

Networking Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 7, 2020

Ran. Chen
Zheng. Zhang
ZTE Corporation
Senthil. Dhanaraj
Huawei
Fengwei. Qin
China Mobile
November 4, 2019

PCEP Extensions for BIER-TE
draft-chen-pce-bier-06

Abstract

Bit Index Explicit Replication (BIER)-TE shares architecture and packet formats with BIER as described in [RFC8279]. BIER-TE forwards and replicates packets based on a BitString in the packet header, but every BitPosition of the BitString of a BIER-TE packet indicates one or more adjacencies as described in [I-D.ietf-bier-te-arch]. BIER-TE Path can be derived from a Path Computation Element (PCE).

This document specifies extensions to the Path Computation Element Protocol (PCEP) that allow a PCE to compute and initiate the path for the BIER-TE.

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 May 7, 2020.

Copyright Notice

Copyright (c) 2019 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. Conventions used in this document	3
3. Overview of PCEP Operation in BIER Networks	3
4. Object Formats	3
4.1. The OPEN Object	3
4.1.1. The BIER-TE PCE Capability sub-TLV	3
4.2. The RP/SRP Object	4
4.3. END-POINTS object	4
4.4. ERO Object	4
4.4.1. BIER-ERO Subobject	5
4.4.2. BIER-ERO Processing	6
5. Security Considerations	6
6. IANA Considerations	6
6.1. PCEP Objects	6
6.1.1. BIER-TE-PCE-CAPABILITY Sub-TLV Type Indicators	7
6.1.2. New Path Setup Type	7
6.1.3. BIER-ERO Subobject	7
6.1.4. PCEP-Error Objects and Types	7
7. Normative references	8
Authors' Addresses	9

1. Introduction

Bit Index Explicit Replication (BIER)-TE shares architecture and packet formats with BIER as described in [RFC8279]. BIER-TE forwards and replicates packets based on a BitString in the packet header, but every BitPosition of the BitString of a BIER-TE packet indicates one or more adjacencies as described in [I-D.ietf-bier-te-arch]. BIER-TE Path can be derived from a Path Computation Element (PCE).

[RFC8231] specifies a set of extensions to PCEP that allow a PCE to compute and recommend network paths in compliance with [RFC4657] and defines objects and TLVs for MPLS-TE LSPs.

This document uses a PCE for computing one or more BIER-TE paths taking into account various constraints and objective functions.

2. Conventions used in this document

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 RFC2119.

3. Overview of PCEP Operation in BIER Networks

BIER-TE forwards and replicates packets based on a BitString in the packet header, and every BitPosition of the BitString of a BIER-TE packet indicates one or more adjacencies as described in [I-D.ietf-bier-te-arch]. In a PCEP session, An ERO object specified in [RFC5440] can be extended to carry a BIER-TE path consists of one or more BIER-ERO subobject(s). BIER-TE computed by a PCE can be represented in the following forms:

- o An ordered set of adjacencies BitString(s) in which each bit represents that the adjacencies to which the BFR should replicate packets to in the domain.

In this document, we define a set of PCEP protocol extensions, including a new PCEP capability, a new Path Setup Type (PST), a new BIER END-POINT Object, new ERO subobjects, new PCEP error codes and procedures.

4. Object Formats

4.1. The OPEN Object

4.1.1. The BIER-TE PCE Capability sub-TLV

[RFC8408] defines the PATH-SETUP-TYPE-CAPABILITY TLV for use in the OPEN object. The PATH-SETUP-TYPE-CAPABILITY TLV contains an optional list of sub-TLVs which are intended to convey parameters that are associated with the path setup types supported by a PCEP speaker.

This document defines a new Path Setup Type (PST) for BIER as follows:

- o PST = TBD2: Path is setup using BIER Traffic Engineering technique.

A PCEP speaker MUST indicate its support of the function described in this document by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this new PST included in the PST list.

This document also defines the BIER-TE-PCE-CAPABILITY sub-TLV. PCEP speakers use this sub-TLV to exchange BIER capability. If a PCEP

speaker includes PST=TBD1 in the PST List of the PATH-SETUP-TYPE-CAPABILITY TLV then it MUST also include the BIER-TE-PCE- CAPABILITY sub-TLV inside the PATH-SETUP-TYPE-CAPABILITY TLV.

The format of the BIER-TE-PCE-CAPABILITY sub-TLV is shown in the following figure:

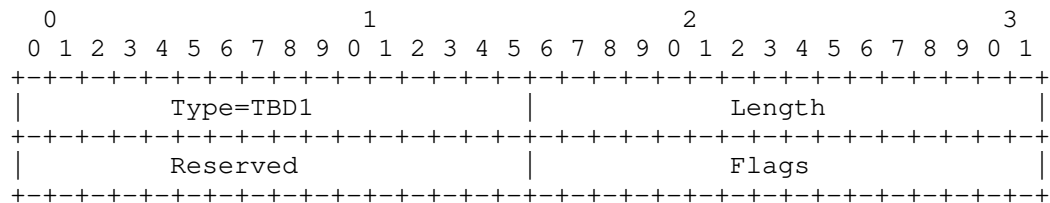


Figure 1 BIER-TE-PCE-CAPABILITY sub-TLV format

The code point for the TLV type is to be defined by IANA.

Length: 4 bytes.

The "Reserved" (2 octet) and "Flags" (2 octet) fields are currently unused, and MUST be set to zero on transmission and ignored on reception.

4.2. The RP/SRP Object

In order to setup an BIER-TE, a new PATH-SETUP-TYPE TLV MUST be contained in RP/SRP object. This document defines a new Path Setup Type (PST=TBD2) for BIER-TE.

4.3. END-POINTS object

The END-POINTS object which is defined in [RFC8306] is used in a PCReq message to specify the BIER information of the path for which a path computation is requested. To represent the end points for a BIER path efficiently, we reuse the P2MP END-POINTS object body for IPv4 (Object-Type 3) and END-POINTS object body for IPv6 (Object-Type 4) which is defined in [RFC8306].

4.4. ERO Object

BIER-TE consists of one or more adjacencies BitStrings where every BitPosition of the BitString indicates one or more adjacencies, as described in ([RFC8279]).

The ERO object specified in [RFC5440] is used to encode the path of a TE LSP through the network. The ERO is carried within a PCRep message to provide the computed TE LSP if the path computation was successful. In order to carry BIER-TE explicit paths, this document defines a new ERO subobjects referred to as "BIER-ERO subobjects" whose formats are specified in the following section. An BIER-ERO subobjects carrying a adjacencies BitStrings consists of one or more BIER-ERO subobject(s).

4.4.1. BIER-ERO Subobject

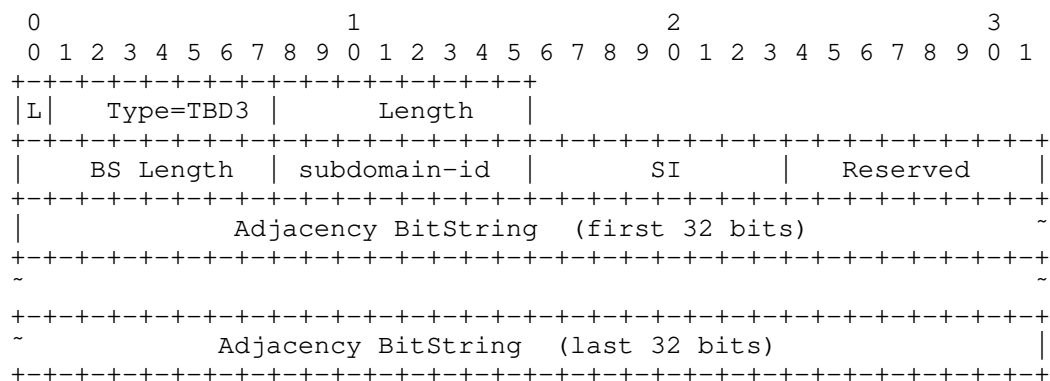


Figure 3

The 'L' Flag: Indicates whether the subobject represents a loose-hop in the LSP[RFC3209]. If the bit is not set, the subobject represents a strict hop in the explicit route.

Type: TBD3

Length: 1 octet ([RFC3209]). Contains the total length of the subobject in octets. The Length MUST be at least 8, and MUST be a multiple of 4.

BS Length: A 1 octet field encodes the length in bits of the BitString as per [RFC8296], the maximum length of the BitString is 5, it indicates the length of BitString is 1024. It is used to refer to the number of bits in the BitString.

subdomain-id: Unique value identifying the BIER subdomain. 1 octet.

SI: Set Identifier (Section 1 of [RFC8279] used in the encapsulation for this BIER subdomain for this BitString length, 1 octet.

The "Reserved" (1 octets) fields are currently unused, and MUST be set to zero on transmission and ignored on reception.

Adjacency BitString: a variable length field encoding the Adjacency BitString where every BitPosition of the BitString indicates one or more adjacencies. the length of this field is according the BS length. The minimum value of this field is 64 bits, and the maximum value of this field is 1024 bits.

Notice:

The maximum value of BS Length is limited to the 1024 bits, in case the BIER-ERO Subobject is too long.

4.4.2. BIER-ERO Processing

The ERO and SR-ERO subobject processing remains as per [RFC5440].

If a PCC receives an BIER-ERO subobject in which either BitStringLength or Adjacency BitString or SI is absent, it MUST consider the entire BIER-ERO subobject invalid and send a PCErr message with Error-Type = 10 ("Reception of an invalid object"), Error-Value = TBD5 ("BitStringLength is absent ") or Error-Value = TBD6 ("Adjacency BitString is absent") or Error-Value = TBD7 ("SI is absent ").

If a PCC receives an BIER-ERO subobject in which BitStringLength values are not chosen from: 64, 128, 256, 512, 1024, as it described in ([RFC8279]). The PCC MUST send a PCErr message with Error-Type =10 ("Reception of an invalid object") and Error-Value = TBD8 ("Invalid BitStringLength").

