SRv6 SID (optional, 16 octets)               //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//     SRv6 Endpoint Behavior and SID Structure (optional)     //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 15

o Length is variable and set to length of SRv6 Endpoint Behavior and SID Structure (optional) + 58.~~.~~

o Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

o SR Algorithm: 1 octet specifying SR Algorithm as described in ~~section ~~Section 3.1.1 in [RFC8402] when A-Flag as defined in Section 2.4.4.2.12 is present. SR Algorithm is used by SRPM as described in ~~section ~~Section 4 in [I-D.ietf-spring-segment-routing-policy]. When A-Flag is not ~~encoded~~set, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

Commenté [BMI59]: MUST?

o Local Interface ID: 4 octets of interface index as defined in [RFC8664].

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

Commenté [BMI60]: Update the description.

o Remote Interface ID: 4 octets of interface index as defined in [RFC8664]. The value MAY be set to zero when the local node address and interface identifiers are sufficient to describe the link.

o IPv6 Remote Node Address: a 16-octet IPv6 address. The value MAY be set to zero when the local node address and interface identifiers are sufficient to describe the link.

Commenté [BMI61]: Update the description

o SRv6 SID: optional, a 16-octet IPv6 address.

o SRv6 Endpoint Behavior and SID Structure: Optional, as defined in

Section 2.4.4.2.13.

The TLV 11 defined for the advertisement of Segment Type J in the earlier versions of this document has been deprecated to avoid backward compatibility issues.

Commenté [BMI62]: Do we still need to include this in the final version?

2.4.4.2.11. Segment Type K

The Type K Segment Sub-TLV encodes an adjacency local address, an adjacency remote address, and an optional SRv6 SID. 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  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//               Local IPv6 Address (16 octets)                //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//               Remote IPv6 Address  (16 octets)              //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                SRv6 SID (optional, 16 octets)               //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//     SRv6 Endpoint Behavior and SID Structure (optional)     //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: 16

o Length is variable.

o Flags: 1 octet of flags as defined in Section 2.4.4.2.12.

o SR Algorithm: 1 octet specifying SR Algorithm as described in ~~section~~ Section 3.1.1 in [RFC8402] when A-Flag as defined in Section 2.4.4.2.12 is present. SR Algorithm is used by SRPM as described in section 4 in [I-D.ietf-spring-segment-routing-policy]. When A-Flag is not ~~encoded~~set, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

Commenté [BMI63]: MUST ?

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

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

o SRv6 SID: optional, a 16-octet IPv6 address.

Commenté [BMI64]: Please update the description

o SRv6 Endpoint Behavior and SID Structure: Optional, as defined in Section 2.4.4.2.13.

The TLV 12 defined for the advertisement of Segment Type K in the earlier versions of this document has been deprecated to avoid backward compatibility issues.

Commenté [BMI65]: Do we need to Include this?

2.4.4.2.12. Segment Flags

The Segment Types sub-TLVs described above ~~MAY~~ may contain the following flags in the "Flags" field defined in Section 6.8:

Commenté [BMI66]: Inappropriate use of normative language.

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

Commenté [BMI67]: Indicate whether multiple flags are allowed unless this is discarded in future flag definitions.

where:

V-Flag: This flag, when set, is used by SRPM for "SID verification" as described in Section 5.1 ~~in~~of [I-D.ietf-spring-segment-routing-policy].

A-Flag: This flag, when set, indicates the presence of SR Algorithm id in the "SR Algorithm" field applicable to various Segment Types. SR Algorithm is used by SRPM as described in ~~section~~ Section 4 ~~in~~ of [I-D.ietf-spring-segment-routing-policy].

S-Flag: This flag, when set, indicates the presence of the SR-MPLS or SRv6 SID depending on the segment type.

B-Flag: This flag, when set, indicates the presence of the SRv6 Endpoint Behavior and SID Structure encoding specified in Section 2.4.4.2.13.

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

The following applies to the Segment Flags:

o V-Flag applies to all Segment Types.

o A-Flag applies to Segment Types C, D, I, J, and K. If A-Flag appears with Segment Types A, B, E, F, G, and H, it MUST be ignored.

o S-Flag applies to Segment Types C, D, E, F, G, H, I, J, and K. If S-Flag appears with Segment Types A or B, it MUST be ignored.

o B-Flag applies to Segment Types B, I, J, and K. If B-Flag appears with Segment Types A, C, D, E, F, G, and H, it MUST be ignored.

