

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 12, 2022

H. Chen
M. McBride
Futurewei
G. Mishra
Verizon Inc.
Y. Liu
China Mobile
A. Wang
China Telecom
L. Liu
Fujitsu
X. Liu
Volta Networks
July 11, 2021

PCE for BIER-TE Ingress Protection
draft-chen-pce-bier-te-ingress-protect-00

Abstract

This document describes extensions to Path Computation Element (PCE) communication Protocol (PCEP) for protecting the ingress of a BIER-TE 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 12, 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Terminologies	3
3. BIER-TE Path Ingress Protection Example	3
4. Behavior around Ingress Failure	4
4.1. Source Detect	5
4.2. Backup Ingress Detect	5
4.3. Both Detect	5
5. Extensions to PCEP	5
5.1. Capability for Ingress Protection	6
5.1.1. Capability for Ingress Protection with Backup Ingress	6
5.1.2. Capability for Ingress Protection with Traffic Source	7
5.2. BIER-TE Path Ingress Protection	8
5.2.1. Extensions for Backup Ingress	8
5.2.2. Extensions for Traffic Source	11
6. IANA Considerations	14
7. Security Considerations	14
8. Acknowledgements	14
9. References	14
9.1. Normative References	14
9.2. Informative References	14
Authors' Addresses	15

1. Introduction

The fast protection of a transit node of a "Bit Index Explicit Replication" (BIER) Traffic Engineering (BIER-TE) path or tunnel is described in [I-D.chen-bier-te-frr]. [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 a BIER-TE path or tunnel.

This document fills that gap and specifies protocol extensions to Path Computation Element (PCE) communication Protocol (PCEP) for the fast protection of the ingress node of a BIER-TE path or tunnel. Ingress node and ingress, fast protection and protection as well as BIER-TE path and BIER-TE tunnel will be used exchangeably in the following sections.

2. Terminologies

The following terminologies are used in this document.

PCE: Path Computation Element

PCEP: PCE communication Protocol

PCC: Path Computation Client

BIER: Bit Index Explicit Replication

CE: Customer Edge

PE: Provider Edge

TE: Traffic Engineering

3. BIER-TE Path Ingress Protection Example

Figure 1 shows an example of protecting ingress PE1 of a BIER-TE path, which is from ingress PE1 to egress nodes PE3 and PE4. This primary BIER-TE path is represented by *** in the figure. The ingress of the primary BIER-TE path is called primary ingress.

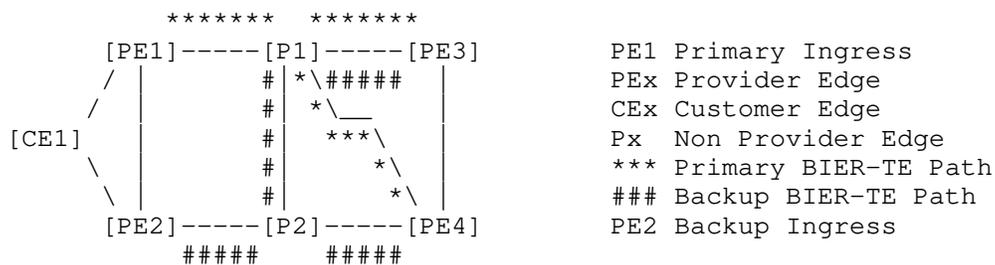


Figure 1: Protecting Ingress PE1 of BIER-TE Path

The backup BIER-TE path is from ingress PE2 to egress nodes PE3 and PE4, which is represented by ### in the figure. The ingress of the backup BIER-TE path is called backup ingress.

In normal operations, CE1 sends the packets with a multicast group and source to ingress PE1, which imports/encapsulates the packets into the BIER-TE path through adding a BIER-TE header. The header contains the BIER-TE path from ingress PE1 to egress nodes PE3 and PE4.

When CE1 detects the failure of ingress PE1 using a failure detection mechanism such as BFD, it switches the traffic to backup ingress PE2, which imports the traffic from CE1 into the backup BIER-TE path. When the traffic is imported into the backup path, it is sent to the egress nodes PE3 and PE4 along the path.

