Workgroup: Network Working Group
Internet-Draft:
draft-chen-idr-sr-ingress-protection-06
Published: 18 April 2022
Intended Status: Standards Track
Expires: 20 October 2022
Authors: H. Chen (Futurewei)
M. Toy (Verizon)
A. Wang (China Telecom)
Z. Li (China Mobile)
L. Liu (Fujitsu)
X. Liu (Volta Networks)

SR Path Ingress Protection

Abstract

This document describes extensions to Border Gateway Protocol (BGP) for protecting the ingress node of a Segment Routing (SR) tunnel or path.

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 20 October 2022.

Copyright Notice

Table of Contents

1. Introduction

The fast protection of a transit node of a Segment Routing (SR) path or tunnel is described in [I-D.bashandy-rtgwg-segment-routing-ti-lfa] and [I-D.hu-spring-segment-routing-proxy-forwarding]. [RFC8424] presents extensions to RSVP-TE for the fast protection of the ingress node of a traffic engineering (TE) Label Switching Path (LSP). However, these documents do not discuss any protocol extensions for the fast protection of the ingress node of an SR path or tunnel.

This document fills that void and specifies protocol extensions to Border Gateway Protocol (BGP) for the fast protection of the ingress node of an SR path or tunnel. Ingress node and ingress, fast protection and protection as well as SR path and SR tunnel will be used exchangeably in the following sections.

2. Terminologies

The following terminologies are used in this document.

SR:
Segment Routing

SRv6: SR for IPv6

SRH: Segment Routing Header

SID: Segment Identifier

CE: Customer Edge

PE: Provider Edge

LFA: Loop-Free Alternate

TI-LFA: Topology Independent LFA

TE: Traffic Engineering

BFD: Bidirectional Forwarding Detection

VPN: Virtual Private Network

L3VPN: Layer 3 VPN

FIB: Forwarding Information Base

PLR: Point of Local Repair

BGP: Border Gateway Protocol

IGP: Interior Gateway Protocol

OSPF: Open Shortest Path First

IS-IS: Intermediate System to Intermediate System

3. SR Path Ingress Protection Example

To protect against the failure of the (primary) ingress node of a (primary) SR path, a backup ingress node is configured or selected and is different from the (primary) ingress node. A backup SR path from the backup ingress node is computed and installed. Primary ingress and ingress as well as primary SR path and SR path will be used exchangeably.

Figure 1 shows an example of protecting ingress PE1 of a SR path, which is from ingress PE1 to egress PE3.

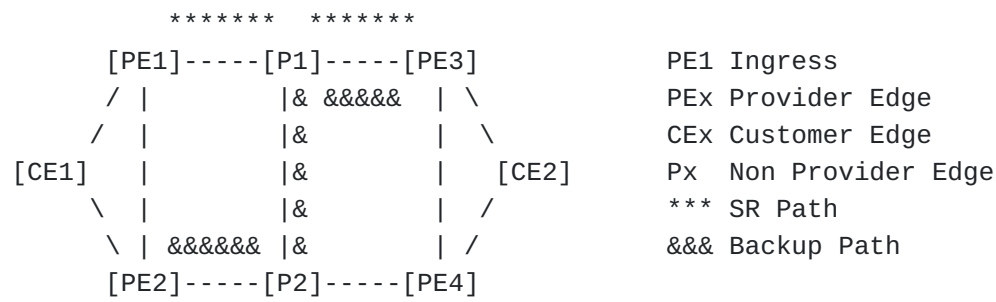

Figure 1: Protecting Ingress PE1 of SR Path PE1-P1-PE3

In normal operations, CE1 sends the traffic with destination PE3 to ingress PE1, which imports the traffic into the SR path.

When CE1 detects the failure of ingress PE1, it switches the traffic to backup ingress PE2, which imports the traffic from CE1 into a backup SR path. The backup path is from the backup ingress PE2 to the egress PE3. When the traffic is imported into the backup path, it is sent to the egress PE3 along the path.

4. Behavior after Ingress Failure

After the failure of the ingress of an SR path happens, there are a couple of different ways to detect the failure. In each way, there may be some specific behavior for the traffic source (e.g., CE1) and the backup ingress (e.g., PE2).

