IDR Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 4, 2021

H. Chen
Futurewei
Z. Li
Huawei
Z. Li
China Mobile
Y. Fan
Casa Systems
M. Toy
Verizon
L. Liu
Fujitsu

October 31, 2020

BGP Extensions for IDs Allocation
draft-wu-idr-bgp-segment-allocation-ext-06

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

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 May 4, 2021.

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 enginerring 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.

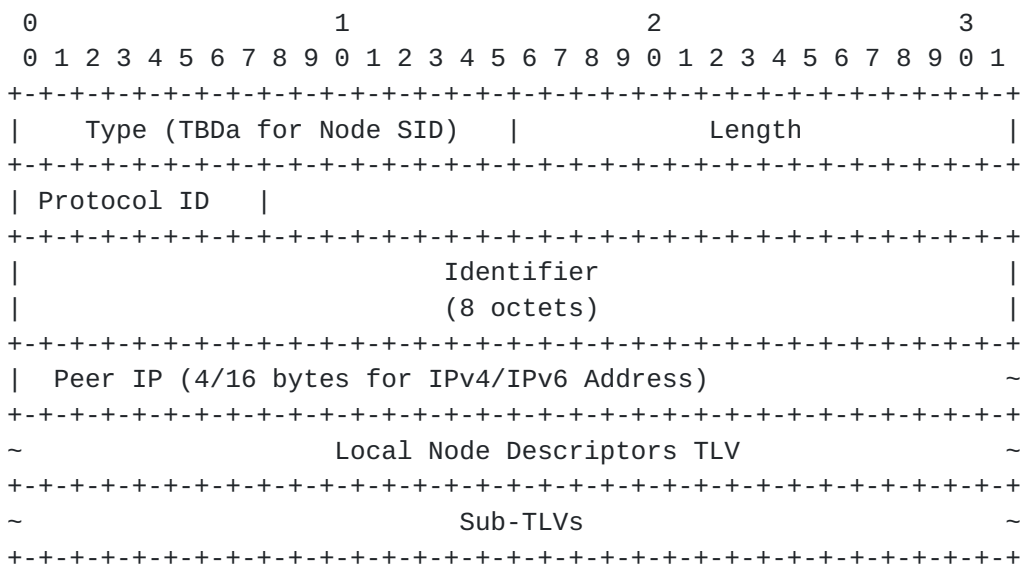
3. Protocol Extensions

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 [RFC7752].



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 [RFC7752], can be reused.

Sub-TLVs may be some of the followings:

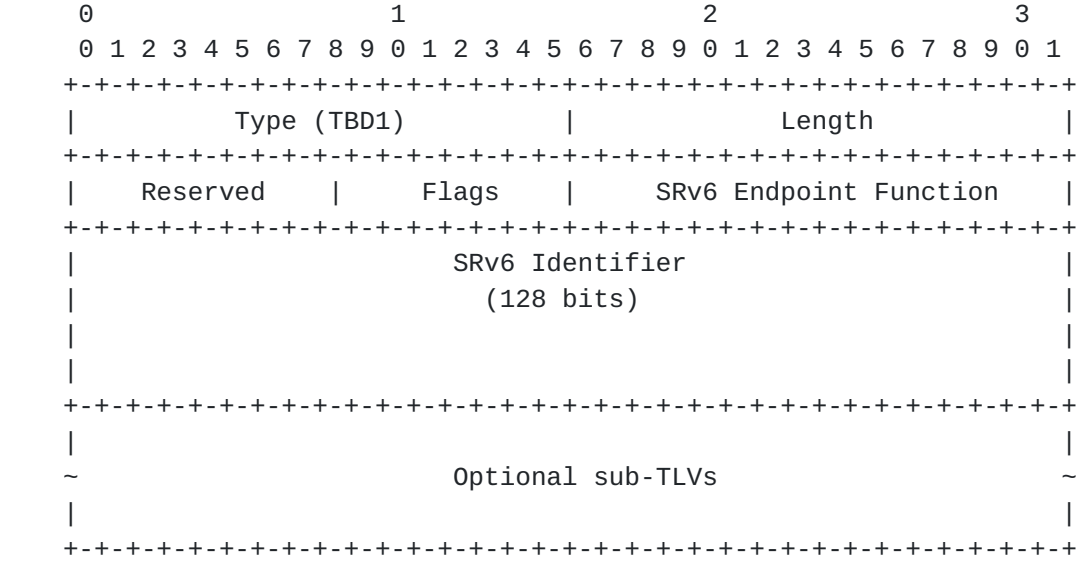
SR-Capabilities TLV (1034): It contains the Segment Routing Global Base (SRGB) range(s) allocated for the node.