Given the traffic source (e.g., CE1), ingress (e.g., PE1) and egresses (e.g., PE3 and PE4) of the primary BIER-TE path, the PCC computes a backup ingress (e.g., PE2), a backup BIER-TE path from the backup ingress to the egresses, and sends the backup BIER-TE path to the PCC of the backup ingress. It also sends the backup ingress, primary ingress and the traffic description to the PCC of the traffic source (e.g., CE1).

When the PCC of the traffic source receives the backup ingress, primary ingress and traffic description, it sets up the fast detection of the primary ingress failure and the switch over target backup ingress. This setup lets the traffic source node switch the traffic (to be sent to the primary ingress) to the backup ingress when it detects the failure of the primary ingress.

When the PCC of the backup ingress receives the backup BIER-TE path, it adds a forwarding entry into its BIFT. This entry encapsulates the packets from the traffic source in the backup BIER-TE path. This makes the backup ingress send the traffic received from the traffic source to the egress nodes via the backup BIER-TE path.

4. Behavior around Ingress Failure

This section describes the behavior of some nodes connected to the ingress before and after the ingress fails. These nodes are the traffic source (e.g., CE1) and the backup ingress (e.g., PE2). It presents three ways in which these nodes work together to protect the ingress. The first way is called source detect, where the traffic source is responsible for fast detecting the failure of the ingress. The second way is called backup ingress detect, in which the backup ingress is responsible for fast detecting the failure of the ingress. The third way is called both detect, where both the traffic source and the backup ingress are responsible for fast detecting the failure of the ingress.

4.1. Source Detect

In normal operations, i.e., before the failure of the ingress, the traffic source sends the traffic to the ingress of the primary BIER-TE path. The backup ingress (e.g., PE2) is ready to import the traffic from the traffic source into the backup BIER-TE path installed.

When the traffic source detects the failure of the ingress, it switches the traffic to the backup ingress, which delivers the traffic to the egress nodes of the BIER-TE path via the backup BIER-TE path.

4.2. Backup Ingress Detect

The traffic source (e.g., CE1) always sends the traffic to both the ingress (e.g., PE1) of the primary BIER-TE path and the backup ingress (e.g., PE2).

The backup ingress does not import any traffic from the traffic source into the backup BIER-TE path in normal operations. When it detects the failure of the ingress of the primary BIER-TE path, it imports the traffic from the source into the backup BIER-TE path.

For the backup ingress to fast detect the failure of the primary ingress, it SHOULD directly connect to the primary ingress. When a PCE computes a backup ingress and a backup BIER-TE path, it SHOULD consider this.

4.3. Both Detect

In normal operations, i.e., before the failure of the ingress, the traffic source sends the traffic to the ingress of the primary BIER-TE path. When it detects the failure of the ingress, it switches the traffic to the backup ingress.

The backup ingress does not import any traffic from the traffic source into the backup BIER-TE path in normal operations. When it detects the failure of the ingress of the primary BIER-TE path, it imports the traffic from the source into the backup BIER-TE path.

5. Extensions to PCEP

A PCC runs on each of the edge nodes such as PEs and CEs of a network normally. A PCE runs on a server as a controller to communicate with PCCs. The PCE and the PCCs running on backup ingress PEs and traffic source CEs work together to support protection for the ingress of a BIER-TE path.

5.1. Capability for Ingress Protection

5.1.1. Capability for Ingress Protection with Backup Ingress

When a PCE and a PCC running on a backup ingress establish a PCEP session between them, they exchange their capabilities of supporting protection for the ingress node of a BIER-TE path/tunnel.

A new sub-TLV called BIER-TE_INGRESS_PROTECTION_CAPABILITY is defined. It is included in the PATH_SETUP_TYPE_CAPABILITY TLV with PST = TBD1 (suggested value 2 for protecting the ingress of a BIER-TE 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.

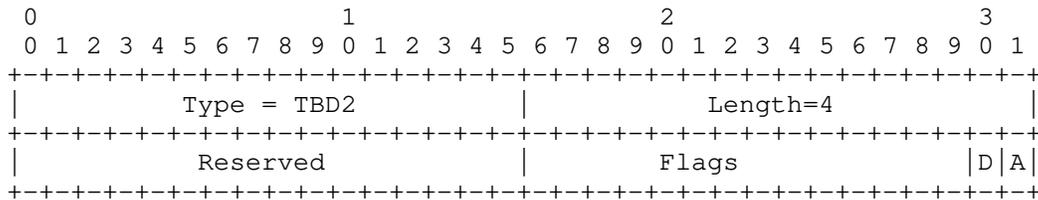


Figure 2: BIER-TE_INGRESS_PROTECTION_CAPABILITY sub-TLV

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 flag bits are defined.