In one way, the traffic source (e.g., CE1) is responsible for fast detecting the failure of the ingress (e.g., PE1) of an SR path. Fast detecting the failure means detecting the failure in a few or tens of milliseconds. The backup ingress (e.g., PE2) is ready to import the traffic from the traffic source into the backup SR path installed.

In normal operations, the source sends the traffic to the ingress of the SR path. When the source detects the failure of the ingress, it switches the traffic to the backup ingress, which delivers the traffic to the egress of the SR path via the backup SR path.

In another way, the backup ingress is responsible for fast detecting the failure of the ingress of an SR path.

In normal operations, the source (e.g., CE1) sends the traffic to the ingress (e.g., PE1) and may send the traffic to the backup ingress (e.g., PE2). It sends the traffic to the backup ingress (e.g., PE2) after the ingress fails.

The backup ingress does not import any traffic from the source into the backup SR path in normal operations. When it detects the failure

of the ingress, it imports the traffic from the source into the backup SR path.

5. Extensions to BGP

For a SR path from a primary ingress node to an egress node, a backup ingress node is selected to protect the failure of the primary ingress node of the SR path. This section describes the extensions to BGP for representing the information for protecting the primary ingress node in a BGP UPDATE message and distributing the information to the backup ingress node. The information includes a SR backup path.

[I-D.ietf-idr-segment-routing-te-policy] specifies a way of representing a SR path in a BGP UPDATE message and distributing the SR path to the ingress node of the SR path.

This is extended to represent the information for protecting the primary ingress by defining a few of new Sub-TLVs.

5.1. SR Path Ingress Protection Sub-TLV

A new Sub-TLV, called SR Path Ingress Protection Sub-TLV, is defined. When a UPDATE message is sent to the backup ingress node for protecting the primary ingress node of a SR path, the message contains this Sub-TLV. Its format is illustrated below.

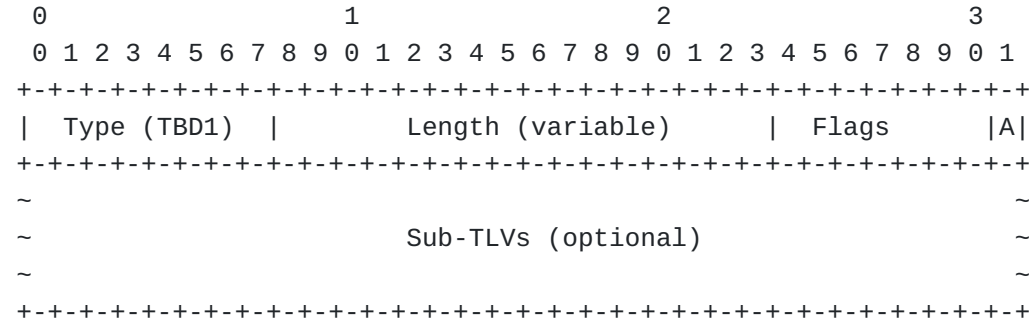

Figure 2: SR Path Ingress Protection Sub-TLV

Type: TBD1 is to be assigned by IANA.

Length: Variable.

Flags: 1 octet. One flag is defined.

Flag A: 1 bit. It is set to

1: request a backup ingress to let the forwarding entry for the backup SR path be Active.

0:
request a backup ingress to let the forwarding entry for the backup SR path be inactive initially and to make the entry be active after detecting the failure of the primary ingress node of the primary SR path.

A few optional Sub-TLVs are defined, which are Primary Ingress Sub-TLV, Service Sub-TLV and Traffic Description Sub-TLV.

5.1.1. Primary Ingress Sub-TLV

A Primary Ingress Sub-TLV indicates the IP address of the primary ingress node of a primary SR path. It has two formats: one for primary ingress node IPv4 address and the other for primary ingress node IPv6 address, which are illustrated below.