Commenté [BMI68]: Having a table would be simple to digest.

2.4.4.2.13. SRv6 SID Endpoint Behavior and Structure

The Segment Types sub-TLVs described above MAY contain the SRv6 Endpoint Behavior and SID Structure [RFC8986] encoding as described below:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Endpoint Behavior       |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    LB Length  |  LN Length    | Fun. Length   |  Arg. Length  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Endpoint Behavior: 2 octets. It carries the SRv6 Endpoint Behavior code point for this SRv6 SID as defined in ~~section~~ Section 9.2 of [RFC8986]. When set with the value 0, the choice of SRv6 Endpoint Behavior is left to the headend.

Commenté [BMI69]: What about other values ?

Reserved: 2 octets of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

Locator Block Length: 1 octet. SRv6 SID Locator Block length in bits.

Locator Node Length: 1 octet. SRv6 SID Locator Node length in bits.

Function Length: 1 octet. SRv6 SID Function length in bits.

Argument Length: 1 octet. SRv6 SID Arguments length in bits.

The total of the locator block, locator node, function, and argument lengths MUST be less than or equal to 128.

2.4.5. Explicit NULL Label Policy Sub-TLV

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

The Explicit NULL Label Policy (ENLP) 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 ENLP Sub-TLV is not present, the decision of whether to push an Explicit NULL label on a given packet is a matter of local configuration.

The ENLP sub-TLV is optional and it MUST NOT appear more than once in the SR Policy encoding.

The contents of this sub-TLV are used by the SRPM as described in ~~section~~ Section 4.1 in [I-D.ietf-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: 14.

Length: 3.

Flags: 1 octet of flags. None are defined ~~at this stage~~in this document. Flags

~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero 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 values:

0: Reserved.

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.

5 - 255: ~~Reserved~~Unassigned.

Commenté [BMI70]: How future values are assigned?

The ENLP reserved values may be used for future extensions and implementations SHOULD ignore the ENLP Sub-TLV with these values. The behavior signaled in this Sub-TLV MAY be overridden by local configuration. ~~The s~~Section 4.1 of [I-D.ietf-spring-segment-routing-policy] describes the behavior on the headend for the handling of the explicit null label.

a mis en forme : Surlignage

2.4.6. 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 contents of this sub-TLV are used by the SRPM as described in S~~s~~ection 2.11 in [I-D.ietf-spring-segment-routing-policy].

Commenté [BMI71]: I guess you meant 2.12.

The Priority sub-TLV is optional and it MUST NOT appear more than once in the SR Policy encoding.

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: 15

Length: 2.

Priority: a 1-octet value.

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

2.4.7. Policy Candidate Path Name Sub-TLV

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

Usage of Policy Candidate Path Name sub-TLV is described in ~~section~~Section 2.6 in [I-D.ietf-spring-segment-routing-policy].

The Policy Candidate Path Name sub-TLV may exceed 255 bytes in length due to a long name. Therefore a 2-octet length is required. According to [RFC9012], the first bit of the sub-TLV codepoint defines the size of the length field. Therefore, for the Policy Candidate Path Name sub-TLV, a code point of 128 or higher is used.

Commenté [BMI72]: See the proposal in 2.4.4

It is RECOMMENDED that the size of the symbolic name for the candidate path is limited to 255 bytes. Implementations MAY choose to truncate long names to 255 bytes when signaling via BGP.

The Policy Candidate Path Name sub-TLV is optional and it MUST NOT appear more than once in the SR Policy encoding.

The Policy Candidate Path 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 Candidate Path Name                     //
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: 129.

Length: Variable.

RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

Policy Candidate Path Name: Symbolic name for the SR Policy candidate path without a NULL terminator as specified in ~~section~~Section 2.6 of [I-D.ietf-spring-segment-routing-policy].

2.4.8. Policy Name Sub-TLV

An operator MAY set the Policy Name sub-TLV to associate a symbolic name with the SR Policy for which the candidate path is being advertised via the SR Policy NLRI.

Usage of Policy Name sub-TLV is described in section 2.1 of [I-D.ietf-spring-segment-routing-policy].

The Policy Name sub-TLV may exceed 255 bytes in length due to a long policy name. Therefore a 2-octet length is required. According to [RFC9012], 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.

Commenté [BMI73]: See the proposal in 2.4.4

It is RECOMMENDED that the size of the symbolic name for the SR Policy is limited to 255 bytes. Implementations MAY choose to truncate long names to 255 bytes when signaling via BGP.

