Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2015

S. Previdi, Ed.
C. Filsfils
S. Ray
K. Patel
Cisco Systems, Inc.
J. Dong
M. Chen
Huawei Technologies
October 25, 2014

Segment Routing Egress Peer Engineering BGPLS Extensions
draft-previdi-idr-bgpls-segment-routing-epe-01

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/.

Copyright Notice

Table of Contents

1. Introduction 3
2. Segment Routing Documents 3
3. BGP Peering Segments 3
4. Peer NLRI-Type 4
4.1. BGP Router ID and Member ASN 4
4.2. EPE Node Descriptors 5
4.3. Peer Attributes 6
5. Peer Node and Peer Adjacency Segments 6
5.1. Peer Node Segment (PeerNode SID) 7
5.2. Peer Adjacency Segment (PeerAdj SID) 8
5.3. Peer Set Segment (PeerSet SID) 9
6. Illustration 9
6.1. Reference Diagram 9
6.1.1. PeerNode for Node D 10
6.1.2. PeerNode for Node H 11
6.1.3. PeerNode for Node E 11
6.1.4. PeerAdj for Node E, Link 1 12
6.1.5. PeerAdj for Node E, Link 2 12
7. BGPLS-EPE TLV/Sub-TLV Code Points Summary 12
8. IANA Considerations 13
9. Manageability Considerations 13
10. Security Considerations 13
11. Acknowledgements 14
12. References 14

12.1. Normative References 14
12.2. Informative References 14
Authors' Addresses 15

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.

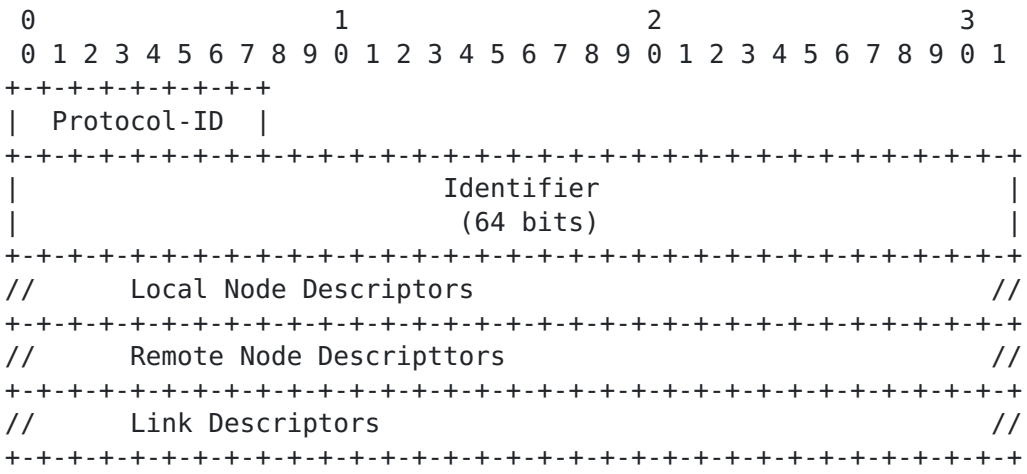
This document defines the BGP EPE Peering Segments: Peer Node, Peer Adjacency and Peer Set.

4. Peer NLRI-Type

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

The use of a new NLRI-Type allows to separate and differentiate the NLRIs carrying BGP-EPE descriptors from the NLRIs carrying IGP link-state information as defined in [I-D.ietf-idr-ls-distribution]. The Peer NLRI Type uses descriptors and attributes already defined in [I-D.ietf-idr-ls-distribution] in addition to new TLVs defined in the following sections of this document.

The format of the Peer NLRI Type is as follows:



Node Descriptors and Link Descriptors are defined in [I-D.ietf-idr-ls-distribution].

4.1. BGP Router ID and Member ASN

Two new Node Descriptors Sub-TLVs are defined in this document:

- BGP Router Identifier (BGP Router-ID):

  Type: TBA (suggested value 516).

  Length: 4 octets

  Value: 4 octet unsigned integer representing the BGP Identifier as defined in [RFC4271] and [RFC6286].