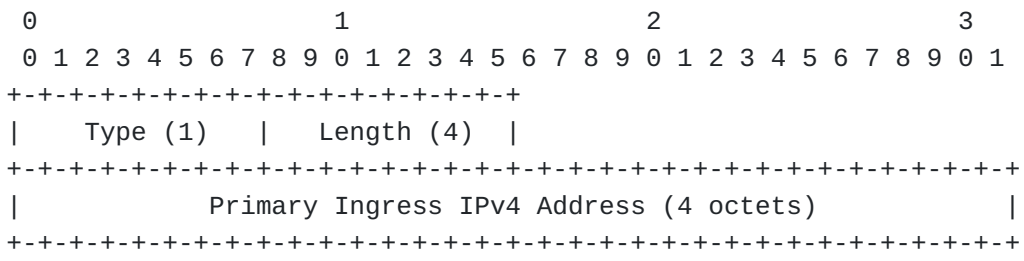

Figure 3: Primary Ingress IPv4 Address Sub-TLV

Type: Its value (1 suggested) is to be assigned by IANA.

Length: 4.

Primary Ingress IPv4 Address: 4 octets. It represents an IPv4 host address of the primary ingress node of a primary SR path.

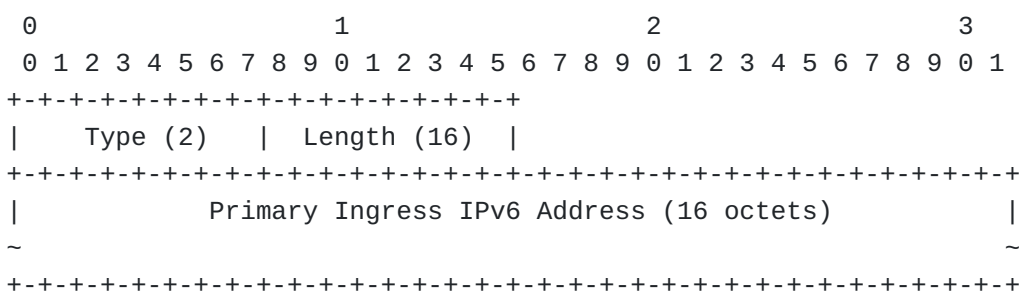

Figure 4: Primary Ingress IPv6 Address Sub-TLV

Type: Its value (2 suggested) is to be assigned by IANA.

Length: 16.

Primary Ingress IPv6 Address:
16 octets. It represents an IPv6 host address of the primary ingress node of a primary SR path.

5.1.2. Service Sub-TLV

A Service Sub-TLV contains a service ID or label to be added into a packet to be carried by a SR path. It has three formats: the first one for the service identified by a label, the second one for the service identified by a service identifier (ID) of 32 bits, and the third one for the service identified by a service identifier (ID) of 128 bits. Their formats are illustrated below.

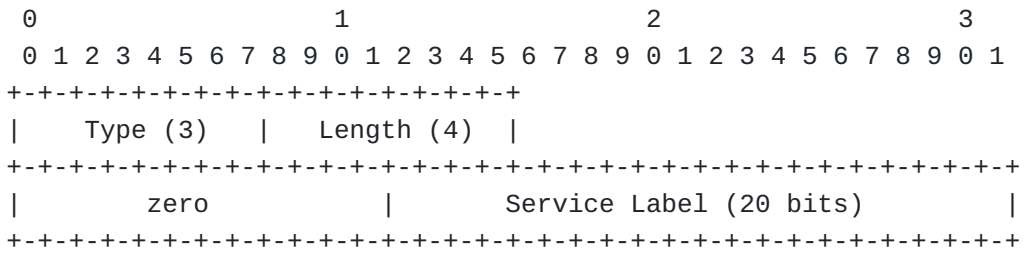

Figure 5: Service Label Sub-TLV

Type: Its value (3 suggested) is to be assigned by IANA.

Length: 4.

Service Label: the least significant 20 bits. It represents a label of 20 bits.

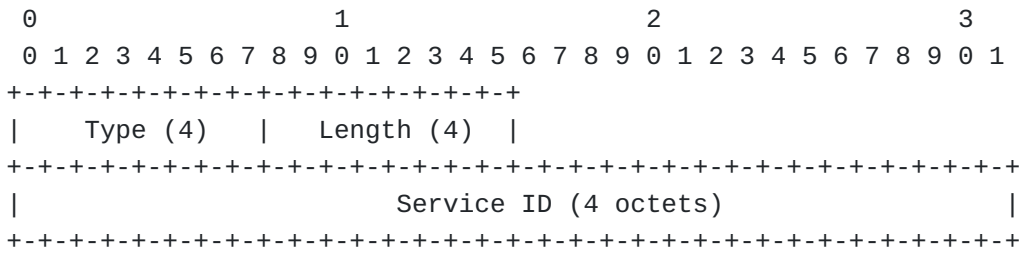

Figure 6: 32 Bits Service ID Sub-TLV

Type: Its value (4 suggested) is to be assigned by IANA.

