      LSP Ping/Traceroute for Prefix SID in Presence of Multi-Algorithm/Multi-
                              Topology Networks
                     draft-iqbal-spring-mpls-ping-algo-02

Abstract

   [RFC8287] defines the extensions to MPLS LSP Ping and Traceroute for
   Segment Routing IGP-Prefix and IGP-Adjacency Segment Identifier
   (SIDs) with an MPLS data plane. The machinery defined in [RFC8287]
   works well in single topology, single algorithm deployments where
   each Prefix SID is only associated with a single IP prefix. In
   multi-topology networks, or networks deploying multiple algorithms
   for the same IP Prefix, MPLS echo request needs to carry additional
   information in the Target FEC Stack sub-TLVs to properly validate IGP
   Prefix SID.

   This document updates [RFC8287] by modifying IPv4 and IPv6 IGP-Prefix
   Segment ID FEC sub-TLVs to also include algorithm identification
   while maintaining backwards compatibility. This document also
   introduces new Target FEC Stack sub-TLVs for Prefix SID validation in
   multi-topology networks.

Status of This Memo

This Internet-Draft will expire on August 27, 2021.

Copyright Notice

Table of Contents

1. Introduction

   [RFC8287] defines the extensions to MPLS LSP Ping and Traceroute for
   Segment Routing IGP-Prefix SID and IGP-Adjacency SID with an MPLS
   data plane. [RFC8287] proposes 3 Target FEC Stack Sub-TLVs to carry
   this information. [I-D.ietf-lsr-flex-algo] introduces the concept of
   Flexible Algorithm that allows IGPs (ISIS, OSPFv2 and OSPFv3) to
   compute constraint-based path over an MPLS network. The constraint-
   based paths enables the IGP of a router to associate one or more
   Segment Routing Prefix-SID with a particular Flexible Algorithm, and
   steer packets along the constraint-based paths. Multiple Flexible
   Algorithms are assigned to the same IPv4/IPv6 Prefix while each
   utilizing a different MPLS Prefix SID label. Similarly, operators
   may deploy same IP prefix across multiple topologies in the network
   using IGP Multi-topology ID (MT-ID). As Flexible-Algorithm based
   deployments in particular, and multi-topology networks in general,
   become more common, existing OAM machinery requires updates to
   correctly diagnose network faults.

   Segment Routing architecture [RFC8402] defines the context for IGP
   Prefix SID as a unique tuple comprised of prefix, topology, and
   algorithm>. Existing MPLS Ping/Traceroute machinery for SR Prefix
   SIDs, defined in [RFC8287], carries prefix, prefix length, and IGP
   protocol. To correctly identify and validate a Prefix-SID, the
   validating device also requires algorithm and topology identification
   to be supplied in the FEC Stack sub-TLV. This document extends SR-
   IGP IPv4 and IPv6 Prefix SID FECs to validate a particular algorithm
   in a single-topology network, while maintaining backwards
   compatibility with existing implementations of [RFC8287]. It also
   introduces new Target FEC Stack sub-TLVs to perform MPLS Ping and
   Traceroute for IGP Prefix SIDs in multi-topology, multi-algorithm
   deployments.

2. Conventions

   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].

The term "Must Be Zero" (MBZ) is used in object descriptions for
reserved fields. These fields MUST be set to zero when sent and
ignored on receipt.

Since this document refers to the MPLS Time to Live (TTL) far more
frequently than the IP TTL, the authors have chosen the convention of
using the unqualified "TTL" to mean "MPLS TTL" and using "IP TTL" for
the TTL value in the IP header.

3. Motivation

In presence of multiple algorithms, a single IGP Prefix may be
associated with zero or more IGP Prefix SIDs in addition to the
default (Shortest Path First) Prefix SID. Each Prefix SID will have
a distinct Prefix SID label and may possibly have a distinct set of
next-hops based on associated constraint-based path calculation
criteria. This means that to reach the same destination, an non-
default algorithm IGP-Prefix SID may take a different path than
default IGP Prefix SID algorithm.

```
          R3------R6
         /          \
        /            \
   R1----R2           R7----R8
         \           /
          \         /
          R4------R5
```

Figure above, which is a simplification of the diagram used in
[RFC8287] illustrates this point through an example. Node Segment
IDs for R1, R2, R3, R4, R5, R6, R7, and R8 for the default algorithm
are 5001, 5002, 5003, 5004, 5005, 5006, 5007, and 5008, respectively.
Nodes R1, R2, R4, R5, R7, and R8 also participate in Flexible
Algorithm 128. Their corresponding Node Segment IDs for the
algorithm are 5801, 5802, 5804, 5805, 5807, and 5808, respectively.