- Confederation Member ASN (Member-ASN)

Type: TBA (suggested value 517).

Length: 4 octets

Value: 4 octet unsigned integer representing the Member ASN inside the Confederation.[RFC5065].

4.2. EPE Node Descriptors

The following Node Descriptors Sub-TLVs MUST appear in the Peer NLRI Type as Local Node Descriptors:

- BGP Router ID, which contains the BGP Identifier of the local BGP EPE node.
- Autonomous System Number, which contains the local ASN or local confederation identifier (ASN) if confederations are used.
- BGP-LS Identifier.

The following Node Descriptors Sub-TLVs MAY appear in the Peer NLRI Type as Local Node Descriptors:

- Member-ASN, which contains the ASN of the confederation member (when BGP confederations are used).
- Node Descriptors as defined in [I-D.ietf-idr-ls-distribution].

The following Node Descriptors Sub-TLVs MUST appear in the Peer NLRI Type as Remote Node Descriptors:

- BGP Router ID, which contains the BGP Identifier of the peer node.
- Autonomous System Number, which contains the peer ASN or the peer confederation identifier (ASN), if confederations are used.

The following Node Descriptors Sub-TLVs MAY appear in the Peer NLRI Type as Remote Node Descriptors:

- Member-ASN, which contains the ASN of the confederation member (when BGP confederations are used).
- Node Descriptors as defined in defined in [I-D.ietf-idr-ls-distribution].

4.3. Peer Attributes

The following BGPLS Attributes TLVs are used in the Peer NLRI Type:

TLV Code Point	Description	Length
1099	Adjacency Segment Identifier (Adj-SID)	variable
1100	LAN Adjacency Segment Identifier (Adj-LAN SID)	variable
TBA	Peer Set SID	variable

Adj-SID and Adj-LAN-SID are defined in [I-D.gredler-idr-bgp-ls-segment-routing-extension].

Peer Set SID is a new attribute:

Type: TBD (suggested value 1036)
Length: variable
Value: the SID representing the group the peer is part of

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

The Adj-SID, Adj-LAN-SID and Peer Set SID SubTLVs MUST be present when BGPLS is used for the use case described in [I-D.filsfils-spring-segment-routing-central-epe] and MAY be omitted for other use cases.

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

5. Peer Node and Peer Adjacency Segments

In this section the following Peer Segments are defined:

- Peer Node Segment (Peer Node SID)
- Peer Adjacency Segment (Peer Adj SID)
- Peer Set Segment (Peer Set SID)

5.1. Peer Node Segment (PeerNode SID)

The PeerNode 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 SID is advertised with a Peer Type NLRI, where:

- Local Node Descriptors contains

  Local BGP Router ID of the EPE enabled egress PE.
  Local ASN.
  BGPLS Identifier.

- Remote Node Descriptors contains

  Peer BGP Router ID (i.e.: the peer BGP ID used in the BGP session).
  Peer ASN.

- Link Descriptors Sub-TLVs, 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 as defined in Section 4.3.
- 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.

5.2. Peer Adjacency Segment (PeerAdj SID)

The PeerAdj 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 address of the peer connected through the interface to which the segment is related.

The PeerAdj SID is advertised with a Peer Type NLRI, where:

- Local Node Descriptors contains

  Local BGP Router ID of the EPE enabled egress PE.
  Local ASN.
  BGPLS Identifier.

- Remote Node Descriptors contains

  Peer BGP Router ID (i.e.: the peer BGP ID used in the BGP session).
  Peer ASN.

- Link Descriptors Sub-TLVs, 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 as defined in Section 4.3.

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. Peer Set Segment (PeerSet SID)

The PeerSet 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 SID is advertised within a Peer Type NLRI (describing a PeerNode or PeerAdj) as a BGPLS attribute.

The PeerSet Attribute contains an Adj-SID TLV, defined in Section 4.3 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.

