IDR R. Chen
Internet-Draft ZTE Corporation
Intended status: Standards Track J. Dong
Expires: 26 August 2024 Huawei
D. Zhao
ZTE Corporation
L. Gong
China mobile
Y. Zhu
China Telecom
R. Pang
China Unicom
23 February 2024

SR Policies Extensions for NRP in BGP-LS
draft-chen-idr-bgp-ls-sr-policy-nrp-05

Abstract

This document defines a new TLV which enable the headend to report the configuration and the states of SR policies carrying NRP information by using BGP-LS.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Segment Routing Policy [RFC9256] is an ordered list of segments (i.e. instructions) that represent a source-routed policy. Packet flows are steered into a SR Policy on a node where it is instantiated called a headend node. The packets steered into an SR Policy carry an ordered list of segments associated with that SR Policy.

[I-D.ietf-teas-ietf-network-slices] provides the definition of IETF network slice for use within the IETF and discusses the general framework for requesting and operating IETF Network Slices, their characteristics, and the necessary system components and interfaces.It also introduces the concept Network Resource Partition (NRP), which is a subset of the resources and associated policies in the underlay network.

[I-D.ietf-teas-ns-ip-mpls] introduces the notion of a Slice-Flow Aggregate which comprises of one or more IETF network slice traffic streams. It also describes the Network Resource Partition (NRP) and the NRP Policy that can be used to instantiate control and data plane behaviors on select topological elements associated with the NRP that supports a Slice-Flow Aggregate.

[I-D.ietf-idr-bgp-ls-sr-policy] describes a mechanism to distribute SR policy information to external components using BGP-LS.

[I-D.ietf-idr-sr-policy-nrp] defines the extensions to BGP SR policy to specify the NRP which the SR Policy candidate path is associated with.

This document defines a new TLV which enable the headend to report the configuration and the states of an SR policy carrying the NRP information by using BPG-LS.

2. 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.

3. Carrying NRP TLV in BGP-LS

In order to collect configuration and states of the NRP SR policy, this document defines a new SR Policy state TLV which enable the headend to report the state at the SR Policy CP level.

This TLV is carried in the optional non-transitive BGP Attribute "LINK_STATE Attribute" defined in [RFC7752]associated with the SR Policy CP NLRI type.

This TLV is optional and only one this TLV is advertised for a given CP. If multiple TLVs are present, then the first one is considered valid and the rest are ignored as describe in [I-D.ietf-idr-bgp-ls-sr-policy].

The 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    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          NRP ID (4 octets)                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

Type: TBD1.

Length: 6 octets.

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

NRP ID: 4-octet domain significant identifier of Network Resource Partition.

4. Scalability Considerations

The mechanism specified in this document defines the headend to report configuration and states of an SR policy carrying the NRP information by using BPG-LS. BGP-LS SR Policy is used to report the SR Policy attributes and status. As a new attribute of the SR Policy, NRP will not increase the number of the SR Policy reported by BGP-LS.

5. Acknowledgements

TBD.

6. IANA Considerations

IANA maintains a registry called "Border Gateway Protocol - Link State (BGP-LS) Parameters" with a sub-registry called "BGP-LS NLRI and Attribute TLVs". The following TLV codepoints are suggested (for early allocation by IANA):

Codepoint	Description	Reference
TBD	NRP TLV	This document

7. Security Considerations

Procedures and protocol extensions defined in this document do not affect the BGP security model. See the "Security Considerations"section of [RFC4271] for a discussion of BGP security.Security considerations for acquiring and distributing BGP-LS information are discussed in [RFC7752].

8. Informative References

[I-D.ietf-idr-bgp-ls-sr-policy]
Previdi, S., Talaulikar, K., Dong, J., Gredler, H., and J. Tantsura, "Advertisement of Segment Routing Policies using BGP Link-State", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-ls-sr-policy-03, 5 November 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ls-sr-policy-03>.

[I-D.ietf-idr-sr-policy-nrp]
Dong, J., Hu, Z., and R. Pang, "BGP SR Policy Extensions for Network Resource Partition", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-nrp-00, 17 December 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-nrp-00>.

[I-D.ietf-teas-ietf-network-slices]
Farrel, A., Drake, J., Rokui, R., Homma, S., Makhijani, K., Contreras, L. M., and J. Tantsura, "A Framework for Network Slices in Networks Built from IETF Technologies", Work in Progress, Internet-Draft, draft-ietf-teas-ietf-network-slices-25, 14 September 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-ietf-network-slices-25>.

[I-D.ietf-teas-ns-ip-mpls]
Saad, T., Beeram, V. P., Dong, J., Wen, B., Ceccarelli, D., Halpern, J. M., Peng, S., Chen, R., Liu, X., Contreras, L. M., Rokui, R., and L. Jalil, "Realizing Network Slices in IP/MPLS Networks", Work in Progress, Internet-Draft, draft-ietf-teas-ns-ip-mpls-03, 26 November 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-ns-ip-mpls-03>.

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

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

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[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>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

Authors' Addresses

Ran Chen
ZTE Corporation
Nanjing
China
Email: chen.ran@zte.com.cn

Jie Dong
Huawei
Beijing
China
Email: jie.dong@huawei.com

Detao Zhao
ZTE Corporation
Nanjing
China
Email: zhao.detao@zte.com.cn

Liyan Gong
China mobile
Beijing
China
Email: gongliyan@chinamobile.com

Yongqing Zhu
China Telecom
Guangzhou
China
Email: zhuyq8@chinatelecom.cn

Ran Pang
China Unicom
Beijing
China
Email: pangran@chinaunicom.cn

IDR Working Group C. Lin
Internet Draft New H3C Technologies
Intended status: Standards Track Z. Li
Expires: September 13, 2024 China Mobile
R. Pang
China Unicom
K. Talaulikar
Cisco Systems
M. Chen
New H3C Technologies
March 17, 2024

Segment Routing BGP Egress Peer Engineering over Layer 2 Bundle
draft-lin-idr-sr-epe-over-l2bundle-05

Abstract

There are deployments where the Layer 3 interface on which a BGP peer session is established is a Layer 2 interface bundle. In order to allow BGP-EPE to control traffic flows on individual member links of the underlying Layer 2 bundle, BGP Peering SIDs need to be allocated to individual bundle member links, and advertisement of such BGP Peering SIDs in BGP-LS is required. This document describes how to support Segment Routing BGP Egress Peer Engineering over Layer 2 bundle.

Table of Contents

1. Introduction

Segment Routing (SR) leverages the source routing paradigm. A node steers a packet through an ordered list of instructions called "segments". Segment Routing can be instantiated on both MPLS and IPv6 data planes, which are referred to as SR-MPLS and SRv6.

BGP Egress Peer Engineering (BGP-EPE) allows an ingress Provider Edge (PE) router within the domain to use a specific egress PE and a specific external interface/neighbor to reach a particular destination.

The SR architecture [RFC8402] defines three types of BGP Peering Segments that may be instantiated at a BGP node:

- o Peer Node Segment (PeerNode SID): instruction to steer to a specific peer node

- o Peer Adjacency Segment (PeerAdj SID): instruction to steer over a specific local interface towards a specific peer node

- o Peer Set Segment (PeerSet SID): instruction to load-balance to a set of specific peer nodes

[RFC9087] illustrates a centralized controller-based BGP-EPE solution involving SR path computation using the BGP Peering Segments. A centralized controller learns the BGP Peering SIDs via Border Gateway Protocol - Link State (BGP-LS) and then uses this information to program a BGP-EPE policy. [RFC9086] defines the extension to BGP-LS for advertisement of BGP Peering Segments along with their BGP peering node information.