Length: 4.

Service ID: 4 octets. It represents a Service Identifier (ID) of 32 bits.

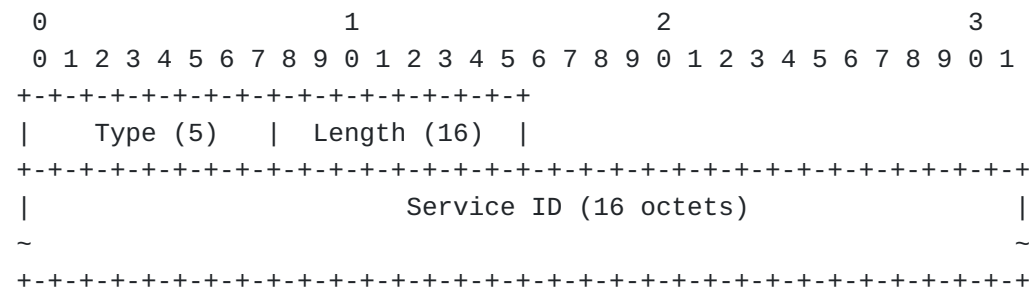

Figure 7: 128 Bits Service ID Sub-TLV

Type: Its value (5 suggested) is to be assigned by IANA.

Length: 16.

Service ID: 16 octets. It represents a Service Identifier (ID) of 128 bits.

5.1.3. Traffic Description Sub-TLVs

A Traffic Description Sub-TLV describes the traffic to be imported into a backup SR path. Five Traffic Description Sub-TLVs are defined. Two of them are FEC Sub-TLVs and the others are interface Sub-TLVs.

Two FEC Sub-TLVs are IPv4 and IPv6 FEC Sub-TLVs. Their formats are illustrated below.

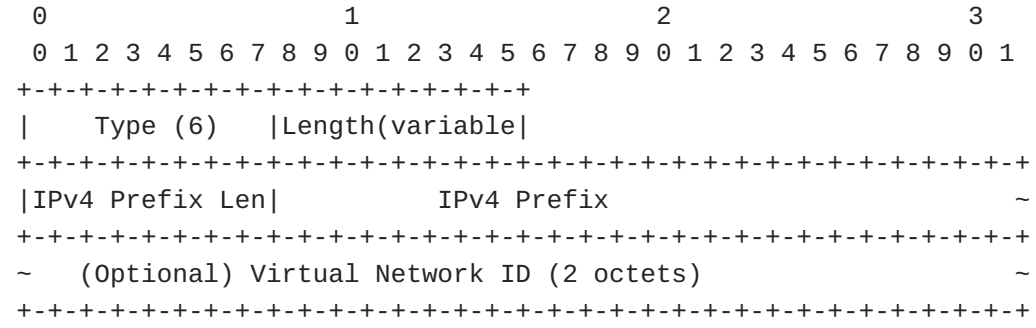

Figure 8: IPv4 FEC Sub-TLV

Type:
Its value (6 suggested) is to be assigned by IANA.

Length: Variable.

IPv4 Prefix Len: Indicates the length of the IPv4 Prefix.

IPv4 Prefix: IPv4 Prefix rounded to octets.

Virtual Network ID: 2 octets. This is optional. It indicates the ID of a virtual network.

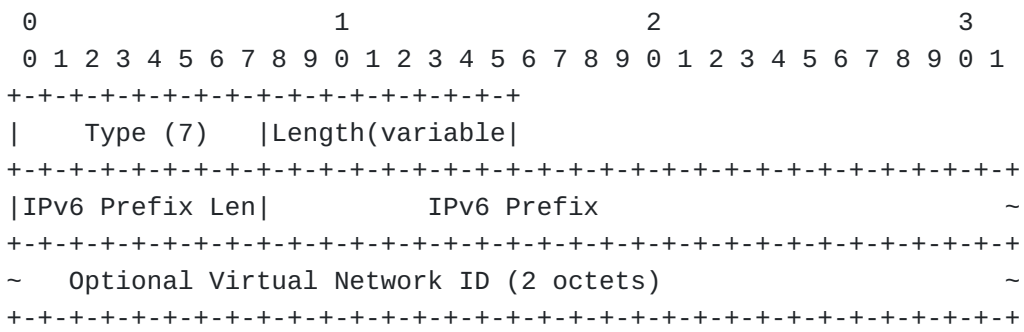

Figure 9: IPv6 FEC Sub-TLV

Type: Its value (7 suggested) is to be assigned by IANA.

