SR Path Ingress Protection  
draft-chen-pce-sr-ingress-protection-00

Abstract

This document describes protocol extensions and procedures for protecting the ingress node of a Segment Routing (SR) 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 January 7, 2020.
1. Introduction

Fast protection of a transit node of a Segment Routing (SR) path is described in [I-D.bashandy-rtgwg-segment-routing-ti-lfa] and [I-D.hu-spring-segment-routing-proxy-forwarding]. However, these documents do not discuss the procedures for fast protection of the ingress node of an SR path.

This document fills that void and specifies protocol extensions and procedures for fast protection of the ingress node of an SR path. Ingress node and ingress as well as fast protection and protection will be used exchangeably.
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

Figure 1 shows an example of protecting ingress PE1 of a SR path, which is from ingress PE1 to egress 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. In one option, this backup path is from the backup ingress PE2 to ingress PE1's next hop (or endpoint) node P1, where the traffic is "merged" into the SR path, and then sent to egress PE3.

In another option, 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 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, both the backup ingress and the traffic source are concurrently 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). It switches the traffic to the backup ingress (e.g., PE2) when it detects the failure of the ingress.

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 PCE

PCC runs on each of the edge nodes of a network normally. PCE runs on a server as a controller to communicate with PCCs. PCE and PCCs work together to support protection for the ingress of a SR path.

5.1. Capability for SR Path Ingress Protection

When a PCE and a PCC establish a PCEP session between them, they exchange their capabilities of supporting protection for the ingress node of an SR path/tunnel.

A new sub-TLV called SR_INGRESS_PROTECTION_CAPABILITY is defined. It is included in the PATH_SETUP_TYPE_CAPABILITY TLV with PST = TBD1 (suggested value 2 for backup SR path/tunnel) in the OPEN object, which is exchanged in Open messages when a PCC and a PCE establish a PCEP session between them. Its format is illustrated below.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Type = TBD2          |           Length=4            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Reserved            |        Flags              |D|A|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 2: SR_INGRESS_PROTECTION_CAPABILITY sub-TLV

Type: TBD2 is to be assigned by IANA.

Length: 4.

Reserved: 2 octets. Must be set to zero in transmission and ignored on reception.

Flags: 2 octets. Two flags are defined.

- D flag: A PCC sets this flag to 1 to indicate that it is able to detect its adjacent node's failure quickly.
A flag: A PCE sets this flag to 1 to request a PCC to let the forwarding entry for the backup SR path/tunnel be Active.

A PCC, which supports ingress protection for a SR tunnel/path, sends a PCE an Open message containing SR_INGRESS_PROTECTION_CAPABILITY sub-TLV. This sub-TLV indicates that the PCC is capable of supporting the ingress protection for a SR tunnel/path.

A PCE, which supports ingress protection for a SR tunnel/path, sends a PCC an Open message containing SR_INGRESS_PROTECTION_CAPABILITY sub-TLV. This sub-TLV indicates that the PCE is capable of supporting the ingress protection for a SR tunnel/path.

Assume that both a PCC and a PCE support SR_PCE_CAPABILITY, that is that each of the Open messages sent by the PCC and PCE contains PATH-SETUP-TYPE-CAPABILITY TLV with a PST list containing PST=1 and a SR-PCE-CAPABILITY sub-TLV.

If a PCE receives an Open message without a SR_INGRESS_PROTECTION_CAPABILITY sub-TLV from a PCC, then the PCE MUST not send the PCC any request for ingress protection of a SR path/tunnel.

If a PCC receives an Open message without a SR_INGRESS_PROTECTION_CAPABILITY sub-TLV from a PCE, then the PCC MUST ignore any request for ingress protection of a SR path/tunnel from the PCE.

If a PCC sets D flag to zero, then the PCE SHOULD send the PCC an Open message with A flag set to one. When the PCE sends the PCC a message for initiating a backup SR path/tunnel, the PCC SHOULD let the forwarding entry for the backup SR path/tunnel be Active.

5.2. SR Path Ingress Protection

A new sub-TLV called SR_INGRESS_PROTECTION is defined. When a PCE sends a PCC a PCInitiate message for initiating a backup SR path/tunnel to protect the primary ingress node of a primary SR path/tunnel, the message contains this TLV in the RP/SRP object. Its format is illustrated below.
Type: TBD3 is to be assigned by IANA.

Length: Variable.

Reserved: 2 octets. Must be set to zero in transmission and ignored on reception.

Flags: 2 octets. One flag is defined.

- A flag: A PCE sets this flag to 1 to request a PCC to let the forwarding entry for the backup SR path/tunnel be Active.

Four optional sub-TLVs are defined.

5.2.1. Traffic-Description sub-TLV

A Traffic-Description sub-TLV describes the traffic to be imported into a backup SR path/tunnel. Its format is illustrated below.

Figure 4: Traffic-Description sub-TLV

Type: TBD4 is to be assigned by IANA.

Length: Variable.
Two optional sub-TLVs are defined. One is FEC sub-TLV and the other interface sub-TLV.

A FEC sub-TLV describes the traffic to be imported into the backup SR path/tunnel. It is an IP prefix with an optional virtual network ID. It has two formats: one for IPv4 and the other for IPv6, which are illustrated below.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type = TBD5           |        Length (variable)      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|IPv4 Prefix Len|          IPv4 Prefix                          ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~   (Optional) Virtual Network ID (2 octets)                    ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Figure 5: IPv4 FEC sub-TLV**

Type: TBD5 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.