Now consider an MPLS LSP Traceroute request to validate the path to
reach node R8 through Flexible Algorithm 128. The TTL of the first
echo request packet expires at node R2 with incoming label 5808.
Node R2 attempts to validate IGP-Prefix SID Target FEC stack sub-TLV
from the echo request. However, this TFS sub-TLV does not contain
information identifying the algorithm. As a result, R2 will attempt
validation with default algorithm which expects the echo packet to
arrive with Prefix SID label 5008. The validation fails, and node R2
responds with error code 10 resulting in a false negative.

Carrying algorithm identification in the Target FEC Stack sub-TLV of
MPLS echo request will help avoid such false negatives. It will also

help detect forwarding deviations such as when the packet for a
particular destination is incorrectly forwarded to a device that is
participating in the default algo but does not participate in a given
Flexible Algorithm.

The above problem statement can also be extended to apply in Multi-
Topology networks. In such networks, the Target FEC Stack sub-TLV
MUST carry Multi-Topology ID (MT-ID) in addition to prefix, its
length, IGP identification, and algorithm.

4. Algorithm Identification for IGP-Prefix SID Sub-TLVs

Section 5 of [RFC8287] defines 3 different Segment ID Sub-TLVs that
will be included in Target FEC Stack TLV defined in [RFC8029]. This
section updates IPv4 IGP-Prefix Segment ID Sub-TLV and IPv6 IGP-
Prefix Segment ID Sub-TLV to also include an additional field
identifying the algorithm.

