Routing area
Internet-Draft
Intended status: Standards Track
Expires: August 27, 2020

S. Hegde
K. Arora
M. Srivastava
Juniper Networks Inc.
S. Ninan
Individual Contributor
X. Xu
Alibaba Inc.

February 24, 2020

Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) Egress Peer Engineering Segment Identifiers (SIDs) with MPLS Data Planes

draft-hegde-mpls-spring-epe-oam-05

Abstract

Egress Peer Engineering (EPE) is an application of Segment Routing to Solve the problem of egress peer selection. The Segment Routing based BGP-EPE solution allows a centralized controller, e.g. a Software Defined Network (SDN) controller to program any egress peer. The EPE solution requires a node to program the PeerNode SID describing a session between two nodes, the PeerAdj SID describing the link (one or more) that is used by sessions between peer nodes, and the PeerSet SID describing an arbitrary set of sessions or links between a local node and its peers. This document provides new sub-TLVs for EPE Segment Identifiers (SID) that would be used in the MPLS Target stack TLV (Type 1), in MPLS Ping and Traceroute procedures.

Table of Contents

1. Introduction

Egress Peer Engineering (EPE) as defined in [I-D.ietf-spring-segment-routing-central-epe] is an effective mechanism to select the egress peer link based on different criteria. The EPE-SIDs provide means to represent egress peer links. Many network deployments have built their networks consisting of multiple Autonomous Systems either for ease of operations or as a result of network mergers and acquisitons. The inter-AS links connecting the two Autonomous Systems could be traffic engineered using EPE-SIDs in this case as well. It is important to be able to validate the control plane to forwarding plane synchronization for these SIDs so that any anomaly can be detected easily by the operator.

This document provides Target Forwarding Equivalence Class (FEC) stack TLV definitions for EPE-SIDs. Other procedures for mpls Ping

and Traceroute as defined in [RFC8287] section 7 and clarified by [RFC8690] are applicable for EPE-SIDs as well.

2. Theory of Operation

[I-D.ietf-idr-bgpls-segment-routing-epe] provides mechanisms to advertise the EPE-SIDs in BGP-LS. These EPE-SIDs may be used to build Segment Routing paths as described in [I-D.ietf-spring-segment-routing-policy]. Data plane monitoring for such paths which consist of EPE-SIDs will use extensions defined in this document to build the Taget FEC stack TLV. The MPLS Ping and Traceroute procedures MAY be initaited by the head-end of the Segment Routing path or a centralized topology-aware data plane monitoring system as described in [RFC8403]. The node initiating the data plane monitoring may acquire the details of EPE-SIDs through BGP-LS advertisements as described in [I-D.ietf-idr-bgpls-segment-routing-epe]. The procedures to operate e-BGP sessions in a scenario with unnumbered interfaces is not very well defined and hence out of scope for this document. During AS migration scenario procedures described in [RFC7705] may be in force. In these scenarios, if the local and remote AS fields in the FEC as described in Section 4carries the global AS and not the "local AS" as defined in [RFC7705], the FEC validation procedures may fail.

3. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14, [RFC2119], [RFC8174] when, and only when, they appear in all capitals, as shown here.

4. FEC Definitions

As described in [RFC8287] sec 5, 3 new type of sub-TLVs for the Target FEC Stack TLV are defined for the Target FEC stack TLV corresponding to each label in the label stack. If a malformed FEC sub-TLV is received, then a return code of 1, "Malformed echo request received" as defined in [RFC8029] SHOULD be sent.

4.1. PeerAdj SID Sub-TLV

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Type = TBD                     |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Local AS Number (4  octets)                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote As Number (4 octets)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Local BGP router ID (4 octets)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote BGP Router ID (4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Local Interface address (4/16 octets)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote Interface address (4/16 octets)           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: PeerAdj SID Sub-TLV

Type : TBD

Length : variable based on ipv4/ipv6 interface address

Local AS Number :

4 octet unsigned integer representing the Member ASN inside the Confederation.[RFC5065]. The AS number corresponds to the AS to which PeerAdj SID advertising node belongs to.

Remote AS Number :

4 octet unsigned integer representing the Member ASN inside the Confederation.[RFC5065]. The AS number corresponds to the AS of the remote node for which the PeerAdj SID is advertised.

Local BGP Router ID :

4 octet unsigned integer of the advertising node representing the BGP Identifier as defined in [RFC4271] and [RFC6286].

Remote BGP Router ID :

4 octet unsigned integer of the receiving node representing the BGP Identifier as defined in [RFC4271] and [RFC6286].