There are deployments where the Layer 3 interface on which a BGP peer session is established is a Layer 2 interface bundle (L2 Bundle), for instance, a Link Aggregation Group (LAG) [IEEE802.1AX]. BGP-EPE may wish to control traffic flows on individual member links of the underlying Layer 2 bundle. In order to do so, BGP Peering SIDs need to be allocated to individual bundle member links, and advertisement of such BGP Peering SIDs in BGP-LS is required.

This document describes how to support Segment Routing BGP Egress Peer Engineering over Layer 2 bundle.

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. Problem Statement

In the network depicted in Figure 1, B and C establish BGP peer session on a Layer 2 bundle. Assume that, the link delays of the members are different because they are over different transport paths, and member link 1 has the lowest delay.

The operator of AS1 wishes to apply a BGP-EPE policy to steer the time-sensitive traffic from AS1 to AS2 via member link 1 of the Layer 2 bundle.

```
                 L2 Bundle      +--------+
              /---member 1---\  |        |
            --+---member 2---+--C   AS2  |
+--------+ /  \---member 3---/  |        |
|        |/                     +--------+
A   AS1  B
|        |\                     +--------+
+--------+ \                    |        |
            --------------------D   AS3  |
                                |        |
                                +--------+
```

Figure 1: BGP-EPE over L2 Bundle

The existing Peer Adjacency SID can be allocated to the Layer 3 interface between B and C, which is a Layer 2 interface bundle. If steered by that Peer Adjacency SID, the traffic will be forwarded by load balancing among all the bundle member links. So, the existing mechanism cannot meet the requirement of steering traffic flows via individual member link.

3. Advertising Peer Adjacency Segment for L2 Bundle Member in BGP-LS

BGP peering segments are generally advertised in BGP-LS from a BGP node along with its peering topology information, in order to enable computation of efficient BGP-EPE policies and strategies.

When a BGP peer session is established over a Layer 2 interface bundle, an implementation MAY allocate one or more Peer Adjacency Segments for each member link. If so, it SHOULD advertise the Peer Adjacency Segments of bundle members in BGP-LS, using the method defined in this section.

3.1. MPLS-SR

For SR-MPLS, Section 5.2 of [RFC9086] described the BGP-LS advertisement of the PeerAdj SID for L3 link.

In order to advertise the PeerAdj SIDs for L2 bundle members in BGP-LS, the L2 Bundle Member Attributes TLVs [RFC9085] MUST also be included in the Link Attributes. Each L2 Bundle Member Attributes TLV identifies an L2 bundle member, and includes the PeerAdj SID TLV [RFC9086] to advertise the PeerAdj SID for the associated L2 bundle member.

This document updates [RFC9085] and [RFC9086] to allow the PeerAdj SID TLV to be included as a sub-TLV of the L2 Bundle Member Attributes TLV.

Note that the inclusion of a L2 Bundle Member Attributes TLV implies that the identified link is a member of the L2 bundle and that the member link is operationally up. If any member link fails, an implementation MUST withdraw the L2 Bundle Member Attributes TLV in BGP-LS, along with the Peer Adjacency Segments for the failed member link.

3.2. SRv6

For SRv6, according to Section 4.1 of [RFC9514], the advertisement of L3 link BGP EPE Peer Adjacency SID is the same as for SR-MPLS, except for using the SRv6 End.X SID TLV [RFC9514] instead of the PeerAdj SID TLV [RFC9086].

Similarly, when advertising the SRv6 BGP Peer Adjacency SIDs for L2 bundle members, the L2 Bundle Member Attributes TLVs [RFC9085] MUST also be included in the Link Attributes. The SRv6 End.X SID TLV [RFC9514] MUST be carried in the L2 Bundle Member Attributes TLV to advertise the SRv6 Peer Adjacency SID for the associated L2 bundle member.

4. Manageability Considerations

The manageability considerations described in [RFC9552] and [RFC9086] also apply to this document.

The operator MUST be provided with the options of configuring, enabling, and disabling the advertisement of Peer Adjacency Segment for L2 Bundle member links, as well as control of which information is advertised to which internal or external peer.

5. Security Considerations

The security considerations described in [RFC9552] and [RFC9086] also apply to this document.

This document does not introduce any new security consideration.

6. IANA Considerations

This document has no IANA actions.

7. References

7.1. Normative References

[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>.

[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>.

[RFC9085] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Gredler, H., and M. Chen, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing", RFC 9085, DOI 10.17487/RFC9085, August 2021, <https://www.rfc-editor.org/info/rfc9085>.

[RFC9086] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Patel, K., Ray, S., and J. Dong, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing BGP Egress Peer Engineering", RFC 9086, DOI 10.17487/RFC9086, August 2021, <https://www.rfc-editor.org/info/rfc9086>.

[RFC9514] Dawra, G., Filsfils, C., Talaulikar, K., Ed., Chen, M., Bernier, D., and B. Decraene, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing over IPv6 (SRv6)", RFC 9514, DOI 10.17487/RFC9514, December 2023, <https://www.rfc-editor.org/info/rfc9514>.

[RFC9552] K. Talaulikar, "Distribution of Link-State and Traffic Engineering Information Using BGP", RFC 9552, DOI 10.17487/RFC9552, December 2023, <https://www.rfc-editor.org/info/rfc9552>.

7.2. Informative References

[IEEE802.1AX] IEEE, "IEEE Standard for Local and metropolitan area networks -- Link Aggregation", IEEE 802.1AX, <https://ieeexplore.ieee.org/document/7055197>.

[RFC8668] Ginsberg, L., Ed., Bashandy, A., Filsfils, C., Nanduri, M., and E. Aries, "Advertising Layer 2 Bundle Member Link Attributes in IS-IS", RFC 8668, DOI 10.17487/RFC8668, December 2019, <https://www.rfc-editor.org/info/rfc8668>.

[RFC9087] Filsfils, C., Ed., Previdi, S., Dawra, G., Ed., Aries, E., and D. Afanasiev, "Segment Routing Centralized BGP Egress Peer Engineering", RFC 9087, DOI 10.17487/RFC9087, August 2021, <https://www.rfc-editor.org/info/rfc9087>.

Appendix A. Example

This section shows an example of how Node B in Figure 1 allocates and advertises Peer Adjacency Segments for L2 bundle members.

B allocates a PeerAdj SID for the Layer 2 interface bundle to peer C, along with a PeerAdj SID for each member link. B programs its forwarding table accordingly:

PeerAdj SID		Outgoing Interface
IF on SR-MPLS Data Plane	IF on SRv6 Data Plane	
1010	A::A0	L2 Bundle to C
1011	A::A1	Member link 1 to C
1012	A::A2	Member link 2 to C
1013	A::A3	Member link 3 to C

B signals the related BGP-LS Link NLRI and Link Attributes including the PeerAdj SID for L3 parent link to the BGP-EPE controller, as specified in Section 5.2 of [RFC9086]. In addition, B also advertises L2 Bundle Member Attribute TLVs carrying the PeerAdj SIDs for L2 bundle members.