4.1. IPv4 IGP-Prefix Segment ID Sub-TLV

The Sub-TLV format for IPv4 IGP-Prefix Segment ID MUST be set as
shown in the below TLV format:

```
    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
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                         IPv4 prefix                           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Prefix Length  |    Protocol   |     Algo      |   Reserved    |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Algo field MUST be set to 0 if the default algorithm is used. Algo
field is set to 1 if Strict Shortest Path First (Strict-SPF)
algorithm is used. For Flex-Algo, the Algo field MUST be set with
the algorithm value (values can be 128-255).

4.2. IPv6 IGP-Prefix Segment ID Sub-TLV

The Sub-TLV format for IPv6 IGP-Prefix Segment ID MUST be set as
shown in the below TLV format:

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    |                                                               |
    |                        IPv6 prefix                            |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |Prefix Length  |    Protocol   |     Algo      |   Reserved    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   Algo field MUST be set to 0 if the default algorithm is used. Algo
   field is set to 1 if Strict Shortest Path First (Strict-SPF)
   algorithm is used. For Flex-Algo, the Algo field MUST be set with
   the algorithm value (values can be 128-255).

5. Multi-topology Support for IGP Prefix SID

   IGP Prefix SID TLVs defined above assume a single-topology network
   for path validation. For Multi-Topology networks, this section
   introduces new Multi-Topology IGP IPv4 Prefix SID and Multi-Topology
   IGP IPv6 Prefix SID sub-TLVs in the Target FEC Stack TLV of MPLS echo
   request. These sub-TLVs carry MT-ID for OSPF and IS-IS protocols as
   specified in [RFC4915] and [RFC5120] respectively.

5.1. Multi-topology IPv4 IGP-Prefix Segment ID Sub-TLV

   The Sub-TLV format for Multi-topology IPv4 IGP-Prefix Segment ID MUST
   be set as shown in the below TLV format:

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                        IPv4 prefix                            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |Prefix Length  |    Protocol   |     Algo      |   Reserved    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |        MT-ID               |               MBZ               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

   MT-ID identifies the Multi-Topology ID associated with the Prefix
   SID. MT-ID is set in trailing 12 bits of the field when the Protocol
   is set to IS-IS. Leading 4-bits of the MT-ID MUST be all zeroes for
   IS-IS. MT-ID is set to trailing 8 bits when the protocol is
   specified as OSPF. The leading octet MUST be set to all zeroes for
   OSPF. MBZ MUST be set to all zeroes.

The Protocol field MUST be set 1 if the responder MUST perform FEC
validation using OSPF as the IGP protocol and MT-ID is an OSPF Multi-
Topology ID. Protocol is set to 2 if the responder MUST perform FEC
validation using IS-IS as the IGP protocol, and the MT-ID is IS-IS
Multi-Topology ID. Protocol MUST not be set to 0 when using Multi-
Topology IPv4 IGP Prefix SID sub-TLV.

5.2. Multi-Topology IPv6 IGP-Prefix Segment ID Sub-TLV

The Sub-TLV format for IPv6 IGP-Prefix Segment ID MUST be set as
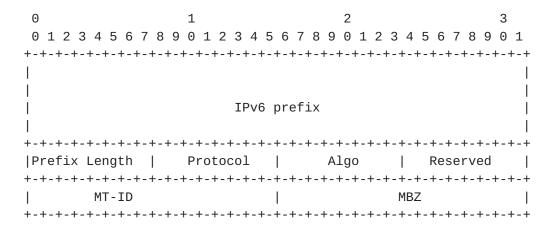shown in the below TLV format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                                                               |
|                        IPv6 prefix                            |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Prefix Length  |     Protocol    |      Algo     |   Reserved   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         MT-ID                    |                MBZ          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

MT-ID identifies the Multi-Topology ID associated with the Prefix
SID. MT-ID is set in trailing 12 bits of the field when the Protocol
is set to IS-IS. Leading 4-bits of the MT-ID MUST be all zeroes for
IS-IS. MT-ID is trailing 8 bits when the protocol is specified as
OSPF. The leading octet MUST be set to all zeroes for OSPF. MBZ
MUST be set to all zeroes.

The Protocol field MUST be set 1 if the responder MUST perform FEC
validation using OSPF as the IGP protocol and MT-ID is an OSPF Multi-
Topology ID. Protocol is set to 2 if the responder MUST perform FEC
validation using IS-IS as the IGP protocol, and the MT-ID is IS-IS
Multi-Topology ID. Protocol MUST not be set to 0 when using Multi-
Topology IPv6 IGP Prefix SID sub-TLV.

6. Procedures

The below section describes LSP Ping and Traceroute procedures beyond
the text specified in LSP

6.1. Single-Topology Networks

An array of network operators may deploy flexible algorithms in their
network for constraint-based shortest paths, without deploying multi-
topology. The updated FEC definitions for IGP Prefix SID allows
operator to achieve LSP Ping and Traceroute in these networks while
maintaining backwards compatibility with existing devices in the
network. Below text highlights the handling procedures and initiator
and responder for the updated FEC definitions.

6.1.1. Initiator Node Procedures

A node initiating LSP echo request packet for the Node Segment ID