Length: Variable.

IPv6 Prefix Len: Indicates the length of the IPv6 Prefix.

IPv6 Prefix: IPv6 Prefix rounded to octets.

Virtual Network ID: 2 octets. This is optional. It indicates the ID of a virtual network.

An Interface sub-TLV indicates the interface from which the traffic is received and imported into the backup SR path/tunnel. It has three formats: one for interface index, the other two for IPv4 and IPv6 address, which are illustrated below.

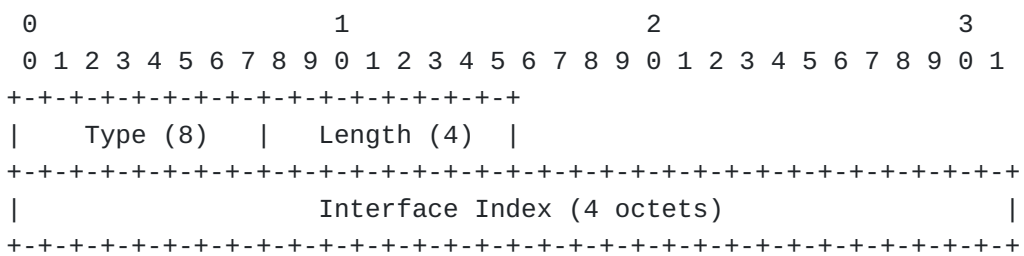

Figure 10: Interface Index Sub-TLV

Type: Its value (8 suggested) is to be assigned by IANA.

Length: 4.

Interface Index: 4 octets. It indicates the index of an interface.

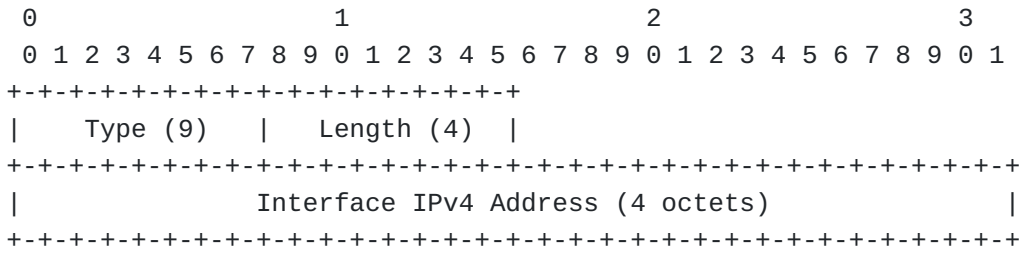

Figure 11: Interface IPv4 Address Sub-TLV

Type: Its value (9 suggested) is to be assigned by IANA.

Length: 4.

Interface IPv4 Address: 4 octets. It represents the IPv4 address of an interface.

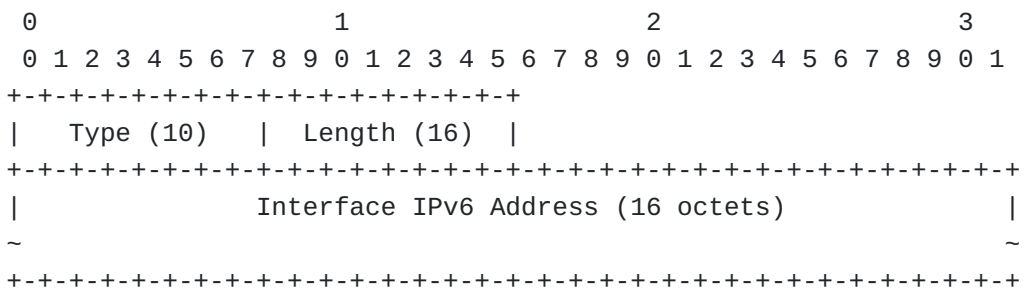

Figure 12: Interface IPv6 Address Sub-TLV

Type: Its value (10 suggested) is to be assigned by IANA.