5. Security Considerations

TBD.

6. IANA Considerations

6.1. PCEP Objects

IANA has made the following Object-Type allocations from the "PCEP Objects" sub-registry.

6.1.1. BIER-TE-PCE-CAPABILITY Sub-TLV Type Indicators

vlaue	Meaning	Reference
-----	-----	-----
TBD1	BIER-TE-PCE-CAPABILITY	This Document

6.1.2. New Path Setup Type

vlaue	Meaning	Reference
-----	-----	-----
TBD2	Path is setup using BIER Traffic Engineering technique	This Document

6.1.3. BIER-ERO Subobject

This document defines a new subobject type for the BIER explicit route object (ERO), The code points for subobject types of these objects is maintained in the RSVP parameters registry.

Object	Sub-Object	Sub-Object Type
-----	-----	-----
EXPLICIT_ROUTE	BIER-ERO (PCEP-specific)	TBD3

6.1.4. PCEP-Error Objects and Types

IANA is requested to allocate code-points in the "PCEP-ERROR Object Error Types and Values" subregistry for the following new error-types and error-values:

Error-Type	Meaning	Reference
-----	-----	-----
10	Reception of an invalid object	RFC5440
	Error-value = TBD4 BitStringLength is absent	This document
	Error-value = TBD5 Adjacency BitString is absent	This document
	Error-value = TBD6 SI is absent	This document
	Error-value = TBD7 Invalid BitStringLength	This document

7. Normative references

- [I-D.ietf-bier-te-arch]
Eckert, T., Cauchie, G., and M. Menth, "Traffic Engineering for Bit Index Explicit Replication (BIER-TE)", draft-ietf-bier-te-arch-05 (work in progress), November 2019.
- [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>>.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<https://www.rfc-editor.org/info/rfc3209>>.
- [RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <<https://www.rfc-editor.org/info/rfc4657>>.
- [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>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.
- [RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January 2018, <<https://www.rfc-editor.org/info/rfc8296>>.
- [RFC8306] Zhao, Q., Dhody, D., Ed., Palleti, R., and D. King, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 8306, DOI 10.17487/RFC8306, November 2017, <<https://www.rfc-editor.org/info/rfc8306>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.

Authors' Addresses

Ran Chen
ZTE Corporation

Email: chen.ran@zte.com.cn

Zheng Zhang
ZTE Corporation

Email: zhang.zheng@zte.com.cn

Senthil Dhanaraj
Huawei

Email: senthil.dhanaraj.ietf@gmail.com

Fengwei Qin
China Mobile

Email: qinfengwei@chinamobile.com

Networking Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 10, 2022

R. Chen
Zh. Zhang
ZTE Corporation
H. Chen
S. Dhanaraj
Futurewei
F. Qin
China Mobile
A. Wang
China Telecom
July 9, 2021

PCEP Extensions for BIER-TE
draft-chen-pce-bier-09

Abstract

Bit Index Explicit Replication (BIER)-TE shares architecture and packet formats with BIER as described in [RFC8279]. BIER-TE forwards and replicates packets based on a BitString in the packet header, but every BitPosition of the BitString of a BIER-TE packet indicates one or more adjacencies as described in [I-D.ietf-bier-te-arch]. BIER-TE Path can be derived from a Path Computation Element (PCE).

This document specifies extensions to the Path Computation Element Protocol (PCEP) that allow a PCE to compute and initiate the path for the BIER-TE.

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 10, 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	3
2. Conventions used in this document	3
3. Overview of PCEP Operation in BIER Networks	3
4. Object Formats	3
4.1. The OPEN Object	4
4.1.1. The BIER-TE PCE Capability sub-TLV	4
4.2. The RP/SRP Object	5
4.3. END-POINTS object	5
4.4. Objective Functions	5
4.5. ERO Object	5
4.5.1. BIER-TE-ERO Subobject	5
4.6. RRO Object	7
5. Procedures	7
5.1. Exchanging the BIER-TE Capability	7
5.2. BIER-TE-ERO Processing	8
5.3. BIER-TE-RRO Processing	8
6. IANA Considerations	8
6.1. PCEP Objects	8
6.1.1. BIER-TE-PCE-CAPABILITY Sub-TLV Type Indicators	9
6.1.2. New Path Setup Type	9
6.1.3. Objective Functions	9
6.1.4. BIER-TE-ERO and RRO Subobjects	9
6.1.5. PCEP-Error Objects and Types	10
7. Security Considerations	10
8. Acknowledgements	10
9. Normative references	10
Authors' Addresses	12

1. Introduction

Bit Index Explicit Replication (BIER)-TE shares architecture and packet formats with BIER as described in [RFC8279]. BIER-TE forwards and replicates packets based on a BitString in the packet header, but every BitPosition of the BitString of a BIER-TE packet indicates one or more adjacencies as described in [I-D.ietf-bier-te-arch]. BIER-TE Path can be derived from a Path Computation Element (PCE).

[RFC8231] specifies a set of extensions to PCEP that allow a PCE to compute and recommend network paths in compliance with [RFC4657] and defines objects and TLVs for MPLS-TE LSPs.

This document uses a PCE for computing one or more BIER-TE paths taking into account various constraints and objective functions.

2. Conventions used in this document

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 RFC2119.

3. Overview of PCEP Operation in BIER Networks

BIER-TE forwards and replicates packets based on a BitString in the packet header, and every BitPosition of the BitString of a BIER-TE packet indicates one or more adjacencies as described in [I-D.ietf-bier-te-arch]. In a PCEP session, An ERO object specified in [RFC5440] can be extended to carry a BIER-TE path consists of one or more BIER-TE-ERO subobject(s). BIER-TE computed by a PCE can be represented in the following forms:

- o An ordered set of adjacencies BitString(s) in which each bit represents that the adjacencies to which the BFR should replicate packets to in the domain.

In this document, we define a set of PCEP protocol extensions, including a new PCEP capability, a new Path Setup Type (PST), reuse BIER END-POINT Object, a new Objective Functions subobjects, a new ERO subobjects, a new RRO subobjects, a new PCEP error codes and procedures.

4. Object Formats

4.1. The OPEN Object

4.1.1. The BIER-TE PCE Capability sub-TLV

[RFC8408] defines the PATH-SETUP-TYPE-CAPABILITY TLV for use in the OPEN object. The PATH-SETUP-TYPE-CAPABILITY TLV contains an optional list of sub-TLVs which are intended to convey parameters that are associated with the path setup types supported by a PCEP speaker.

This document defines a new Path Setup Type (PST) for BIER-TE as follows:

- o PST = TBD2: Path is setup using BIER-TE technique.

A PCEP speaker MUST indicate its support of the function described in this document by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this new PST included in the PST list.

This document also defines the BIER-TE-PCE-CAPABILITY sub-TLV. PCEP speakers use this sub-TLV to exchange BIER capability. If a PCEP speaker includes PST=TBD2 in the PST List of the PATH-SETUP-TYPE-CAPABILITY TLV then it MUST also include the BIER-TE-PCE-CAPABILITY sub-TLV inside the PATH-SETUP-TYPE-CAPABILITY TLV.

The format of the BIER-TE-PCE-CAPABILITY sub-TLV is shown in the following figure:

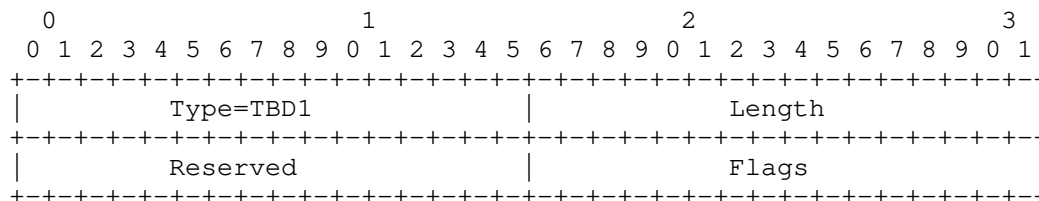


Figure 1 BIER-TE-PCE-CAPABILITY sub-TLV format

The code point for the TLV type is to be defined by IANA.

Length: 4 bytes.

The "Reserved" (2 octet) and "Flags" (2 octet) fields are currently unused, and MUST be set to zero on transmission and ignored on reception.

4.2. The RP/SRP Object

In order to setup an BIER-TE, a new PATH-SETUP-TYPE TLV MUST be contained in RP/SRP object. This document defines a new Path Setup Type (PST=TBD2) for BIER-TE.

4.3. END-POINTS object

The END-POINTS object which is defined in [RFC8306] is used in a PCReq message to specify the BIER information of the path for which a path computation is requested. To represent the end points for a BIER path efficiently, we reuse the P2MP END-POINTS object body for IPv4 (Object-Type 3) and END-POINTS object body for IPv6 (Object-Type 4) which is defined in [RFC8306].

4.4. Objective Functions

[RFC5541] defines a mechanism to specify an objective function (OF) that is used by a PCE when it computes a path. For a BIER-TE path, a new OF is defined.

Objective Function Code: TBD3

Name: Minimum Bit Sets (MBS)

Description: Find a path represented by BitPositions that has the minimum number of bit sets.

4.5. ERO Object

BIER-TE consists of one or more adjacencies BitStrings where every BitPosition of the BitString indicates one or more adjacencies, as described in ([RFC8279]).

The ERO object specified in [RFC5440] is used to encode the path of a TE LSP through the network. The ERO is carried within a PCRep message to provide the computed TE LSP if the path computation was successful. In order to carry BIER-TE explicit paths, this document defines a new ERO subobjects referred to as "BIER-TE-ERO subobjects" whose formats are specified in the following section. An BIER-TE-ERO subobjects carrying a adjacencies BitStrings consists of one or more BIER-TE-ERO subobject(s).

