IDR

Internet-Draft

Intended status: Standards Track

Expires: April 20, 2018

S. Previdi, Ed. C. Filsfils A. Lindem A. Sreekantiah Cisco Systems H. Gredler RtBrick Inc. October 17, 2017

Segment Routing Prefix SID extensions for BGP draft-ietf-idr-bgp-prefix-sid-07

Abstract

Segment Routing (SR) architecture allows a node to steer a packet flow through any topological path and service chain by leveraging source routing. The ingress node prepends a SR header to a packet containing a set of segment identifiers (SID). Each SID represents a topological or a service-based instruction. Per-flow state is maintained only at the ingress node of the SR domain.

This document defines a new optional, transitive BGP attribute for announcing BGP Prefix Segment Identifiers (BGP Prefix-SID) 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 RFC 2119 [RFC2119] only when they appear in all upper case. They may also appear in lower or mixed case as English words, without any normative meaning.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of \underline{BCP} 78 and \underline{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 April 20, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to <u>BCP 78</u> and the IETF Trust's Legal Provisions Relating to IETF Documents (https://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

<u>1</u> . Introduction	. 3
$\underline{\textbf{2}}. BGP-Prefix-SID $. 4
<u>2.1</u> . MPLS BGP Prefix SID	. 4
2.2. IPv6 Prefix Segment	. 5
$\underline{3}$. BGP-Prefix-SID Attribute	. 5
<u>3.1</u> . Label-Index TLV	. 6
3.2. IPv6 SID	
3.3. Originator SRGB TLV	
$\underline{4}$. Receiving BGP-Prefix-SID Attribute	. 9
<u>4.1</u> . MPLS Dataplane: Labeled Unicast	. 9
<u>4.2</u> . IPv6 Dataplane	. 10
$\underline{5}$. Announcing BGP-Prefix-SID Attribute	. 10
<u>5.1</u> . MPLS Dataplane: Labeled Unicast	. 10
<u>5.2</u> . IPv6 Dataplane	. 11
$\underline{6}$. Error Handling of BGP-Prefix-SID Attribute	. 11
$\underline{7}$. IANA Considerations	. 12
$\underline{\textbf{8}}$. Manageability Considerations	. 12
$\underline{9}.$ Security Considerations	. 13
$\underline{10}$. Contributors	. 13
$\underline{11}$. Acknowledgements	
$\underline{12}$. References	. 14
$\underline{12.1}$. Normative References	. 14
<u>12.2</u> . Informative References	
Authors' Addresses	. 15

1. Introduction

Segment Routing (SR) architecture leverages the source routing paradigm. A group of inter-connected nodes that use SR forms a SR domain. A segment represents either a topological instruction such as "go to prefix P following shortest path" or a service instruction (e.g.: "pass through deep packet inspection"). Other types of segments may be defined in the future.

A segment is identified through a Segment Identifier (SID). Typically, the ingress node of the SR domain prepends a SR header containing segments identifiers (SIDs) to an incoming packet.

As described in [I-D.ietf-spring-segment-routing], when SR is applied to the MPLS dataplane ([I-D.ietf-spring-segment-routing-mpls]) the SID consists of a label while when SR is applied to the IPv6 dataplane the SID consists of an IPv6 address.

A BGP-Prefix Segment (and its BGP Prefix-SID), is a BGP segment attached to a BGP prefix. A BGP Prefix-SID is always a global SID ([I-D.ietf-spring-segment-routing]) within the SR/BGP domain (i.e., the set of Autonomous Systems under a common administration and control and where SR is used) and identifies an instruction to forward the packet over the ECMP-aware best-path computed by BGP to the related prefix. The BGP Prefix-SID is the identifier of the BGP prefix segment. In this document, we always refer to the BGP Segment by the BGP Prefix-SID.

This document describes the BGP extension to signal the BGP Prefix-SID. Specifically, this document defines a new BGP attribute known as the BGP Prefix-SID attribute and specifies the rules to originate, receive and handle error conditions of the new attribute.