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

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

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

If both a PCC and a PCE support BIER-TE_INGRESS_PROTECTION_CAPABILITY, each of the Open messages sent by the PCC and PCE contains PATH-SETUP-TYPE-CAPABILITY TLV with a PST list containing PST=TBD1 and a BIER-TE-INGRESS_PROTECTION_CAPABILITY sub-TLV.

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

If a PCC receives an Open message without a BIER-TE_INGRESS_PROTECTION_CAPABILITY sub-TLV from a PCE, then the PCC MUST ignore any request for ingress protection of a BIER-TE 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 and the fast detection of the failure of the primary ingress MUST be done by the traffic source. When the PCE sends the PCC a message for initiating a backup BIER-TE path, the PCC MUST let the forwarding entry for the backup BIER-TE path be Active.

5.1.2. Capability for Ingress Protection with Traffic Source

When a PCE and a PCC running on a traffic source node establish a PCEP session between them, they exchange their capabilities of supporting protection for the ingress node of a BIER-TE path/tunnel.

The PCECC-CAPABILITY sub-TLV defined in [I-D.ietf-pce-pcep-extension-for-pce-controller] is included in the OPEN object in the PATH-SETUP-TYPE-CAPABILITY TLV, which is exchanged in Open messages when a PCC and a PCE establish a PCEP session between them.

A new flag bit P is defined in the Flags field of the PCECC-CAPABILITY sub-TLV:

- o P flag (for Ingress Protection): if set to 1 by a PCEP speaker, the P flag indicates that the PCEP speaker supports and is willing to handle the PCECC based central controller instructions for ingress protection. The bit MUST be set to 1 by both a PCC and a

PCE for the PCECC ingress protection instruction download/report on a PCEP session.

5.2. BIER-TE Path Ingress Protection

This section specifies the extensions to PCEP for the backup ingress and the traffic source. The extensions let the traffic source

S1: fast detect the failure of the primary ingress and switch the traffic to the backup ingress when the traffic source detects the failure of the primary ingress, or

S2: always send the traffic to both the primary ingress and the backup ingress.

The extensions let the backup ingress

B1: always import the traffic received from the traffic source with possible service ID into the backup BIER-TE path, or

B2: import the traffic with possible service ID into the backup BIER-TE path when the backup ingress detects the failure of the primary ingress.

The following lists the combinations of S_i and B_i ($i = 1,2$) for different ways of failure detects.

Source Detect: S_1 and B_1 .

Backup Ingress Detect: S_2 and B_2 .

Both Detect: S_1 and B_2 .

5.2.1. Extensions for Backup Ingress

For the packets from the traffic source, if the primary ingress (i.e., the ingress of the primary BIER-TE path) encapsulates the packets with a service ID or label into the BIER-TE path, the backup ingress MUST have this service ID or label and encapsulates the packets with the service ID or label into the backup BIER-TE path when the primary ingress fails.

If the backup ingress is requested to detect the failure of the primary ingress, it MUST have the information about the primary ingress such as the address of the primary ingress.

A new TLV called BIER-TE_INGRESS_PROTECTION TLV is defined to transfer the information about the primary ingress and/or the service

ID or label. When a PCE sends the PCC of a backup ingress a PCInitiate message for initiating a backup BIER-TE path/tunnel to protect the primary ingress of a primary BIER-TE path/tunnel, the message contains this TLV in the RP/SRP object. Its format is illustrated below.

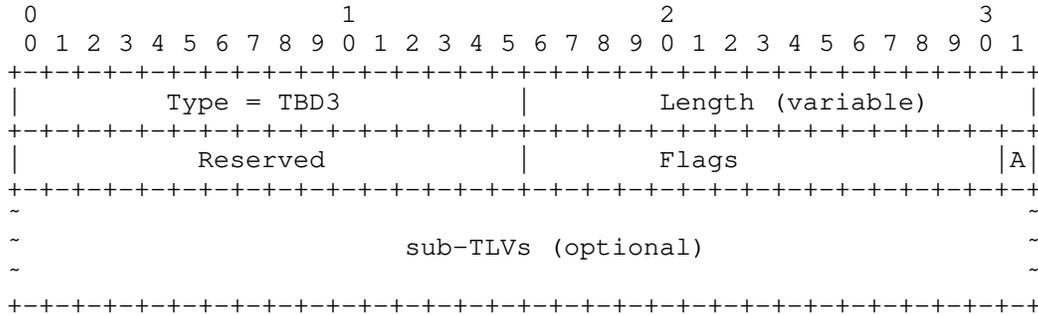


Figure 3: BIER-TE_INGRESS_PROTECTION TLV

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 bit is defined.