For MPLS-SR, the Link Attributes are as follows:

o PeerAdj SID TLV (Label-1010)

o L2 Bundle Member Attribute TLV (Link Local Identifier describing the member link 1)

   * PeerAdj SID TLV (Label-1011)

  * (Optional) Min/Max Unidirectional Link Delay TLV (Delay of member link 1)

- L2 Bundle Member Attribute TLV (Link Local Identifier describing the member link 2)

  * PeerAdj SID TLV (Label-1012)

  * (Optional) Min/Max Unidirectional Link Delay TLV (Delay of member link 2)

- L2 Bundle Member Attribute TLV (Link Local Identifier describing the member link 3)

  * PeerAdj SID TLV (Label-1013)

  * (Optional) Min/Max Unidirectional Link Delay TLV (Delay of member link 3)

For SRv6, the Link Attributes are as follows:

- SRv6 End.X SID TLV (SID-A::A0)

- L2 Bundle Member Attribute TLV (Link Local Identifier describing the member link 1)

  * SRv6 End.X SID TLV (SID-A::A1)

  * (Optional) Min/Max Unidirectional Link Delay TLV (Delay of member link 1)

- L2 Bundle Member Attribute TLV (Link Local Identifier describing the member link 2)

  * SRv6 End.X SID TLV (SID-A::A2)

  * (Optional) Min/Max Unidirectional Link Delay TLV (Delay of member link 2)

- L2 Bundle Member Attribute TLV (Link Local Identifier describing the member link 3)

  * SRv6 End.X SID TLV (SID-A::A3)

  * (Optional) Min/Max Unidirectional Link Delay TLV (Delay of member link 3)

Acknowledgements

The authors would like to thank Sasha Vainshtein for his review and comments of this document.

Authors' Addresses

Changwang Lin
New H3C Technologies
China
Email: linchangwang.04414@h3c.com

Zhenqiang Li
China Mobile
China
Email: lizhenqiang@chinamobile.com

Ran Pang
China Unicom
China
Email: pangran@chinaunicom.cn

Ketan Talaulikar
Cisco Systems
India
Email: ketant.ietf@gmail.com

Mengxiao Chen
New H3C Technologies
China
Email: chen.mengxiao@h3c.com

IDR Working Group C. Lin
Internet Draft New H3C Technologies
Intended status: Standards Track W. Jiang
Expires: June 5, 2024 Y. Liu
China Mobile
M. Chen
H. Li
New H3C Technologies
December 8, 2023

BGP Extensions of SR Policy for Headend Behavior
draft-lin-idr-sr-policy-headend-behavior-03

Abstract

This document defines extensions to Border Gateway Protocol (BGP) to distribute SR policies carrying headend behavior.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

Segment routing (SR) [RFC8402] is a source routing paradigm that explicitly indicates the forwarding path for packets at the ingress node. The ingress node steers packets into a specific path according to the Segment Routing Policy (SR Policy) as defined in [RFC9256]. In order to distribute SR policies to the headend, [I-D.ietf-idr-segment-routing-te-policy] specifies a mechanism by using BGP.

As described in [RFC9256], a headend can steer a packet flow into an SR Policy in various ways, including BSID steering, per-destination steering, per-flow steering, and policy-based steering. Moreover, [I-D.jiang-idr-ts-flowspec-srv6-policy] describes a way by using BGP FlowSpec to steer packets into an SRv6 Policy.

[RFC8986] defines End.B6.Encaps behavior and End.B6.Encaps.Red behavior for SRv6 BSID. [I-D.filsfils-spring-srv6-net-pgm-insertion] extends the SRv6 BSID behaviors with End.B6.Insert and End.B6.Insert.Red. When receiving packets with an active SID matching a local BSID of these kinds, the headend will perform corresponding behaviors. Different BSID behaviors are suitable for different scenarios. For example, comparing with End.B6.Encaps, End.B6.Encaps.Red reduces the size of the SRH by excluding the first SID, which can be useful for the devices with lower capacity of SID depths, like the switches in data center network. End.B6.Insert inserts a new SRH in between the IPv6 Header and the received SRH rather than pushing a new IPv6 header, which can be applied to express scalable traffic-engineering policies across multiple domains.

The SRv6 Binding SID sub-TLV is defined in [I-D.ietf-idr-segment-routing-te-policy] to signal the SRv6 BSID information along with SR Policies. It enables the specified SRv6 BSID behavior to be instantiated on the headend node. However, if the packets are steering into an SR Policy in some other way than using BSID, the headend behavior is not specified during the distributing of SR Policy by BGP. The network operator has to use additional tools, like NETCONF, to signal the headend behavior.

This document defines extensions to Border Gateway Protocol (BGP) to distribute SR policies carrying headend behavior. So that the headend can be instructed to perform specific behavior when packets are steered into the SR policy without BSID.

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. Headend Behavior in SR Policy

As defined in [I-D.ietf-idr-segment-routing-te-policy], 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
        SRv6 Binding SID
        Preference
        Priority
        Policy Name
        Policy Candidate Path Name
        Explicit NULL Label Policy (ENLP)
        Segment List
           Weight
          Segment
          Segment
          ...
        ...
```

SR policy with headend behavior is expressed as follows:

```
SR Policy SAFI NLRI: <Distinguisher, Policy-Color, Endpoint>
  Attributes:
    Tunnel Encaps Attribute (23)
      Tunnel Type: SR Policy
        Binding SID
        SRv6 Binding SID
        Preference
        Priority
        Policy Name
        Policy Candidate Path Name
        Explicit NULL Label Policy (ENLP)
        Headend Behavior
        L2 Headend Behavior
        Segment List
           Weight
          Segment
          Segment
          ...
        ...
```

2.1. Headend Behavior Sub-TLV

The Headend Behavior sub-TLV encodes the default headend behavior associated with the candidate path for L3 traffic. When the headend steers L3 packets into that SR Policy and the associated candidate path is active, the specific headend behavior should be performed by default. In the case of BSID steering, the behavior defined by the BSID overrides the default headend behavior.

The Headend Behavior sub-TLV is optional, and MUST NOT appear more than once in the SR Policy encoding.

The Headend Behavior 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           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Headend Behavior       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: to be assigned by IANA.

o Length: 4.

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

o Headend Behavior: a 2-octet value. The following values are defined.

* TBD: H.Encaps. A headend behavior defined in [RFC8986].

* TBD: H.Encaps.Red. A headend behavior defined in [RFC8986].

* TBD: H.Insert. A headend behavior defined in [I-D.filsfils-spring-srv6-net-pgm-insertion].

* TBD: H.Insert.Red. A headend behavior defined in [I-D.filsfils-spring-srv6-net-pgm-insertion].

2.2. L2 Headend Behavior Sub-TLV

The L2 Headend Behavior sub-TLV encodes the default headend behavior associated with the candidate path for L2 traffic. When the headend steers L2 packets into that SR Policy and the associated candidate path is active, the specific headend behavior should be performed by default.

The L2 Headend Behavior sub-TLV is optional, and MUST NOT appear more than once in the SR Policy encoding.

The L2 Headend Behavior 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           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      L2 Headend Behavior      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where:

o Type: to be assigned by IANA.

o Length: 4.

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

o L2 Headend Behavior: a 2-octet value. The following values are defined.

  * TBD: H.Encaps.L2. A headend behavior defined in [RFC8986].

  * TBD: H.Encaps.L2.Red. A headend behavior defined in [RFC8986].

3. Extensions of BGP-LS

[I-D.ietf-idr-bgp-ls-sr-policy] describes a mechanism to collect the SR policy information that is locally available in a node and advertise it into BGP-LS updates. Extensions of BGP-LS for headend behavior of SR Policy will be included in the future version of this draft.

4. Security Considerations

Procedures and protocol extensions defined in this document do not affect the security considerations discussed in [I-D.ietf-idr-segment-routing-te-policy].

5. IANA Considerations

Headend Behavior Sub-TLV (TBD)

L2 Headend Behavior Sub-TLV (TBD)

6. References

6.1. Normative References

[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>.

[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>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[I-D.ietf-idr-segment-routing-te-policy] Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., Rosen, E., Jain, D., and S. Lin, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-26, 23 October 2023, <http://www.ietf.org/internet-drafts/draft-ietf-idr-segment-routing-te-policy-26.txt>.

6.2. Informative References

[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>.

[I-D.jiang-idr-ts-flowspec-srv6-policy] Jiang, W., Liu, Y., Chen, S., and S. Zhuang, "Traffic Steering using BGP Flowspec with SRv6 Policy", Work in Progress, Internet-Draft, draft-jiang-idr-ts-flowspec-srv6-policy-07, 23 March 2022, <http://www.ietf.org/internet-drafts/draft-jiang-idr-ts-flowspec-srv6-policy-07.txt>.

[I-D.filsfils-spring-srv6-net-pgm-insertion] Filsfils, C., Camarillo, P., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "SRv6 NET-PGM extension: Insertion", Work in Progress, Internet-Draft, draft-filsfils-spring-srv6-net-pgm-insertion-09, 16 August 2023, <http://www.ietf.org/internet-drafts/draft-filsfils-spring-srv6-net-pgm-insertion-09.txt>.

[I-D.ietf-idr-bgp-ls-sr-policy] Previdi, S., Talaulikar, K., Dong, J., Gredler, H., and J. Tantsura, "Advertisement of Segment Routing Policies using BGP Link-State", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-ls-sr-policy-03, 5 November 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ls-sr-policy-03>.

Authors' Addresses

Changwang Lin
New H3C Technologies

Email: linchangwang.04414@h3c.com

Wenying Jiang
China Mobile

Email: jiangwenying@chinamobile.com

Yisong Liu
China Mobile

Email: liuyisong@chinamobile.com

Mengxiao Chen
New H3C Technologies

Email: chen.mengxiao@h3c.com

Hao Li
New H3C Technologies

Email: lihao@h3c.com

IDR
Internet-Draft
Intended status: Experimental
Expires: 18 September 2024
Y. Liu
S. Peng
ZTE
G. Mishra
Verizon Inc.
17 March 2024

Advertising SID Algorithm Information in BGP
draft-peng-idr-segment-routing-te-policy-attr-09

Abstract

This document defines new Segment Types and proposes extensions for BGP to provide algorithm information for SR-MPLS Adjacency-SIDs when delivering SR Policy via BGP.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Segment Routing (SR) [RFC8402] allows a headend node to steer a packet flow along any path. [RFC9256] details the concepts of SR Policy and steering into an SR Policy. These apply equally to the MPLS and IPv6 data plane instantiations of Segment Routing with their respective representations of segments as SR-MPLS SID and SRv6 SID as described in [RFC8402].

[I-D.ietf-idr-sr-policy-safi] 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. It defines a new BGP address family (SAFI), i.e., SR Policy SAFI NLRI. In UPDATE messages of that address family, the NLRI identifies an SR Policy Candidate Path, and the attributes encode the segment lists and other details of that SR Policy Candidate Path.

11 segment-descriptor types (from type A all the way to type K) for SR segments are defined [RFC9256] section 4. [I-D.ietf-idr-sr-policy-safi] specifies the encoding for segment types A and B in BGP SR Policy SAFI. And the encoding for the remaining 9 types are specified in [I-D.ietf-idr-bgp-sr-segtypes-ext].

As specified in [RFC9256], the SR algorithm can be optionally specified for Segment Types C(IPv4 Node and SID), D(IPv6 Node and SID for SR-MPLS), I(IPv6 Node and SID for SRv6), J(IPv6 Node, index for remote and local pair, and SID for SRv6), and K(IPv6 Local/Remote addresses and SID for SRv6). That is, currently the algorithm can be carried along with SR-MPLS prefix SID, SRv6 prefix SID and SRv6 adjacency SID when delivering SR Policy.

[I-D.ietf-lsr-algorithm-related-adjacency-sid] complements that, besides the SR-MPLS prefix SID, the algorithm can be also included as part of an SR-MPLS Adjacency-SID advertisement, in scenarios where multiple algorithm share the same link resource. In this case, an SR-MPLS Policy advertised to the headend may also contain algorithm specific Adjacency-SID.

This document defines new Segment Types and proposes extensions for BGP to provide algorithm information for SR-MPLS Adjacency-SIDs when delivering SR Policy via BGP.

2. 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.

3. New Segment Types for SR-MPLS Adjacency with optional SR Algorithm

This section defines four new Segment types and the corresponding Segment Sub-TLVs of Segment List Sub-TLV to provide algorithm information for SR-MPLS Adjacency-SIDs.

The processing procedures for SID with algorithm specified in [RFC9256] and [I-D.ietf-idr-bgp-sr-segtypes-ext] are still applicable for the new segment types. When the algorithm is not specified for the SID types above which optionally allow for it, the headend SHOULD use the Strict Shortest Path algorithm if available; otherwise, it SHOULD use the default Shortest Path algorithm.

3.1. Type L: IPv4 Node Address and Local Interface ID with optional SR Algorithm for SR-MPLS

This type allows for identification of an Adjacency SID or BGP Peer Adjacency SID (as defined in [RFC8402]) SR-MPLS label for point-to-point links including IP unnumbered links. The headend is required to resolve the specified IPv4 Local Node Address to the node originating it and then use the Local Interface ID to identify the point-to-point link whose adjacency is being referred to. The Local Interface ID link descriptor follows semantics as specified in [RFC9552]. This type can also be used to indicate indirection into a layer 2 interface (i.e., without IP address) like a representation of an optical transport path or a layer 2 Ethernet port or circuit at the specified node. The SR Algorithm (refer to Section 3.1.1 of [RFC8402]) MAY also be provided.

The encoding for Type L Segment Sub-TLV 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)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 IPv4 Node Address (4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SR-MPLS SID (optional, 4 octets)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: TBD1

Length: Specifies the length of the value field (i.e., not including Type and Length fields) in terms of octets. The value MUST be 14 when the SR-MPLS SID is present else it MUST be 10.

Flags: 1 octet of flags as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext].

SR Algorithm: 1 octet specifying SR Algorithm as described in Section 3.1.1 of [RFC8402]) when A-Flag as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext] is present. SR Algorithm is used by SRPM as described in Section 4 of [RFC9256]). When A-Flag is not encoded, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

Local Interface ID: 4 octets of interface index of local interface (refer TLV 258 of [RFC9552]).

IPv4 Node Address: a 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 of [I-D.ietf-idr-sr-policy-safi].

3.2. Type M: IPv4 Addresses for link endpoints as Local, Remote pair with optional SR Algorithm for SR-MPLS

This type allows for identification of an Adjacency SID or BGP Peer Adjacency SID (as defined in [RFC8402]) SR-MPLS label for links. The headend is required to resolve the specified Local IPv4 Address to the node originating it and then use the Remote IPv4 Address to identify the link adjacency being referred to. The Local and Remote Address pair link descriptors follow semantics as specified in [RFC9552]. The SR Algorithm (refer to Section 3.1.1 of [RFC8402]) MAY also be provided.

The format of Type M Segment Sub-TLV 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 IPv4 Address (4 octets)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 Remote IPv4 Address  (4 octets)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                 SR-MPLS SID (optional, 4 octets)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: TBD2