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.



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.

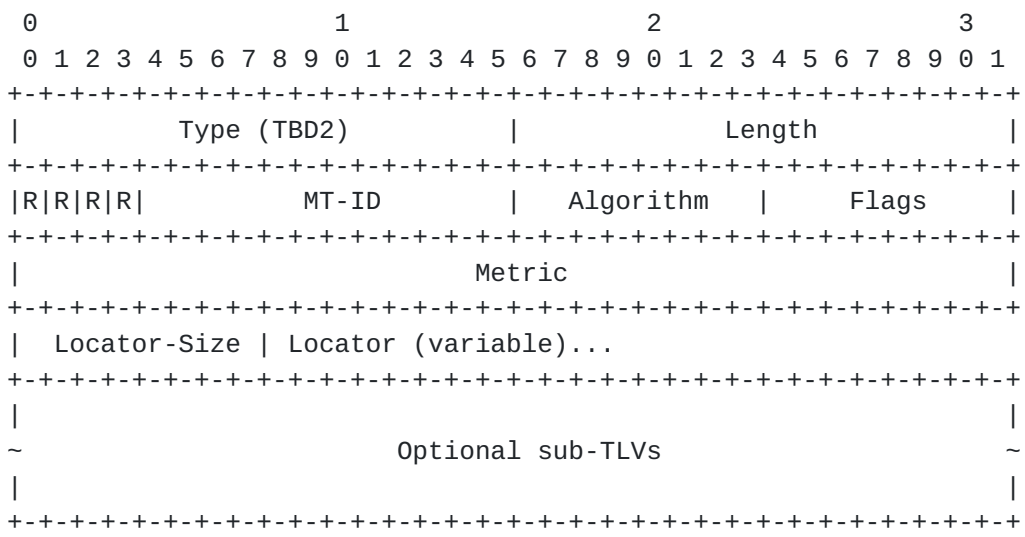
SRv6 Endpoint Function: 2 octets. The function associated with SRv6 SID.

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.



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 [I-D.ietf-lsr-isis-srv6-extensions].

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

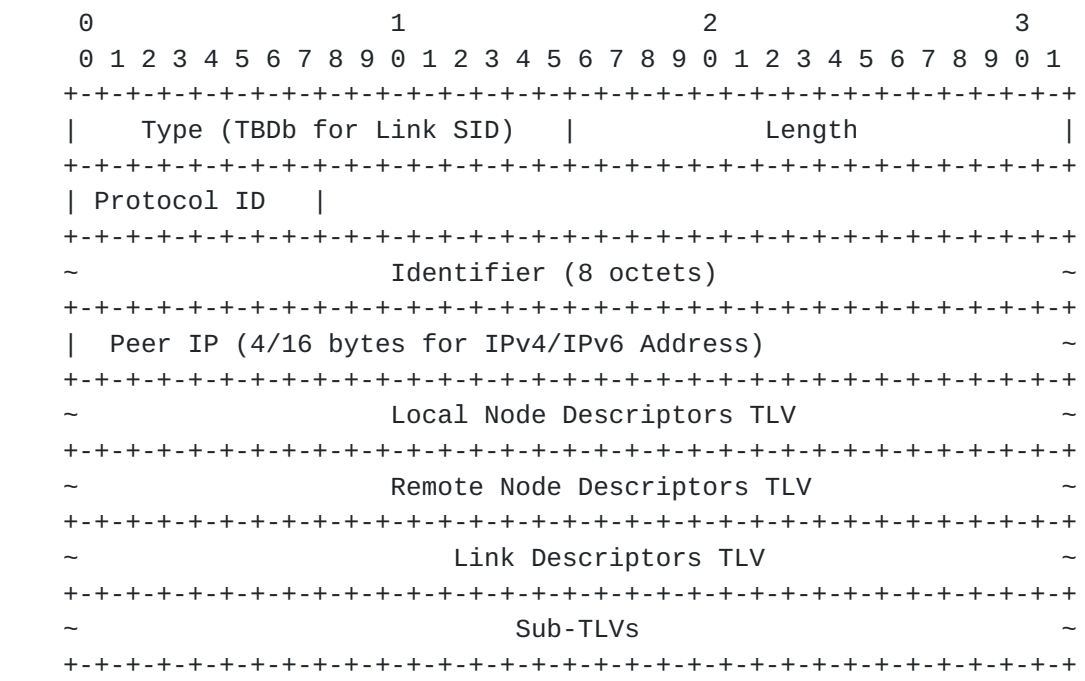
Locator-Size: 1 octet. Number of bits in the Locator field (1 to 128).