4.5.1. BIER-TE-ERO Subobject

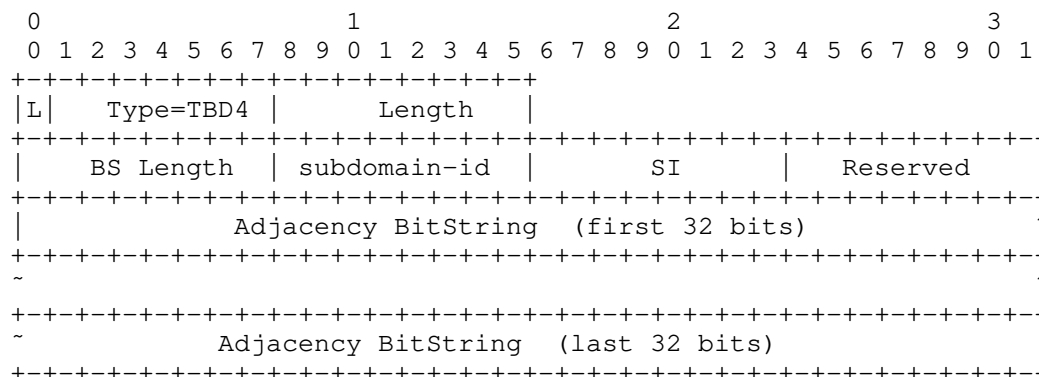


Figure 3

The 'L' Flag: Indicates whether the subobject represents a loose-hop in the LSP[RFC3209]. If the bit is not set, the subobject represents a strict hop in the explicit route.

Type: TBD4

Length: 1 octet ([RFC3209]). Contains the total length of the subobject in octets. The Length MUST be at least 8, and MUST be a multiple of 4.

BS Length: A 1 octet field encodes the length in bits of the BitString as per [RFC8296], the maximum length of the BitString is 5, it indicates the length of BitString is 1024. It is used to refer to the number of bits in the BitString.

subdomain-id: Unique value identifying the BIER subdomain. 1 octet.

SI: Set Identifier (Section 1 of [RFC8279] used in the encapsulation for this BIER subdomain for this BitString length, 1 octet.

The "Reserved" (1 octets) fields are currently unused, and MUST be set to zero on transmission and ignored on reception.

Adjacency BitString: a variable length field encoding the Adjacency BitString where every BitPosition of the BitString indicates one or more adjacencies. the length of this field is according the BS length. The minimum value of this field is 64 bits, and the maximum value of this field is 1024 bits.

Notice:

The maximum value of BS Length is limited to the 1024 bits, in case the BIER-TE-ERO Subobject is too long.

4.6. RRO Object

An RRO contains one or more subobjects called "BIER-TE-RRO subobjects", whose format is shown below:

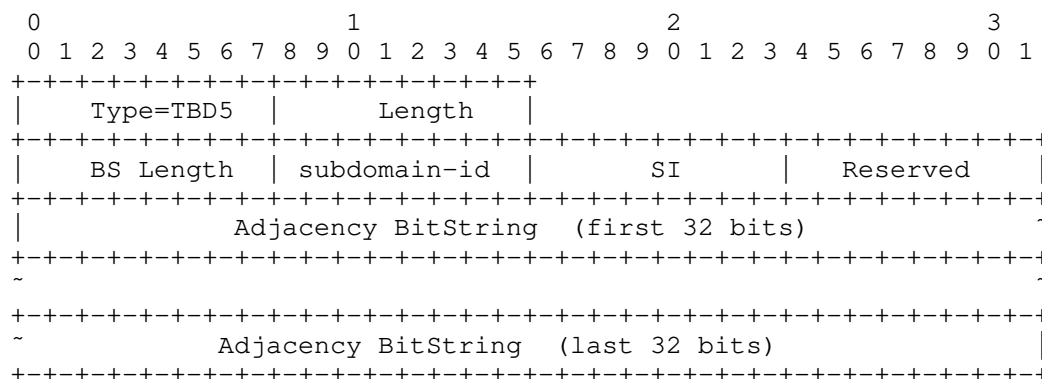


Figure 4

The format of the BIER-TE-RRO subobject is the same as that of the BIER-TE-ERO subobject, but without the L-Flag.

For the integrity of the protocol, we define a new BIER-TE-RRO object, but its actual value is consistent with ERO. The PCC reports an BIER-TE to a PCE by sending a PCRpt message with RRO object.

5. Procedures

5.1. Exchanging the BIER-TE Capability

A PCC indicates that it is capable of supporting the head-end functions for BIER-TE by including the BIER-TE-PCE-CAPABILITY sub-TLV in the Open message that it sends to a PCE. A PCE indicates that it is capable of computing BIER-TE by including the BIET-TE-PCE-CAPABILITY sub-TLV in the Open message that it sends to a PCC.

If a PCEP speaker receives a PATH-SETUP-TYPE-CAPABILITY TLV with a PST list containing PST=TBD2, and supports that path setup type, then it checks for the presence of the SR-PCE-CAPABILITY sub-TLV. If that sub-TLV is absent, then the PCEP speaker MUST send a PCErr message

with Error-Type = 10 ("Reception of an invalid object") and Error-value = TBD6("Missing PCE-BIER-TE-CAPABILITY sub-TLV") and MUST then close the PCEP session. If a PCEP speaker receives a PATH-SETUP-TYPE- CAPABILITY TLV with a BIER-TE-PCE-CAPABILITY sub-TLV, but the PST list does not contain PST=TBD2, then the PCEP speaker MUST ignore the BIER-TE-PCE-CAPABILITY sub-TLV.

5.2. BIER-TE-ERO Processing

If a PCC does not support the BIER-TE PCE Capability and thus cannot recognize the BIER-TE-ERO or BIER-TE-RRO subobjects, The ERO and BIER-TE-ERO subobject processing remains as per [RFC5440].

If a PCC receives an BIER-TE-ERO subobject in which either BitStringLength or Adjacency BitString or SI is absent, it MUST consider the entire BIER-TE-ERO subobject invalid and send a PCErr message with Error-Type = 10 ("Reception of an invalid object"), Error-Value = TBD7 ("BitStringLength is absent ") or Error-Value = TBD8 ("Adjacency BitString is absent") or Error-Value = TBD9 ("SI is absent").

If a PCC receives an BIER-TE-ERO subobject in which BitStringLength values are not chosen from: 64, 128, 256, 512, 1024, as it described in ([RFC8279]). The PCC MUST send a PCErr message with Error-Type =10 ("Reception of an invalid object") and Error-Value = TBD10 ("Invalid BitStringLength").

When a PCEP speaker detects that all subobjects of ERO are not of type TBD4, and if it does not handle such ERO, it MUST send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD11 ("Non-identical ERO subobjects") as per [RFC8664].

5.3. BIER-TE-RRO Processing

The syntax checking rules that apply to the BIER-TE-RRO subobject are identical to those of the BIER-TE-ERO subobject

The actual value of BIER-TE-RRO subobject is consistent with ERO. The PCC reports an BIER-TE to a PCE by sending a PCRpt message with RRO object.

6. IANA Considerations

6.1. PCEP Objects

IANA has made the following Object-Type allocations from the "PCEP Objects" sub-registry.

6.1.1. BIER-TE-PCE-CAPABILITY Sub-TLV Type Indicators

Value	Meaning	Reference
TBD1	BIER-TE-PCE-CAPABILITY	This Document

6.1.2. New Path Setup Type

Value	Meaning	Reference
TBD2	Path is setup using BIER TE technique	This Document

6.1.3. Objective Functions

Value	Meaning	Reference
TBD3	Minimum Bit Sets (MBS)	This Document

6.1.4. BIER-TE-ERO and RRO Subobjects

This document defines a new subobject type for the PCEP explicit route object (ERO) and a new subobject type for the PCEP RRO. The code points for subobject types of these objects are maintained in the RSVP parameters registry, under the EXPLICIT_ROUTE and ROUTE_RECORD objects, respectively.

Object	Subobject	Subobject Type
EXPLICIT_ROUTE	BIER-TE-ERO (PCEP specific)	TBD4
ROUTE_RECORD	BIER-TE-RRO (PCEP specific)	TBD5

6.1.5. PCEP-Error Objects and Types

IANA is requested to allocate code-points in the "PCEP-ERROR Object Error Types and Values" subregistry for the following new error-types and error-values:

Error-Type	Meaning	Error-value
10	Reception of an invalid object	
		TBD6: Missing PCE-BIER-TE-CAPABILITY subobjects
		TBD7: BitStringLength is absent
		TBD8: Adjacency BitString is absent
		TBD9: SI is absent
		TBD10: Invalid BitStringLength
		TBD11: Non-identical ERO subobjects

7. Security Considerations

The security considerations described in [RFC5440], [RFC8231], [RFC8281] and [RFC8408] are applicable to this specification. No additional security measures are required.

8. Acknowledgements

The authors thank Dhruv Dhody, Benchong Xu, Chun Zhu, and Zhaohui Zhang and many others for their suggestions and comments.

9. Normative references

[I-D.ietf-bier-te-arch]
 Eckert, T., Cauchie, G., and M. Menth, "Tree Engineering for Bit Index Explicit Replication (BIER-TE)", draft-ietf-bier-te-arch-09 (work in progress), October 2020.

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<https://www.rfc-editor.org/info/rfc3209>>.
- [RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <<https://www.rfc-editor.org/info/rfc4657>>.
- [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>>.
- [RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, DOI 10.17487/RFC5541, June 2009, <<https://www.rfc-editor.org/info/rfc5541>>.
- [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>>.
- [RFC8279] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast Using Bit Index Explicit Replication (BIER)", RFC 8279, DOI 10.17487/RFC8279, November 2017, <<https://www.rfc-editor.org/info/rfc8279>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8296] Wijnands, IJ., Ed., Rosen, E., Ed., Dolganow, A., Tantsura, J., Aldrin, S., and I. Meilik, "Encapsulation for Bit Index Explicit Replication (BIER) in MPLS and Non-MPLS Networks", RFC 8296, DOI 10.17487/RFC8296, January 2018, <<https://www.rfc-editor.org/info/rfc8296>>.