Length: Specifies the length of the value field (i.e., not including Type and Length fields) in terms of octets. The value MUST be 14 when the SR-MPLS SID is present else it MUST be 10.

Flags: 1 octet of flags as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext].

SR Algorithm: 1 octet specifying SR Algorithm as described in Section 3.1.1 of [RFC8402]) when A-Flag as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext] is present. SR Algorithm is used by SRPM as described in Section 4 of [RFC9256]). When A-Flag is not encoded, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

Local IPv4 Address: a 4-octet IPv4 address representing the local link address of the node.

Remote IPv4 Address: a 4-octet IPv4 address representing the link address of the neighbor node.

SR-MPLS SID: optional, 4-octet field containing label, TC, S and TTL as defined in Section 2.4.4.2.1 of [I-D.ietf-idr-sr-policy-safi].

3.3. Type N: IPv6 Node Addresses and Interface ID for link endpoints as Local, Remote pair, with optional SR Algorithm for SR-MPLS

This type allows for identification of an Adjacency SID or BGP Peer Adjacency SID (as defined in [RFC8402]) label for links including those with only Link-Local IPv6 addresses. The headend is required to resolve the specified IPv6 Node Address to the node originating it and then use the Local Interface ID to identify the point-to-point link whose adjacency is being referred to. For other than point-to-point links, additionally the specific adjacency over the link needs to be resolved using the IPv6 Remote Node Address and Interface ID. The Local and Remote pair of Node Address and Interface ID link descriptor follows semantics as specified in [RFC9552]. This type can also be used to indicate indirection into a layer 2 interface (i.e., without IP address) like a representation of an optical transport path or a layer 2 Ethernet port or circuit at the specified node. The SR Algorithm (refer to Section 3.1.1 of [RFC8402]) MAY also be provided.

The format of Type N Segment Sub-TLV 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)         //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SR-MPLS SID (optional, 4 octets)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: TBD3

Length: Specifies the length of the value field (i.e., not including Type and Length fields) in terms of octets. The value MUST be 46 when the SR-MPLS SID is present else it MUST be 42.

Flags: 1 octet of flags as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext].

SR Algorithm: 1 octet specifying SR Algorithm as described in Section 3.1.1 of [RFC8402]) when A-Flag as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext] is present. SR Algorithm is used by SRPM as described in Section 4 of [RFC9256]). When A-Flag is not encoded, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

Local Interface ID: 4 octets of interface index of local interface (refer TLV 258 of [RFC9552]).

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

Remote Interface ID: 4 octets of interface index of remote interface (refer TLV 258 of [RFC9552]). 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 of [I-D.ietf-idr-sr-policy-safi].

3.4. Type O: IPv6 Addresses for link endpoints as Local, Remote pair, with optional SR Algorithm for SR-MPLS

This type allows for identification of an Adjacency SID or BGP Peer Adjacency SID (as defined in [RFC8402]) label for links with Global IPv6 addresses. The headend is required to resolve the specified Local IPv6 Address to the node originating it and then use the Remote IPv6 Address to identify the link adjacency being referred to. The Local and Remote IPv6 Address pair link descriptors follow semantics as specified in [RFC9552]. The SR Algorithm (refer to Section 3.1.1 of [RFC8402]) MAY also be provided.

TThe format of Type O Segment Sub-TLV 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)                //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                SR-MPLS SID (optional, 4 octets)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type: TBD4

Length: Specifies the length of the value field (i.e., not including Type and Length fields) in terms of octets. The value MUST be 38 when the SR-MPLS SID is present else it MUST be 34.

Flags: 1 octet of flags as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext].

SR Algorithm: 1 octet specifying SR Algorithm as described in Section 3.1.1 of [RFC8402]) when A-Flag as defined in [I-D.ietf-idr-bgp-sr-segtypes-ext] is present. SR Algorithm is used by SRPM as described in Section 4 of [RFC9256]). When A-Flag is not encoded, this field SHOULD be set to zero on transmission and MUST be ignored on receipt.

Local IPv6 Address: a 16-octet IPv6 address representing the local link address of the node.

Remote IPv6 Address: a 16-octet IPv6 address representing the link address of the neighbor node.

SR-MPLS SID: optional, 4-octet field containing label, TC, S and TTL as defined in Section 2.4.4.2.1 of [I-D.ietf-idr-sr-policy-safi].

4. IANA Considerations

This document requests codepoint allocations for new Sub-TLVs of the "Segment List sub-TLV" under the "BGP Tunnel Encapsulation".

Value	Description	Reference
TBD1	Segment Type L sub-TLV	This document
TBD2	Segment Type M sub-TLV	This document
TBD3	Segment Type N sub-TLV	This document
TBD4	Segment Type O sub-TLV	This document

5. Security Considerations

Procedures and protocol extensions defined in this document do not affect the security considerations discussed in [RFC9256] and [I-D.ietf-idr-sr-policy-safi].

6. Acknowledgement

The authors would like to thank Ketan Talaulikar, Nat Kao and Zhenqiang Li for their comments and suggestions.

7. References

7.1. Normative References

[I-D.ietf-idr-bgp-sr-segtypes-ext]
Talaulikar, K., Filsfils, C., Previdi, S., Mattes, P., and D. Jain, "Segment Routing Segment Types Extensions for BGP SR Policy", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-sr-segtypes-ext-03, 4 March 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-sr-segtypes-ext-03>.

[I-D.ietf-idr-sr-policy-safi]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-sr-

policy-safi-02, 16 March 2024,
<https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-safi-02>.

[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>.

[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>.

[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>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[RFC9552] Talaulikar, K., Ed., "Distribution of Link-State and Traffic Engineering Information Using BGP", RFC 9552, DOI 10.17487/RFC9552, December 2023, <https://www.rfc-editor.org/info/rfc9552>.

7.2. Informative References

[I-D.ietf-lsr-algorithm-related-adjacency-sid]
Peng, S., Chen, R., Talaulikar, K., and P. Psenak, "Algorithm Related IGP-Adjacency SID Advertisement", Work in Progress, Internet-Draft, draft-ietf-lsr-algorithm-related-adjacency-sid-06, 5 December 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-lsr-algorithm-related-adjacency-sid-06>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[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>.

[RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <https://www.rfc-editor.org/info/rfc8665>.

[RFC8666] Psenak, P., Ed. and S. Previdi, Ed., "OSPFv3 Extensions for Segment Routing", RFC 8666, DOI 10.17487/RFC8666, December 2019, <https://www.rfc-editor.org/info/rfc8666>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

[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>.

Authors' Addresses

Yao Liu
ZTE
Nanjing
China
Email: liu.yao71@zte.com.cn

Shaofu Peng
ZTE
Nanjing
China
Email: peng.shaofu@zte.com.cn

Gyan Mishra
Verizon Inc.
Email: gyan.s.mishra@verizon.com

IDR Working Group H. Chen
Internet-Draft Futurewei
Intended status: Standards Track Z. Li
Expires: 29 September 2024 Huawei
Z. Li
China Mobile
Y. Fan
Casa Systems
M. Toy
Verizon
L. Liu
Fujitsu
28 March 2024

BGP Extensions for IDs Allocation
draft-wu-idr-bgp-segment-allocation-ext-14

Abstract

This document describes extensions to the BGP for IDs allocation. The IDs are SIDs for segment routing (SR), including SR for IPv6 (SRv6). They are distributed to their domains if needed.

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].

Status of This Memo

Table of Contents

1. Introduction

In a network with a central controller, the controller has the link state information of the network, including the resource such as traffic engineering and SIDs information. It is valuable for the controller to allocate and manage the resources including SIDs of the network in a centralized way, especially for the SIDs representing network resources [I-D.ietf-teas-enhanced-vpn].

When BGP as a controller allocates an ID, it is natural and beneficial to extend BGP to send it to its corresponding network elements.

PCE may be extended to send IDs to their corresponding network elements after the IDs are allocated by a controller. However, when BGP is already deployed in a network, using PCE for IDs will need to deploy an extra protocol PCE in the network. This will increase the CapEx and OpEx.

Yang may be extended to send IDs to their corresponding network elements after the IDs are allocated by a controller. However, Yang progress may be slow. Some people may not like this.

There may not be these issues when BGP is used to send IDs. In addition, BGP may be used to distribute IDs into their domains easily when needed. It is also fit for the dynamic and static allocation of IDs.

This document proposes extensions to the BGP for sending Segment Identifiers (SIDs) for segment routing (SR) including SRv6 to their corresponding network elements after SIDs are allocated by the controller. If needed, they will be distributed into their network domains.

2. Terminology

The following terminology is used in this document.

SR: Segment Routing.

SRv6: SR for IPv6

SID: Segment Identifier.

IID: Indirection Identifier.

SR-Path: Segment Routing Path.

SR-Tunnel: Segment Routing Tunnel.

RR: Route Reflector.

MPP: MPLS Path Programming.

NAI: Node or Adjacency Identifier.

TED: Traffic Engineering Database.

3. Protocol Extensions

A new AFI and SAFI are defined: the Identifier AFI and the SID SAFI whose codepoints are to be assigned by IANA. A few new NLRI TLVs are defined for the new AFI/SAFI, which are Node, Link and Prefix SID NLRI TLVs. When a SID for a node, link or prefix is allocated by the controller, it may be sent to a network element in a UPDATE message containing a MP_REACH NLRI with the new AFI/SAFI and the SID NLRI TLV. When the SID is withdrawn by the controller, a UPDATE message containing a MP_UNREACH NLRI with the new AFI/SAFI and the SID NLRI TLV may be sent to the network element.

3.1. Node SID NLRI TLV

The Node SID NLRI TLV is used to represent the IDs such as SID associated with a node. Its format is illustrated in the Figure below, which is similar to the corresponding one defined in [I-D.ietf-idr-rfc7752bis].

