IDR
Internet-Draft
Intended status: Standards Track
Expires: November 26, 2015

K. Patel
S. Previdi
C. Filsfils
A. Sreekantiah
Cisco Systems
S. Ray
Unaffiliated

May 25, 2015

Segment Routing Prefix SID extensions for BGP
draft-keyupate-idr-bgp-prefix-sid-02

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 "segments". Each segment represents a topological or a service-based instruction. Per-flow state is maintained only at the ingress node of the SR domain.

The Segment Routing architecture can be implemented using MPLS with no changes to the forwarding plane. It requires minor extensions to the existing routing protocols.

This document describes the BGP extension for announcing BGP Prefix Segment Identifier (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].

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 http://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 November 26, 2015.

Table of Contents

1. Introduction

Segment Routing (SR) architecture leverages the source routing paradigm. A group of inter-connected nodes that use SR forms a SR domain. The ingress node of the SR domain prepends a SR header containing "segments" to an incoming packet. Each segment represents a topological instruction (such as "go to prefix P following shortest path") or a service instruction ("pass through deep packet inspection"). By inserting the desired sequence of instructions, the

ingress node is able to steer a packet via any topological path and/or service chain; per-flow state is maintained only at the ingress node of the SR domain.

Each segment is identified by a Segment Identifier (SID). By using MPLS labels as SIDs, the SR architecture can be implemented using the existing MPLS dataplane.

A BGP-Prefix Segment (aka BGP-Prefix-SID), is a BGP segment attached to a BGP prefix. A BGP-Prefix-SID is always global within the SR/BGP domain 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.

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

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in [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.

3. Segment Routing Documents

The main reference for this document is the SR architecture defined in [I-D.ietf-spring-segment-routing].

The Segment Routing Egress Peer Engineering architecture is described in [I-D.filsfils-spring-segment-routing-central-epe].

The Segment Routing Egress Peer Engineering BGPLS extensions are described in [I-D.previdi-idr-bgpls-segment-routing-epe].

A practical use case of the BGP Prefix SID is illustrated in [I-D.filsfils-spring-segment-routing-msdc].

4. 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). This Segment is realized on a MPLS dataplane in the following way:

According to [I-D.ietf-spring-segment-routing], each BGP speaker is configured with a label block called Segment Routing Global Block (SRGB). The SRGB could be different on different speakers.

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.

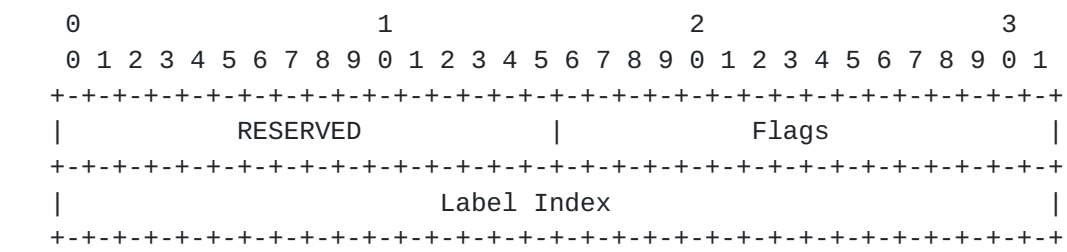
The index L_I is a 32 bit offset in the SRGB. Each BGP speaker derives its local MPLS label, L, by adding L_I to the start value of its own SRGB, and programs L in its MPLS dataplane as its incoming/local label for the prefix.

If the BGP speakers are configured with the same SRGB start value, they will all program the same MPLS label value for a given prefix P. This has the effect of having a single label value for prefix P across all BGP speakers despite the MPLS paradigm of "local label" is preserved.

In order to advertise the SRGB label index of a given prefix P, a new extension to BGP is needed. This extension is described in subsequent sections.

5. BGP-Prefix-SID Label Index Attribute

BGP Prefix Label Index is a new optional, transitive BGP path attribute. The attribute type code for BGP Label Index attribute is to be assigned by IANA (suggested value: 40). The value field of the Label Index attribute has following format:



where:

- RESERVED: 16 bit field. SHOULD be unset on transmission and MUST be ignored on reception.

- Flags: 16 bits of flags. None are defined in this document. Flags SHOULD be unset on transmission an MUST be ignored at reception.

- Label Index: 32 bit value representing the index value in the SRGB space.

Using the BGP protocol Label index as an offset, a label value for a given prefix is computed from a BGP SRGB. The BGP SRGB protocol label block is configured explicitly on each BGP Speaker enabled with BGP-Prefix-SID extensions.

6. Receiving BGP-Prefix-SID Label Index Attribute

It is assumed that a BGP speaker is configured with an SRGB=[GB_S, GB_E]. Given a label index L_I, we call L = L_I + GB_S as the derived label. A BGP Label Index 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.