- [RFC8306] Zhao, Q., Dhody, D., Ed., Palleti, R., and D. King,
"Extensions to the Path Computation Element Communication
Protocol (PCEP) for Point-to-Multipoint Traffic
Engineering Label Switched Paths", RFC 8306,
DOI 10.17487/RFC8306, November 2017,
<<https://www.rfc-editor.org/info/rfc8306>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J.
Hardwick, "Conveying Path Setup Type in PCE Communication
Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408,
July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.
- [RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W.,
and J. Hardwick, "Path Computation Element Communication
Protocol (PCEP) Extensions for Segment Routing", RFC 8664,
DOI 10.17487/RFC8664, December 2019,
<<https://www.rfc-editor.org/info/rfc8664>>.

Authors' Addresses

Ran Chen
ZTE Corporation

Email: chen.ran@zte.com.cn

Zheng Zhang
ZTE Corporation

Email: zhang.zheng@zte.com.cn

Huaimo Chen
Futurewei

Email: huaimo.chen@futurewei.com

Senthil Dhanaraj
Futurewei

Email: senthil.dhanaraj.ietf@gmail.com

Fengwei Qin
China Mobile

Email: qinfengwei@chinamobile.com

Aijun Wang
China Telecom

Email: wangaj3@chinatelecom.cn

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 26, 2020

H. Chen
Futurewei
M. Toy
Verizon
A. Wang
China Telecom
Z. Li
China Mobile
L. Liu
Fujitsu
X. Liu
Volta Networks
October 24, 2019

SR Path Ingress Protection
draft-chen-pce-sr-ingress-protection-02

Abstract

This document describes extensions to Path Computation Element (PCE) communication Protocol (PCEP) 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 April 26, 2020.

Copyright Notice

Copyright (c) 2019 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. SR Path Ingress Protection Example	3
4. Behavior after Ingress Failure	4
5. Extensions to PCEP	5
5.1. Capability for SR Path Ingress Protection	5
5.2. SR Path Ingress Protection	6
5.2.1. Traffic-Description sub-TLV	7
5.2.2. Primary-Ingress sub-TLV	10
5.2.3. Service sub-TLV	11
6. Security Considerations	12
7. Acknowledgements	12
8. IANA Considerations	12
9. References	12
9.1. Normative References	12
9.2. Informative References	13
Authors' Addresses	14

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 Path Computation Element (PCE) communication Protocol (PCEP) 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

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

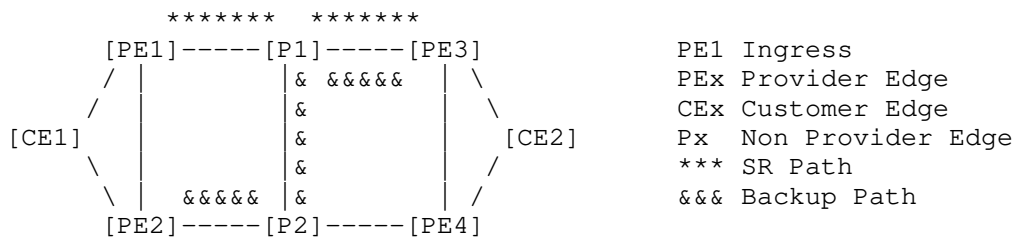


Figure 1: Protecting Ingress PE1 of SR Path

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 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 PCEP

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.

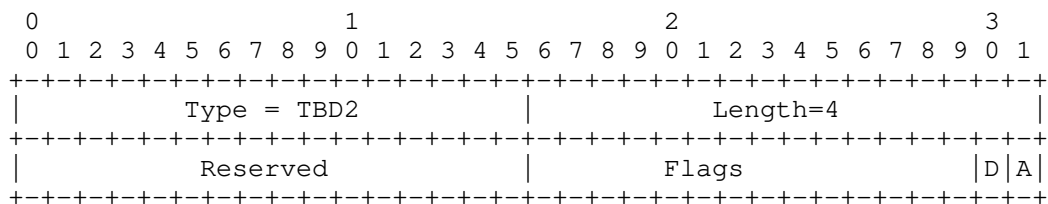


Figure 2: SR_INGRESS_PROTECTION_CAPABILITY sub-TLV

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.

- o D flag: A PCC sets this flag to 1 to indicate that it is able to detect its adjacent node's failure quickly.
- o 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.

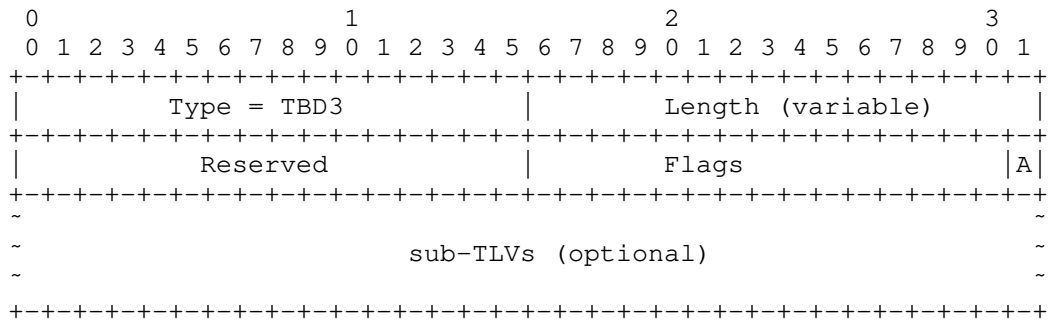


Figure 3: SR_INGRESS_PROTECTION sub-TLV

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 is defined.

- o 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.

Three 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.

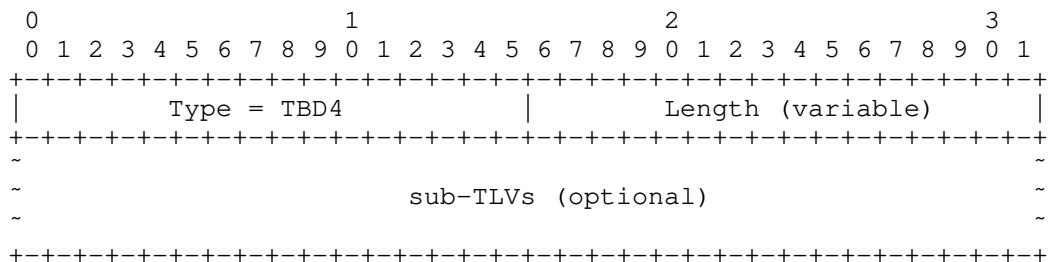


Figure 4: Traffic-Description sub-TLV

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.

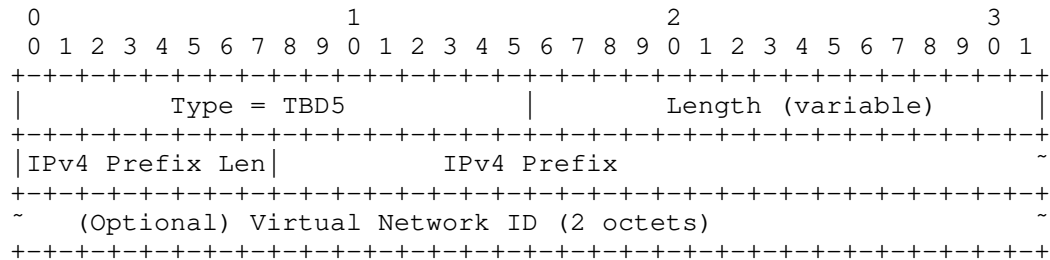


Figure 5: IPv4 FEC sub-TLV

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.

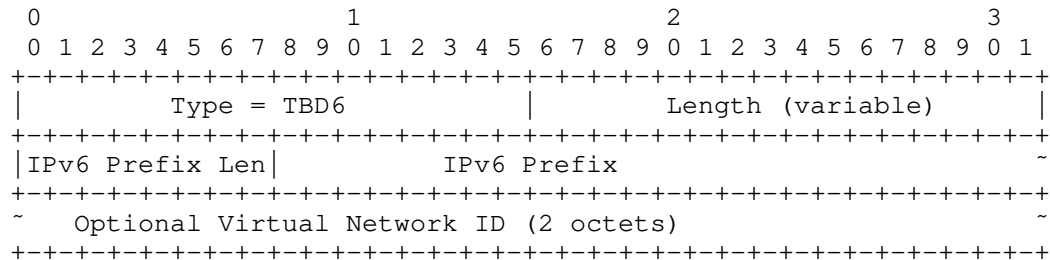


Figure 6: IPv6 FEC sub-TLV

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.

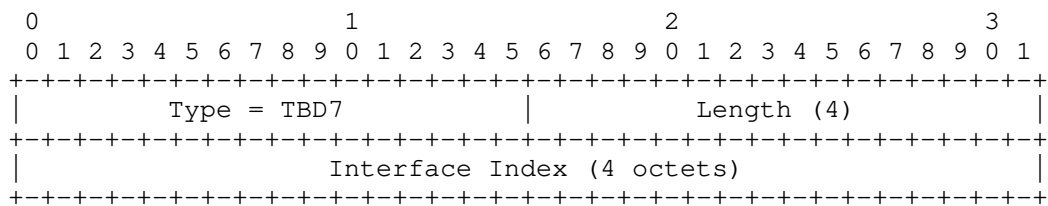


Figure 7: Interface Index sub-TLV

Type: TBD7 is to be assigned by IANA.

Length: 4.