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 [RFC7752].



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 [RFC7752], can be reused.

The Sub-TLVs may be some of the followings:

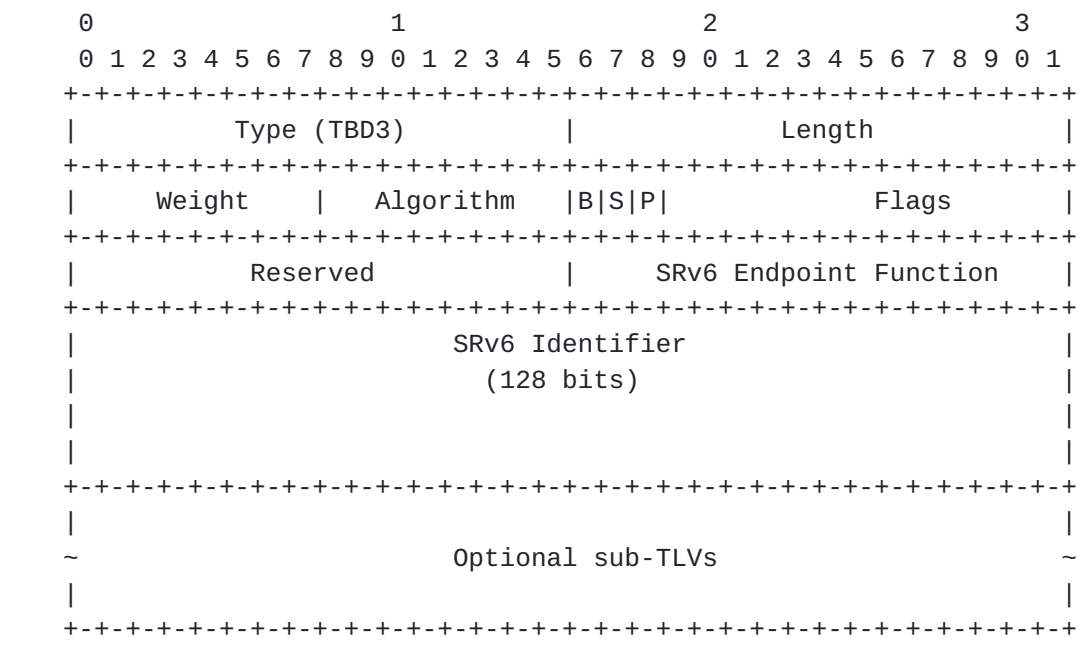
Adj-SID TLV (1099): It contains the Segment Identifier (SID) allocated for the link/adjacency.

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.



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.