Local Interface Address :

In case of PeerAdj SID Local interface address corresponding to the PeerAdj SID should be apecified in this field. For IPv4,this field is 4 octets; for IPv6, this field is 16 octets. Link Local IPv6 addresses are allowed in this field.

Remote Interface Address :

In case of PeerAdj SID Remote interface address corresponding to the PeerAdj SID should be apecified in this field. For IPv4,this field is 4 octets; for IPv6, this field is 16 octets.Link Local IPv6 addresses are allowed in this field.

4.2. PeerNode SID Sub-TLV

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Type = TBD                     |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Local AS Number (4  octets)                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote As Number (4 octets)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Local BGP router ID (4 octets)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote BGP Router ID (4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   No.of IPv4 interface pairs  |   No.of IPv6 interface pairs  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Local Interface address1 (4/16 octets)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote Interface address1 (4/16 octets)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Local Interface address2 (4/16 octets)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              ......                                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: PeerNode SID Sub-TLV

Type : TBD

Length : variable based on ipv4/ipv6 interface address. There could be multiple pairs of local and remote interface pairs. The length includes all the pairs.

Local AS Number :

4 octet unsigned integer representing the Member ASN inside the Confederation.[RFC5065]. The AS number corresponds to the AS to which PeerNode SID advertising node belongs to.

Remote AS Number :

4 octet unsigned integer representing the Member ASN inside the Confederation.[RFC5065]. The AS number corresponds to the AS of the remote node for which the PeerNode SID is advertised.

Local BGP Router ID :

4 octet unsigned integer of the advertising node representing the BGP Identifier as defined in [RFC4271] and [RFC6286].

Remote BGP Router ID :

4 octet unsigned integer of the receiving node representing the BGP Identifier as defined in [RFC4271] and [RFC6286].

Number of IPv4 interface pairs:

Total number of IPV4 local and remote interface address pairs.

Number of IPv6 interface pairs:

Total number of IPV6 local and remote interface address pairs.

There can be multiple Layer 3 interfaces on which a peerNode SID loadbalances the traffic. All such interfaces local/remote address MUST be included in the FEC.

When a PeerNode SID load-balances over few interfaces with IPv4 only address and few interfaces with IPv6 address then the FEC definition should list all IPv4 address pairs together followed by IPv6 address pairs.

Local Interface Address :

In case of PeerNode SID, the interface local address ipv4/ipv6 which corresponds to the PeerNode SID MUST be specified. For IPv4,this field is 4 octets; for IPv6, this field is 16 octets.Link Local IPv6 addresses are allowed in this field.

Remote Interface Address :

In case of PeerNode SID, the interface remote address ipv4/ipv6 which corresponds to the PeerNode SID MUST be specified. For IPv4,this field is 4 octets; for IPv6, this field is 16 octets.Link Local IPv6 addresses are allowed in this field.

4.3. PeerSet SID Sub-TLV

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Type = TBD                     |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Local AS Number (4  octets)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Local BGP router ID (4 octets)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   No.of elements in set       |          Reserved             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote As Number (4 octets)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote BGP Router ID (4 octets)                  |
++-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-++
|   No.of IPv4 interface pairs  |   No.of IPv6 interface pairs  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Local Interface address1 (4/16 octets)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote Interface address1 (4/16 octets)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Local Interface address2 (4/16 octets)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              ......                                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: PeerSet SID Sub-TLV

Type : TBD

Length : variable based on ipv4/ipv6 interface address and number of elements in the set

Local AS Number :

4 octet unsigned integer representing the Member ASN inside the Confederation.[RFC5065]. The AS number corresponds to the AS to which PeerSet SID advertising node belongs to.

Remote AS Number :

4 octet unsigned integer representing the Member ASN inside the Confederation.[RFC5065]. The AS number corresponds to the AS of the remote node for which the PeerSet SID is advertised.

Advertising BGP Router ID :

4 octet unsigned integer of the advertising node representing the BGP Identifier as defined in [RFC4271] and [RFC6286].

Receiving BGP Router ID :

4 octet unsigned integer of the receiving node representing the BGP Identifier as defined in [RFC4271] and [RFC6286].

No.of elements in set:

Number of remote ASes, the set SID load-balances on.

PeerSet SID may be associated with a number of PeerNode SIDs and PeerAdj SIDs. Link address details of all these SIDs should be included in the peerSet SID FEC so that the data-plane can be correctly verified on the remote node.

Number of IPv4 interface pairs:

Total number of IPV4 local and remote interface address pairs.

Number of IPv6 interface pairs:

Total number of IPV6 local and remote interface address pairs.