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

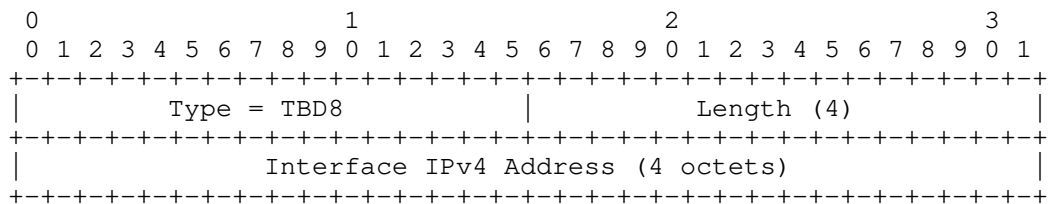


Figure 8: Interface IPv4 Address sub-TLV

Type: TBD8 is to be assigned by IANA.

Length: 4.

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

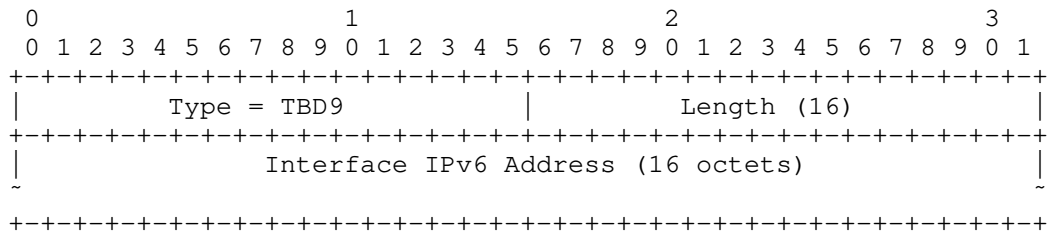


Figure 9: Interface IPv6 Address sub-TLV

Type: TBD9 is to be assigned by IANA.

Length: 16.

Interface IPv6 Address: 16 octets. It represents the IPv6 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.

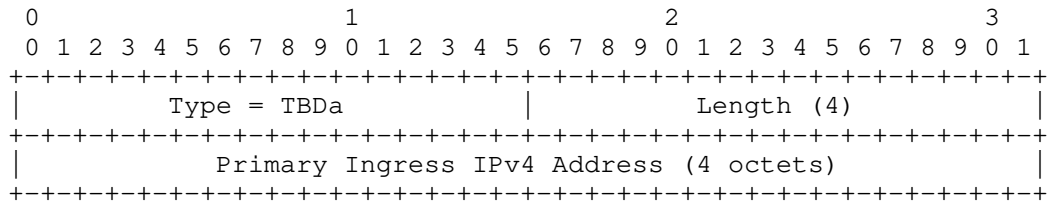


Figure 10: Primary Ingress IPv4 Address sub-TLV

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.

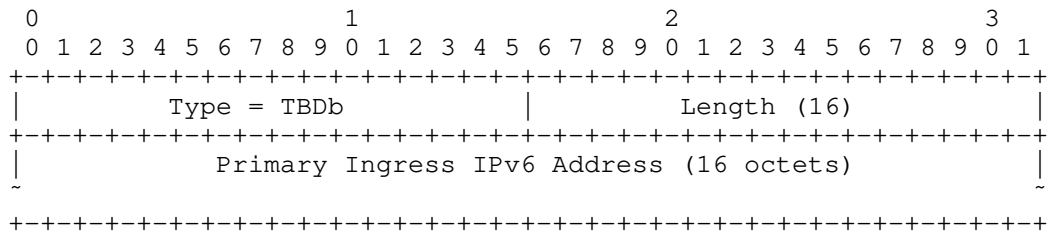


Figure 11: Primary Ingress IPv6 Address sub-TLV

Type: TBDb 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 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.

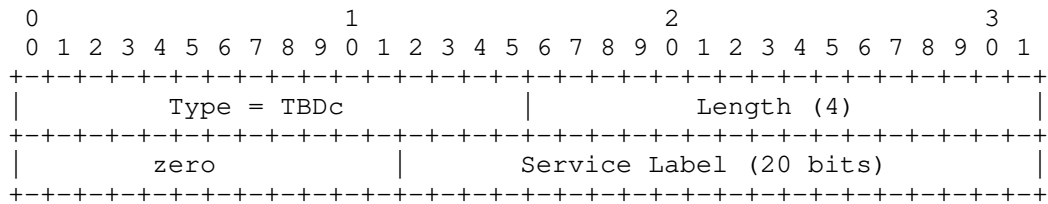


Figure 12: Service Label sub-TLV

Type: TBDC is to be assigned by IANA.

Length: 4.

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

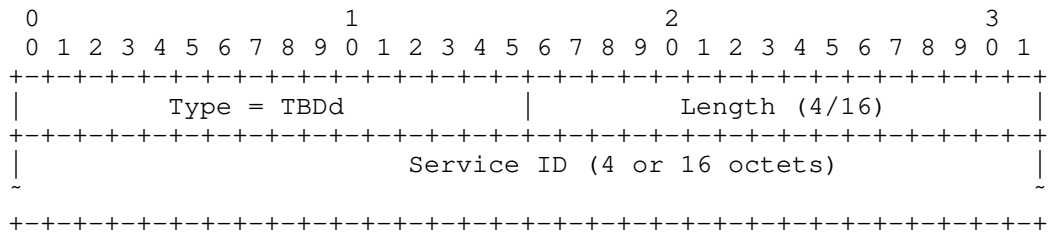


Figure 13: Service ID sub-TLV

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.

6. Security Considerations

TBD

7. Acknowledgements

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

8. IANA Considerations

TBD

9. References

9.1. Normative References

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

Psenak, P., Filsfils, C., Bashandy, A., Decraene, B., and Z. Hu, "IS-IS Extensions to Support Routing over IPv6 Dataplane", draft-bashandy-isis-srv6-extensions-05 (work in progress), March 2019.

[I-D.hu-spring-segment-routing-proxy-forwarding]

Hu, Z., Chen, H., Yao, J., Bowers, C., and Y. Zhu, "SR-TE Path Midpoint Protection", draft-hu-spring-segment-routing-proxy-forwarding-04 (work in progress), July 2019.

- [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-05 (work in progress), August 2019.
- [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>>.

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]

Sivabalan, S., Filsfils, C., Tantsura, J., Hardwick, J., Previdi, S., and C. Li, "Carrying Binding Label/Segment-ID in PCE-based Networks.", draft-sivabalan-pce-binding-label-sid-07 (work in progress), July 2019.

[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>>.

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

Lei Liu
Fujitsu
USA

Email: liulei.kddi@gmail.com

Xufeng Liu
Volta Networks
McLean, VA
USA

Email: xufeng.liu.ietf@gmail.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 17 May 2022

H. Chen
M. McBride
Futurewei
M. Toy
G. Mishra
Verizon Inc.
A. Wang
China Telecom
Z. Li
Y. Liu
China Mobile
B. Khasanov
Yandex LLC
L. Liu
Fujitsu
X. Liu
Volta Networks
13 November 2021

Path Ingress Protections
draft-chen-pce-sr-ingress-protection-07

Abstract

This document describes extensions to Path Computation Element (PCE) communication Protocol (PCEP) for fast protecting the ingress nodes of two types of paths or tunnels, which are Segment Routing (SR) paths and Bit Index Explicit Replication Tree/Traffic Engineering (BIER-TE) paths. The extensions comprise a foundation for protecting the ingress nodes of different types of paths. Based on this, the ingress protection of a new type of paths can be easily supported.

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 17 May 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	3
1.1. Terminologies	3
2. Path Ingress Protection Examples	4
2.1. SR Path Ingress Protection Example	4
2.2. BIER-TE Path Ingress Protection Example	5
3. Behavior around Ingress Failure	6
3.1. Source Detect	6
3.2. Backup Ingress Detect	6
3.3. Both Detect	7
4. Extensions to PCEP	7
4.1. Capabilities for Ingress Protection	7
4.1.1. Capability for Ingress Protection with Backup Ingress	7
4.1.2. Capability for Ingress Protection with Traffic Source	9
4.2. Extensions for Backup Ingress and Traffic Source	10
4.2.1. Extensions for Backup Ingress	10
4.2.2. Extensions for Traffic Source	16
5. Security Considerations	19
6. Acknowledgements	19
7. IANA Considerations	19
8. References	19
8.1. Normative References	19
8.2. Informative References	19
Authors' Addresses	20

1. Introduction

The fast protection of a transit node in each type of paths or tunnels have been proposed. For example, the fast protection of a transit node in a Segment Routing (SR) path or tunnel is described in [I-D.ietf-rtgwg-segment-routing-ti-lfa]. 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 an SR path/tunnel, a BIER-TE path/tunnel, or other type of paths/tunnels.

This document fills that void and specifies protocol extensions to Path Computation Element (PCE) communication Protocol (PCEP) [RFC5440] and [RFC9050] for fast protecting the ingress nodes of two types of paths: SR paths and BIER-TE paths. The extensions comprise a foundation for protecting the ingress nodes of different types of paths. Based on this, the ingress protection of a new type of paths can be easily supported.

Ingress node and ingress, fast protection and protection, path ingress protection and ingress protection, SR path and SR tunnel, as well as BIER-TE path and BIER-TE tunnel will be used exchangeably in the following sections.

1.1. Terminologies

The following terminologies are used in this document.

PCE: Path Computation Element or Path Computation Element server

PCEP: PCE communication Protocol

PCC: Path Computation Client

BIER: Bit Index Explicit Replication

BIFT: Bit Index Forwarding Table

CE: Customer Edge

PE: Provider Edge

TE: Traffic Engineering

SR: Segment Routing
 LFA: Loop-Free Alternate
 TI-LFA: Topology Independent LFA
 BFD: Bidirectional Forwarding Detection
 VPN: Virtual Private Network
 L3VPN: Layer 3 VPN
 FIB: Forwarding Information Base

2. Path Ingress Protection Examples

This section shows two examples of path ingress protection. One is SR path ingress protection, and the other is BIER-TE path ingress protection.

