PMS/Head-end based MPLS Ping and Traceroute in Inter-AS SR Networks

draft-ninan-spring-mpls-inter-as-oam-00

Abstract

Segment Routing (SR) architecture leverages source routing and tunneling paradigms and can be directly applied to the use of a Multiprotocol Label Switching (MPLS) data plane. Segment Routing also provides an easy and efficient way to provide inter connectivity in a large scale network as described in [RFC8604]. [RFC8287] illustrates the problem and defines extensions to perform LSP Ping and Traceroute for Segment Routing IGP-Prefix and IGP-Adjacency Segment Identifiers (SIDs) with an MPLS data plane. It is useful to have the LSP Ping and traceroute procedures when an SRend-to-end path spans across multiple ASes. This document describes mechanisms to facilitate LSP ping and traceroute in inter-AS SR networks in an efficient manner with simple OAM protocol extension.

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 https://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 January 5, 2020.
Table of Contents

1. Introduction .................................................. 2
2. Reverse Path label stack TLV ................................. 4
   2.1. Reverse Path label stack TLV definition ................. 4
   2.2. SRv6 Dataplane ........................................... 5
3. Detailed Procedures ............................................. 5
   3.1. Sending an Echo-Request .................................. 5
   3.2. Receiving an Echo-Request ............................... 6
   3.3. Sending an Echo-Reply .................................... 6
4. Detailed Example ................................................ 6
   4.1. Procedures for Segment Routing LSP ping ............... 7
   4.2. Procedures for Segment Routing LSP Traceroute ...... 7
5. Security Considerations ......................................... 8
6. IANA Considerations ............................................ 8
7. Acknowledgments ................................................ 8
8. References ..................................................... 8
   8.1. Normative References ..................................... 8
   8.2. Informative References ................................... 8
Authors’ Addresses ............................................... 9

1. Introduction
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 acquisitions. Segment Routing can be deployed in such scenarios to provide end to end paths, traversing multiple Autonomous systems (AS). These paths consist of Segment Identifiers (SID) of different type as per [RFC8402].

[I-D.ietf-spring-segment-routing-mpls] specifies the forwarding plane behaviour to allow Segment Routing to operate on top of MPLS data plane. [I-D.ietf-spring-segment-routing-central-epe] describes BGP peering SIDs, which will help in steering packet from one Autonomous system to another. Using above SR capabilities, paths which span across multiple Autonomous systems can be created.

For example Figure 1 describes a topology consisting of inter-AS network consisting of ASes AS1 and AS2. Both AS1 and AS2 are Segment Routing enabled and the EPE links have EPE labels configured and advertised via [I-D.ietf-idr-bgppls-segment-routing-epe]. Controller or head-end can build end-to-end Traffic-Engineered path Node-SIDs, Adjacency-SIDs and EPE-SIDs. It is advantageous for operations to be able to perform LSP ping and traceroute procedures on these inter-AS SR paths. LSP ping/traceroute procedures use IP connectivity for Echo-reply to reach the head-end. In inter-AS networks, IP connectivity may not be there from each router in the path. For example in Figure 1 P3 and P4 may not have IP connectivity for PE1.

[RFC8403] describes mechanisms to carry out the mpls ping/traceroute from a PMS. It is possible to build GRE tunnels or static routes to each router in the network to get IP connectivity for the return.
path. This mechanism is operationally very heavy and requires PMS to be capable of building huge number of GRE tunnels, which may not be feasible.

It is not possible to carry out LSP ping and Traceroute functionality on these paths to verify basic connectivity and fault isolation using existing LSP ping and Traceroute mechanism ([RFC8287] and [RFC8029]). This is because, there exists no IP connectivity to source address of ping packet, which is in a different AS, from the destination of Ping/Traceroute.

[RFC7743] describes a Echo-relay based solution based on advertising a new Relay Node Address Stack TLV containing stack of Echo-relay ip addresses. This mechanism requires the return ping packet to reach the control plane on every relay node. This document describes a mechanism which is efficient and simple and can be easily deployed in SR networks.