MUST identify and include the algorithm associated with the IGP
Prefix SID in the Target FEC Stack sub-TLV. If the initiating node
is not aware of the algorithm, the default algorithm (id 0) of
Shortest Path First is assumed.

6.1.2. Responder Node Procedures

This section updates the procedures defined in Section 7.4 of
[RFC8287] for IPv4/IPv6 IGP Prefix SID FEC. If the algorithm is 0,
the procedures from [RFC8287] do not require any change. For any
other algorithm value, if the responding node is validating the FEC
stack, it MUST also validate the IGP Prefix SID advertisement for the
algorithm defined in Algo field.

If the responding node is including IGP Prefix SID FEC in the FEC
stack due to FEC Stack Change operation, it MUST also include
algorithm associated with the Prefix SID.

6.2. Multi-Topology Networks

In presence of Multi-Topology networks, the operators can use the new
Multi-Topology IGP IPv4/IPv6 Prefix SID FEC definitions to achieve
path validation and fault isolation. Below text describes handling
procedures for Multi-Topology networks for initiator and responder.
The procedures defined in [RFC8287] are still applicable and the text
below updates them instead of replacing them.

6.2.1. Initiator Node Procedures

A node initiating LSP echo request packet for Single-Topology network
MAY use Multi-Topology IGP IPv4/IPv6 Prefix SID defined above. A
node initiating LSP echo request for Multi-Topology networks MUST use
Multi-Topology IGP IPv4/IPv6 Prefix SID defined above. The node MUST
identify and include both the IGP MT-ID and the algorithm associated
with the IGP prefix SID in addition to prefix, prefix length, and the

protocol. If the initiating node is not aware of the algorithm, the
default algorithm (id 0) of Shortest Path First is assumed. The
protocol MUST be set to 1 if the responding node is running OSPF, and
2 if the responding node is running IS-IS.

6.2.2. Responding Node Procedures

This section updates the procedures defined in Section 7.4 of
[RFC8287] for Multi-Topology IPv4/IPv6 IGP Prefix SID FEC. Upon
reception of the sub-TLV, responding node MUST validate that Protocol
field is not 0 to correctly parse MT-ID. In addition to procedures
defined in [RFC8287], if responding node is validating the FEC Stack,
it MUST validate the IGP Prefix SID advertisement for the algorithm
and the MT-ID described in the incoming FEC sub-TLV.

If the responding node is including Multi-Topology IGP Prefix SID FEC
in the FEC stack due to a FEC Stack Change operation, it MUST also
include the algorithm and MT-ID associated with the Prefix SID, and
set the Protocol to 1 or 2, based on the corresponding IGP.

7. IANA Considerations

7.1. New Target FEC tack Sub-TLV

IANA is requested to assign two new Sub-TLVs from "Sub-TLVs for TLV
Types 1, 16 and 21" sub-registry from the "Multi-Protocol Label
Switching (MPLS) Label Switched Paths (LSPs) Ping Parameters" (IANA-
MPLS-LSP-PING) registry.

Sub-Type	Sub-TLV Name	Reference
TBD1	Multi-topology IPv4 IGP-Prefix Segment ID	This document
TBD2	Multi-topology IPv6 IGP-Prefix Segment ID	This document

7.2. Algorithm in the Segment ID Sub-TLV

IANA is requested to create a new "Algorithm in the Segment ID Sub-
TLV" registry under the "Multi-Protocol Label Switching (MPLS) Label
Switched Paths (LSPs) Ping Parameters" registry. The initial entries
are requested as below:

Value	Meaning	Reference
0	Default Algorithm	This document
1	Strict Shortest Path First (Strict-SPF)	This document

8. Security Considerations

This document updates [RFC8287] and does not introduce any security
considerations.

9. Acknowledgements

TBA.

10. Contributors

TBA

11. References

11.1. 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>.

[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>.

11.2. Informative References

[I-D.ietf-lsr-flex-algo]
            Psenak, P., Hegde, S., Filsfils, C., Talaulikar, K., and
            A. Gulko, "IGP Flexible Algorithm", draft-ietf-lsr-flex-
            algo-13 (work in progress), October 2020.

[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", draft-ietf-spring-segment-routing-mpls-22
            (work in progress), May 2019.

   [RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P.
              Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF",
              RFC 4915, DOI 10.17487/RFC4915, June 2007,
              <https://www.rfc-editor.org/info/rfc4915>.

   [RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi
              Topology (MT) Routing in Intermediate System to
              Intermediate Systems (IS-ISs)", RFC 5120,
              DOI 10.17487/RFC5120, February 2008,
              <https://www.rfc-editor.org/info/rfc5120>.

   [RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L.,
              Decraene, B., Litkowski, S., and R. Shakir, "Segment
              Routing Architecture", RFC 8402, DOI 10.17487/RFC8402,
              July 2018, <https://www.rfc-editor.org/info/rfc8402>.

Authors' Addresses

   Nagendra Kumar (editor)
   Cisco Systems, Inc.

   Email: naikumar@cisco.com

   Zafar Ali
   Cisco Systems, Inc.

   Email: zali@cisco.com

   Carlos Pignataro
   Cisco Systems, Inc.

   Email: cpignata@cisco.com

   Faisal Iqbal
   Arista Networks

   Email: faisal.ietf@gmail.com

   Deepti
   Juniper Networks Inc.

   Email: deeptir@juniper.net

Shraddha
Juniper Networks Inc.

Email: shraddha@juniper.net