2.1. 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 and represented by *** in the figure.

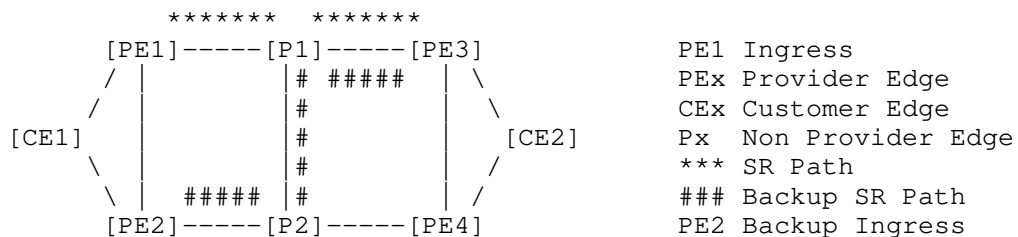


Figure 1: Protecting Ingress PE1 of SR Path

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 and represented by ### in the figure. When the traffic is imported into the backup path, it is sent to the egress PE3 along the path.

2.2. BIER-TE Path Ingress Protection Example

Figure 2 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.

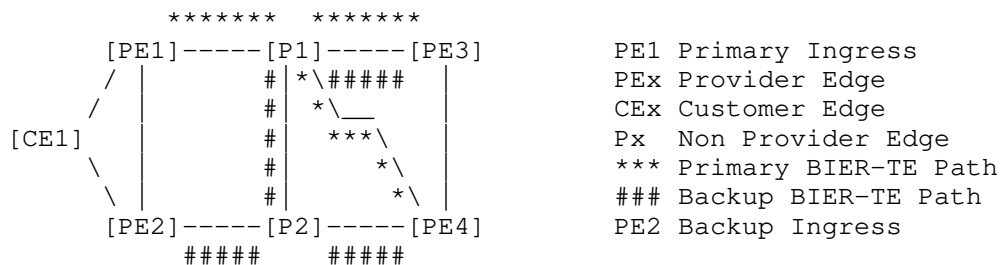


Figure 2: Protecting Ingress PE1 of BIER-TE Path

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 PCE 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 information about the backup ingress, the primary ingress and the traffic to the PCC of the traffic source (e.g., CE1).

When the PCC of the traffic source receives the information about the backup ingress, the primary ingress and the traffic, 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.

3. 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.

3.1. Source Detect

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

3.2. Backup Ingress Detect

The traffic source (e.g., CE1) always sends the traffic to both the ingress (e.g., PE1) of the primary path such as 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 path such as the backup BIER-TE path in normal operations. When it detects the failure of the ingress of the primary path, it imports the traffic from the source into the backup 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 path, it SHOULD consider this.

3.3. 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 path such as 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 path such as the backup BIER-TE path in normal operations. When it detects the failure of the ingress of the primary path, it imports the traffic from the source into the backup path.

4. Extensions to PCEP

A PCC runs on each of the edge nodes such as PEs of a network normally. A 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 path. The path is a SR path, a BIER-TE path, or a path of another type.

4.1. Capabilities for Ingress Protection

4.1.1. 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 each of different types of paths.

A new sub-TLV called INGRESS_PROTECTION_CAPABILITY is defined. It is included in the PATH_SETUP_TYPE_CAPABILITY TLV with PST = TBD1 (suggested value 2 for path ingress protection) 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.

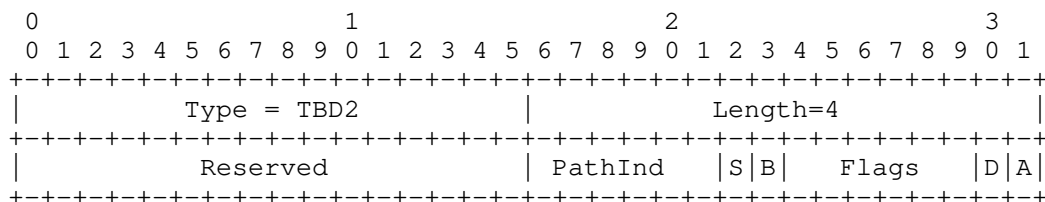


Figure 3: INGRESS_PROTECTION_CAPABILITY sub-TLV

Type: TBD2 is to be assigned by IANA.

Length: 4.

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

PathInd: 1 octet. Indicators for the types of paths whose ingress protections are supported. Two indicators are defined.

- o S : S = 1 indicating that the ingress protection of a SR path is supported.
- o B : B = 1 indicating that the ingress protection of a BIER-TE path is supported.

Flags: 1 octet. Two flags are defined.

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

A PCC, which supports ingress protection for different types of paths, sends a PCE an Open message containing INGRESS_PROTECTION_CAPABILITY sub-TLV. This sub-TLV indicates that the PCC is capable of supporting the ingress protection for the types of paths.

For example, if a PCC supports ingress protection for SR path and BIER-TE path, the PCC sends a PCE an Open message containing INGRESS_PROTECTION_CAPABILITY sub-TLV with S = 1 and B = 1.

A PCE, which supports ingress protection for different types of paths, sends a PCC an Open message containing INGRESS_PROTECTION_CAPABILITY sub-TLV. This sub-TLV indicates that the PCE is capable of supporting the ingress protection for the types of paths.

If both a PCC and a PCE support 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 an INGRESS_PROTECTION_CAPABILITY sub-TLV.

If a PCE receives an Open message from a PCC without a INGRESS_PROTECTION_CAPABILITY sub-TLV indicating PCC's support for the ingress protection of a type of paths, then the PCE MUST not send the PCC any request for ingress protection of the type of paths.

If a PCC receives an Open message from a PCE without a INGRESS_PROTECTION_CAPABILITY sub-TLV indicating PCE's support for the ingress protection of a type of paths, then the PCC MUST ignore any request for ingress protection of the type of paths 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 path, the PCC MUST let the forwarding entry for the backup path be Active.

4.1.2. 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 ingress protection.

The PCECC-CAPABILITY sub-TLV defined in [RFC9050] 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:

- * 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.

4.2. Extensions for Backup Ingress and Traffic Source

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 path, or

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

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

Source Detect: S1 and B1.

Backup Ingress Detect: S2 and B2.

Both Detect: S1 and B2.

4.2.1. Extensions for Backup Ingress

For the packets from the traffic source, if the primary ingress (i.e., the ingress of the primary path) encapsulates the packets with a service ID or label into the path, the backup ingress MUST have this service ID or label and encapsulates the packets with the service ID or label into the backup 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 sub-TLV called INGRESS_PROTECTION is defined. When a PCE sends a PCC a PCInitiate message for initiating a backup path to protect the primary ingress node of a primary path, the message contains this TLV in the RP/SRP object. Its format is illustrated below.

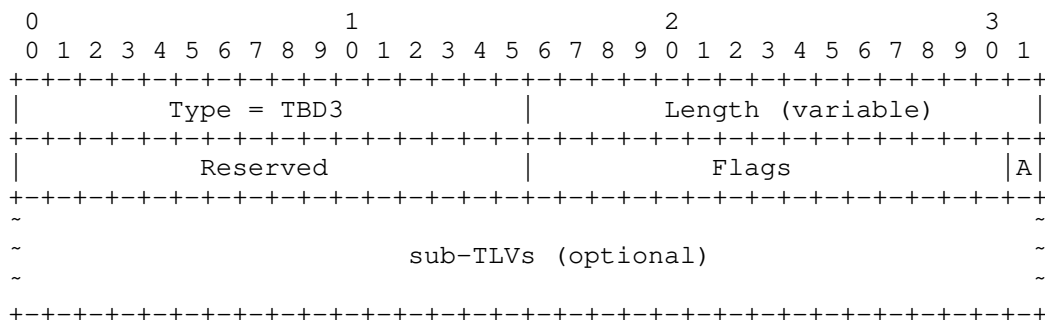


Figure 4: INGRESS_PROTECTION sub-TLV

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 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 path 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 path 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.

Three optional sub-TLVs are defined: Primary-Ingress sub-TLV, Service sub-TLV, and Traffic-Description sub-TLV. The Traffic-Description sub-TLV describes the traffic to be imported into the backup SR path. 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.

4.2.1.1. Primary-Ingress sub-TLV

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

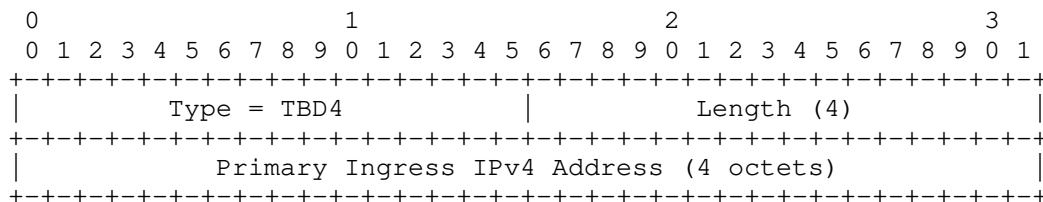


Figure 5: Primary Ingress IPv4 Address sub-TLV

Type: TBD4 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 path.

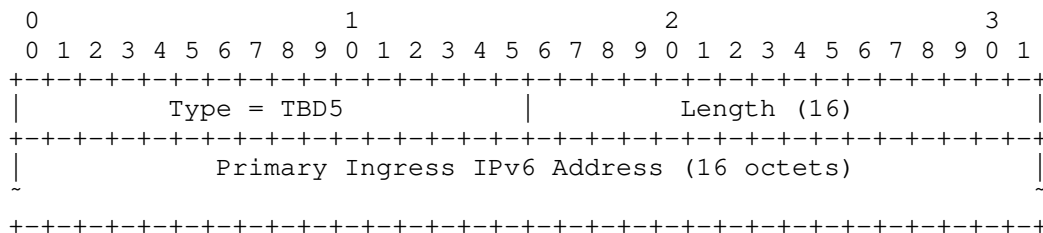


Figure 6: Primary Ingress IPv6 Address sub-TLV

Type: TBD5 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 path.