```
 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 (TBDa for Node SID)   |            Length             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Protocol ID   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                           Identifier                          |
|                           (8 octets)                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Peer IP (4/16 bytes for IPv4/IPv6 Address)                   ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                   Local Node Descriptors TLV                  ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                           Sub-TLVs                            ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type (TBDa): It is to be assigned by IANA.

Length: It is the length of the value field in bytes.

Peer IP: 4/16 octet value indicates an IPv4/IPv6 peer. When receiving a UPDATE message, a BGP speaker processes it only if the peer IP is the IP address of the BGP speaker or 0.

Protocol-ID, Identifier, and Local Node Descriptor: defined in [I-D.ietf-idr-rfc7752bis], can be reused.

Sub-TLVs may be some of the followings:

SR-Capabilities TLV (1034): It contains the Segment Routing Global Base (SRGB) range(s) allocated for the node.

SR Local Block TLV (1036): The SR Local Block (SRLB) TLV contains the range(s) of SIDs/labels allocated to the node for local SIDs.

SRv6 SID Node TLV (TBD1): A new TLV, called SRv6 Node SID TLV, contains an SRv6 SID and related information.

SRv6 Locator TLV (TBD2): A new TLV, called SRv6 Locator TLV, contains an SRv6 locator and related information.

The format of SRv6 SID Node TLV is illustrated below.

```
 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 (TBD1)         |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Reserved   |     Flags     |    SRv6 Endpoint Function     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        SRv6 Identifier                        |
|                          (128 bits)                           |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
~                        Optional sub-TLVs                      ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

SRv6 Node SID TLV

Type: TBD1 for SRv6 Node SID TLV is to be assigned by IANA.

Length: Variable.

Flags: 1 octet. No flags are defined now.

SRv6 Endpoint Function: 2 octets. The function associated with SRv6 SID.

SRv6 Identifier: 16 octets. IPv6 address representing SRv6 SID.

Reserved: MUST be set to 0 while sending and ignored on receipt.

SRv6 node SID inherits the topology and algorithm from its locator.

The format of SRv6 locator TLV is illustrated below.

```
     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 (TBD2)         |            Length             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |R|R|R|R|        MT-ID          |   Algorithm   |     Flags     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                            Metric                             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Locator-Size | Locator (variable)...
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    ~                       Optional sub-TLVs                       ~
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

SRv6 Locator TLV

Type: TBD2 for SRv6 Locator TLV is to be assigned by IANA.

Length: Variable.

MT-ID: Multitopology Identifier as defined in [RFC5120].

Algorithm: 1 octet. Associated algorithm.

Flags: 1 octet. As described in [RFC9352].

Metric: 4 octets. As described in [RFC5305].

Locator-Size: 1 octet. Number of bits in the Locator field (1 to 128).

Locator: 1 to 16 octets. SRv6 Locator encoded in the minimum number of octets for the given Locator-Size.

Reserved: MUST be set to 0 while sending and ignored on receipt.

3.2. Link SID NLRI TLV

The Link SID NLRI TLV is used to represent the IDs such as SID associated with a link. Its format is illustrated in the Figure below, which is similar to the corresponding one defined in [I-D.ietf-idr-rfc7752bis].

```
      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 (TBDb for Link SID)   |            Length             |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     | Protocol ID   |
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     ~                     Identifier (8 octets)                     ~
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     |  Peer IP (4/16 bytes for IPv4/IPv6 Address)                   ~
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     ~                     Local Node Descriptors TLV                ~
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     ~                     Remote Node Descriptors TLV               ~
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     ~                         Link Descriptors TLV                  ~
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
     ~                             Sub-TLVs                          ~
     +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type (TBDb): It is to be assigned by IANA.

Length: It is the length of the value field in bytes.

Peer IP: 4/16 octet value indicates an IPv4/IPv6 peer.

Protocol-ID, Identifier, Local Node Descriptors, Remote Node Descriptors and Link Descriptors: defined in [I-D.ietf-idr-rfc7752bis], can be reused.

The Sub-TLVs may be some of the followings:

Adj-SID TLV (1099): It contains the Segment Identifier (SID) allocated for the link/adjacency.

LAN Adj-SID TLV (1100): It contains the Segment Identifier (SID) allocated for the adjacency/link to a non-DR router on a broadcast, NBMA, or hybrid link.

SRv6 Adj-SID TLV (TBD3): A new TLV, called SRv6 Adj-SID TLV, contains an SRv6 Adj-SID and related information.

SRv6 LAN Adj-SID TLV (TBD4): A new TLV, called SRv6 LAN Adj-SID TLV, contains an SRv6 LAN Adj-SID and related information.

The format of an SRv6 Adj-SID TLV is illustrated below.