The Policy Name sub-TLV is optional and it MUST NOT appear more than once in the SR Policy encoding.

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: TBD2

Commenté [BMI74]: The temporary assigned value is 130

Length: Variable.

RESERVED: 1 octet of reserved bits. ~~SHOULD~~ MUST be set to zero on transmission and MUST be ignored on receipt.

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

3. Color Extended Community

The Color Extended Community [RFC9012] is used to steer traffic corresponding to BGP routes into an SR Policy with matching color value. The Color Extended Community MAY be carried in any BGP UPDATE message whose AFI/SAFI is 1/1 (IPv4 Unicast), 2/1 (IPv6 Unicast), 1/4 (IPv4 Labeled Unicast), 2/4 (IPv6 Labeled Unicast), 1/128 (VPN-IPv4 Labeled Unicast), 2/128 (VPN-IPv6 Labeled Unicast), or 25/70 (Ethernet VPN, usually known as EVPN). Use of the Color Extended Community in BGP UPDATE messages of other AFI/SAFIs is outside the scope of this document.

Two bits from the Flags field of the Color Extended Community are used as follows to support the requirements of Color-Only steering as specified in Section 8.8 of [I-D.ietf-spring-segment-routing-policy]:

```
                     1
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|~~C ~~C|O|        RESERVED          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The CO bits together form the Color-Only Type field which indicates the various matching criteria between BGP NH and SR Policy endpoint in addition to the matching of the color value. Following types are defined:

- Type 0: Specific Endpoint Match: Request match for the endpoint that is the BGP NH

- Type 1: Specific or Null Endpoint Match: Request match for either the endpoint that is the BGP NH or a null endpoint (e.g., like a default gateway)

- Type 2: Specific, Null, or Any Endpoint Match: Request match for either the endpoint that is the BGP NH or with a null or any endpoint

- Type 3: reserved for future use and SHOULD NOT be used. Upon reception, an implementation MUST treat it like Type 0.

The details of the SR Policy steering mechanisms based on these Color-Only types are specified in section 8.8 of [I-D.ietf-spring-segment-routing-policy].

One or more Color Extended Communities MAY be associated with a BGP route update. Sections 8.4.1, 8.5.1, and 8.8.2 of [I-D.ietf-spring-segment-routing-policy] specify the steering behaviors over SR Policies when multiple Color Extended Communities are associated with a BGP route.

4. SR Policy Operations

As already mentioned ~~described in this document~~in Section 1, BGP is not the actual consumer of an SR Policy NLRI. BGP is in charge of the origination and propagation of the SR Policy NLRI but its installation and use are outside the scope of BGP. The details of SR Policy installation and use are specified in [I-D.ietf-spring-segment-routing-policy].

4.1. Advertisement of SR Policies

Typically, but not limited to, an SR Policy is computed by a controller or a path computation engine (PCE) and originated by a BGP speaker on its behalf.

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

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

Moreover, one or more route targets SHOULD be attached to the advertisement, where each route target identifies one or more intended ~~head-end~~headends for the advertised SR Policy update.

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 a case, the NO_ADVERTISE community MUST be attached to the SR Policy update.

Commenté [BMI75]: Add a pointer to RFC1997

4.2. Reception of an SR Policy NLRI

On reception of an SR Policy NLRI, a BGP speaker first determines if it is acceptable and then 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 MUST first, determine if it is~~'s~~ acceptable. The following rules apply in addition to the validation described in Section 5:

- 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).

- The SR Policy update MUST have either the NO_ADVERTISE community or at least one route target extended community in IPv4-address format or both. If a router supporting this specification receives an SR Policy update with no route target extended communities and no NO_ADVERTISE community, the update MUST be considered as malformed.

- 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).

A router that receives an SR Policy update that is not valid according to these criteria MUST treat the update as malformed and the SR Policy candidate path MUST NOT be passed to the SRPM.

4.2.2. Usable SR Policy NLRI

An SR Policy update that has been determined to be acceptable is further evaluated for its usability by the receiving node.

An SR Policy NLRI update without any route target extended community but having the NO_ADVERTISE community is considered usable.

If one or more route targets are present, then at least one route target MUST match the BGP Identifier of the receiver 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 none matches the local BGP Identifier, then, while the SR Policy NLRI is acceptable, it is not usable on the receiver node.