2. Reverse Path label stack TLV

Segment Routing networks statically assign the labels to nodes and PMS/Head-end knows entire database. The return path can be built from PMS/Head-end by stacking labels for the return path. A new TLV "Reverse Path label stack TLV" is defined. Each TLV contains a list of labels which may be a prefix/adjacency/binding SID/EPE SID. MPLS Echo -request should contain this TLV, which defines reverse path to reach source from the destination.

The new Reverse Path label stack TLV is an optional TLV. This TLV is carried in the Echo-Request message. This optional TLV MAY appear in the Echo-request message in any order before or after Target FEC Stack TLV. The Reverse Path label stack TLV is defined as below. Each MPLS Echo-request SHOULD contain this TLV in inter-AS cases, which will enable remote end to send the reply to source.

2.1. Reverse Path label stack TLV definition
Type: TBD

Length: Length of TLV including TLV header

No.  Of elements in set:

Ordered set of Labels in the Reverse Path label stack

Label :

Represents 20 bit label. This field should be used to build the return packet. The first label in the label stack represents the top most label that should be encoded in the return packet.

2.2.  SRv6 Dataplane

A future version of this document will address the SRv6 Dataplane.

3.  Detailed Procedures

3.1.  Sending an Echo-Request

LSP ping initiator MUST add a Return Path Label Stack TLV in the Echo-request message. The return label stack MUST correspond to the return path from the egress. The Reverse Path Label Stack TLV is an ordered list of labels. The first label corresponds to the top label that the reponder should use to construct the Echo-reply.
3.2. Receiving an Echo-Request

When a receiver does not understand the Reverse Path Label Stack TLV, it SHOULD silently ignore the TLV and proceed with normal processing as described in [RFC8029]. When a Reverse Path Label Stack TLV is received, and the responder supports processing it, it MUST use the labels in Reverse Path Label Stack TLV to build the echo-reply. The responder MUST follow the normal FEC validation procedures as described in [RFC8029] and [RFC8287] and this document does not suggest any change to those procedures. When the Echo-reply has to be sent out the Reverse Path label Stack TLV is used to construct the MPLS packet to send out.

3.3. Sending an Echo-Reply

The Echo-Reply message is sent as mpls packet with a mpls label stack. The Echo-Reply message MUST be constructed as described in the [RFC8029]. An MPLS packet is constructed with Echo-reply in the payload. The top label MUST be the first label from the Reverse Path Label Stack TLV. The remaining labels MUST follow the order from the Reverse Path Label Stack. The responder MAY check the reachability of the top label in its own LFIB before sending the Echo-Reply.

4. Detailed Example

An example topology is given in Figure 1. This will be used in below sections to explain LSP Ping and Traceroute procedures. The PMS/Head-end has complete view of topology. PE1, P1, P2, ASBR1 and ASBR2 are in AS1. Similarly ASBR3, ASBR4, P3, P4 and PE4 are in AS2.

AS1 and AS2 are Segment Routing enabled. IGP like OSPF/ISIS are used to flood SIDs in each Autonomous System. The ASBR1, ASBR2, ASBR3, ASBR4 advertise BGP EPE SIDs for the inter-AS links. Topology of AS1 and AS2 are advertised via BGP-LS to the controller/PMS or Head-end node. The EPE-SIDs are also advertised via BGP-LS as described in [I-D.ietf-idr-bgp-ls-segment-routing-epe]

The description in the document uses below notations for Segment Identifiers(SIDs).

Node SIDs : N-PE1, N-P1, N-ASBR1 etc.

Adjacency SIDs : Adj-PE1-P1, Adj-P1-P2 etc.

EPE SIDS : EPE-ASBR2-ASBR3, EPE-ASBR1-ASBR4, EPE-ASBR3-ASBR2 etc.

Let us consider a traffic engineered path built from PE1 to PE4 with label stack as below. N-P1, N-ASBR1, EPE-ASBR1-ASBR4, N-PE4 for
following procedures. This stack may be programmed by controller/PMS or Head-end router PE1 may have imported the whole topology information from BGP-LS and computed the inter-AS path.