A flag bit: it is set to 1 or 0 by PCE.

- o 1 is to request the backup ingress to let the forwarding entry for the backup BIER-TE path/tunnel be Active always. In this case, the traffic source detects the failure of the primary ingress and switches the traffic to the backup ingress when it detects the failure.
- o 0 is to request the backup ingress to detect the failure of the primary ingress and let the forwarding entry for the backup BIER-TE path/tunnel be Active when the primary ingress fails. In this case, the TLV includes the primary ingress address in a Primary-Ingress sub-TLV. The traffic source can send the traffic to both the primary ingress and the backup ingress. It may switch the traffic to the backup ingress from the primary ingress when it detects the failure of the primary ingress.

Two optional sub-TLVs are defined. One is Service sub-TLV. The other is Primary-Ingress sub-TLV. The Multicast Flow Specification TLV for IPv4 or IPv6, which is defined in [I-D.ietf-pce-pcep-flowspec], is used as a sub-TLV to indicate the traffic to be imported into the backup BIER-TE path.

5.2.1.1. Service sub-TLV

A Service sub-TLV contains a service label such as VPN service label or ID to be added into a packet to be carried by a BIER-TE 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.

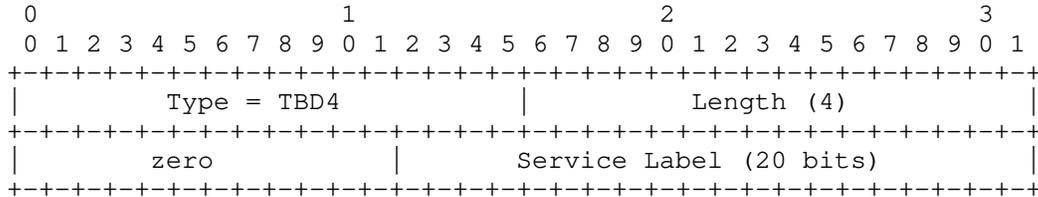


Figure 4: Service Label sub-TLV

Type: TBD4 is to be assigned by IANA.

Length: 4.

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

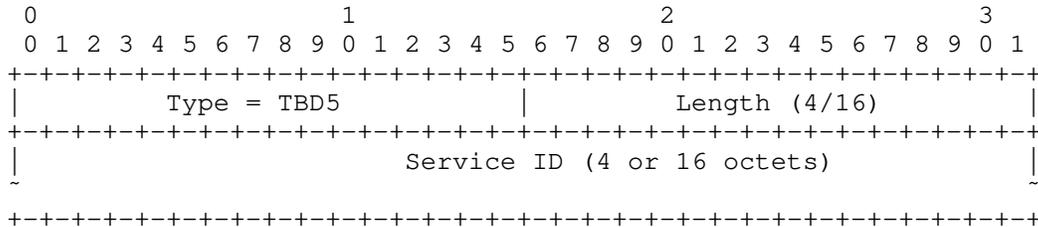


Figure 5: Service ID sub-TLV

Type: TBD5 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.

5.2.1.2. Primary-Ingress sub-TLV

A Primary-Ingress sub-TLV indicates the IP address of the primary ingress node of a primary BIER-TE 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.

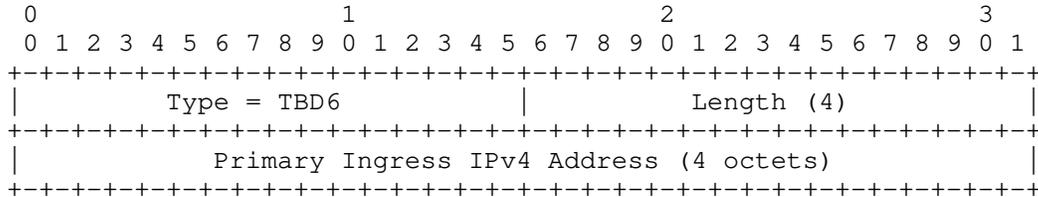


Figure 6: Primary Ingress IPv4 Address sub-TLV

Type: TBD6 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 BIER-TE path/tunnel.

