Workgroup: Routing area
Internet-Draft:
draft-ietf-mpls-egress-tlv-for-nil-fec-04
Published: 30 March 2022
Intended Status: Standards Track
Expires: 1 October 2022
Authors: D. Rathi, Ed.
Juniper Networks Inc.
K. Arora
Juniper Networks Inc.
S. Hegde
Juniper Networks Inc.
Z. Ali
Cisco Systems, Inc.
N. Nainar
Cisco Systems, Inc.

Egress TLV for Nil FEC in Label Switched Path Ping and Traceroute Mechanisms

Abstract

MPLS ping and traceroute mechanism as described in RFC 8029 and related extensions for SR as defined in RFC 8287 is very useful to precisely validate the control plane and data plane synchronization. There is a possibility that all intermediate or transit nodes may not have been upgraded to support these validation procedures. A simple mpls ping and traceroute mechanism comprises of ability to traverse any path without having to validate the control plane state. RFC 8029 supports this mechanism with Nil FEC. The procedures described in RFC 8029 are mostly applicable when the Nil FEC is used as intermediate FEC in the label stack. When all labels in label stack are represented using single Nil FEC, it poses some challenges.

This document introduces new TLV as additional extension to exisiting Nil FEC and describes mpls ping and traceroute procedures using Nil FEC with this additional extensions to overcome these challenges.

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.

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 1 October 2022.

Table of Contents

1. Introduction

Segment routing supports the creation of explicit paths using adjacency- sids, node-sids, and anycast-sids. In certain usecases,

the TE paths are built using mechanisms described in [I.D-ietf-spring-segment-routing-policy] by stacking the labels that represent the nodes and links in the explicit path. When the SR-TE paths are built by the controller, the head-end routers may not have the complete database of the network and may not be aware of the FEC associated with labels that are used in the label stack. A very useful Operations And Maintenance (OAM) requirement is to be able to ping and trace these paths. A simple mpls ping and traceroute mechanism comprises of ability to traverse the SR-TE path without having to validate the control plane state.

MPLS ping and traceroute mechanism as described in [RFC8029] and related extensions for SR as defined in [RFC8287] is very useful to precisely validate the control plane and data plane synchronization. It also provides ability to traverse multiple ECMP paths and validate each of the ECMP paths. Use of Target FEC requires all nodes in the network to have implemented the validation procedures. All intermediate nodes may not have been upgraded to support validation procedures. In such cases, it is useful to have ability to traverse the paths using ping and traceroute without having to obtain the Forwarding Equivalence Class (FEC) for each label. [RFC8029] supports this mechanism with FECs like Nil FEC and Generic FEC.

Generic IPv4 and IPv6 FEC are used when the protocol that is advertising the label is unknown. The information that is carried in Generic FEC is the IPv4 or IPv6 prefix and prefix length. Thus Generic FEC types perform an additional control plane validation. But the details of generic FEC and validation procedures are not very detailed in the [RFC8029]. The use-case mostly specifies inter-AS VPNs as the motivation. Certain aspects of Segment Routing such as anycast SIDs requires clear guidelines on how the validation procedure should work. Also Generic FEC may not be widely supported and if transit routers are not upgraded to support validation of generic FEC, traceroute may fail. on other hand, Nil FEC consists of the label and there is no other associated FEC information. NIL FEC is used to traverse the path without validation for cases where the FEC is not defined or routers are not upgraded to support the FECs. Thus it can be used to check any combination of segments on any data path. The procedures described in [RFC8029] are mostly applicable when the Nil FEC is used where the Nil FEC is an intermediate FEC in the label stack. When all labels in label-stack are represented using single Nil FEC, it poses some challenges.

Section 2 discusses the problems associated with using single Nil FEC in a MPLS ping/traceroute procedure and Section 3 and Section 4 discusses simple extensions needed to solve the problem.

2. Problem with Nil FEC

The purpose of Nil FEC as described in [RFC8029] is to ensure hiding of transit tunnel information and in some cases to avoid false negatives when the FEC information is unknown.

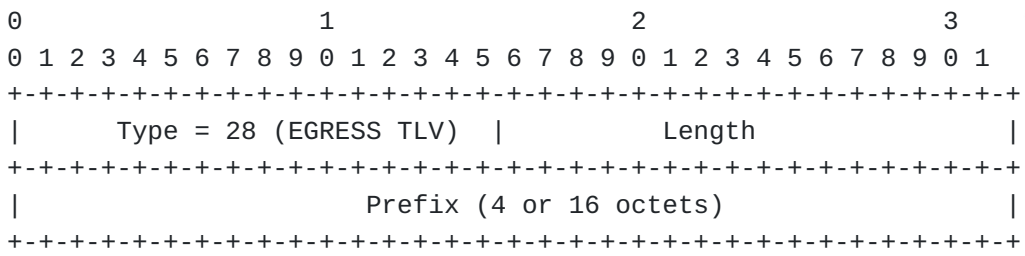
This draft uses single NIL FEC to represent complete label stack in MPLS ping/traceroute packet irrespective of number of segments in the label-stack. When router in the label-stack path receives MPLS ping/traceroute packets, there is no definite way to decide on whether it is the intended egress router since Nil FEC does not carry any information. So there is high possibility that the packet may be mis-forwarded to incorrect destination but the ping/ traceroute might still return success.