When a BGP speaker receives a path from a neighbor with an acceptable BGP Label Index attribute, it SHOULD 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 the later section of the document. A BGP speaker MAY log an error for further analysis.

When a BGP speaker receives a path from a neighbor with an unacceptable BGP Label Index attribute, for the purpose of label allocation, it SHOULD treat the path as if it came without a Label Index attribute. A BGP speaker MAY choose to assign a local (also called dynamic) label (non-SRGB) for such a prefix. A BGP speaker MAY log an error for further analysis.

A BGP speaker receiving a BGP Label index attribute from a 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.

A BGP speaker receiving a prefix with a Label index 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 a hint to the receiving node on which local/incoming label he SHOULD use.

7. Announcing BGP-Prefix-SID Label Index Attribute

A BGP speaker that originates a prefix attaches the Label Index attribute when it advertises the prefix to its neighbors. The value of the Label Index is determined by configuration.

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

In all cases, the label field of the NLRI ([RFC3107], [RFC4364]) MUST be set to the label programmed in the MPLS dataplane for the given prefix. If the prefix is that of a local interface of the speaker, this label is the usual MPLS label (such as implicit or explicit NULL label).

The BGP Label Index attribute SHOULD only be announced with BGP Prefixes carried in a labeled address-family (SAFI value 4 or SAFI value 128). Since the BGP Label index value must be unique within an SR domain, by default an implementation SHOULD NOT advertise the BGP Label Index attribute outside an Autonomous System unless it is explicitly configured to do so. To contain distribution of the BGP Label Index attribute beyond its intended scope of applicability, attribute filtering MAY be deployed.

8. Error Handling of BGP-Prefix-SID Label Index Attribute

When a BGP Speaker receives a BGP Update message containing more than one, or a malformed BGP Label Index attribute, it MUST ignore the received BGP Label Index attributes and not pass it to other BGP peers. (see [I-D.ietf-idr-error-handling], Section 7). This is equivalent to the -attribute discard- action specified in [[I-D.ietf-idr-error-handling]. A BGP speaker MAY log an error when discarding an attribute for further analysis.

When a BGP Speaker receives a BGP Label Index attribute that is attached to prefixes belonging to SAFI value other than 4 or 128, it MUST quietly ignore the received attribute and not pass it to other BGP peers. A BGP speaker MAY log an error for further analysis.

When a BGP speaker receives an unacceptable Label Index attribute, it MAY log an error for further analysis.

9. IANA Considerations

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

10. Security Considerations

This document introduces no new security considerations above and beyond those already specified in [RFC4271] and [RFC3107].

11. Acknowledgements

The authors would like to thanks Satya Mohanty and Acee Lindem for their contribution to this document.

12. Change Log

Initial Version: Sep 21 2014

13. References

13.1. Normative References

[I-D.ietf-idr-error-handling]
Chen, E., Scudder, J., Mohapatra, P., and K. Patel, "Revised Error Handling for BGP UPDATE Messages", draft-ietf-idr-error-handling-19 (work in progress), April 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[RFC4271] Rekhter, Y., Li, T., and S. Hares, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.

13.2. Informative References

[I-D.filsfils-spring-segment-routing-central-epe]
Filsfils, C., Previdi, S., Patel, K., Aries, E., shaw@fb.com, s., Ginsburg, D., and D. Afanasiev, "Segment Routing Centralized Egress Peer Engineering", draft-filsfils-spring-segment-routing-central-epe-03 (work in progress), January 2015.

[I-D.filsfils-spring-segment-routing-msdc]
Filsfils, C., Previdi, S., Mitchell, J., Black, B., Afanasiev, D., Ray, S., and K. Patel, "BGP-Prefix Segment in large-scale data centers", draft-filsfils-spring-segment-routing-msdc-01 (work in progress), April 2015.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., Litkowski, S., Horneffer, M., Shakir, R., Tantsura, J., and E. Crabbe, "Segment Routing Architecture", draft-ietf-spring-segment-routing-02 (work in progress), May 2015.

[I-D.previdi-idr-bgpls-segment-routing-epe]
Previdi, S., Filsfils, C., Ray, S., Patel, K., Dong, J., and M. Chen, "Segment Routing Egress Peer Engineering BGP-LS Extensions", draft-previdi-idr-bgpls-segment-routing-epe-03 (work in progress), April 2015.

Authors' Addresses

Keyur Patel
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95124 95134
USA

Email: keyupate@cisco.com

Stefano Previdi
Cisco Systems
Via Del Serafico, 200
Rome 00142
Italy

Email: sprevidi@cisco.com

Clarence Filsfils
Cisco Systems
Brussels
Belgium

Email: cfilsfils@cisco.com

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

Email: asreekan@cisco.com

Saikat Ray
Unaffiliated

Email: raysaikat@gmail.com