When the SR Policy tunnel type includes any sub-TLV that is unrecognized or unsupported, the update SHOULD NOT be considered usable. An implementation MAY provide an option for ignoring unsupported sub-TLVs.

4.2.3. Passing a usable SR Policy NLRI to the SRPM

Commenté [BMI76]: I would add a note saying that how this is passed to the SRPM is implementation-specific.

Once BGP on the receiving node has determined that the SR Policy NLRI is usable, it passes the SR Policy candidate path to the SRPM. Note that, along with the candidate path details, BGP also passes the originator information for breaking ties in the candidate path selection process as described in S~~s~~ection 2.4 in [I-D.ietf-spring-segment-routing-policy].

When an update for an SR Policy NLRI results in its becoming unusable, BGP speaker MUST delete its corresponding SR Policy candidate path from the SRPM.

The SRPM applies the rules defined in ~~section~~ Section 2 in [I-D.ietf-spring-segment-routing-policy] to determine whether the SR Policy candidate path is valid and to select the best candidate path among the valid ones for a given SR Policy.

4.2.4. Propagation of an SR Policy

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

SR Policy NLRIs that have the NO_ADVERTISE community attached to them MUST NOT be propagated.

By default, a BGP node receiving an SR Policy NLRI MUST NOT propagate it to any EBGP neighbor. An implementation MAY provide an explicit configuration to override this and enable the propagation of acceptable SR Policy NLRIs to specific EBGP neighbors.

A BGP node advertises a received SR Policy NLRI to its IBGP neighbors according to normal IBGP propagation rules.

By default, a BGP node receiving an SR Policy NLRI SHOULD NOT remove route target extended community before propagation. An implementation MAY provide support for configuration to filter and/or remove route target extended community before propagation.

5. Error Handling

This section describes the error handling actions, as described in [RFC7606], that are to be performed for the handling of the BGP update messages for BGP SR Policy SAFI.

A BGP Speaker MUST perform the following syntactic validation of the SR Policy NLRI to determine if it is malformed. This includes the validation of the length of each NLRI and the total length of the MP_REACH_NLRI and MP_UNREACH_NLRI attributes.

When the error determined allows for the router to skip the malformed NLRI(s) and continue the processing of the rest of the update message, then it MUST handle such malformed NLRIs as 'Treat-as-withdraw'. In other cases, where the error in the NLRI encoding results in the inability to process the BGP update message (e.g.,, length related encoding errors), then the router SHOULD handle such malformed NLRIs as 'AFI/SAFI disable' when other AFI/SAFI besides SR Policy are being advertised over the same session. Alternately, the router MUST perform 'session reset' when the session is only being used for SR Policy or when it 'AFI/SAFI disable' action is not possible.

a mis en forme : Surlignage

The validation of the TLVs/sub-TLVs introduced in this document and defined in their respective sub-sections of Section 2.4 MUST be performed to determine if they are malformed or invalid. The validation of the Tunnel Encapsulation Attribute itself and the other TLVs/sub-TLVs specified in [RFC9012] MUST be done as described in that document. In case of any error detected, either at the attribute or its TLV/sub-TLV level, the "treat-as-withdraw" strategy MUST be applied. This is because an SR Policy update without a valid Tunnel Encapsulation Attribute (comprising of all valid TLVs/sub-TLVs) is not usable.

An SR Policy update that is determined to be not acceptable, and therefore malformed, based on rules described in Section 4.2.1 MUST be handled by the "treat-as-withdraw" strategy.

The validation of the individual fields of the TLVs/sub-TLVs defined in Section 2.4 are beyond the scope of BGP as they are handled by the SRPM as described in the individual TLV/sub-TLV sub-sections. A BGP implementation MUST NOT perform semantic verification of such fields nor consider the SR Policy update to be invalid or not acceptable/usable based on such validation.

An implementation SHOULD log ~~an error for~~ any errors found during the above validation for further analysis.

xx. 6. IANA Considerations

Commenté [BMI77]: What about manageability considerations?

This document requests codepoint allocations in the following existing registries:

- Subsequent Address Family Identifiers (SAFI) Parameters registry

- BGP Tunnel Encapsulation Attribute Tunnel Types registry under the BGP Tunnel Encapsulation registry

- BGP Tunnel Encapsulation Attribute sub-TLVs registry under the BGP Tunnel Encapsulation registry