As described in [I-D.ietf-spring-segment-routing-msdc], the BGP Prefix-SID attribute defined in this document can be attached to prefixes from AFI/SAFI:

Multiprotocol BGP labeled IPv4/IPv6 Unicast ([RFC3107]).

Multiprotocol BGP ([RFC4760]) unlabeled IPv6 Unicast.

[I-D.ietf-spring-segment-routing-msdc] describes use cases where the Prefix-SID is used for the above AFI/SAFI.

It has to be noted that:

o A BGP Prefix-SID MAY be global between domains when the interconnected domains agree on the SID allocation scheme.

Alternatively, when interconnecting domains, the ASBRs of each domain will have to handle the advertisement of unique SIDs. The mechanisms for such interconnection are outside the scope of the protocol extensions defined in this document.

o As described in [I-D.ietf-spring-segment-routing-msdc], a BGP Prefix-SID MAY be attached to a prefix. In addition, each prefix will likely have a different as_path attribute. This implies that each prefix is advertised individually, reducing the ability to pack BGP advertisements (when sharing common attributes).

2. BGP-Prefix-SID

The BGP Prefix-SID attached to a BGP prefix P represents the instruction "go to Prefix P" along its BGP bestpath (potentially ECMP-enabled).

2.1. MPLS BGP Prefix SID

The BGP Prefix-SID is realized on the MPLS dataplane ([I-D.ietf-spring-segment-routing-mpls]) in the following way:

As described in [I-D.ietf-spring-segment-routing-msdc] the operator assigns a globally unique "index", L_I, to a locally sourced prefix of a BGP speaker N which is advertised to all other BGP speakers in the SR domain.

According to [I-D.ietf-spring-segment-routing], each BGP speaker is configured with a label block called the Segment Routing Global Block (SRGB). While [I-D.ietf-spring-segment-routing] recommends to use the same SRGB across all the nodes within the SR domain, the SRGB of a node is a local property and could be different on different speakers. The drawbacks of the use case where BGP speakers have different SRGBs are documented in [I-D.ietf-spring-segment-routing] and [I-D.ietf-spring-segment-routing-msdc].

If traffic-engineering within the SR domain is required, each node may also be required to advertise topological information and Peering SID's for each of its links and peers. This information is required in order to perform the explicit path computation and to express any explicit path into a list of SIDs. The advertisement of topological information and Peer segments (Peer SIDs) is assumed to be done through [I-D.ietf-idr-bgpls-segment-routing-epe].

If the BGP speakers are not all configured with the same SRGB, and if traffic-engineering within the SR domain is required, each node

may be required to advertise its local SRGB in addition to the topological information.

This documents assumes that BGP-LS is the preferred method for collecting both topological, peer segments (Peer SIDs) and SRGB information through [RFC7752],

[I-D.ietf-idr-bgpls-segment-routing-epe] and [I-D.ietf-idr-bgp-ls-segment-routing-ext]. However, as an optional alternative for the advertisement of the local SRGB without the topology nor the peer SID's, hence without applicability for TE, the Originator SRGB TLV of the prefix-SID attribute, is specified in Section 3.3 of this document.

As defined in [I-D.ietf-spring-segment-routing-mpls], the index L_I is an offset in the SRGB. Each BGP speaker derives its local MPLS label, L, by adding L_I to the start value of its own SRGB, and programs L in its MPLS dataplane as its incoming/local label for the prefix. It has to be noted that while SRGBs and SIDs are advertised using 32 bit values, the derived label is to be considered as the 20 right-most bits. See Section 4.1 for more details.

The outgoing label for the prefix is found in the NLRI of the Multiprotocol BGP labeled IPv4/IPv6 Unicast prefix advertisement. The index L_I is only used as a hint to derive the local/incoming label.

<u>Section 3.1</u> of this document specifies the Label-Index TLV of the BGP Prefix-SID attribute; this TLV can be used to advertise the label index of a given prefix.