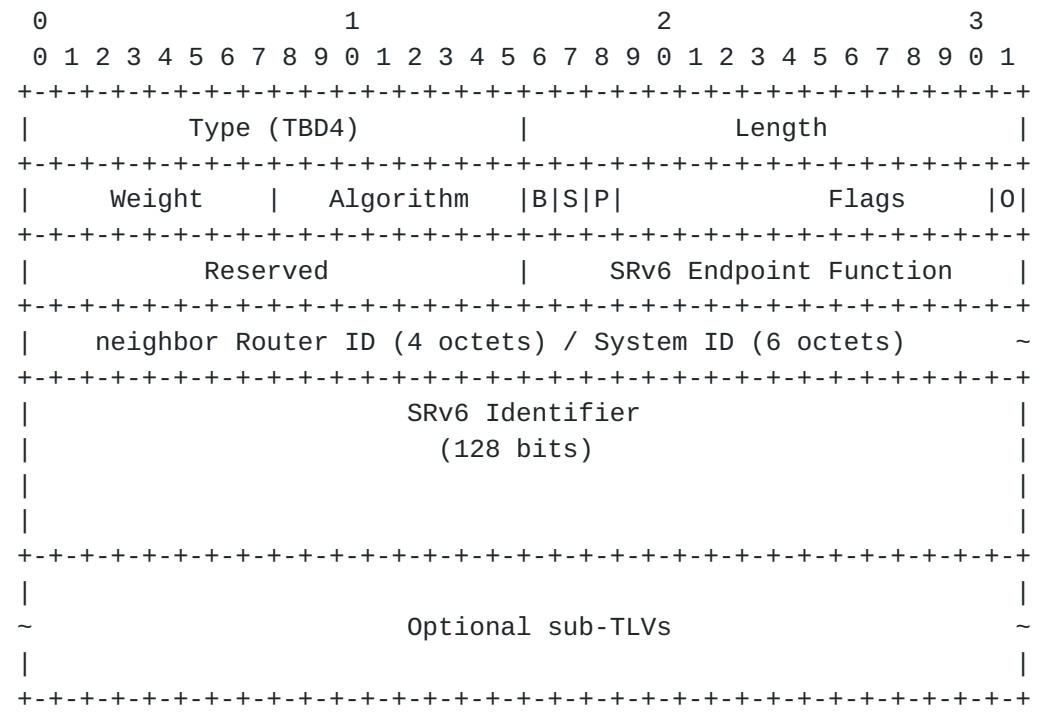
Flags: 2 octets. Three flags are defined in [I-D.ietf-lsr-isis-srv6-extensions].

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.



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 [I-D.ietf-lsr-isis-srv6-extensions]. 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.

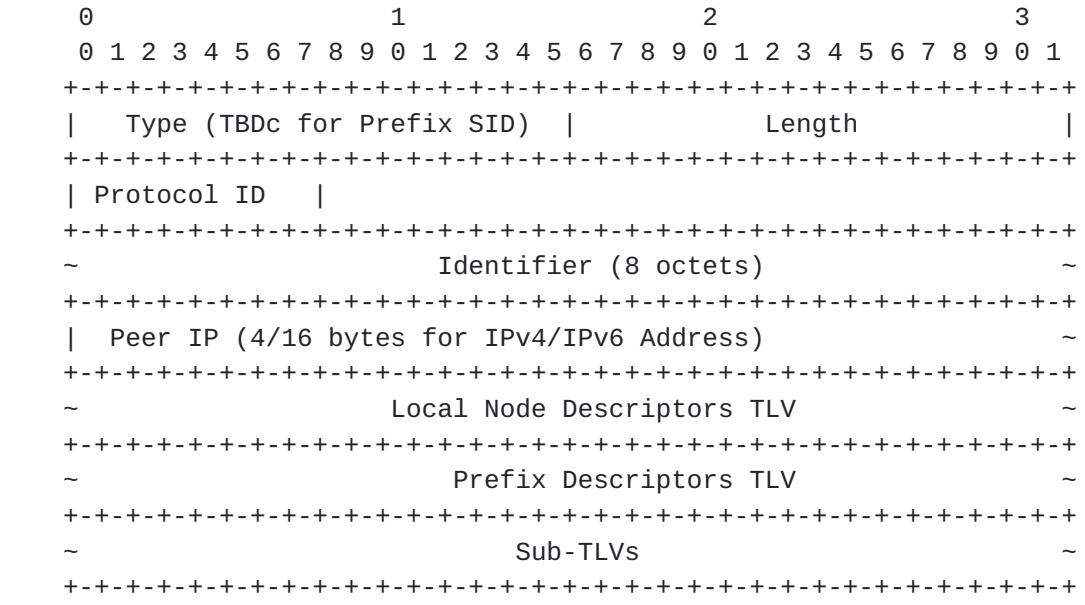
SRv6 Endpoint Function: 2 octets. The function associated with SRv6 SID.