4.2.1.2. Service sub-TLV

A Service sub-TLV contains a service ID or label to be added into a packet to be carried by a path. 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.

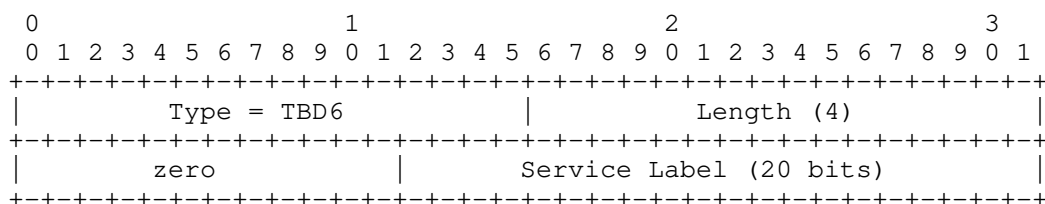


Figure 7: Service Label sub-TLV

Type: TBD6 is to be assigned by IANA.

Length: 4.

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

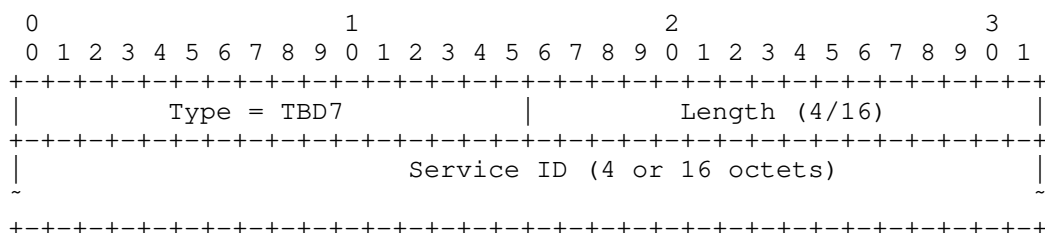


Figure 8: Service ID sub-TLV

Type: TBD7 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.

4.2.1.3. Traffic-Description sub-TLV

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

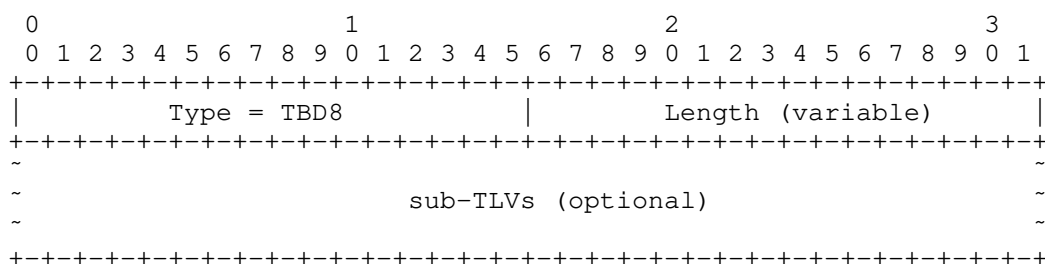


Figure 9: Traffic-Description sub-TLV

Type: TBD8 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 path. 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.

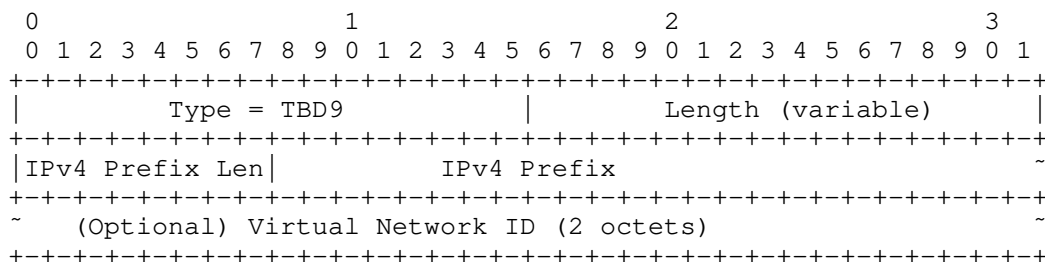


Figure 10: IPv4 FEC sub-TLV

Type: TBD9 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.

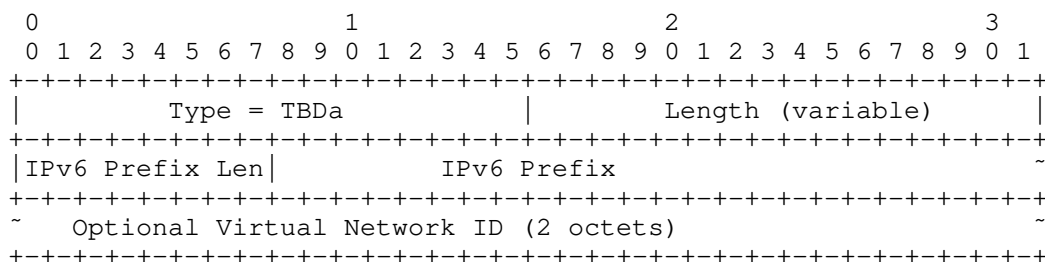


Figure 11: IPv6 FEC sub-TLV

Type: TBDA 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 path. It has three formats: one for interface index, the other two for IPv4 and IPv6 address, which are illustrated below.

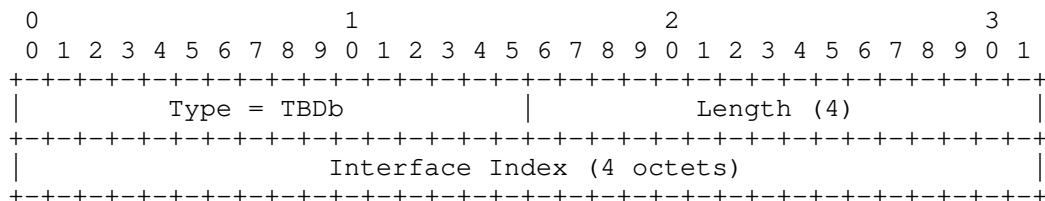


Figure 12: Interface Index sub-TLV

Type: TBDb is to be assigned by IANA.

Length: 4.

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

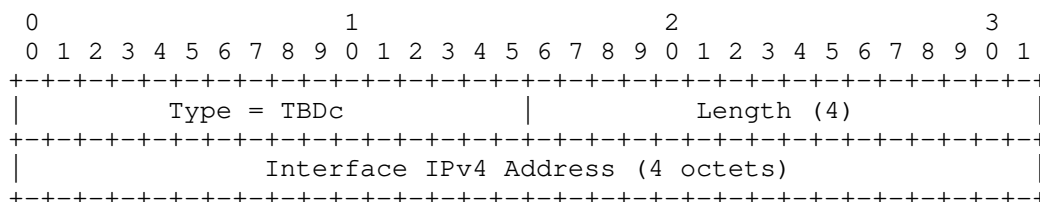


Figure 13: Interface IPv4 Address sub-TLV

Type: TBDc is to be assigned by IANA.

Length: 4.

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

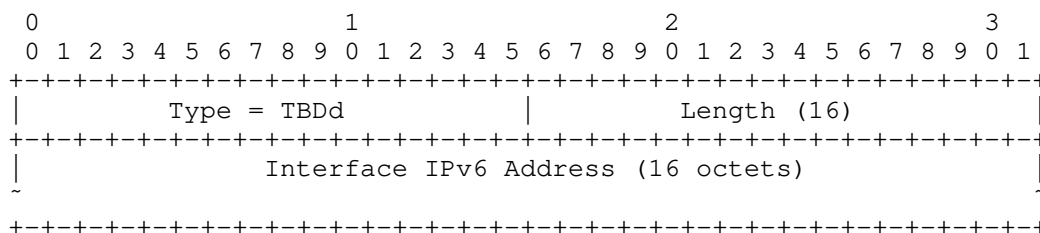


Figure 14: Interface IPv6 Address sub-TLV

Type: TBDd is to be assigned by IANA.

Length: 16.

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

4.2.2. 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 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 [RFC9050] for a PCE as a controller to send instructions for LSPs to a PCC. This document defines a new object-type (TBDt) for 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.

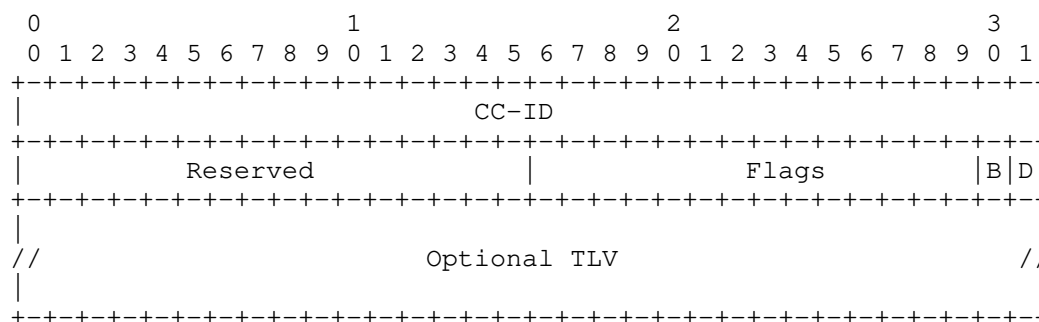


Figure 15: INGRESS-PROTECTION Object Body

CC-ID: It is the same as described in [RFC9050].

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, Traffic-Description TLV 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 Traffic-Description sub-TLV defined above is used as a TLV to contain the information about the traffic for a SR path 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 for a BIER-TE path in the object. A new TLV, called backup ingress TLV, is defined to contain the information about the backup ingress in the object.