To avoid this problem, there is a need to add additional information in the MPLS ping and traceroute packet along with Nil FEC to do minimal validation on egress/destination router and sends proper information to ingress router on success and failure. This additional information should help to report transit router information to ingress/initiator router that can be used by offline application to validate the traceroute path.

Thus addition of egress information in ping/traceroute packet will help in validating Nil-FEC on each receiving router on label-stack path to ensure the correct destination. It can be used to check any combination of segments on any path without upgrading transit nodes.

3. Egress TLV

The Egress object is a TLV that MAY be included in an MPLS Echo Request message. Its an optional TLV and should appear before FEC-stack TLV in the MPLS Echo Request packet. In case multiple Nil FEC is present in Target FEC Stack TLV, Egress TLV should be added corresponding to the ultimate egress of the label-stack. It can be use for any kind of path with Egress TLV added corresponding to the end-point of the path. Explicit Path can be created using node-sid, adj-sid, binding-sid etc, EGRESS TLV prefix will be derived from path egress/destination and not based on labels used in the path to reach the destination. The format is as specified below:



Type : 28 (Section 7.1)

Length : variable based on IPV4/IPV6 prefix. Length excludes the length of Type and length field. Length will be 4 octets for IPv4 and 16 octets for IPv6.

Prefix : This field carries the valid IPv4 prefix of length 4 octets or valid IPv6 Prefix of length 16 octets. It can be obtained from egress of Nil FEC corresponding to last label in the label-stack or SR-TE policy endpoint field [I.D-ietf-idr-segment-routing-te-policy].

4. Procedure

This section describes aspects of LSP Ping and Traceroute operations that require further considerations beyond [RFC8029].

4.1. Sending Egress TLV in MPLS Echo Request

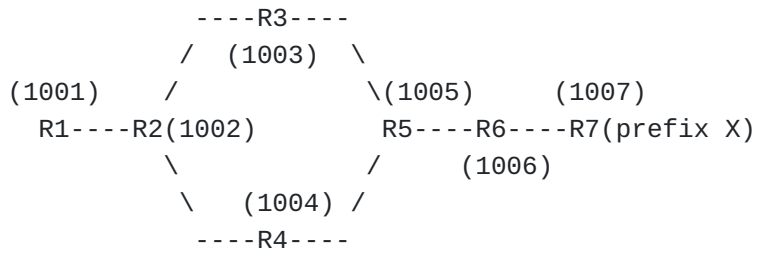
As stated earlier, when the sender node builds a Echo Request with target FEC Stack TLV, Egress TLV SHOULD appear before Target FEC-stack TLV in MPLS Echo Request packet.

Ping

When the sender node builds a Echo Request with target FEC Stack TLV that contains a single NiL FEC corresponding to the last segment of the SR-TE path, sender node MUST add a Egress TLV with prefix obtained from SR-TE policy endpoint field [I.D-ietf-idr-segment-routing-te-policy] to indicate the egress for this Nil FEC in the Echo Request packet. In case endpoint is not specified or is equal to 0, sender MUST use the prefix corresponding to last segment of the SR-TE path as prefix for Egress TLV.

Traceroute

When the sender node builds a Echo Request with target FEC Stack TLV that contains a single NiL FEC corresponding to complete segment-list of the SR-TE path, sender node MUST add a Egress TLV with prefix obtained from SR-TE policy endpoint field [I.D-ietf-idr-segment-routing-te-policy] to indicate the egress for this Nil FEC in the Echo Request packet. Some implementations may send multiple NilFEC but it is not really required. In case headend sends multiple Nil FECs the last one should have the egress TLV. When the label stack becomes zero, all Nil FEC TLVs are removed and egress TLV MUST be validated from last Nil FEC In case endpoint is not specified or is equal to 0 (as in case of color-only SR-TE policy), sender MUST use the prefix corresponding to the last segment endpoint of the SR-TE path i.e. ultimate egress as prefix for Egress TLV.



Consider the SR-TE policy configured with label-stack as 1002, 1004 , 1007 and end point/destination as prefix X on ingress router R1 to reach egress router R7. Segment 1007 belongs to R7 that has prefix X locally configured on it.