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 [RFC7752].



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 [RFC7752], can be reused.

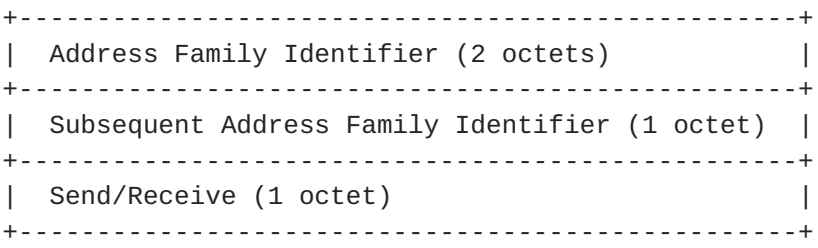
Sub-TLVs may be some of the followings:

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:



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>.

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 [RFC7752].

6. Acknowledgements

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

7. References

7.1. Normative References

[I-D.ietf-idr-flowspec-path-redirect]
Velde, G., Patel, K., and Z. Li, "Flowspec Indirection-id Redirect", draft-ietf-idr-flowspec-path-redirect-11 (work in progress), May 2020.

[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions-25 (work in progress), May 2019.

[I-D.ietf-lsr-isis-srv6-extensions]
Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extension to Support Segment Routing over IPv6 Dataplane", draft-ietf-lsr-isis-srv6-extensions-11 (work in progress), October 2020.

[I-D.ietf-rtgwg-bgp-routing-large-dc]
Lapukhov, P., Premji, A., and J. Mitchell, "Use of BGP for routing in large-scale data centers", draft-ietf-rtgwg-bgp-routing-large-dc-11 (work in progress), June 2016.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-15 (work in progress), January 2018.

[I-D.ietf-spring-segment-routing-ldp-interop]
Bashandy, A., Filsfils, C., Previdi, S., Decraene, B., and S. Litkowski, "Segment Routing interworking with LDP", draft-ietf-spring-segment-routing-ldp-interop-15 (work in progress), September 2018.

[I-D.li-ospf-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak, "OSPFv3 Extensions for SRv6", draft-li-ospf-ospfv3-srv6-extensions-07 (work in progress), November 2019.

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

[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>.

7.2. Informative References

[I-D.gredler-idr-bgp-ls-segment-routing-extension]
Gredler, H., Ray, S., Previdi, S., Filsfils, C., Chen, M., and J. Tantsura, "BGP Link-State extensions for Segment Routing", draft-gredler-idr-bgp-ls-segment-routing-extension-02 (work in progress), October 2014.

[I-D.ietf-idr-bgpls-segment-routing-epe]
Previdi, S., Talaulikar, K., Filsfils, C., Patel, K., Ray, S., and J. Dong, "BGP-LS extensions for Segment Routing BGP Egress Peer Engineering", draft-ietf-idr-bgpls-segment-routing-epe-19 (work in progress), May 2019.

[I-D.ietf-teas-enhanced-vpn]
Dong, J., Bryant, S., Li, Z., Miyasaka, T., and Y. Lee, "A Framework for Enhanced Virtual Private Networks (VPN+) Service", draft-ietf-teas-enhanced-vpn-06 (work in progress), July 2020.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA
USA

Email: Huaimo.chen@futurewei.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
USA

Email: yfan@casa-systems.com

Mehmet Toy
Verizon
USA

Email: mehmet.toy@verizon.com

Lei Liu
Fujitsu

USA

Email: liulei.kddi@gmail.com