c

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: November 27, 2014

S. Previdi, Ed.
C. Filsfils
S. Ray
K. Patel
Cisco Systems, Inc.
May 26, 2014

Segment Routing Egress Peer Engineering BGPLS Extensions
draft-previdi-idr-bgpls-segment-routing-epe-00

Abstract

Segment Routing (SR) leverages source routing. A node steers a packet through a controlled set of instructions, called segments, by prepending the packet with an SR header. A segment can represent any instruction, topological or service-based. SR allows to enforce a flow through any topological path and service chain while maintaining per-flow state only at the ingress node of the SR domain.

The Segment Routing architecture can be directly applied to the MPLS dataplane with no change on the forwarding plane. It requires minor extension to the existing link-state routing protocols.

This document outline a BGPLS extension for exporting BGP egress point topology information (including its peers, interfaces and peering ASs) in a way that is exploitable in order to compute efficient Egress Point Engineering policies and strategies.

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 27, 2014.

Table of Contents

1. Introduction

Segment Routing (SR) leverages source routing. A node steers a packet through a controlled set of instructions, called segments, by prepending the packet with an SR header. A segment can represent any instruction, topological or service-based. SR allows to enforce a flow through any topological path and service chain while maintaining per-flow state only at the ingress node of the SR domain.

The Segment Routing architecture can be directly applied to the MPLS dataplane with no change on the forwarding plane. It requires minor extension to the existing link-state routing protocols.

This document outline a BGPLS extension for exporting BGP egress point topology information (including its peers, interfaces and peering ASs) in a way that is exploitable in order to compute efficient Egress Point Engineering policies and strategies.

2. Segment Routing Documents

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

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

3. BGP Peering Segments

As defined in [draft-filsfils-spring-segment-routing-epe], an EPE enabled Egress PE node MAY advertise segments corresponding to its attached peers. These segments are called BGP peering segments or BGP Peering SIDs. They enable the expression of source-routed inter-domain paths.

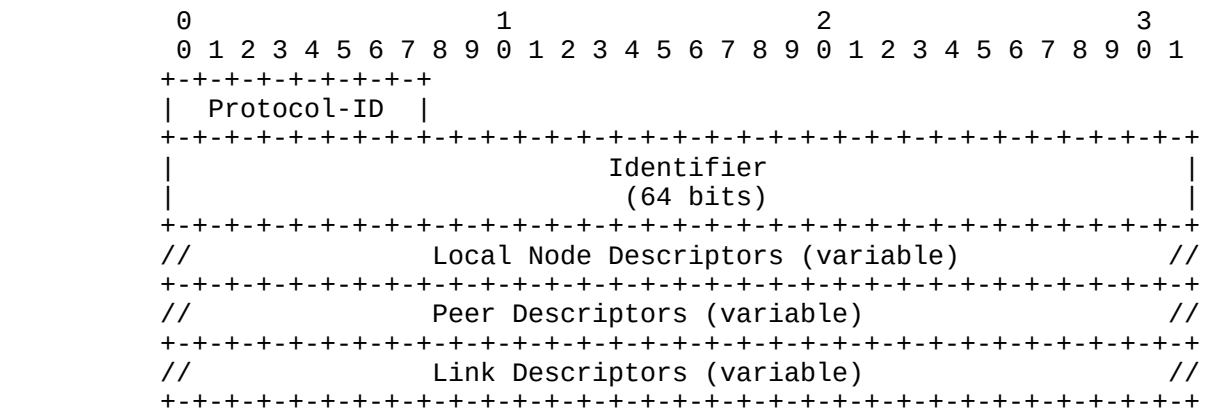
An ingress border router of an AS may compose a list of segments to steer a flow along a selected path within the AS, towards a selected egress border router C of the AS and through a specific peer. At minimum, a BGP Peering Engineering policy applied at an ingress PE involves two segments: the Node SID of the chosen egress PE and then the BGP Peering Segment for the chosen egress PE peer or peering interface.

Hereafter, we will define three types of BGP peering segments/SID’s: PeerNodeSID, PeerAdjSID and PeerGroupSID.

4. Peering Segment NLRI-Type

This section described a new NLRI-Type in the BGP-LS specification ([I-D.ietf-idr-ls-distribution]). The new NLRI-Type (5) is called the Peer NLRI-Type and describes the connectivity of a BGP Egress router.

The format of the Peer NLRI Type is as follows:



Where:

Local Node Descriptors: as defined in [I-D.ietf-idr-ls-distribution] Section 3.2.1.2.

Link Descriptors: as defined in [I-D.ietf-idr-ls-distribution] Section 3.2.2.

4.1. Peer Descriptors

The following Sub-TLVs are allowed to be used as Peer Descriptors:

Sub-TLV Code Point	Description	Length
512	Peer Autonomous System	4
513	BGP-LS Identifier	4

Peer Descriptors Sub-TLVs are defined in [I-D.ietf-idr-ls-distribution].