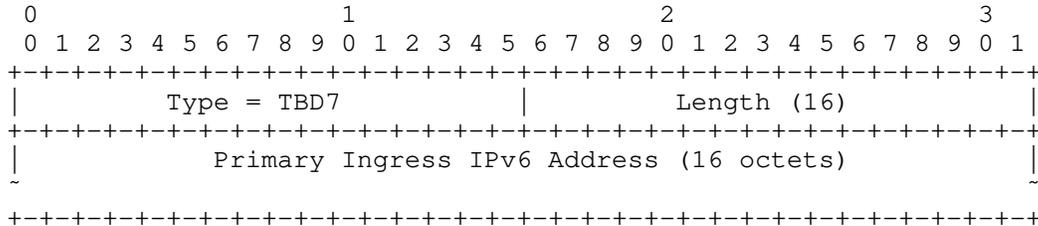


Figure 7: Primary Ingress IPv6 Address sub-TLV

Type: TBD7 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 BIER-TE path/tunnel.

5.2.2. Extensions for Traffic Source

If the traffic source is requested to detect the failure of the primary ingress and switch the traffic (to be sent to the primary ingress) to the backup ingress when the primary ingress fails, it MUST have the information about the backup ingress, the primary

ingress and the traffic. This information may be transferred via a CCI object for BIER-TE-INGRESS-PROTECTION to the PCC of the traffic source node from a PCE.

If the traffic source PCC does not accept the request from the PCE or support the extensions, the PCE SHOULD have the information about the behavior of the traffic source configured such as whether it detects the failure of the primary ingress. Based on the information, the PCE instructs the backup ingress accordingly.

The Central Control Instructions (CCI) Object is defined in [I-D.ietf-pce-pcep-extension-for-pce-controller] for a PCE as a controller to send instructions for LSPs to a PCC. This document defines a new object-type (TBDt) for BIER-TE ingress protection based on the CCI object. The body of the object with the new object-type is illustrated below. The object may be in PCRpt, PCUpd, or PCInitiate message.

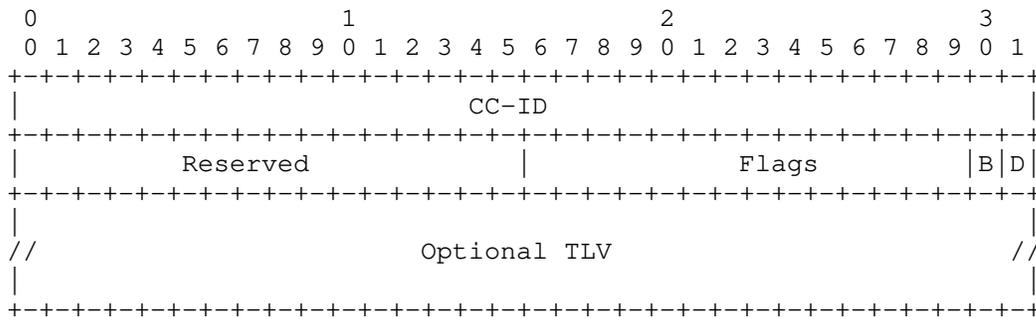


Figure 8: BIER-TE-INGRESS-PROTECTION Object Body

CC-ID: It is the same as described in [I-D.ietf-pce-pcep-extension-for-pce-controller].

Flags: Two flag bits D and B are defined as follows:

D: D = 1 instructs the PCC of the traffic source to Detect the failure of the primary ingress and switch the traffic to the backup ingress when it detects the failure.

B: B = 1 instructs the PCC of the traffic source to send the traffic to Both the primary ingress and the backup ingress.

Optional TLV: Primary ingress TLV, backup ingress TLV and/or Multicast Flow Specification TLV.

The primary ingress sub-TLV defined above is used as a TLV to contain the information about the primary ingress in the object. The Multicast Flow Specification TLV for IPv4 or IPv6, which is defined in [I-D.ietf-pce-pcep-flowspec], is used to contain the information about the traffic in the object. A new TLV, called backup ingress TLV, is defined to contain the information about the backup ingress in the object.