- Color Extended Community Flags registry under the BGP Tunnel Encapsulation registry

This document also requests the creation of the following new registries:

- SR Policy Segment List Sub-TLVs under the BGP Tunnel Encapsulation registry

- SR Policy Binding SID Flags under the BGP Tunnel Encapsulation registry

a mis en forme : Surlignage

- SR Policy Segment Flags under the BGP Tunnel Encapsulation registry

- Color Extended Community Color-Only Types registry under the BGP Tunnel Encapsulation registry

a mis en forme : Surlignage

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

This document introduces a SAFI in the registry "Subsequent Address Family Identifiers (SAFI) Parameters" that has been assigned a codepoint by IANA as follows:

Commenté [BMI78]: The action is to update the existing entry. Please update the text accordingly.

Codepoint	Description	Reference
73	SR Policy SAFI	This document

6.2. Existing Registry: BGP Tunnel Encapsulation Attribute Tunnel Types

This document introduces a Tunnel-Type in the registry "BGP Tunnel Encapsulation Attribute Tunnel Types" that has been assigned a codepoint by IANA as follows:

Commenté [BMI79]: Idem

Codepoint	Description	Reference
15	SR Policy	This document

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

This document defines sub-TLVs in the registry "BGP Tunnel Encapsulation Attribute sub-TLVs" that has been assigned codepoints by IANA as follows via the early allocation process:

Commenté [BMI80]: Idem

Also, the action should be to update the reference to point to this document rather that the one currently in the registry.

Codepoint	Description	Reference
12	Preference sub-TLV	This document
13	Binding SID sub-TLV	This document
14	ENLP sub-TLV	This document
15	Priority sub-TLV	This document
20	SRv6 Binding SID sub-TLV	This document
128	Segment List sub-TLV	This document
129	Policy Candidate Path Name sub-TLV	This document
130	Policy Name sub-TLV	This document

6.4. Existing Registry: Color Extended Community Flags

This document requests allocations in the registry called "Color Extended Community Flags" under the "BGP Tunnel Encapsulation" registry.

The following bits have been assigned by IANA via the early allocation process to form the Color-Only Types field:

Bit Position	Description	Reference
0-1	Color-only Types Field	This document

6.5. New Registry: SR Policy Segment List Sub-TLVs

This document requests the creation of a new registry called "SR Policy Segment List Sub-TLVs" under the "BGP Tunnel Encapsulation" registry. The allocation policy of this registry is "Standards Action" according to [RFC8126].

Following initial Sub-TLV codepoints are assigned by this document:

Value	Description	Reference
0	Reserved	This document
1	Segment Type A sub-TLV	This document
2	Deprecated	This document
3	Segment Type C sub-TLV	This document
4	Segment Type D sub-TLV	This document
5	Segment Type E sub-TLV	This document
6	Segment Type F sub-TLV	This document
7	Segment Type G sub-TLV	This document
8	Segment Type H sub-TLV	This document
9	Weight sub-TLV	This document
10	Deprecated	This document
11	Deprecated	This document
12	Deprecated	This document
13	Segment Type B sub-TLV	This document
14	Segment Type I sub-TLV	This document
15	Segment Type J sub-TLV	This document
16	Segment Type K sub-TLV	This document
17-255	Unassigned	

Commenté [BMI81]: Any reason why this value is not open for assignements?

Commenté [BMI82]: Can be released for future assignments. If so, please indicate so in the text.

Commenté [BMI83]: Are those open for reassignment?

6.6. New Registry: SR Policy Binding SID Flags

This document requests the creation of a new registry called "SR Policy Binding SID Flags" under the "BGP Tunnel Encapsulation" registry. The allocation policy of this registry is "Standards Action" ~~according to~~(Section 4.9 of [RFC8126]).

The following flags are defined:

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

6.7. New Registry: SR Policy SRv6 Binding SID Flags

This document requests the creation of a new registry called "SR Policy SRv6 Binding SID Flags" under the "BGP Tunnel Encapsulation" registry. The allocation policy of this registry is "Standards Action" (Section 4.9 of ~~according to~~ [RFC8126]).

The following flags are defined:

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

2	SRv6 Endpoint Behavior & SID Structure Flag (B-Flag)	This document
3-7	Unassigned	

6.8. New Registry: SR Policy Segment Flags

This document requests the creation of a new registry called "SR Policy Segment Flags" under the "BGP Tunnel Encapsulation" registry. The allocation policy of this registry is "Standards Action" (Section 4.9 of ~~according to~~ [RFC8126]).

The following Flags are defined:

Bit	Description	Reference
0	Segment Verification Flag (V-Flag)	This document
1	SR Algorithm Flag (A-Flag)	This document
2	SID Specified Flag (S-Flag)	This document
3	SRv6 Endpoint Behavior & SID Structure Flag (B-Flag)	This document
