Stream: Internet Engineering Task Force (IETF)
RFC: 9703
Category: Standards Track
Published: December 2024
ISSN: 2070-1721

Authors:
S. Hegde
Juniper Networks Inc.
M. Srivastava
Juniper Networks Inc.
K. Arora
Individual Contributor
S. Ninan
Ciena
X. Xu
China Mobile

RFC 9703
Label Switched Path (LSP) Ping/Traceroute for Segment Routing (SR) Egress Peer Engineering (EPE) Segment Identifiers (SIDs) with MPLS Data Plane

Abstract

Egress Peer Engineering (EPE) is an application of Segment Routing (SR) that solves the problem of egress peer selection. The SR-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 the node or the SDN controller to program 1) the PeerNode Segment Identifier (SID) describing a session between two nodes, 2) the PeerAdj SID describing the link or links that are used by the sessions between peer nodes, and 3) the PeerSet SID describing any connected interface to any peer in the related group. This document provides new sub-TLVs for EPE-SIDs that are used in the Target FEC Stack TLV (Type 1) in MPLS Ping and Traceroute procedures.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9703.

Copyright Notice

Table of Contents

1. Introduction

Egress Peer Engineering (EPE), as defined in [RFC9087], is an effective mechanism that is used to select the egress peer link based on different criteria. In this scenario, egress peers may belong to a completely different ownership. The EPE-SIDs provide the means to represent egress peer nodes, links, sets of links, and sets of nodes. Many network deployments have built their networks consisting of multiple Autonomous Systems (ASes) either for the ease of operations or as a result of network mergers and acquisitions. The inter-AS links connecting any two ASes could be traffic-engineered using EPE-SIDs in this case, where there is single ownership but different AS numbers. It is important to validate the control plane to forwarding plane synchronization for these SIDs so that any anomaly can be easily detected by the network operator. EPE-SIDs may also be used in an ingress Segment Routing (SR) policy [RFC9256] to choose exit points where the remote AS has a completely different ownership. This scenario is out of scope for this document.

```
     +---------+      +------+
     |         |      |      |
     |    H    B------D      G
     |         | +---/| AS2  |\  +------+
     |         |/     +------+ \ |      |---L/8
     A   AS1   C---+            \|      |
     |         |\\  \  +------+ /| AS4  |---M/8
     |         | \\  +-E      |/ +------+
     |    X    |  \\   |      K
     |         |   +===F AS3  |
     +---------+       +------+
```

Figure 1: Reference Diagram

In Figure 1, EPE-SIDs are configured on AS1 towards AS2 and AS3 and advertised in the Border Gateway Protocol - Link State (BGP-LS) [RFC9086]. In certain cases, the EPE-SIDs advertised by the control plane may not be in synchronization with the label programmed in the data plane. For example, on C, a PeerAdj SID could be advertised to indicate it is for the link C->D. Due to some software anomaly, the actual data forwarding on this PeerAdj SID could be happening over the C->E link. If E had relevant data paths for further forwarding the packet, this kind of anomaly would go unnoticed by the network operator. A detailed example of a correctly programmed state and an incorrectly programmed state along with a description of how the incorrect state can be detected is described in Appendix A. A Forwarding Equivalence Class (FEC) definition for the EPE-SIDs will detail the control plane association of the SID. The data plane validation of the SID will be done during the MPLS Traceroute procedure. When there is a multi-hop External BGP (EBGP) session between the ASBRs, a PeerNode SID is advertised, and the traffic **MAY** be load-balanced between the interfaces connecting the two nodes. In Figure 1, C and F could have a PeerNode SID advertised. When the Operations, Administration, and Maintenance (OAM) packet is received on F, it needs to be validated that the packet came from one of the two interfaces connected to C.

This document provides Target Forwarding Equivalence Class (FEC) Stack TLV definitions for EPE-SIDs. This solution requires the node constructing the Target FEC Stack TLV to determine the types of SIDs along the path of the LSP. Other procedures for MPLS Ping and Traceroute, as defined in Section 7 of [RFC8287] and clarified in [RFC8690], are applicable for EPE-SIDs as well.