4.1. Procedures for Segment Routing LSP ping

To perform LSP ping procedure on an SR-Path from PE1 to PE4 consisting of label stack [N-P1,N-ASBR1,EPE-ASBR1-ASBR4, N-PE4], The remote end (PE4) needs IP connectivity to head end (PE1) for the Segment Routing ping to succeed, because Echo-reply needs to travel back to PE1 from PE4. But in typical deployment scenario there will be no ip route from PE4 to PE1 as they belong to different ASes.

PE1 adds Reverse Path from PE4 to PE1 in the MPLS Echo-request using multiple labels in "Reverse Path Label Stack TLV" as defined above. An example return path label stack for PE1 to PE4 for LSP ping is [N-ASBR4, EPE-ASBR4-ASBR1, N-PE1]. An implementation may also build a Reverse Path Label stack consisting of labels to reach its own AS. Once the label stack is popped-off the Echo-reply message will be exposed. The further packet forwarding will be based on ip lookup. An example Reverse Path Label Stack for this case could be [N-ASBR4, EPE-ASBR4-ASBR1].

On receiving MPLS Echo-request PE4 first validates FEC in the Echo-request and calculates label stack to send the response from PE4 to PE1 using "Return label stack TLV". PE4 builds the Echo-reply packet with the mpls label stack constructed out of Reverse Path Label Stack TLV and sends out the Echo-reply to PE1. This label stack can successfully steer reply back to Head-end node (PE1).

4.2. Procedures for Segment Routing LSP Traceroute

As described in the procedures for LSP ping, the return label stack may be sent from head-end in which case the LSP Traceroute procedures are similar to mpls-ping. The head-end constructs the Reverse Path Label Stack TLV and the egress node uses the Reverse Path Label Stack to construct the Echo-reply packet header. Head-end/PMS is aware of the return path from every node visited in the network and builds the Reverse Path Label Stack for every visited node accordingly.

For Example:

For the same traffic engineered path PE1 to PE4 mentioned in above sections, let us assume there is no return path available from the nodes ASBR2 to PE1. During the Traceroute procedure, when PE1 has to visit ASBR2, it builds Return Path Label Stack TLV and includes label to the border-node which has the route to, PE1. In this example the Return Path Label Stack TLV will contain [EPE-ASBR2-ASBR1]. Further
down the traceroute procedure when P3 or P4 node is being visited, PE1 build the Return Path Label Stack TLV containing [N-ASBR2, EPE-ASBR2-ASBR1]. The Echo-reply will be an mpls packet with this label stack and will be forwarded to PE1.

5. Security Considerations

TBD

6. IANA Considerations

Multiprotocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters TLVs Registry

Reverse Path label stack TLV : TBD

7. Acknowledgments

8. References

8.1. Normative References

[I-D.ietf-spring-segment-routing-central-epe]


8.2. Informative References

[I-D.ietf-idr-bgpls-segment-routing-epe]

[I-D.ietf-mpls-interas-lspping]
[I-D.ietf-spring-segment-routing-mpls]
Bashandy, A., Filsfils, C., Previdi, S., Decraene, B.,
Litkowski, S., and R. Shakir, "Segment Routing with MPLS data plane",

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels",
BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997,

Swallow, Ed., "Relayed Echo Reply Mechanism for Label Switched Path (LSP) Ping",
RFC 7743, DOI 10.17487/RFC7743, January 2016,

[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,

Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture",
RFC 8402, DOI 10.17487/RFC8402, July 2018,

Kumar, "A Scalable and Topology-Aware MPLS Data-Plane Monitoring System",
RFC 8403, DOI 10.17487/RFC8403, July 2018,

Henderickx, W., and D. Cooper, "Interconnecting Millions of Endpoints with Segment Routing",
RFC 8604, DOI 10.17487/RFC8604, June 2019,

Authors' Addresses

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

Email: shraddha@juniper.net