In order to advertise the label index of a given prefix P and, optionally, the SRGB, a new extension to BGP is needed: the BGP Prefix-SID attribute. This extension is described in subsequent sections.

2.2. IPv6 Prefix Segment

As illustrated in [I-D.ietf-spring-segment-routing-msdc], when SR is used over an IPv6 dataplane, the BGP Prefix-SID consists of an IPv6 address assigned to the BGP speaker.

3. BGP-Prefix-SID Attribute

The BGP Prefix-SID attribute is an optional, transitive BGP path attribute. The attribute type code 40 has been assigned by IANA (see Section 7).

The BGP Prefix-SID attribute is defined here to be a set of elements encoded as "Type/Length/Value" (i.e., a set of TLVs). The following TLVs are defined:

- o Label-Index TLV
- o IPv6 SID TLV
- o Originator SRGB TLV

Label-Index and Originator SRGB TLVs are used only when SR is applied to the MPLS dataplane.

IPv6 SID TLV is used only when SR is applied to the IPv6 dataplane.

3.1. Label-Index TLV

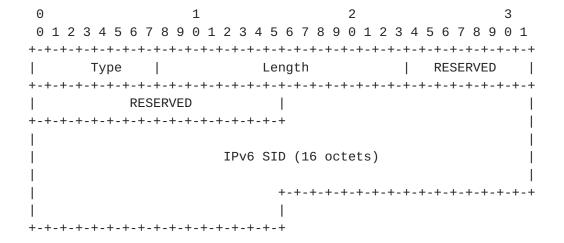
The Label-Index TLV MUST be present in the Prefix-SID attribute attached to Labeled IPv4/IPv6 unicast prefixes ([RFC3107]) and has the following format:

where:

- o Type is 1.
- o Length: is 7, the total length of the value portion of the TLV.
- o RESERVED: 8 bit field. MUST be clear on transmission an MUST be ignored at reception..
- o Flags: 16 bits of flags. None is defined by this document. The flag field MUST be clear on transmission and MUST be ignored at reception.
- o Label Index: 32 bit value representing the index value in the SRGB space.

3.2. IPv6 SID

The IPv6-SID TLV MAY be present in the Prefix-SID attribute attached to MP-BGP unlabeled IPv6 unicast prefixes ([RFC4760]) and has the following format:

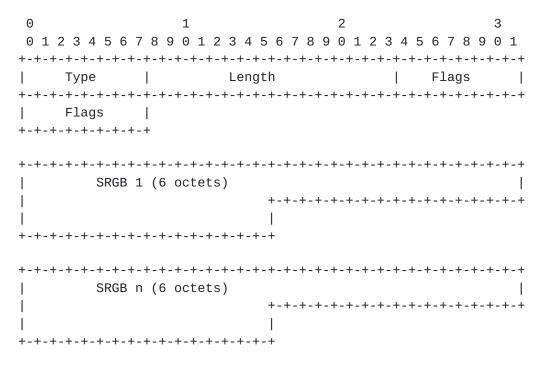


where:

- o Type is 2.
- o Length: is 19, the total length of the value portion of the TLV.
- o RESERVED: 24 bit field for future use. MUST be clear on transmission an MUST be ignored at reception.
- o IPv6 SID: 16 octets.

3.3. Originator SRGB TLV

The Originator SRGB TLV is an optional TLV and has the following format:



where:

- o Type is 3.
- o Length is the total length of the value portion of the TLV: 2 + multiple of 6.
- o Flags: 16 bits of flags. None is defined in this document. Flags MUST be clear on transmission an MUST be ignored at reception.
- o SRGB: 3 octets of base followed by 3 octets of range. Note that the SRGB field MAY appear multiple times. If the SRGB field appears multiple times, the SRGB consists of multiple ranges.

The Originator SRGB TLV contains the SRGB of the node originating the prefix to which the BGP Prefix-SID is attached. The Originator SRGB TLV MUST NOT be changed during the propagation of the BGP update.