2. Theory of Operation

[RFC9086] provides mechanisms to advertise the EPE-SIDs in BGP-LS. These EPE-SIDs may be used to build SR paths and may be communicated using extensions described in [SR-SEGTYPES] and [SR-BGP-POLICY] or Path Computation Element Protocol (PCEP) extensions as defined in [RFC8664]. Data plane monitoring for such paths that consist of EPE-SIDs will use extensions defined in this document to build the Target FEC Stack TLV. The MPLS Ping and Traceroute procedures **MAY** be initiated by the head-end of the SR path or a centralized topology-aware data plane monitoring system, as described in [RFC8403]. The extensions in [SR-SEGTYPES], [SR-BGP-POLICY], and [RFC8664] do not define how to acquire and carry the details of the SID that can be used to construct the FEC. Such extensions are out of scope for this document. The node initiating the data plane monitoring may acquire the details of EPE-SIDs through BGP-LS advertisements, as described in [RFC9086]. There may be other possible mechanisms that can be used to learn the definition of the SID from the controller. Details of such mechanisms are out of scope for this document.

The EPE-SIDs are advertised for inter-AS links that run EBGP sessions. [RFC9086] does not define the detailed procedures of how to operate EBGP sessions in a scenario with unnumbered interfaces. Therefore, these scenarios are out of scope for this document. Anycast and multicast addresses are not in the scope of this document. During the 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 4) carry the globally configured AS Number and not the "local AS" (as defined in [RFC7705]), then the FEC validation procedures may fail.

As described in Section 1, this document defines Target FEC Stack TLVs for EPE-SIDs that can be used in detecting MPLS data plane failures [RFC8029]. This mechanism applies to paths created across ASes of cooperating administrations. If the ping or traceroute packet enters a non-cooperating AS domain, it might be dropped by the routers in the non-cooperating domain. Although a complete path validation cannot be done across non-cooperating domains, it still provides useful information that the ping or traceroute packet entered a non-cooperating domain.

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

In this document, three new sub-TLVs are defined for the Target FEC Stack TLV (Type 1), the Reverse-Path Target FEC Stack TLV (Type 16), and the Reply Path TLV (Type 21); see Table 1.

4.1. 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 = 39                      |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Local AS Number (4 octets)                       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote AS Number (4 octets)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Local BGP Router ID (4 octets)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote BGP Router ID (4 octets)                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: PeerNode SID Sub-TLV

Type: 2 octets

Value: 39

Length: 2 octets

Value: 16

Local AS Number: 4 octets. The unsigned integer representing the AS number [RFC6793] of the AS to which the PeerNode SID advertising node belongs. If Confederations [RFC5065] are in use, and if the remote node is a member of a different Member-AS within the local Confederation, this is the Member-AS Number inside the Confederation and not the Confederation Identifier.

Remote AS Number: 4 octets. The unsigned integer representing the AS number [RFC6793] of the AS of the remote node for which the PeerNode SID is advertised. If Confederations [RFC5065] are in use, and if the remote node is a member of a different Member-AS within the local Confederation, this is the Member-AS Number inside the Confederation and not the Confederation Identifier.

Local BGP Router ID: 4 octets. The unsigned integer representing the BGP Identifier of the PeerNode SID advertising node as defined in [RFC4271] and [RFC6286].

Remote BGP Router ID: 4 octets. The unsigned integer representing the BGP Identifier of the remote node as defined in [RFC4271] and [RFC6286].

When there is a multi-hop EBGP session between two ASBRs, a PeerNode SID is advertised for this session, and traffic can be load-balanced across these interfaces. An EPE controller that performs bandwidth management for these links should be aware of the links on which the traffic will be load-balanced. As per [RFC8029], the node advertising the EPE-SIDs will send a Downstream Detailed Mapping (DDMAP) TLV specifying the details of the next-hop interfaces. Based on this information, the controller **MAY** choose to verify the actual forwarding state with the topology information that the controller has. On the router, the validation procedures will include the received DDMAP validation, as specified in [RFC8029], to verify the control state and the forwarding state synchronization on the two routers. Any discrepancies between the controller's state and the forwarding state will not be detected by the procedures described in this document.

4.2. 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 = 38                      |          Length               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Adj type      |            RESERVED                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              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 3: PeerAdj SID Sub-TLV

Type: 2 octets

Value: 38

Length: 2 octets

Value: Variable based on the IPv4/IPv6 interface address. Length excludes the length of the Type and Length fields. For IPv4 interface addresses, the length will be 28 octets. In the case of an IPv6 address, the length will be 52 octets.

Adj type: 1 octet

Value: Set to 1 when the Adjacency Segment is IPv4. Set to 2 when the Adjacency Segment is IPv6.

RESERVED: 3 octets. **MUST** be zero when sending and ignored on receiving.

Local AS Number: 4 octets. The unsigned integer representing the AS number [RFC6793] of the AS to which the PeerAdj SID advertising node belongs. If Confederations [RFC5065] are in use, and if the remote node is a member of a different Member-AS within the local Confederation, this is the Member-AS Number inside the Confederation and not the Confederation Identifier.

