Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 19, 2015

S. Previdi, Ed.
C. Filsfils
Cisco Systems, Inc.
S. Ray
Individual Contributor
K. Patel
Cisco Systems, Inc.
J. Dong
M. Chen
Huawei Technologies
June 17, 2015

Segment Routing Egress Peer Engineering BGP-LS Extensions
draft-ietf-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 BGP-LS 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 December 19, 2015.

Table of Contents

1. Introduction 3
2. Segment Routing Documents 3
3. BGP Peering Segments 3
4. Link NLRI for EPE Connectivity Description 4
4.1. BGP Router ID and Member ASN 5
4.2. EPE Node Descriptors 5
4.3. Link Attributes 6
5. Peer Node and Peer Adjacency Segments 8
5.1. Peer Node Segment 8
5.2. Peer Adjacency Segment 9
5.3. Peer Set Segment 10
6. Illustration 10
6.1. Reference Diagram 10
6.1.1. Peer Node Segment for Node D 12
6.1.2. Peer Node Segment for Node H 13
6.1.3. Peer Node Segment for Node E 13
6.1.4. Peer Adj Segment for Node E, Link 1 13
6.1.5. Peer Adj Segment for Node E, Link 2 14
7. BGP-LS EPE TLV/Sub-TLV Code Points Summary 14
8. IANA Considerations 15
9. Manageability Considerations 15

10. Security Considerations 15
11. Contributors 15
12. Acknowledgements 15
13. References 15
13.1. Normative References 16
13.2. Informative References 16
Authors' Addresses 17

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 BGP-LS 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.

This document defines new types of segments: a Peer Node segment describing the BGP session between two nodes; a Peer Adjacency Segment describing the link (one or more) that is used by the BGP session; the Peer Set Segment describing an arbitrary set of sessions or links between the local BGP node and its peers.

2. 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].

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.

Each BGP session MUST be described by a Peer Node Segment. The description of the BGP session MAY be augmented by additional Adjacency Segments. Finally, each Peer Node Segment and Peer Adjacency Segment MAY be part of the same group/set so to be able to group EPE resources under a common Peer-Set Segment Identifier (SID).

Therefore, when the extensions defined in this document are applied to the use case defined in [I-D.filsfils-spring-segment-routing-central-epe]:

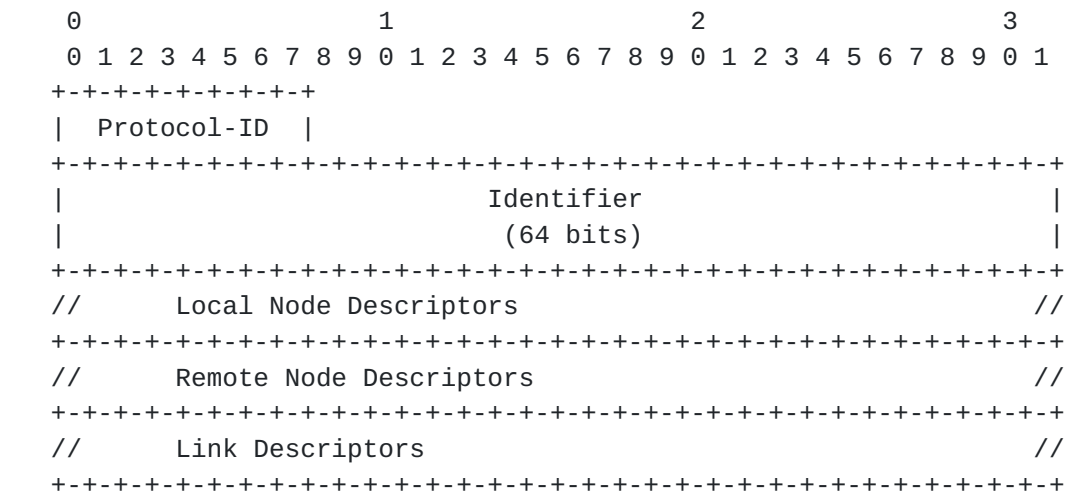
- One Peer Node Segment MUST be present.
- One or more Peer Adjacency Segments MAY be present.
- Each of the Peer Node and Peer Adjacency Segment MAY use the same Peer-Set.

4. Link NLRI for EPE Connectivity Description

This section describes the NLRI used for describing the connectivity of the BGP Egress router. The connectivity is based on links and remote peers/ASs and therefore the existing Link-Type NLRI (defined in [I-D.ietf-idr-ls-distribution]) is used. A new Protocol ID is used (codepoint to be assigned by IANA, suggested value 7).