5.2.2.1. Backup-Ingress TLV

A Backup-Ingress TLV indicates the IP address of the ingress node of a backup BIER-TE path/tunnel. It has two formats: one for backup ingress node IPv4 address and the other for backup ingress node IPv6 address, which are illustrated below. They have the same format as the Primary-Ingress sub-TLVs.

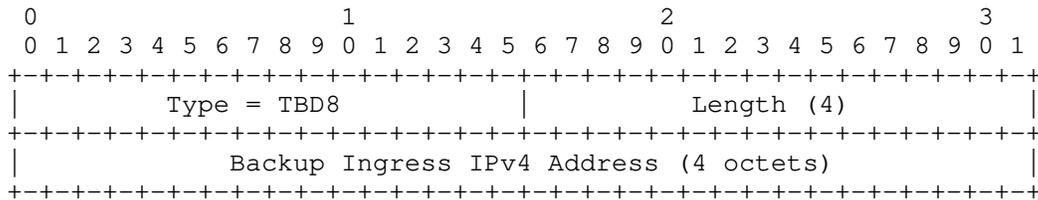


Figure 9: Backup Ingress IPv4 Address TLV

Type: TBD8 is to be assigned by IANA.

Length: 4.

Backup Ingress IPv4 Address: 4 octets. It represents an IPv4 host address of the backup ingress.

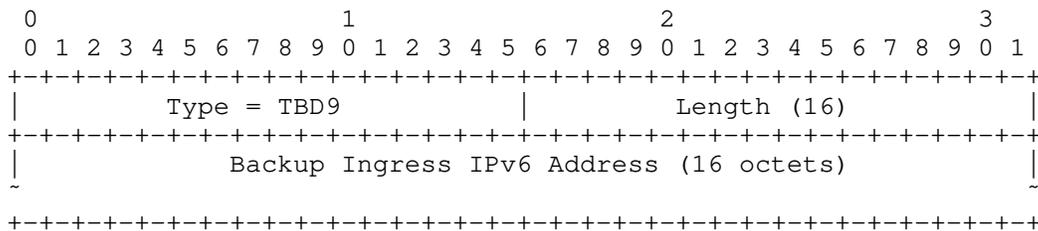


Figure 10: Backup Ingress IPv6 Address TLV

Type: TBD9 is to be assigned by IANA.

Length: 16.

Backup Ingress IPv6 Address: 16 octets. It represents an IPv6 host address of the backup ingress node.

6. IANA Considerations

TBD

7. Security Considerations

TBD

8. Acknowledgements

TBD

9. References

9.1. 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>>.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, DOI 10.17487/RFC8231, September 2017, <<https://www.rfc-editor.org/info/rfc8231>>.

9.2. Informative References

- [I-D.chen-bier-te-frr] Chen, H., McBride, M., Liu, Y., Wang, A., Mishra, G. S., Fan, Y., Liu, L., and X. Liu, "BIER-TE Fast ReRoute", draft-chen-bier-te-frr-00 (work in progress), February 2021.

- [I-D.ietf-pce-pcep-extension-for-pce-controller]
Li, Z., Peng, S., Negi, M. S., Zhao, Q., and C. Zhou,
"PCEP Procedures and Protocol Extensions for Using PCE as
a Central Controller (PCECC) of LSPs", draft-ietf-pce-
pcep-extension-for-pce-controller-14 (work in progress),
March 2021.
- [I-D.ietf-pce-pcep-flowspec]
Dhody, D., Farrel, A., and Z. Li, "PCEP Extension for Flow
Specification", draft-ietf-pce-pcep-flowspec-12 (work in
progress), October 2020.
- [RFC8402] Filtsfils, 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>>.
- [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>>.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA
USA

Email: Huaimo.chen@futurewei.com

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring MD 20904
USA

Phone: 301 502-1347
Email: gyan.s.mishra@verizon.com

Yisong Liu
China Mobile

Email: liuyisong@chinamobile.com

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

Email: wangaj3@chinatelecom.cn

Lei Liu
Fujitsu

USA

Email: liulei.kddi@gmail.com

Xufeng Liu
Volta Networks

McLean, VA
USA

Email: xufeng.liu.ietf@gmail.com