The originator SRGB describes the SRGB of the node where the BGP Prefix SID is attached. It is used to build segment routing policies when different SRGB's are used in the fabric ([I-D.ietf-spring-segment-routing-msdc]).

The originator SRGB may only appear on Prefix-SID attribute attached to prefixes of SAFI 4 (labeled unicast, [RFC3107]).

4. Receiving BGP-Prefix-SID Attribute

A BGP speaker receiving a BGP Prefix-SID attribute from an EBGP neighbor residing outside the boundaries of the SR domain, SHOULD discard the attribute unless it is configured to accept the attribute from the EBGP neighbor. A BGP speaker MAY log an error for further analysis when discarding an attribute.

4.1. MPLS Dataplane: Labeled Unicast

A Multiprotocol BGP labeled IPv4/IPv6 Unicast ([RFC3107]) session type is required.

A BGP speaker MAY be locally configured with an SRGB=[SRGB_Start, SRGB_End]. The preferred method for deriving the SRGB is a matter of local node configuration.

Given a label_index L_I, we call L = L_I + SRGB_Start as the derived label. A BGP Prefix-SID attribute is called "unacceptable" for a speaker M if the derived label value L lies outside the SRGB configured on M. Otherwise the Label Index attribute is called "acceptable" to speaker M.

The mechanisms through which a given label_index value is assigned to a given prefix are outside the scope of this document. The label-index value associated with a prefix is locally configured at the BGP node originating the prefix.

The Prefix-SID attribute MUST contain the Label-Index TLV and MAY contain the Originator SRGB TLV. A BGP Prefix-SID attribute received without a Label-Index TLV MUST be considered as "unacceptable" by the receiving speaker.

If multiple prefixes are received with the same label_index value, all these prefixes MUST have their BGP Prefix-SID attribute considered as "unacceptable" by the receiving speaker.

When a BGP speaker receives a path from a neighbor with an acceptable BGP Prefix-SID attribute, it MUST program the derived label as the local label for the prefix in its MPLS dataplane. In case of any error, a BGP speaker MUST resort to the error handling rules specified in <u>Section 6</u>. A BGP speaker MAY log an error for further analysis.

When a BGP speaker receives a path from a neighbor with an unacceptable BGP Prefix-SID attribute or when a BGP speaker receives a path from a neighbor with a BGP Prefix-SID attribute but is unable to process it (it does not have the capability or local policy

disables the capability), it MUST treat the path as if it came without a Prefix-SID attribute. For the purposes of local label allocation, a BGP speaker MUST assign a local (also called dynamic) label (non-SRGB) for such a prefix as per classic Multiprotocol BGP labeled IPv4/IPv6 Unicast ([RFC3107]) operation. A BGP speaker MAY log an error for further analysis.

The outgoing label is always programmed as per classic Multiprotocol BGP labeled IPv4/IPv6 Unicast (RFC3107 [RFC3107]) operation.

Specifically, a BGP speaker receiving a prefix with a Prefix-SID attribute and a label NLRI field of implicit-null from a neighbor MUST adhere to standard behavior and program its MPLS dataplane to pop the top label when forwarding traffic to the prefix. The label NLRI defines the outbound label that MUST be used by the receiving node. The Label Index gives the information to the receiving node on which local/incoming label the BGP speaker SHOULD use.

4.2. IPv6 Dataplane

When an SR IPv6 BGP speaker receives a IPv6 Unicast BGP Update with a prefix having the BGP Prefix-SID attribute attached, it checks whether the IPv6 SID TLV is present. If present, then the receiver assumes that the originator supports SR on the IPv6 dataplane.

The Originator SRGB MUST be ignored on reception.

A BGP speaker receiving a BGP Prefix-SID attribute from an EBGP neighbor residing outside the boundaries of the SR domain, SHOULD discard the attribute unless it is configured to accept the attribute from the EBGP neighbor. A BGP speaker MAY log an error for further analysis when discarding an attribute.