In Ping Echo Request, with target FEC Stack TLV that contains a single NiL FEC corresponding to 1007, should add Egress TLV for endpoint/destination prefix X with type as EGRESS-TLV, length depends on if X is IPv4 or IPv6 address and prefix as X.

In Traceroute Echo Request, with target FEC Stack TLV that contains a single NiL FEC corresponding to complete label-stack (1002, 1004, 1007) or multiple Nil-FEC corresponding to each label in label-stack, should add single Egress TLV for endpoint/destination prefix X with type as EGRESS-TLV, length depends on if X is IPv4 or IPv6 address and prefix as X. In case X is not present or is set to 0 (as in case of color-only SR-TE policy), sender should use endpoint of segment 1007 as prefix for Egress TLV.

4.2. Receiving Egress TLV in MPLS Echo Request

No change in the processing for Nil FEC as defined in [RFC8029] in Target FEC stack TLV Node that receives an MPLS echo request.

Additional processing done for Egress TLV on receiver node as follows:

1. If the Label-stack-depth is greater than 0 and the Target FEC Stack sub-TLV at FEC-stack-depth is Nil FEC, set Best-return-code to 8 ("Label switched at stack-depth") and Best-return-subcode to Label-stack-depth to report transit switching in MPLS Echo Reply message.

2. If the Label-stack-depth is 0 and the Target FEC Stack sub-TLV at FEC-stack-depth is Nil FEC then do the look up for an exact match of the EGRESS TLV prefix to any of locally configured interfaces or loopback addresses.

2a. If EGRESS TLV prefix look up succeeds, set Best-return-code to 36 ("Replying router is an egress for the prefix in EGRESS-TLV") (Section 7.2) and Best-return-subcode to 1 to report egress ok in MPLS Echo Reply message.

2b. If EGRESS TLV prefix look up fails, set the Best-return-code to 10, "Mapping for this FEC is not the given label at stack-depth" and Best-return-subcode to 1.

5. Backward Compatibility

The extension proposed in this document is backward compatible with procedures described in [RFC8029]. Router that does not support EGRESS-TLV, will ignore it and use current NIL-FEC procedures described in [RFC8029].

When the egress node in the path does not support the extensions proposed in this draft egress validation will not be done and Best-return-code as 3 ("Replying router is an egress for the FEC at stack-depth") and Best-return- subcode as 1 to report egress ok will be set in MPLS Echo Reply message.

When the transit node in the path does not support the extensions proposed in this draft Best-return-code as 8 ("Label switched at stack-depth") and Best-return-subcode as Label-stack-depth to report transit switching will be set in MPLS Echo Reply message.

6. Security Considerations

TBD

7. IANA Considerations

The code points in section Section 7.1 and Section 7.2 have been assigned by [IANA] by early allocation on 2021-11-08.

7.1. New TLV

[IANA] need to assign new value for EGRESS TLV in the "Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" in "TLVs" sub-registry.

Value	Description	Reference
28	EGRESS TLV	Section 3
		This document

Table 1: TLVs Sub-Registry

7.2. New Return code

[IANA] need to assign new Return Code for "Replying router is an egress for the prefix in EGRESS-TLV" in the "Multi-Protocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" in "Return Codes" sub-registry.

Value	Description	Reference
36	Replying router is an egress	Section 4.2
	for the prefix in EGRESS-TLV	This document

Table 2: Return code Sub-Registry

8. Acknowledgements

TBD.

9. References

9.1. Normative References

[I.D-ietf-idr-segment-routing-te-policy]
Filsfils, C., Ed., Previdi, S., Ed., Talaulikar, K., Mattes, P., Rosen, E., Jain, D., and S. Lin, "Advertising Segment Routing Policies in BGP", draft-ietf-idr-segment-routing-te-policy-09, work in progress, May 2020, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-09>.

[I.D-ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Bogdanov, A., Mattes, P., and D. Voyer, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-08, work in progress, July 2020, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-segment-routing-policy-08>.

[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>.

[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>.

9.2. Informative References

[IANA] IANA, "Multiprotocol Label Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters", <http://www.iana.org/assignments/mpls-lsp-ping-parameters>.

Authors' Addresses

Deepti N. Rathi (editor)
Juniper Networks Inc.
Exora Business Park
Bangalore 560103
KA
India

Email: deeptirathi.ietf@gmail.com

Kapil Arora
Juniper Networks Inc.
Exora Business Park
Bangalore 560103
KA
India

Email: kapilaro@juniper.net

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

Email: shraddha@juniper.net

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

Nagendra Kumar Nainar
Cisco Systems, Inc.

Email: naikumar@cisco.com