4-7	Unassigned	

6.9. New Registry: Color Extended Community Color-Only Types

This document requests the creation of a new registry called "Color Extended Community Color-Only Types" under the "BGP Tunnel Encapsulation" registry for assignment of codepoints (values 0 through 3) in the Color-Only Type field of the Color Extended Community Flags field. The allocation policy of this registry is "Standards Action" ~~according to~~(Section 4.9 of [RFC8126]).

The following types are defined:

Type	Description	Reference
0	Specific Endpoint Match	This document
1	Specific or Null Endpoint Match	This document
2	Specific, Null, or Any Endpoint Match	This document
3	~~Unallocated & reserved for future~~Unassigned	This document

7. Security Considerations

Commenté [BMI84]: Cite Section 15 of RFC9012 as well.

The security mechanisms of the base BGP security model apply to the extensions described in this document as well. See the Security Considerations section of [RFC4271] for a discussion of BGP security. Also, refer to [RFC4272] and [RFC6952] for analysis of security issues for BGP.

The BGP SR Policy extensions specified in this document enable traffic engineering and service programming use-cases within ~~the~~ an SR domain as described in [I-D.ietf-spring-segment-routing-policy]. SR operates within a trusted SR domain [RFC8402] and its security considerations also apply to BGP sessions when carrying SR Policy information. The SR Policies distributed by BGP are expected to be used entirely within this trusted SR domain, i.e., within a single AS or between multiple AS/domains within a single provider network. Therefore, precaution is necessary to ensure that the SR Policy information advertised via BGP sessions is limited to nodes in a secure manner within this trusted SR domain. BGP peering sessions for address-families other than SR Policy SAFI may be set up to routers outside the SR domain. The isolation of BGP SR Policy SAFI peering sessions may be used to ensure that the SR Policy information is not advertised by accident or error to an EBGP peering session outside the SR domain.

Additionally, it may be considered that the export of SR Policy information, as described in this document, constitutes a risk to confidentiality of mission-critical or commercially sensitive information about the network (more specifically endpoint/node addresses, SR SIDs, and the SR Policies deployed). BGP peerings are not automatic and require configuration; thus, it is the responsibility of the network operator to ensure that only trusted nodes (that include both routers and controller applications) within the SR domain are configured to receive such information.

8. Acknowledgments

The authors of this document would like to thank Shyam Sethuram, John Scudder, Przemyslaw Krol, Alex Bogdanov, Nandan Saha, Bruno Decraene, Gurusiddesh Nidasesi, Kausik Majumdar, Zafar Ali, Swadesh Agarwal, Jakob Heitz, Viral Patel, Peng Shaofu, Cheng Li, Martin Vigoureux, and John Scudder for their comments and review of this document. The authors would like to thank Sue Hares for her detailed shepherd review that helped in improving the document.

9. Contributors

Eric Rosen
Juniper Networks
US

Email: erosen@juniper.net

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

Peng Shaofu
ZTE Corporation
China

Email: peng.shaofu@zte.com.cn

10. References

10.1. Normative References

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-22 (work in progress), March 2022.

[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>.

[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>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <https://www.rfc-editor.org/info/rfc8660>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

[RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <https://www.rfc-editor.org/info/rfc8754>.

[RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <https://www.rfc-editor.org/info/rfc8986>.

[RFC9012] Patel, K., Van de Velde, G., Sangli, S., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", RFC 9012, DOI 10.17487/RFC9012, April 2021, <https://www.rfc-editor.org/info/rfc9012>.

10.2. Informational References

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, DOI 10.17487/RFC4272, January 2006, <https://www.rfc-editor.org/info/rfc4272>.

[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>.

[RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, DOI 10.17487/RFC6952, May 2013, <https://www.rfc-editor.org/info/rfc6952>.

Authors' Addresses

Stefano Previdi
Huawei Technologies

IT

Email: stefano@previdi.net

Clarence Filsfils
Cisco Systems
Brussels
BE

Email: cfilsfil@cisco.com

Ketan Talaulikar (editor)
Arrcus Inc
India

Email: ketant.ietf@gmail.com

Paul Mattes
Microsoft
One Microsoft Way
Redmond, WA 98052
USA

Email: pamattes@microsoft.com

Dhanendra Jain
Google

Email: dhanendra.ietf@gmail.com
Steven Lin
Google

Email: stevenlin@google.com