Remote AS Number: 4 octets. The unsigned integer representing the AS number [RFC6793] of the remote node's AS for which the PeerAdj SID is advertised. If Confederations [RFC5065] are in use, and if the remote node is a member of a different Member-AS within the local Confederation, this is the Member-AS Number inside the Confederation and not the Confederation Identifier.

Local BGP Router ID: 4 octets. The unsigned integer representing the BGP Identifier of the PeerAdj SID advertising node as defined in [RFC4271] and [RFC6286].

Remote BGP Router ID: 4 octets. The unsigned integer representing the BGP Identifier of the remote node as defined in [RFC4271] and [RFC6286].

Local Interface Address: 4 octets or 16 octets. In the case of PeerAdj SID, the local interface address corresponding to the PeerAdj SID should be specified in this field. For IPv4, this field is 4 octets; for IPv6, this field is 16 octets. Link-local IPv6 addresses are not in the scope of this document.

Remote Interface Address: 4 octets or 16 octets. In the case of PeerAdj SID, the remote interface address corresponding to the PeerAdj SID should be specified in this field. For IPv4, this field is 4 octets; for IPv6, this field is 16 octets. Link-local IPv6 addresses are not in the scope of this document.

[RFC9086] mandates sending a local interface ID and remote interface ID in the link descriptors and allows a value of 0 in the remote descriptors. It is useful to validate the incoming interface for an OAM packet, but if the remote descriptor is 0, this validation is not possible. Optional link descriptors of local and remote interface addresses are allowed as described in Section 4.2 of [RFC9086]. In this document, it is **RECOMMENDED** to send these optional descriptors and use them to validate incoming interfaces. When these local and remote interface addresses are not available, an ingress node can send 0 in the local and/or remote interface address field. The receiver **SHOULD** skip the validation for the incoming interface if the address field contains 0.

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 = 40                     |          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)                  |
++-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-++


 One element in set consists of the details below
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote AS Number (4 octets)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Remote BGP Router ID (4 octets)                  |
++-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-++
```

Figure 4: PeerSet SID Sub-TLV

Type: 2 octets

Value: 40

Length: 2 octets

Value: Expressed in octets and is a variable based on the number of elements in the set. The length field does not include the length of Type and Length fields.

Local AS Number: 4 octets. The unsigned integer representing the AS number [RFC6793] of the AS to which the PeerSet SID advertising node belongs. If Confederations [RFC5065] are in use, and if the remote node is a member of a different Member-AS within the local Confederation, this is the Member-AS Number inside the Confederation and not the Confederation Identifier.

Local BGP Router ID: 4 octets. The unsigned integer representing the BGP Identifier of the PeerSet SID advertising node, as defined in [RFC4271] and [RFC6286].

No. of elements in set: 2 octets. The number of remote ASes over which the set SID performs load-balancing.

Reserved: 2 octets. **MUST** be zero when sent and ignored when received.

Remote AS Number: 4 octets. The unsigned integer representing the AS number [RFC6793] of the remote node's AS for which the PeerSet SID is advertised. If Confederations [RFC5065] are in use, and if the remote node is a member of a different Member-AS within the local Confederation, this is the Member-AS Number inside the Confederation and not the Confederation Identifier.

Remote BGP Router ID: 4 octets. The unsigned integer representing the BGP Identifier of the remote node as defined in [RFC4271] and [RFC6286].

PeerSet SID may be associated with a number of PeerNode SIDs and PeerAdj SIDs. The remote AS number and the Router ID of each of these PeerNode SIDs and PeerAdj SIDs **MUST** be included in the FEC.

5. EPE-SID FEC Validation

When a remote ASBR of the EPE-SID advertisement receives the MPLS OAM packet with the top FEC being the EPE-SID, it **MUST** perform validity checks on the content of the EPE-SID FEC sub-TLV. The basic length check should be performed on the received FEC.

```
PeerAdj SID sub-TLV
-----------
If Adj type = 1, Length should be 28 octets
If Adj type = 2, Length should be 52 octets

PeerNode SID sub-TLV
-------------
Length = (20 + No. of IPv4 interface pairs * 8 +
          No. of IPv6 interface pairs * 32) octets

PeerSet SID sub-TLV
-----------
Length = (9 + No. of elements in the set *
         (8 + No. of IPv4 interface pairs * 8 +
          No. of IPv6 interface pairs * 32) octets
```

Figure 5: Length Validation

If a malformed FEC sub-TLV is received, then a return code of 1, "Malformed echo request received", as defined in [RFC8029] **MUST** be sent. The section below is appended to the procedure given in step 4a of Section 7.4 of [RFC8287].

5.1. EPE-SID FEC Validation Rules

This is an example of Segment Routing IGP-Prefix, IGP-Adjacency SID, and EPE-SID validations. Note that the term "receiving node" in this section corresponds to the node that receives the OAM message with the Target FEC Stack TLV.