Length: 16.

Interface IPv6 Address: 16 octets. It represents the IPv6 address of an interface.

6. Backup Ingress Behavior

When a backup ingress node receives a UPDATE message containing the information for protecting the primary ingress node of a SR path, it

installs a forwarding entry in its FIB based on the information. The information is encoded in a SR policy of the following structure:

```
SR Policy SAFI NLRI: <Distinguisher, Policy-Color, Endpoint>
Attributes:
    Tunnel Encaps Attribute (23)
        Tunnel Type (15): SR Policy
            SR Path Ingress Protection Sub-TLV
                Primary Ingress Sub-TLV
                Service Sub-TLV
                Traffic Description Sub-TLV
            Preference Sub-TLV
            Binding SID Sub-TLV
            Explicit NULL Label Policy (ENLP) Sub-TLV
            Priority Sub-TLV
            Policy Name Sub-TLV
            Segment List Sub-TLV
                Weight Sub-TLV
                Segment Sub-TLV
                Segment Sub-TLV
                ...
            ...
```

Where:

- SR Policy SAFI NLRI is defined in [I-D.ietf-idr-segment-routing-te-policy].
- Tunnel Encapsulation Attribute is defined in [I-D.ietf-idr-tunnel-encaps].
- Tunnel Type of SR Policy is defined in [I-D.ietf-idr-segment-routing-te-policy].
- SR Path Ingress Protection, Primary Ingress, Service and Traffic Description Sub-TLVs are defined in this document.
- Preference, Binding SID, ENLP, Priority, Policy Name, Segment List, Weight and Segment Sub-TLVs are defined in [I-D.ietf-idr-segment-routing-te-policy].

After receiving a SR policy with a SR Path Ingress Protection Sub-TLV, the backup ingress node will install one or more candidate paths into its "BGP table". Another module such as SRPM will choose one or more paths and install the forwarding entries for them in the data plane.

The forwarding entries for the paths installed in the data plane will be set to be inactive if the flag A in the SR Path Ingress Protection Sub-TLV is zero. When the primary ingress node fails,

these forwarding entries are set to be active. The failure of the primary ingress may be detected by the backup ingress node through using a mechanism such as BFD. The IP address of the primary ingress in the Primary Ingress Sub-TLV may be used for detecting the failure of the primary ingress node.

If the flag A in the SR Path Ingress Protection Sub-TLV is one, then the forwarding entries for the paths installed in the data plane will be set to be active.

When there is a Service Sub-TLV in the SR Path Ingress Protection Sub-TLV, the ID or Label in the Service Sub-TLV will be included in the forwarding entries. When a packet is imported into a backup SR path using the forwarding entries, the service ID or Label is pushed first and then the sequence of segments represented in the Segment List Sub-TLV.

7. Security Considerations

Protocol extensions defined in this document do not affect the BGP security other than those as discussed in the Security Considerations section of [RFC5575].

8. Acknowledgements

The authors of this document would like to thank Dhruv Dhody for the comments.

9. IANA Considerations

9.1. BGP Tunnel Encapsulation Attribute Sub-TLVs

Under Existing Registry Name: "BGP Tunnel Encapsulation Attribute Sub-TLVs", IANA is requested to assign a new Sub-TLV value for SR Path Ingress Protection as follows:

Value	sub-TLV Name	Reference
TBD1	SR Path Ingress Protection Sub-TLV	This Document

9.2. Ingress Protection Information Sub-TLVs

A new registry called "Ingress Protection Information Sub-TLVs" is defined in this document. IANA is requested to create and maintain new registry:

- Ingress Protection Information Sub-TLVs

Initial values for the registry are given below. The future assignments are to be made through IETF Review [RFC5226].

Value	sub-TLV Name	Reference
0	Reserved	
1	Primary Ingress IPv4 Address Sub-TLV	This Document
2	Primary Ingress IPv6 Address Sub-TLV	This Document
3	Service Label Sub-TLV	This Document
4	32 Bits Service ID Sub-TLV	This Document
5	128 Bits Service ID Sub-TLV	This Document
6	IPv4 FEC Sub-TLV	This Document
7	IPv6 FEC Sub-TLV	This Document
8	Interface Index Sub-TLV	This Document
9	Interface IPv4 Address Sub-TLV	This Document
10	Interface IPv6 Address Sub-TLV	This Document
11-255	Unassigned	

10. References

10.1. Normative References

[I-D.ietf-idr-segment-routing-te-policy] Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., Jain, D., and S. Lin, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-17, 14 April 2022, <https://www.ietf.org/archive/id/draft-ietf-idr-segment-routing-te-policy-17.txt>.