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type = TBD6           |        Length (variable)      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|IPv6 Prefix Len|          IPv6 Prefix                          ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~   Optional Virtual Network ID (2 octets)                    ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Figure 6: IPv6 FEC sub-TLV**

Type: TBD6 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 7: Interface Index sub-TLV

Type: TBD7 is to be assigned by IANA.

Length: 4.

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

Figure 8: Interface IPv4 Address sub-TLV

Type: TBD8 is to be assigned by IANA.

Length: 4.

Interface IPv4 Address: 4 octets. It represents the IPv4 address of an interface.
5.2.2. Primary-Ingress sub-TLV

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

```
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = TBDa | Length (4) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Primary Ingress IPv4 Address (4 octets) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
Figure 10: Primary Ingress IPv4 Address sub-TLV
```

Type: TBDa 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 SR path/tunnel.
5.2.3. 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/tunnel. It has two formats: one for the service identified by a label and the other for the service identified by a service identifier (ID) of 32 or 128 bits, which are illustrated below.

```plaintext
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = TBDc | Length (4) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| zero | Service Label (20 bits) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Figure 12: Service Label sub-TLV**

Type: TBDc is to be assigned by IANA.

Length: 4.

Service Label: the least significant 20 bits. It represents a label of 20 bits.
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type = TBDd           |         Length (4/16)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Service ID (4 or 16 octets)            |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
Figure 13: Service ID sub-TLV

Type:  TBDd is to be assigned by IANA.

Length:  4 or 16.

Service ID:  4 or 16 octets.  It represents Identifier (ID) of a service in 4 or 16 octets.

5.2.4.  Downstream-Node sub-TLV

A Downstream-Node sub-TLV gives the IP address of the downstream endpoint node of the primary ingress along the primary SR path/tunnel. It has two formats: one for downstream node IPv4 address and the other for downstream node IPv6 address, which are illustrated below.

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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|         Type = TBDe           |          Length (4)           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Downstream Node IPv4 Address (4 octets)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
Figure 14: Downstream-Node IPv4 Address sub-TLV

Type:  TBDe is to be assigned by IANA.

Length:  4.

Downstream Node IPv4 Address:  4 octets.  It represents an IPv4 host address of the downstream endpoint node of the primary ingress node of a SR path/tunnel.
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 = TBDf           |         Length (16)           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Downstream Node IPv6 Address (16 octets) |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

Figure 15: Downstream-Node IPv6 Address sub-TLV

Type: TBDf is to be assigned by IANA.

Length: 16.

Downstream Node IPv6 Address: 4 octets. It represents an IPv6 host address of the downstream endpoint node of the primary ingress node of a SR path/tunnel.

6. IANA Considerations
   TBD

7. Security Considerations
   TBD

8. Acknowledgements
   TBD

9. References

9.1. Normative References

[I-D.bashandy-isis-srv6-extensions]

[I-D.hu-spring-segment-routing-proxy-forwarding]
[I-D.ietf-isis-segment-routing-extensions]
Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A.,
Gredler, H., and B. Decraene, "IS-IS Extensions for
Segment Routing", draft-ietf-isis-segment-routing-
extensions-25 (work in progress), May 2019.

[I-D.ietf-ospf-segment-routing-extensions]
Psenak, P., Previdi, S., Filsfils, C., Gredler, H.,
Shakir, R., Henderickx, W., and J. Tantsura, "OSPF
Extensions for Segment Routing", draft-ietf-ospf-segment-
routing-extensions-27 (work in progress), December 2018.

[I-D.li-ospf-ospfv3-srv6-extensions]
Li, Z., Hu, Z., Cheng, D., Talaulikar, K., and P. Psenak,
"OSPFv3 Extensions for SRv6", draft-li-ospf-
ospfv3-srv6-extensions-03 (work in progress), March 2019.

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

[RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding
Scope Link State PDUs (LSPs)", RFC 7356,
DOI 10.17487/RFC7356, September 2014,

9.2. Informative References

[I-D.bashandy-rtgwg-segment-routing-ti-lfa]
Bashandy, A., Filsfils, C., Decraene, B., Litkowski, S.,
Francois, P., daniel.voyer@bell.ca, d., Clad, F., and P.
Camarillo, "Topology Independent Fast Reroute using
Segment Routing", draft-bashandy-rtgwg-segment-routing-ti-
lfa-05 (work in progress), October 2018.

[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", draft-hegde-spring-
node-protection-for-sr-te-paths-05 (work in progress),
July 2019.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Sivabalan, S., daniel.voyer@bell.ca, d.,
bogdanov@google.com, b., and P. Mattes, "Segment Routing
Policy Architecture", draft-ietf-spring-segment-routing-
policy-03 (work in progress), May 2019.
[I-D.sivabalan-pce-binding-label-sid]


Authors’ Addresses

Huaimo Chen
Futurewei
Boston, MA
USA
Email: Huaimo.chen@futurewei.com

Mehmet Toy
Verizon
USA
Email: mehmet.toy@verizon.com

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing 102209
China
Email: wangaj.bri@chinatelecom.cn

Zhenqiang Li
China Mobile
32 Xuanwumen West Ave, Xicheng District
Beijing 100053
China
Email: lizhengqiang@chinamobile.com