5. Announcing BGP-Prefix-SID Attribute

The BGP Prefix-SID attribute MAY be attached to labeled BGP prefixes (IPv4/IPv6) [RFC3107] or to IPv6 prefixes [RFC4760]. In order to prevent distribution of the BGP Prefix-SID attribute beyond its intended scope of applicability, attribute filtering SHOULD be deployed.

<u>5.1</u>. MPLS Dataplane: Labeled Unicast

A BGP speaker that originates a prefix attaches the Prefix-SID attribute when it advertises the prefix to its neighbors via Multiprotocol BGP labeled IPv4/IPv6 Unicast ([RFC3107]). The value of the Label-Index in the Label-Index TLV is determined by configuration.

A BGP speaker that originates a Prefix-SID attribute MAY optionally announce Originator SRGB TLV along with the mandatory Label-Index TLV. The content of the Originator SRGB TLV is determined by the configuration.

Since the Label-index value must be unique within an SR domain, by default an implementation SHOULD NOT advertise the BGP Prefix-SID attribute outside an Autonomous System unless it is explicitly configured to do so.

A BGP speaker that advertises a path received from one of its neighbors SHOULD advertise the Prefix-SID received with the path without modification regardless of whether the Prefix-SID was acceptable. If the path did not come with a Prefix-SID attribute, the speaker MAY attach a Prefix-SID to the path if configured to do so. The content of the TLVs present in the Prefix-SID is determined by the configuration.

In all cases, the label field of the advertised NLRI ([RFC3107], [RFC4364]) MUST be set to the local/incoming label programmed in the MPLS dataplane for the given advertised prefix. If the prefix is associated with one of the BGP speakers interfaces, this label is the usual MPLS label (such as the implicit or explicit NULL label).

5.2. IPv6 Dataplane

A BGP speaker that originates an IPv6 prefix with the Prefix-SID attribute, MAY include the IPv6 SID TLV.

A BGP speaker that advertises a path received from one of its neighbors SHOULD advertise the Prefix-SID received with the path without modification regardless of whether the Prefix-SID was acceptable. If the path did not come with a Prefix-SID attribute, the speaker MAY attach a Prefix-SID to the path if configured to do so.

6. Error Handling of BGP-Prefix-SID Attribute

When a BGP Speaker receives a BGP Update message containing a malformed BGP Prefix-SID attribute, it MUST ignore the received BGP Prefix-SID attributes and not pass it to other BGP peers. This is equivalent to the -attribute discard- action specified in [RFC7606]. When discarding an attribute, a BGP speaker MAY log an error for further analysis.

If the BGP Prefix-SID attribute appears more than once in an BGP Update message, then, according to [RFC7606], all the occurrences of

the attribute other than the first one SHALL be discarded and the BGP Update message SHALL continue to be processed.

When a BGP speaker receives an unacceptable Prefix-SID attribute, it MAY log an error for further analysis.

7. IANA Considerations

This document defines a new BGP path attribute known as the BGP Prefix-SID attribute. This document requests IANA to assign a new attribute code type (suggested value: 40) for BGP the Prefix-SID attribute from the BGP Path Attributes registry.

Currently, IANA temporarily assigned the following:

```
40 BGP Prefix-SID (TEMPORARY - registered 2015-09-30, expires 2016-09-30) [draft-ietf-idr-bgp-prefix-sid]
```

This document defines 3 new TLVs for BGP Prefix-SID attribute. These TLVs need to be registered with IANA. We request IANA to create a new registry for BGP Prefix-SID Attribute TLVs as follows:

Under "Border Gateway Protocol (BGP) Parameters" registry, "BGP Prefix-SID attribute Types" Reference: draft-ietf-idr-bgp-prefix-sid Registration Procedure(s): Values 1-254 First Come, First Served, Value 0 and 255 reserved

Value	Туре	Reference		
0	Reserved	this	document	
1	Label-Index	this	document	
2	IPv6 SID	this	document	
3	Originator SRGB	this	document	
4-254	Unassigned			
255	Reserved	this	document	

8. Manageability Considerations

This document defines a new BGP attribute in order to address the use case described in [I-D.ietf-spring-segment-routing-msdc]. It i assumed that the new attribute (BGP Prefix-SID) advertisement is controlled by the operator in order to:

o prevent undesired origination/advertisement of the BGP Prefix-SID attribute. By default, a BGP Prefix-SID attribute SHOULD NOT be originated and attached to a prefix. The operator MUST be capable of explicitly enabling the BGP Prefix-SID origination.

o Prevent any undesired propagation of the BGP Prefix-SID attribute. By default the BGP Prefix-SID is not advertised outside the boundary of an AS. The propagation to other ASS MUST be explicitly configured.