[I-D.ietf-idr-tunnel-encaps] Patel, K., Velde, G. V. D., Sangli, S. R., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", Work in Progress, Internet-Draft, draft-ietf-idr-tunnel-encaps-22, 7 January 2021, <https://www.ietf.org/archive/id/draft-ietf-idr-tunnel-encaps-22.txt>.

[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>.

[RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, DOI 10.17487/RFC7356, September 2014, <https://www.rfc-editor.org/info/rfc7356>.

[RFC8424] Chen, H., Ed. and R. Torvi, Ed., "Extensions to RSVP-TE for Label Switched Path (LSP) Ingress Fast Reroute (FRR) Protection", RFC 8424, DOI 10.17487/RFC8424, August 2018, <https://www.rfc-editor.org/info/rfc8424>.

10.2. Informative References

[I-D.bashandy-rtgwg-segment-routing-ti-lfa]
Bashandy, A., Filsfils, C., Decraene, B., Litkowski, S., Francois, P., Voyer, D., Clad, F., and P. Camarillo, "Topology Independent Fast Reroute using Segment Routing", Work in Progress, Internet-Draft, draft-bashandy-rtgwg-segment-routing-ti-lfa-05, 4 October 2018, <https://www.ietf.org/archive/id/draft-bashandy-rtgwg-segment-routing-ti-lfa-05.txt>.

[I-D.hegde-spring-node-protection-for-sr-te-paths]
Hegde, S., Bowers, C., Litkowski, S., Xu, X., and F. Xu, "Node Protection for SR-TE Paths", Work in Progress, Internet-Draft, draft-hegde-spring-node-protection-for-sr-te-paths-07, 30 July 2020, <https://www.ietf.org/archive/id/draft-hegde-spring-node-protection-for-sr-te-paths-07.txt>.

[I-D.hu-spring-segment-routing-proxy-forwarding]
Hu, Z., Chen, H., Yao, J., Bowers, C., Yongqing, and Yisong, "SR-TE Path Midpoint Restoration", Work in Progress, Internet-Draft, draft-hu-spring-segment-routing-proxy-forwarding-19, 11 April 2022, <https://www.ietf.org/archive/id/draft-hu-spring-segment-routing-proxy-forwarding-19.txt>.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", Work in Progress, Internet-Draft, draft-ietf-spring-segment-routing-policy-22, 22 March 2022, <https://www.ietf.org/archive/id/draft-ietf-spring-segment-routing-policy-22.txt>.

[I-D.sivabalan-pce-binding-label-sid]
Sivabalan, S., Filsfils, C., Tantsura, J., Hardwick, J., Previdi, S., and C. Li, "Carrying Binding Label/Segment-ID in PCE-based Networks.", Work in Progress, Internet-Draft, draft-sivabalan-pce-binding-label-sid-07, 8 July 2019, <https://www.ietf.org/archive/id/draft-sivabalan-pce-binding-label-sid-07.txt>.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, DOI 10.17487/RFC5226, May 2008, <https://www.rfc-editor.org/info/rfc5226>.

[RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", RFC 5462, DOI 10.17487/

RFC5462, February 2009, <https://www.rfc-editor.org/info/rfc5462>.

[RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <https://www.rfc-editor.org/info/rfc5575>.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA,
United States of America

Email: huaimo.chen@futurewei.com

Mehmet Toy
Verizon
United States of America

Email: mehmet.toy@verizon.com

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing
102209
China

Email: wangaj3@chinatelecom.cn

Zhenqiang Li
China Mobile
32 Xuanwumen West Ave, Xicheng District
Beijing
100053
China

Email: lizhengqiang@chinamobile.com

Lei Liu
Fujitsu
United States of America

Email: liulei.kddi@gmail.com

Xufeng Liu
Volta Networks
McLean, VA

United States of America

Email: xufeng.liu.ietf@gmail.com