There can be multiple Layer 3 interfaces on which a peerNode SID loadbalances the traffic. All such interfaces local/remote address MUST be included in the FEC.

When a PeerSet SID load-balances over few interfaces with IPv4 only address and few interfaces with IPv6 address then the Link address

TLV should list all IPv4 address pairs together followed by IPv6 address pairs.

Local Interface Address :

In case of PeerNodeSID/PeerAdj SID, the interface local address ipv4/ ipv6 which corresponds to the PeerNode SID/PeerAdj SID MUST be specified. For IPv4,this field is 4 octets; for IPv6, this field is 16 octets. Link Local IPv6 addresses are allowed in this field.

Remote Interface Address :

In case of PeerNodeSID/PeerAdj SID, the interface remote address ipv4/ipv6 which corresponds to the PeerNode SID/PeerAdj SID MUST be specified. For IPv4,this field is 4 octets; for IPv6, this field is 16 octets. Link Local IPv6 addresses are allowed in this field.

5. IANA Considerations

New Target FEC stack sub-TLV from the "sub-TLVs for TLV types 1,16 and 21" subregistry of the "Multi-Protocol Label switching (MPLs) Label Switched Paths (LSPs) Ping parameters" registry

PeerAdj SID Sub-TLV : TBD

PeerNode SID Sub-TLV : TBD

PeerSet SID Sub-TLV : TBD

6. Security Considerations

The EPE-SIDs are advertised for egress links for Egress Peer Engineering purposes or for inter-As links between co-operating ASes. When co-operating domains are involved, they can allow the packets arriving on trusted interfaces to reach the control plane and get processed. When EPE-SIDs which are created for egress TE links where the neighbor AS is an independent entity, it may not allow packets arriving from external world to reach the control plane. In such deployments mpls OAM packets will be dropped by the neighboring AS that receives the MPLS OAM packet.

7. Acknowledgments

Thanks to Loa Andersson and Alexander Vainshtein for careful review and comments.

8. References

8.1. Normative References

[I-D.ietf-idr-bgpls-segment-routing-epe]
Previdi, S., Talaulikar, K., 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-19 (work in progress), May 2019.

[RFC8029] Kompella, K., Swallow, G., Pignataro, C., Ed., Kumar, N., Aldrin, S., and M. Chen, "Detecting Multiprotocol Label Switched (MPLS) Data-Plane Failures", RFC 8029, DOI 10.17487/RFC8029, March 2017, <https://www.rfc-editor.org/info/rfc8029>.

[RFC8287] Kumar, N., Ed., Pignataro, C., Ed., Swallow, G., Akiya, N., Kini, S., and M. Chen, "Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) IGP-Prefix and IGP-Adjacency Segment Identifiers (SIDs) with MPLS Data Planes", RFC 8287, DOI 10.17487/RFC8287, December 2017, <https://www.rfc-editor.org/info/rfc8287>.

8.2. Informative References

[I-D.ietf-spring-segment-routing-central-epe]
Filsfils, C., Previdi, S., Dawra, G., Aries, E., and D. Afanasiev, "Segment Routing Centralized BGP Egress Peer Engineering", draft-ietf-spring-segment-routing-central-epe-10 (work in progress), December 2017.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-06 (work in progress), December 2019.

[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>.

[RFC7705] George, W. and S. Amante, "Autonomous System Migration Mechanisms and Their Effects on the BGP AS_PATH Attribute", RFC 7705, DOI 10.17487/RFC7705, November 2015, <https://www.rfc-editor.org/info/rfc7705>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8403] Geib, R., Ed., Filsfils, C., Pignataro, C., Ed., and N. Kumar, "A Scalable and Topology-Aware MPLS Data-Plane Monitoring System", RFC 8403, DOI 10.17487/RFC8403, July 2018, <https://www.rfc-editor.org/info/rfc8403>.

[RFC8690] Nainar, N., Pignataro, C., Iqbal, F., and A. Vainshtein, "Clarification of Segment ID Sub-TLV Length for RFC 8287", RFC 8690, DOI 10.17487/RFC8690, December 2019, <https://www.rfc-editor.org/info/rfc8690>.

Authors' Addresses

Shraddha Hegde
Juniper Networks Inc.
Exora Business Park
Bangalore, KA 560103
India

Email: shraddha@juniper.net

Kapil Arora
Juniper Networks Inc.

Email: kapilaro@juniper.net

Mukul Srivastava
Juniper Networks Inc.

Email: msri@juniper.net

Samson Ninan
Individual Contributor

Email: samson.cse@gmail.com

Xiaohu Xu
Alibaba Inc.
Beijing
China

Email: xiaohu.xxh@alibaba-inc.com