The deployment model described in [I-D.ietf-spring-segment-routing-msdc] assumes multiple Autonomous Systems (AS) under a common administration. The BGP Prefix-SID advertisement is therefore applicable to inter-AS context while it is confined within a single SR Domain.

9. Security Considerations

This document introduces a new BGP attribute (BGP Prefix-SID) which inherits the security considerations expressed in: [RFC4271] and [RFC3107].

The BGP Prefix-SID attribute addresses the requirements introduced in [I-D.ietf-spring-segment-routing-msdc] and It has to be noted, as described in Section 8, that this document refer to a deployment model where all nodes are under the same administration. In this context, we assume that the operator doesn't want to leak outside of the domain any information related to internal prefixes and topology. The internal information includes the BGP Prefix-SID. In order to prevent such leaking, the standard BGP mechanisms (filters) are applied on the boundary of the SR domain.

10. Contributors

Keyur Patel Arrcus, Inc. US

Email: Keyur@arrcus.com

Saikat Ray Unaffiliated US

Email: raysaikat@gmail.com

11. Acknowledgements

The authors would like to thanks Satya Mohanty for his contribution to this document.

12. References

12.1. Normative References

- [I-D.ietf-spring-segment-routing]
 Filsfils, C., Previdi, S., Decraene, B., Litkowski, S.,
 and R. Shakir, "Segment Routing Architecture", draft-ietfspring-segment-routing-12 (work in progress), June 2017.
- [I-D.ietf-spring-segment-routing-mpls]
 Filsfils, C., Previdi, S., Bashandy, A., Decraene, B.,
 Litkowski, S., and R. Shakir, "Segment Routing with MPLS
 data plane", <u>draft-ietf-spring-segment-routing-mpls-10</u>
 (work in progress), June 2017.
- [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.
- [RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, DOI 10.17487/RFC3107, May 2001, https://www.rfc-editor.org/info/rfc3107.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", <u>RFC 4364</u>, DOI 10.17487/RFC4364, February 2006, https://www.rfc-editor.org/info/rfc4364.

12.2. Informative References

[I-D.ietf-idr-bgp-ls-segment-routing-ext]

Previdi, S., Psenak, P., Filsfils, C., Gredler, H., and M.

Chen, "BGP Link-State extensions for Segment Routing",

draft-ietf-idr-bgp-ls-segment-routing-ext-03 (work in progress), July 2017.

[I-D.ietf-idr-bgpls-segment-routing-epe]
Previdi, S., 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-13 (work in progress), June 2017.

[I-D.ietf-spring-segment-routing-msdc]
Filsfils, C., Previdi, S., Mitchell, J., Aries, E., and P.
Lapukhov, "BGP-Prefix Segment in large-scale data
centers", draft-ietf-spring-segment-routing-msdc-05
in progress), June 2017.

[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.

[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>.

Authors' Addresses

Stefano Previdi (editor) Cisco Systems TT

Email: stefano@previdi.net

Clarence Filsfils Cisco Systems Brussels Belgium

Email: cfilsfils@cisco.com

Acee Lindem Cisco Systems 170 W. Tasman Drive San Jose, CA 95124 95134 USA

Email: acee@cisco.com

Arjun Sreekantiah Cisco Systems 170 W. Tasman Drive San Jose, CA 95124 95134 USA

Email: asreekan@cisco.com

Hannes Gredler RtBrick Inc.

Email: hannes@rtbrick.com