```
     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 (TBD3)          |            Length             |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |     Weight    |   Algorithm   |B|S|P|             Flags       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |            Reserved           |     SRv6 Endpoint Function    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                        SRv6 Identifier                        |
    |                          (128 bits)                           |
    |                                                               |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    ~                        Optional sub-TLVs                      ~
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

SRv6 Adj-SID TLV

Type: TBD3 for SRv6 Adj-SID TLV is to be assigned by IANA.

Length: Variable.

Weight: 1 octet. The value represents the weight of the SID for the purpose of load balancing.

Algorithm: 1 octet. Associated algorithm.

Flags: 2 octets. Three flags are defined in [RFC9352].

SRv6 Endpoint Function: 2 octets. The function associated with SRv6 SID.

SRv6 Identifier: 16 octets. IPv6 address representing SRv6 SID.

Reserved: MUST be set to 0 while sending and ignored on receipt.

The format of an SRv6 LAN Adj-SID TLV is illustrated below.

```
     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 (TBD4)          |             Length            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |     Weight    |   Algorithm   |B|S|P|            Flags      |O|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Reserved            |    SRv6 Endpoint Function     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |    neighbor Router ID (4 octets) / System ID (6 octets)       ~
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                        SRv6 Identifier                        |
    |                          (128 bits)                           |
    |                                                               |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    ~                        Optional sub-TLVs                      ~
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

SRv6 LAN Adj-SID TLV

Type: TBD4 for SRv6 LAN Adj-SID TLV is to be assigned by IANA.

Length: Variable.

Weight: 1 octet. The value represents the weight of the SID for the purpose of load balancing.

Algorithm: 1 octet. Associated algorithm.

Flags: 2 octets. Three flags B, S and P are defined in [RFC9352]. Flag O set to 1 indicating OSPF neighbor Router ID of 4 octets, set to 0 indicating IS-IS neighbor System ID of 6 octets.

SRv6 Endpoint Function: 2 octets. The function associated with SRv6 SID.

SRv6 Identifier: 16 octets. IPv6 address representing SRv6 SID.

Reserved: MUST be set to 0 while sending and ignored on receipt.

3.3. Prefix SID NLRI TLV

The Prefix SID NLRI TLV is used to represent the IDs such as SID associated with a prefix. Its format is illustrated in the Figure below, which is similar to the corresponding one defined in [I-D.ietf-idr-rfc7752bis].

```
  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 (TBDc for Prefix SID)  |            Length             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Protocol ID   |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ~                      Identifier (8 octets)                    ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |  Peer IP (4/16 bytes for IPv4/IPv6 Address)                   ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ~                   Local Node Descriptors TLV                  ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ~                       Prefix Descriptors TLV                  ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 ~                           Sub-TLVs                            ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Where:

Type (TBDc): It is to be assigned by IANA.

Length: It is the length of the value field in bytes.

Peer IP: 4/16 octet value indicates an IPv4/IPv6 peer.

Protocol-ID, Identifier, Local Node Descriptors and Prefix Descriptors: defined in [I-D.ietf-idr-rfc7752bis], can be reused.

Sub-TLVs may be some of the followings:

Prefix-SID TLV (1158): It contains the Segment Identifier (SID) allocated for the prefix.

Prefix Range TLV (1159): It contains a range of prefixes and the Segment Identifier (SID)s allocated for the prefixes.

3.4. Capability Negotiation

It is necessary to negotiate the capability to support BGP Extensions for sending and receiving Segment Identifiers (SIDs). The BGP SID Capability is a new BGP capability [RFC5492]. The Capability Code for this capability is to be specified by the IANA. The Capability Length field of this capability is variable. The Capability Value field consists of one or more of the following tuples:

```
+---------------------------------------------------+
|  Address Family Identifier (2 octets)             |
+---------------------------------------------------+
|  Subsequent Address Family Identifier (1 octet)   |
+---------------------------------------------------+
|  Send/Receive (1 octet)                           |
+---------------------------------------------------+
```

BGP SID Capability

The meaning and use of the fields are as follows:

Address Family Identifier (AFI): This field is the same as the one used in [RFC4760].

Subsequent Address Family Identifier (SAFI): This field is the same as the one used in [RFC4760].

Send/Receive: This field indicates whether the sender is (a) willing to receive SID from its peer (value 1), (b) would like to send SID to its peer (value 2), or (c) both (value 3) for the <AFI, SAFI>.

If a BGP speaker has not sent the BGP SID Capability in its BGP OPEN message on a particular BGP session, or if it has not received the BGP SID Capability in the BGP OPEN message from its peer on that BGP session, the BGP speaker MUST NOT send on that session any UPDATE message with SID.

When both a local BGP speaker and a peer BGP speaker send the BGP SID Capability in their BGP OPEN messages on a BGP session, the negotiation results from the capability parameters are shown in the table below.

Local Parameter	Peer Parameter	Negotiation Result
Send	Receive or Both	Local speaker can send SID and peer speaker can receive SID.
Receive	Send or Both	Local speaker can receive SID and peer speaker can send SID.
Both	Both	Both local and peer speakers can send and receive SID.

Negotiation Results from Both, Send, Receive Parameters

4. IANA Considerations

This document requests assigning a new AFI in the registry "Address Family Numbers" as follows:

Code Point	Description	Reference
TBDx	Identifier AFI	This document

This document requests assigning a new SAFI in the registry "Subsequent Address Family Identifiers (SAFI) Parameters" as follows:

Code Point	Description	Reference
TBDy	SID SAFI	This document

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

Following TLV code points are defined:

Code Point	Description	Reference
1 (TBDa)	Node SID NLRI	This document
2 (TBDb)	Link SID NLRI	This document
3 (TBDc)	Prefix SID NLRI	This document

This document requests assigning a code-point from the registry "BGP-LS Node Descriptor, Link Descriptor, Prefix Descriptor, and Attribute TLVs" as follows:

TLV Code Point	Description	Reference
TBD1	SRv6 Node SID	This document
TBD2	SRv6 Allocator	This document
TBD3	SRv6 Adj-SID	This document
TBD4	SRv6 LAN Adj-SID	This document

5. Security Considerations

Protocol extensions defined in this document do not affect the BGP security other than those as discussed in the Security Considerations section of [I-D.ietf-idr-rfc7752bis].

6. Acknowledgements

The authors would like to thank Eric Wu, Robert Raszuk, Zhengquiang Li, Susan Hares, and Ketan Talaulikar for their valuable suggestions and comments on this draft.

7. References

7.1. Normative References

[I-D.ietf-idr-rfc7752bis]
Talaulikar, K., "Distribution of Link-State and Traffic Engineering Information Using BGP", Work in Progress, Internet-Draft, draft-ietf-idr-rfc7752bis-17, 25 August 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-rfc7752bis-17>.

[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>.

[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>.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <https://www.rfc-editor.org/info/rfc5120>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <https://www.rfc-editor.org/info/rfc5305>.