4.2. Peer Attributes

The Peer Attributes Sub-TLVs codepoints and lengths are listed in the following table:

TLV Code Point	Description	Length	IS-IS SR TLV /sub-TLV
1099	Adjacency Segment Identifier (Adj-SID)	variable	31 (section 2.3.1)
1100	LAN Adjacency Segment Identifier (Adj-LAN SID)	variable	32 (section 2.3.2)
TBA	Peer Set SID	variable	31 (section 2.3.1)

Sections refer to [I-D.ietf-idr-ls-distribution].

The value of the Adj-SID, Adj-LAN-SID and Peer Set SID Sub-TLV SHOULD be persistent across router restart.

5. Definition of PeerNode and PeerAdj

In this section the following Peer Segments are defined:

PeerNode Segment (PeerNodeSID)

PeerAdj Segment (PeerAdjSID)

PeerSet Segment (PeerSetSID)

5.1. PeerNode Segment (PeerNodeSID)

A BGP PeerNode segment/SID is a local segment. At the BGP node advertising it, its semantics is:

- SR header operation: NEXT (as defined in [I-D.filsfils-spring-segment-routing]).
- Next-Hop: the connected peering node to which the segment is related.

The PeerNode is advertised with a Peering Segment NLRI, where:

- Local Node Descriptor is the IGP node describing the EPE enabled egress PE.

- Peer Descriptor is the ASN of the peer.
- Link Descriptors, as defined in [I-D.ietf-idr-ls-distribution] contain the addresses used by the BGP session:
  - IPv4 Interface Address (Sub-TLV 259) contains the BGP session IPv4 local address.
  - IPv4 Neighbor Address (Sub-TLV 260) contains the BGP session IPv4 peer address.
  - IPv6 Interface Address (Sub-TLV 261) contains the BGP session IPv6 local address.
  - IPv6 Neighbor Address (Sub-TLV 262) contains the BGP session IPv6 peer address.
- Peer Attribute contains the Adj-SID TLV

5.2. PeerAdj Segment (PeerAdjSID)

A BGP PeerAdj segment/SID is a local segment. At the BGP node advertising it, its semantics is:

- SR header operation: NEXT (as defined in [I-D.filsfils-spring-segment-routing]).
- Next-Hop: the peer connected through the interface to which the segment is related.

The PeerAdj is advertised with a Peering Segment NLRI, where:

- Local Node Descriptor is the IGP node describing the EPE enabled egress PE.
- Peer Descriptor is the ip address and ASN of the peer.
- Link Descriptors, as defined in [I-D.ietf-idr-ls-distribution]contain the addresses used by the BGP session:
  - IPv4 Interface Address (Sub-TLV 259) contains the BGP session IPv4 local address.
  - IPv4 Neighbor Address (Sub-TLV 260) contains the BGP session IPv4 peer address.

    * IPv6 Interface Address (Sub-TLV 261) contains the BGP session IPv6 local address.

    * IPv6 Neighbor Address (Sub-TLV 262) contains the BGP session IPv6 peer address.

  o Peer Attribute contains the Adj-SID TLV

In addition, BGPLS Link Attributes, as defined in [I-D.ietf-idr-ls-distribution]MAY be inserted in order to advertise the characteristics of the link.

5.3. PeerSet Segment (PeerSetSID)

A PeerSet segment/SID is a local segment. At the BGP node advertising it, its semantics is:

- SR header operation: NEXT (as defined in [I-D.filsfils-spring-segment-routing]).
- Next-Hop: loadbalance across any connected interface to any peer in the related set.

The PeerSet is advertised in a Peering Segment NLRI (PeerNode or PeerAdj) as a BGPLS attribute.

The PeerSet Attribute contains an Adj-SID TLV, defined in Section 4.2 identifying the Set the PeerNode or PeerAdj is part of.

6. Illustration

6.1. Reference Diagram

The following reference diagram is used throughout this document. The solution is described for IPv4 with MPLS-based segments.

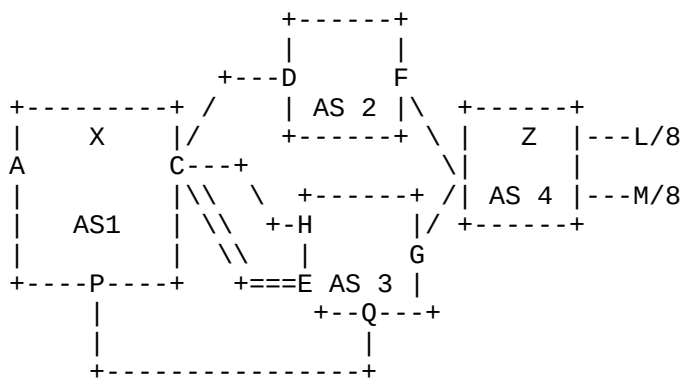


Figure 1: Reference Diagram

IPv4 addressing:

- C's interface to D: 1.0.1.1/24, D's interface: 1.0.1.2/24
- C's interface to H: 1.0.2.1/24, H's interface: 1.0.2.2/24
- C's upper interface to E: 1.0.3.1/24, E's interface: 1.0.3.2/24
- C's lower interface to E: 1.0.4.1/24, E's interface: 1.0.4.2/24
- Loopback of E used for eBGP multi-hop peering to C: 1.0.5.2/32
- C's loopback is 3.3.3.3/32 with SID 64

C's BGP peering:

- Single-hop eBGP peering with neighbor 1.0.1.2 (D)
- Single-hop eBGP peering with neighbor 1.0.2.2 (H)
- Multi-hop eBGP peering with E on ip address 1.0.5.2 (E)

C's resolution of the multi-hop eBGP session to E:

- Static route 1.0.5.2/32 via 1.0.3.2
- Static route 1.0.5.2/32 via 1.0.4.2

Node C configuration is such that:

- A PeerNode segment is allocated to each peer (D, H and E).

- A PeerAdj segment is defined for each recursing interface to a multi-hop peer (CE upper and lower interfaces).
- A PeerSet is defined to include all peers in AS3 (peers H and E).

6.1.1. PeerNode for Node D

Descriptors:

- Node Descriptors (router-ID, ASN): 3.3.3.3 , AS1
- Peer Descriptors (peer ASN): AS2
- Link Descriptors (IPv4 interface address, neighbor IPv4 address): 1.0.1.1, 1.0.1.2

Attributes:

- Adj-SID: 1012

6.1.2. PeerNode for Node H

Descriptors:

- Node Descriptors (router-ID, ASN): 3.3.3.3 , AS1
- Peer Descriptors (peer ASN): AS3
- Link Descriptors (IPv4 interface address, neighbor IPv4 address): 1.0.2.1, 1.0.2.2

Attributes:

- Adj-SID: 1022
- PeerSetSID: 1060
- Link Attributes: see section 3.3.2 of [I-D.ietf-idr-ls-distribution]

6.1.3. PeerNode for Node E

Descriptors:

- Node Descriptors (router-ID, ASN): 3.3.3.3 , AS1
- Peer Descriptors (peer ASN): AS3

- Link Descriptors (IPv4 interface address, neighbor IPv4 address): 3.3.3.3, 1.0.5.2

Attributes:

- Adj-SID: 1052
- PeerSetSID: 1060

6.1.4. PeerAdj for Node E, Link 1

Descriptors:

- Node Descriptors (router-ID, ASN): 3.3.3.3 , AS1
- Peer Descriptors (peer ASN): AS3
- Link Descriptors (IPv4 interface address, neighbor IPv4 address): 1.0.3.1 , 1.0.3.2

Attributes:

- Adj-SID: 1032
- LinkAttributes: see section 3.3.2 of [I-D.ietf-idr-ls-distribution]

6.1.5. PeerAdj for Node E, Link 2

Descriptors:

- Node Descriptors (router-ID, ASN): 3.3.3.3 , AS1
- Peer Descriptors (peer ASN): AS3
- Link Descriptors (IPv4 interface address, neighbor IPv4 address): 1.0.4.1 , 1.0.4.2

Attributes:

- Adj-SID: 1042
- LinkAttributes: see section 3.3.2 of [I-D.ietf-idr-ls-distribution]

7. IANA Considerations

This document defines a new BGPLS NLRI TYPE known as the Peer NLRI Type and a new BGP attribute known as the Peer Set SID TLV. The code points are to be assigned by IANA.

8. Manageability Considerations

TBD

9. Security Considerations

TBD

10. Acknowledgements

TBD

11. References

11.1. Normative References

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

11.2. Informative References

[I-D.filsfils-spring-segment-routing]
Filsfils, C., Previdi, S., Bashandy, A., Decraene, B., Litkowski, S., Horneffer, M., Milojevic, I., Shakir, R., Ytti, S., Henderickx, W., Tantsura, J., and E. Crabbe, "Segment Routing Architecture", draft-filsfils-spring-segment-routing-01 (work in progress), May 2014.

[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-01 (work in progress), May 2014.

[I-D.ietf-idr-ls-distribution]
Gredler, H., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and TE Information using BGP", draft-ietf-idr-ls-distribution-05 (work in progress), May 2014.

Authors' Addresses

Stefano Previdi (editor)
Cisco Systems, Inc.
Via Del Serafico, 200
Rome 00142
Italy

Email: sprevidi@cisco.com

Clarence Filsfils
Cisco Systems, Inc.
Brussels
BE

Email: cfilsfil@cisco.com

Saikat Ray
Cisco Systems, Inc.
170, West Tasman Drive
San Jose, CA 95134
US

Email: sairay@cisco.com

Keyur Patel
Cisco Systems, Inc.
170, West Tasman Drive
San Jose, CA 95134
US

Email: keyupate@cisco.com