The use of a new Protocol-ID allows separation and differentiation between the NLRIs carrying BGP-EPE descriptors from the NLRIs carrying IGP link-state information as defined in[I-D.ietf-idr-ls-distribution]. The Link 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 Link 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 Link NLRI 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.

It has to be noted that [RFC6286] (section 2.1) requires the BGP identifier (router-id) to be unique within an Autonomous System. Therefore, the <ASN, BGP identifier> tuple is globally unique.

The following Node Descriptors Sub-TLVs MAY appear in the Link NLRI 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 Link NLRI 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 Link NLRI 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. Link Attributes

The following BGP-LS Link attributes TLVs are used with the Link NLRI:

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

Adj-SID is defined in [I-D.gredler-idr-bgp-ls-segment-routing-extension] and the same format is used for the Peer-SID and Peer-Set-SID TLVs.

Peer-SID and Peer-Set SID are two new sub-TLVs with the same format as the Adj-SID and whose codepoints are to be assigned by IANA:

Peer-SID: SID representing the peer of the BGP session. The format is the same as defined for the Adj-SID in [I-D.gredler-idr-bgp-ls-segment-routing-extension]. Suggested codepoint value: 1036

Peer-Set-SID: the SID representing the group the peer is part of. The format is the same as defined for the Adj-SID in [I-D.gredler-idr-bgp-ls-segment-routing-extension]. Suggested codepoint value: 1037

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

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

The Adj-SID and Peer-Set-SID SubTLVs MAY be present when BGP-LS 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, BGP-LS 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

The Peer Node Segment describes the BGP session peer (neighbor). It MUST be present when describing an EPE topology as defined in [I-D.filsfils-spring-segment-routing-central-epe]. The Peer Node Segment is encoded within the BGP-LS Link NLRI specified in Section 4.

The Peer Node Segment is a local segment. At the BGP node advertising it, its semantic is:

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

The Peer Node Segment is advertised with a Link NLRI, where:

- Local Node Descriptors contains

  Local BGP Router ID of the EPE enabled egress PE.
  Local ASN.
  BGP-LS 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.

- Link Attribute contains the Peer-SID TLV as defined in Section 4.3.

- In addition, BGP-LS 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

The Peer Adjacency Segment is a local segment. At the BGP node advertising it, its semantic is:

- SR header operation: NEXT (as defined in [I-D.ietf-spring-segment-routing]).

- Next-Hop: the interface peer address.

The Peer Adjacency Segment is advertised with a Link NLRI, where:

- Local Node Descriptors contains

  Local BGP Router ID of the EPE enabled egress PE.
  Local ASN.
  BGP-LS 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:

  * Link Local/Remote Identifiers (Sub-TLV 258) contains the 4-octet Link Local Identifier followed by the 4-octet value 0 indicating the Link Remote Identifier in unknown [RFC5307].

  * IPv4 Neighbor Address (Sub-TLV 260) contains the IPv4 address of the peer interface used by the BGP session.

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

- Link attribute used with the Peer Adjacency SID contains the Adj-SID TLV as defined in Section 4.3.

In addition, BGP-LS 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

The Peer Set Segment is a local segment. At the BGP node advertising it, its semantic is:

- SR header operation: NEXT (as defined in [I-D.ietf-spring-segment-routing]).

- Next-Hop: load balance across any connected interface to any peer in the related set.

The Peer Set Segment is advertised within a Link NLRI (describing a Peer Node Segment or a Peer Adjacency segment) as a BGP-LS attribute.

The Peer Set Attribute contains the Peer-Set-SID TLV, defined in Section 4.3 identifying the set of which the Peer Node Segment or Peer Adjacency Segment is a member.

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.

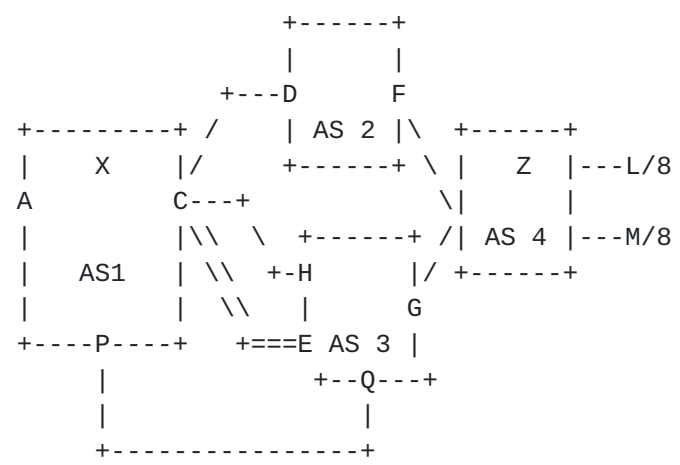


Figure 1: Reference Diagram

IPv4 addressing:

- C's IPv4 address of interface to D: 1.0.1.1/24, D's interface: 1.0.1.2/24
- C's IPv4 address of interface to H: 1.0.2.1/24, H's interface: 1.0.2.2/24
- C's IPv4 address of upper interface to E: 1.0.3.1, E's interface: 1.0.3.2/24
- C's local identifier of upper interface to E: 0.0.0.1.0.0.0.0
- C's IPv4 address of lower interface to E: 1.0.4.1/24, E's interface: 1.0.4.2/24
- C's local identifier of lower interface to E: 0.0.0.2.0.0.0.0
- 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 Peer Node segment is allocated to each peer (D, H and E).
- An Adjacency segment is defined for each recursing interface to a multi-hop peer (CE upper and lower interfaces).
- A Peer Set segment is defined to include all peers in AS3 (peers H and E).

Local BGP-LS Identifier in router C is set to 10000.

The Link NLRI Type is used in order to encode C's connectivity. the Link NLRI uses the new Protocol-ID value (to be assigned by IANA).

6.1.1. Peer Node Segment for Node D

Descriptors:

- Local Node Descriptors (BGP Router-ID, local ASN, BGP-LS Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID, peer ASN): 4.4.4.4, AS2
- Link Descriptors (BGP session IPv4 local address, BGP session IPv4 neighbor address): 1.0.1.1, 1.0.1.2

Attributes:

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

6.1.2. Peer Node Segment 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 (BGP session IPv4 local address, BGP session IPv4 peer address): 1.0.2.1, 1.0.2.2

Attributes:

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

6.1.3. Peer Node Segment for Node E

Descriptors:

- Local Node Descriptors (BGP Router-ID, ASN, BGP-LS Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID, ASN): 5.5.5.5, AS3
- Link Descriptors (BGP session IPv4 local address, BGP session IPv4 peer address): 3.3.3.3, 1.0.5.2

Attributes:

- Peer-SID: 1052
- Peer-Set-SID: 1060

6.1.4. Peer Adj Segment for Node E, Link 1

Descriptors:

- Local Node Descriptors (BGP Router-ID, ASN, BGP-LS Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID, ASN): 5.5.5.5, AS3

- Link Descriptors (local interface identifier, IPv4 peer interface address): 0.0.0.1.0.0.0.0 , 1.0.3.2

Attributes:

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

6.1.5. Peer Adj Segment for Node E, Link 2

Descriptors:

- Local Node Descriptors (BGP Router-ID, ASN, BGP-LS Identifier): 3.3.3.3 , AS1, 10000
- Remote Node Descriptors (BGP Router-ID, ASN): 5.5.5.5, AS3
- Link Descriptors (local interface identifier, IPv4 peer interface address): 0.0.0.2.0.0.0.0 , 1.0.4.2

Attributes:

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

7. BGP-LS EPE TLV/Sub-TLV Code Points Summary

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

Suggested Codepoint	Description	Defined in:
7	Protocol-ID	Section 4
516	BGP Router ID	Section 4.1
517	BGP Confederation Member	Section 4.1
1036	Peer-SID	Section 4.3
1037	Peer-Set-SID	Section 4.3

Table 1: Summary Table of BGP-LS EPE Codepoints

8. IANA Considerations

This document defines:

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

- A new Protocol-ID for EPE: BGP-EPE.

- Two new BGP-LS Attribute Sub-TLVs: the Peer-SID and the Peer-Set-SID.

The codepoints are to be assigned by IANA.

9. Manageability Considerations

TBD

10. Security Considerations

[I-D.ietf-idr-ls-distribution] defines BGP-LS 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 Node Descriptors Sub-TLVs: BGP-Router ID and BGP Confederation Member;

- Peer-SID and Peer-Set-SID attributes

defined in this document.

11. Contributors

Acee Lindem gave a substantial contribution to this document.

12. Acknowledgements

The authors would like to thank Jakob Heitz, Howard Yang and Hannes Gredler for their feedback and comments.

13. References

13.1. 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.

[RFC5307] Kompella, K. and Y. Rekhter, "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.

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

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.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-11 (work in progress), June 2015.

[I-D.ietf-spring-segment-routing]
Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-03 (work in progress), May 2015.

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
Individual Contributor

Email: raysaikat@gmail.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