[RFC5492] Scudder, J. and R. Chandra, "Capabilities Advertisement with BGP-4", RFC 5492, DOI 10.17487/RFC5492, February 2009, <https://www.rfc-editor.org/info/rfc5492>.

[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>.

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[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>.

[RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

[RFC9352] Psenak, P., Ed., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Segment Routing over the IPv6 Data Plane", RFC 9352, DOI 10.17487/RFC9352, February 2023, <https://www.rfc-editor.org/info/rfc9352>.

7.2. Informative References

[I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for NRP-based Enhanced Virtual Private Network", Work in Progress, Internet-Draft, draft-ietf-teas-enhanced-vpn-17, 25 December 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-teas-enhanced-vpn-17>.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA,
United States of America
Email: hchen.ietf@gmail.com

Zhenbin Li
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: lizhenbin@huawei.com

Zhenqiang Li
China Mobile
No. 29 Finance Street, Xicheng District
Beijing
100029
P.R. China
Email: li_zhenqiang@hotmail.com

Yanhe Fan
Casa Systems
United States of America
Email: yfan@casa-systems.com

Mehmet Toy
Verizon
United States of America
Email: mehmet.toy@verizon.com

Lei Liu
Fujitsu
United States of America
Email: liulei.kddi@gmail.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 26 March 2024
23 September 2023

K. Zhang
Z. Hu
J. Dong
Huawei

BGP SR Policy Extensions for template
draft-zhang-idr-sr-policy-template-03

Abstract

Segment Routing(SR) Policies can be advertised using BGP. An SR Policy may has lots of attributes, and as the application and features evolve, the SR Policy may need have more and more attribute attributes. To avoid modifying BGP when attributes are added to an SR Policy, we can define a template. The identifier and content of the template are defined by the receiver of the SR Policy. The advertiser of an SR policy only needs to know the ID of the template. When advertising SR policy, the advertiser carries the template ID in the tunnel encapsulation information of the SR policy. After receiving the SR Policy information, the receiver obtains the corresponding template and content according to the template ID, thereby obtaining abundant constraint configuration information.

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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Status of This Memo

This Internet-Draft will expire on 26 March 2024.

Table of Contents

1. Introduction

[I-D.ietf-idr-segment-routing-te-policy]defines some attributes encoding of the SR Policy path. However, in actual applications, there are many other attributes of SR Policy path. These attributes are valid only on the device where the SR Policy path is installed. Such attributes may include backup protection, Bidirectional Forwarding Detection information, traffic statistics collection, or in-situ Flow Information Telemetry detection information, etc. If these attributes are directly delivered through BGP, the BGP SR Policy protocol may change frequently. This document defines a general method to carry the path attributes of SR Policies.

2. Terminology

SR Policy: An ordered list of segments.

Candidate Path: the unit for signaling of an SR Policy to a headend via protocol extensions like Path Computation Element (PCE) Communication Protocol (PCEP) [RFC8664] [I-D.ietf-pce-segment-routing-policy-cp] or BGP SR Policy [I-D.ietf-idr-segment-routing-te-policy].

SRPM: SR Policy Module.

Template: A collection of attributes sets.

Template ID: The identifier of a template.

3. Template ID defination

To support the attributes extension of SR Policies, this document defines a constraint template identifier. The constraint template ID is valid only for the recipient. The SR policy publisher only needs to carry the template ID when publishing the SR policy. The receiver of the SR Policy may create a template corresponding to the template identifier in advance before receiving the SR Policy, or may define a corresponding template after receiving the template definition of the SR Policy. The template can contain any attributes on the SR Policy path, including but not limited to backup protection, Bidirectional Forwarding Detection information, traffic statistics collection, or in-situ Flow Information Telemetry detection information, etc. After receiving the SR Policy information, the receiver matches the template information based on the template ID and adds attributes to the SR Policy based on the attributes defined in the template.

Template ID is an local identifier, just to use on the headend of the SR Policy. And it is a local configured identifier, need to be unique only on the headend device. We need no further process to coordinate the template ID between multiple routers.

4. SR Policy and Tunnel Encapsulation Attribute Update

As the template ID is defined, the tunnel attribute encapsulation of the BGP SR Policy needs to be updated.

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
        Policy Candidate Path Name
        Explicit NULL Label Policy (ENLP)
        Template ID
        Segment List
           Weight
           Segment
           Segment
           ....
        ....
```

Where Tempate ID indicates the template ID for the SR Policy candidate path.

4.1. Template ID sub-TLV

A new sub-TLV called Template ID sub-TLV is defined. Template ID sub-TLV specifies the template ID of an SR policy candidate path. Each sub-TLV is encoded as shown in Figure 1.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |     Flags     |N|   RESERVED  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   Template ID(4 octets)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
//                  Template Name (optional)                   //
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Figure 1: Template ID Sub-TLV

Type: Template ID, 1 octet, TBD.

Length: 6.

Flags: 1 octet of flags.

Where:

N-Flag: This flag, when set, indicates the presence of the Template Name encoding.

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

Template ID: a 4-octet value.

Template Name: Template MAY also be assigned with a template name, such template name MUST NOT be considered as identifiers for a template. The size of the tempalte name for the template is limited to 255 bytes.

5. SR Policy Operations

5.1. Advertisement of SR Policies

When BGP advertises an SR Policy, different candidate paths of the same SR Policy may have different template IDs or the same template ID, depending on the attributes required by the candidate paths of the SR Policy.

Reflectors just need to advertise the route of SR, no need to process it.

5.2. Reception of an SR Policy

SR Policy is only to be processed on the SR Policy headend, reflectors just need to reflect the route of SR Policy, no need to process it. 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 headend for the advertised policy. This is done by attaching one or more Route Target Extended Communities to the advertisement [RFC4360]. This process is defined in [I-D.ietf-idr-segment-routing-te-policy]. This draft does not add any extra process in this process.

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. The SRPM then determine how to use the template ID in SR Policy. The SRPM find the local configured template by template ID, and get all the attributes that is configured in the template, and then process the candidate path with these attributes. For example, if the template configure seamless bfd, then the SRPM can create sbfd sessions for each Segment List in the candidate path. If there is no template find, the SRPM should ignore the template ID and use the candidate path as there is no template ID.

6. Acknowledgements

TBD.

7. IANA Considerations

This document requests that IANA allocates a new sub-TLV type as defined in Section 4.1 from the "Sub-TLVs for SR Policy" registry as specified.

Value	Description	Reference
TBD	SR Policy Template ID	This document

Figure 2: Figure 2: Template ID sub-TLV

8. Security Considerations

These extensions to BGP SR Policy do not add any new security issues to the existing protocol.

9. References

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-23, 25 July 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-23>.

[I-D.ietf-pce-segment-routing-policy-cp]
Koldychev, M., Sivabalan, S., Barth, C., Peng, S., and H. Bidgoli, "PCEP extension to support Segment Routing Policy Candidate Paths", Work in Progress, Internet-Draft, draft-ietf-pce-segment-routing-policy-cp-12, 24 July 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-segment-routing-policy-cp-12>.

[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>.

[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>.

[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>.

[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>.

Authors' Addresses

Ka Zhang
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: zhangka@huawei.com

Zhibo Hu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: huzhibo@huawei.com

Jie Dong
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: jie.dong@huawei.com