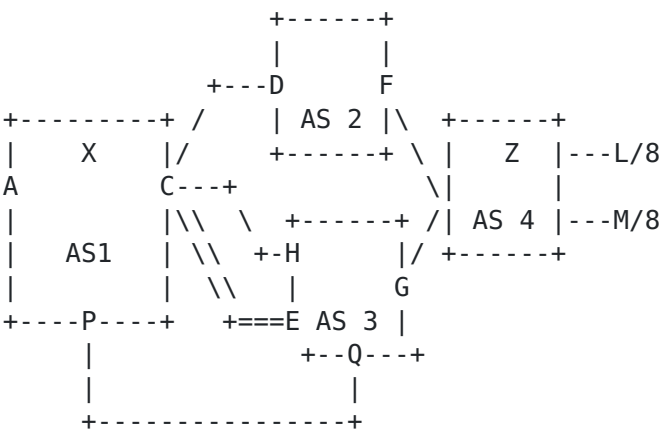


Figure 1: Reference Diagram

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

BGP Router-IDs are C, D, H and E.

- C's BGP Router-ID: 3.3.3.3
- D's BGP Router-ID: 4.4.4.4
- E's BGP Router-ID: 5.5.5.5
- H's BGP Router-ID: 6.6.6.6

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).

Local BGPLS Identifier in router C is set to 10000.

6.1.1. PeerNode for Node D

Descriptors:

- Local Node Descriptors (BGP Router-ID, local ASN, BGPLS Identifier): 3.3.3.3 , AS1, 10000

- Remote Node Descriptors (BGP Router-ID, peer ASN): 4.4.4.4, AS2
- Link Descriptors (IPv4 interface address, neighbor IPv4 address): 1.0.1.1, 1.0.1.2

Attributes:

- Adj-SID: 1012
- Link Attributes: see section 3.3.2 of [I-D.ietf-idr-ls-distribution]

6.1.2. PeerNode for Node H

Descriptors:

- Local Node Descriptors (BGP Router-ID, ASN, BGPL Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID ASN): 6.6.6.6, 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:

- Local Node Descriptors (BGP Router-ID, ASN, BGPLS Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID, ASN): 5.5.5.5, 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:

- Local Node Descriptors (BGP Router-ID, ASN, BGPLS Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID, ASN): 5.5.5.5, 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:

- Local Node Descriptors (BGP Router-ID, ASN, BGPLS Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID, ASN): 5.5.5.5, 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. BGPLS-EPE TLV/Sub-TLV Code Points Summary

The following table contains the TLVs/Sub-TLVs defined in this document.

Suggested Code Point	Description	Defined in:
5	Peer NLRI Type	Section 4
516	BGP Router ID	Section 4.1
517	BGP Confederation Member	Section 4.1
1036	Peer Set SID	Section 4.3

Table 1: Summary Table of BGPLS-EPE TLV/Sub-TLV Codepoints

8. IANA Considerations

This document defines:

Two new Node Descriptors Sub-TLVs: BGP-Router-ID and BGP Confederation Member.

A New NLRI Type for EPE: Peer Type NLRI.

A new BGPLS Attribute Sub-TLV: the Peer Set SID.

The code points are to be assigned by IANA.

9. Manageability Considerations

TBD

10. Security Considerations

[I-D.ietf-idr-ls-distribution] defines BGPLS NLRIs to which the extensions defined in this document apply.

The Security Section of [I-D.ietf-idr-ls-distribution] also applies to the:

new NLRI-Type: Peer NLRI;

new Node Descriptors Sub-TLVs: BGP-Router ID and BGP Confederation Member;

Peer NLRI-Type;

Peer Set SID attribute

defined in this document.

11. Acknowledgements

TBD

12. References

12.1. Normative References

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

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

[RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, August 2007.

[RFC6286] Chen, E. and J. Yuan, "Autonomous-System-Wide Unique BGP Identifier for BGP-4", RFC 6286, June 2011.

12.2. 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-04 (work in progress), July 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-02 (work in progress), July 2014.

[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-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-06 (work in progress), September 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

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: jie.dong@huawei.com

Mach (Guoyi) Chen
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: mach.chen@huawei.com