```
Else, if the Label-stack-depth is 0 and the Target FEC Stack sub-TLV
at FEC stack-depth is 38 (PeerAdj SID sub-TLV), {

    Set the Best-return-code to 10, "Mapping for this FEC is not
    the given label at stack-depth <RSC>" [RFC8029].  Check if
    any below conditions fail:

           -  Validate that the receiving node's BGP Local AS matches
              with the remote AS field in the received PeerAdj SID
              sub-TLV.

           -  Validate that the receiving node's BGP Router-ID
              matches with the Remote Router ID field in the
              received PeerAdj SID sub-TLV.

           -  Validate that there is an EBGP session with a peer
              having a local AS number and BGP Router-ID as
              specified in the local AS number and Local Router-ID
              field in the received PeerAdj SID sub-TLV.

    If the remote interface address is not zero, validate the
    incoming interface.  Set the Best-return-code to 35,
    "Mapping for this FEC is not associated with the incoming
    interface" [RFC8287].  Check if any below conditions fail:

           -  Validate that the incoming interface on which the
              OAM packet was received matches with the remote
              interface specified in the PeerAdj SID sub-TLV.

    If all above validations have passed, set the return code to 3,
    "Replying router is an egress for the FEC at stack-depth <RSC>"
    [RFC8029].
    }

Else, if the Target FEC Stack sub-TLV at FEC stack-depth is 39
     (PeerNode SID sub-TLV), {

    Set the Best-return-code to 10, "Mapping for this FEC is not
    the given label at stack-depth <RSC>" [RFC8029].  Check if any
    below conditions fail:

       -  Validate that the receiving node's BGP Local AS matches
          with the remote AS field in the received PeerNode SID
          FEC sub-TLV.

       -  Validate that the receiving node's BGP Router-ID matches
          with the Remote Router ID field in the received
          PeerNode SID FEC.

       -  Validate that there is an EBGP session with a peer
          having a local AS number and BGP Router-ID as
          specified in the local AS number and Local Router-ID
          field in the received PeerNode SID FEC sub-TLV.

    If all above validations have passed, set the return code to 3,
    "Replying router is an egress for the FEC at stack-depth <RSC>"
    [RFC8029].
```

```
     }
 Else, if the Target FEC Stack sub-TLV at FEC stack-depth is 40
      (PeerSet SID sub-TLV), {

     Set the Best-return-code to 10, "Mapping for this FEC is not
     the given label at stack-depth <RSC>" [RFC8029].  Check if any
     below conditions fail:

        -  Validate that the receiving node's BGP Local AS matches
           with one of the remote AS fields in the received
           PeerSet SID FEC sub-TLV.

        -  Validate that the receiving node's BGP Router-ID matches
           with one of the Remote Router ID fields in the
           received PeerSet SID FEC sub-TLV.

        -  Validate that there is an EBGP session with a peer having
           a local AS number and BGP Router-ID as specified in the
           local AS number and Local Router-ID fields in the received
           PeerSet SID FEC sub-TLV.

     If all above validations have passed, set the return code to 3,
     "Replying router is an egress for the FEC at stack-depth <RSC>"
     [RFC8029].
     }
```

6. IANA Considerations

IANA has allocated three new Target FEC Stack sub-TLVs in the "Sub-TLVs for TLV Types 1, 16, and 21" registry [MPLS-LSP-PING] within the "TLVs" registry of the "Multiprotocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" registry group.

Sub-Type	Sub-TLV Name
38	PeerAdj SID
39	PeerNode SID
40	PeerSet SID

Table 1: Sub-TLVs for TLV Types 1, 16, and 21 Registry

7. Security Considerations

The EPE-SIDs are advertised for egress links for EPE purposes or for inter-AS links between cooperating ASes. When cooperating domains are involved, they can allow the packets arriving on trusted interfaces to reach the control plane and be processed.

When EPE-SIDs are created for egress TE links where the neighbor AS is an independent entity, it may not allow the packets arriving from the external world to reach the control plane. In such deployments, the MPLS OAM packets will be dropped by the neighboring AS that receives the MPLS OAM packet.

In MPLS Traceroute applications, when the AS boundary is crossed with the EPE-SIDs, the Target FEC Stack TLV is changed. [RFC8287] does not mandate that the initiator, upon receiving an MPLS Echo Reply message that includes the Target FEC Stack Change TLV with one or more of the original segments being popped, remove the corresponding FEC(s) from the Target FEC Stack TLV in the next (TTL+1) traceroute request.

If an initiator does not remove the FECs belonging to the previous AS that has traversed, it may expose the internal AS information to the following AS being traversed in the traceroute.

8. References

8.1. Normative References

[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>.

[RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet Autonomous System (AS) Number Space", RFC 6793, DOI 10.17487/RFC6793, December 2012, <https://www.rfc-editor.org/info/rfc6793>.

[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>.

[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>.

[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>.

[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>.

[RFC9086] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Patel, K., Ray, S., and J. Dong, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing BGP Egress Peer Engineering", RFC 9086, DOI 10.17487/RFC9086, August 2021, <https://www.rfc-editor.org/info/rfc9086>.

8.2. Informative References

[MPLS-LSP-PING] IANA, "Sub-TLVs for TLV Types 1, 16, and 21", <https://www.iana.org/assignments/mpls-lsp-ping-parameters>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC5065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 5065, DOI 10.17487/RFC5065, August 2007, <https://www.rfc-editor.org/info/rfc5065>.

[RFC6286] Chen, E. and J. Yuan, "Autonomous-System-Wide Unique BGP Identifier for BGP-4", RFC 6286, DOI 10.17487/RFC6286, June 2011, <https://www.rfc-editor.org/info/rfc6286>.

[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>.

[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>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

[RFC9087] Filsfils, C., Ed., Previdi, S., Dawra, G., Ed., Aries, E., and D. Afanasiev, "Segment Routing Centralized BGP Egress Peer Engineering", RFC 9087, DOI 10.17487/RFC9087, August 2021, <https://www.rfc-editor.org/info/rfc9087>.

[RFC9256] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

[SR-BGP-POLICY] Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-sr-policy-safi-10, 7 November 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-sr-policy-safi-10>.

[SR-SEGTYPES] Talaulikar, K., Filsfils, C., Previdi, S., Mattes, P., and D. Jain, "Segment Routing Segment Types Extensions for BGP SR Policy", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-sr-segtypes-ext-06, 7 November 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-sr-segtypes-ext-06>.

Appendix A. Examples of Programmed States

This section describes examples of both a correctly and an incorrectly programmed state and provides details on how the new sub-TLVs described in this document can be used to validate the correctness. Consider the diagram from Figure 1.

Correctly programmed state:

- C assigns label 16001 and binds it to adjacency C->E
- C signals that label 16001 is bound to adjacency C->E (e.g., via BGP-LS)
- The controller/ingress programs an SR path that has SID/label 16001 to steer the packet on the exit point from C onto adjacency C->E
- Using MPLS Traceroute procedures defined in this document, the PeerAdj SID sub-TLV is populated with entities to be validated by C when the OAM packet reaches it
- C receives the OAM packet and validates that the top label (16001) is indeed corresponding to the entities populated in the PeerAdj SID sub-TLV

Incorrectly programmed state:

- C assigns label 16001 and binds it to adjacency C->D
- The controller learns that PeerAdj SID label 16001 is bound to adjacency C->E (e.g., via BGP-LS) -- this could be a software bug on C or on the controller
- The controller/ingress programs an SR path that has SID/label 16001 to steer the packet on the exit point from C onto adjacency C->E
- Using MPLS Traceroute procedures defined in this document, the PeerAdj SID sub-TLV is populated with entities to be validated by C (including a local/remote interface address of C->E) when the OAM packet reaches it
- C receives the OAM packet and validates that the top label (16001) is NOT bound to C->E as populated in the PeerAdj SID sub-TLV and then responds with the respective error code

Acknowledgments

Thanks to Loa Andersson, Dhruv Dhody, Ketan Talaulikar, Italo Busi, Alexander Vainshtein, and Deepti Rathi for careful reviews and comments. Thanks to Tarek Saad for providing the example described in Appendix A.

Authors' Addresses

Shraddha Hegde
Juniper Networks Inc.
Exora Business Park
Bangalore 560103
Karnataka
India
Email: shraddha@juniper.net

Mukul Srivastava
Juniper Networks Inc.
Email: msri@juniper.net

Kapil Arora
Individual Contributor
Email: kapil.it@gmail.com

Samson Ninan
Ciena
Email: samson.cse@gmail.com

Xiaohu Xu
China Mobile
Beijing
China
Email: xuxiaohu_ietf@hotmail.com