4.2.2.1. Backup-Ingress TLV

A Backup-Ingress TLV indicates the IP address of the ingress node of a backup path. 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.

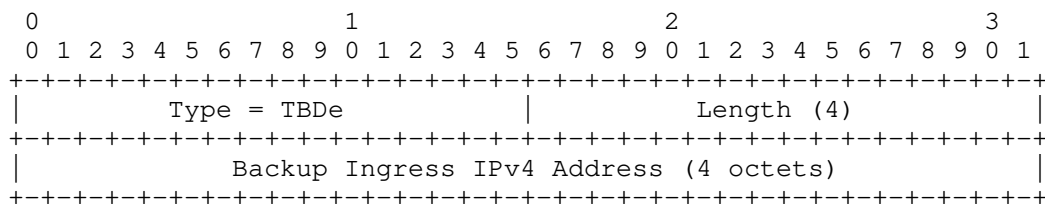


Figure 16: Backup Ingress IPv4 Address TLV

Type: TBDe is to be assigned by IANA.

Length: 4.

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

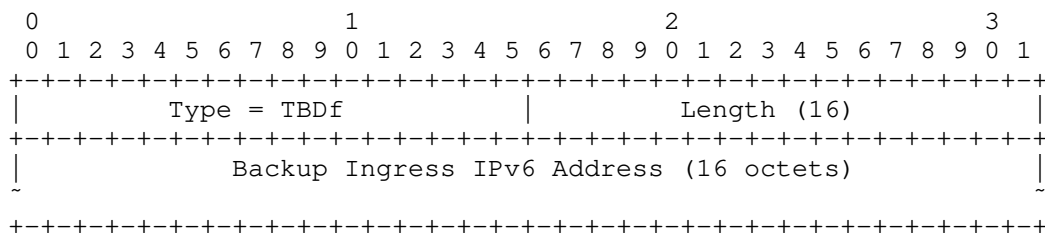


Figure 17: Backup Ingress IPv6 Address TLV

Type: TBdf 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.

5. Security Considerations

TBD

6. Acknowledgements

The authors of this document would like to thank Dhruv Dhody and Robin Li for their reviews and comments.

7. IANA Considerations

TBD

8. References

8.1. 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>>.
- [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>>.
- [RFC9050] Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "Path Computation Element Communication Protocol (PCEP) Procedures and Extensions for Using the PCE as a Central Controller (PCECC) of LSPs", RFC 9050, DOI 10.17487/RFC9050, July 2021, <<https://www.rfc-editor.org/info/rfc9050>>.

8.2. 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", Work in Progress, Internet-Draft, draft-chen-bier-te-frr-01, 23 August 2021, <<https://www.ietf.org/archive/id/draft-chen-bier-te-frr-01.txt>>.

[I-D.ietf-pce-pcep-flowspec]

Dhody, D., Farrel, A., and Z. Li, "PCEP Extension for Flow Specification", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-flowspec-13, 14 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-pce-pcep-flowspec-13.txt>>.

[I-D.ietf-rtgwg-segment-routing-ti-lfa]

Litkowski, S., Bashandy, A., Filsfils, C., Francois, P., Decraene, B., and D. Voyer, "Topology Independent Fast Reroute using Segment Routing", Work in Progress, Internet-Draft, draft-ietf-rtgwg-segment-routing-ti-lfa-07, 29 June 2021, <<https://www.ietf.org/archive/id/draft-ietf-rtgwg-segment-routing-ti-lfa-07.txt>>.

[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>>.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA,
United States of America

Email: Huaimo.chen@futurewei.com

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Mehmet Toy
Verizon Inc.
United States of America

Email: mehmet.toy@verizon.com

Gyan S. Mishra
Verizon Inc.
13101 Columbia Pike
Silver Spring, MD 20904
United States of America

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

Yisong Liu
China Mobile

Email: liuyisong@chinamobile.com

Boris Khasanov
Yandex LLC
Moscow

Email: bhassanov@yahoo.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

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 23, 2020

M. Negi
Z. Li
X. Geng
Huawei Technologies
August 22, 2019

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) for P2MP LSPs
draft-dhody-pce-pcep-extension-pce-controller-p2mp-02

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

The PCE has been identified as an appropriate technology for the determination of the paths of point- to-multipoint (P2MP) TE Label Switched Paths (LSPs).

PCE was developed to derive paths for MPLS P2MP LSPs, which are supplied to the head end (root) of the LSP using PCEP. PCEP has been proposed as a control protocol to allow the PCE to be fully enabled as a central controller.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the P2MP LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the P2MP path while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP protocol extensions for using the PCE as the central controller for P2MP TE LSP.

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 February 23, 2020.

Copyright Notice

Copyright (c) 2019 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	3
1.1. Requirements Language	5
2. Terminology	5
3. Basic PCECC Mode	5
4. Procedures for Using the PCE as the Central Controller (PCECC) for P2MP	5
4.1. Stateful PCE Model	5
4.2. PCECC Capability Advertisement	6
4.3. LSP Operations	6
4.3.1. Basic PCECC LSP Setup	6
4.3.2. Central Control Instructions	7
4.3.2.1. Label Download	7
4.3.2.2. Label Cleanup	8
4.3.3. PCE Initiated PCECC LSP	9
4.3.4. PCECC LSP Update	9
4.3.5. Re Delegation and Cleanup	9
4.3.6. Synchronization of Central Controllers Instructions .	9
4.3.7. PCECC LSP State Report	9
5. PCEP Objects	10
5.1. OPEN Object	10
5.1.1. PCECC Capability sub-TLV	10
5.2. PATH-SETUP-TYPE TLV	10
5.3. CCI Object	10
6. Security Considerations	11

7. Manageability Considerations	11
7.1. Control of Function and Policy	11
7.2. Information and Data Models	11
7.3. Liveness Detection and Monitoring	11
7.4. Verify Correct Operations	11
7.5. Requirements On Other Protocols	11
7.6. Impact On Network Operations	11
8. IANA Considerations	11
8.1. PCECC-CAPABILITY TLV	12
8.2. PCEP-Error Object	12
9. Acknowledgments	12
10. References	12
10.1. Normative References	12
10.2. Informative References	13
Appendix A. Contributor Addresses	16
Authors' Addresses	16

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the path while leveraging the existing PCE technologies as much as possible.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static P2P LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label-forwarding instructions to program and what resources to reserve. The PCE-based controller keeps a view of the network and determines the paths of the end-to-end LSPs, and the controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

[RFC4857] describes how to set up point-to-multipoint (P2MP) Traffic Engineering Label Switched Paths (TE LSPs) for use in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks. The PCE has been identified as a suitable application for the computation of paths for P2MP TE LSPs ([RFC5671]). The extensions of PCEP to request path computation for P2MP TE LSPs are described in [RFC8306]. Further [RFC8623] specify the extensions that are necessary in order for the deployment of stateful PCEs to support P2MP TE LSPs as well as the setup, maintenance and teardown of PCE-initiated P2MP LSPs under the stateful PCE model.

This document extends

[I-D.ietf-pce-pcep-extension-for-pce-controller] to specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static P2MP LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path with an added functionality of a P2MP branch node. As per [RFC4875], a branch node is an LSR that replicates the incoming data on to one or more outgoing interfaces.

[I-D.ietf-teas-pcecc-use-cases] describes the use cases for P2MP in PCECC architecture.

1.1. 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.

2. Terminology

Terminologies used in this document is same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

3. Basic PCECC Mode

As described in [I-D.ietf-pce-pcep-extension-for-pce-controller], in this mode LSPs are provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label forwarding instructions to program and what resources to reserve. The controller uses PCEP to communicate with each router along the path of the end-to-end LSP. Note that the PCE-based controller will take responsibility for managing some part of the MPLS label space for each of the routers that it controls, and may take wider responsibility for partitioning the label space for each router and allocating different parts for different uses. This is also described in section 3.1.2. of [RFC8283]. For the purpose of this document, it is assumed that label range to be used by a PCE is known and set on both PCEP peers. A future extension could add this capability to advertise the range via possible PCEP extensions as well.

This document extends the functionality to include support for central control instruction for replication at the branch nodes.

The rest of processing is similar to the existing stateful PCE mechanism for P2MP.

4. Procedures for Using the PCE as the Central Controller (PCECC) for P2MP

4.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231] and extended for P2MP [RFC8623]. PCE as a central controller (PCECC) reuses existing Active stateful PCE mechanism as much as possible to control the LSP.

[I-D.ietf-pce-pcep-extension-for-pce-controller] extends PCEP messages - PCRpt, PCInitiate, PCUpd message for the Central

Controller's Instructions (CCI) (label forwarding instructions in the context of this document). This documents specify the procedure for additional instruction for branch node needed for P2MP.

4.2. PCECC Capability Advertisement

As per [I-D.ietf-pce-pcep-extension-for-pce-controller], during PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this PST=PCECC included in the PST list.

[I-D.ietf-pce-pcep-extension-for-pce-controller] also defines the PCECC Capability sub-TLV. A new M-bit is added in PCECC-CAPABILITY TLV to indicate support for PCECC-P2MP. A PCC MUST set M-bit in PCECC-CAPABILITY TLV and include STATEFUL-PCE-CAPABILITY TLV with P2MP bits set ([RFC8623]) in OPEN Object to support the PCECC P2MP extensions defined in this document. If M-bit is set in PCECC-CAPABILITY TLV and N-bit in STATEFUL-PCE-CAPABILITY TLV is not set in OPEN Object, PCE SHOULD send a PCerr message with Error-Type=19 (Invalid Operation) and Error-value=TBD (P2MP capability was not advertised) and terminate the session.

4.3. LSP Operations

The PCEP messages pertaining to PCECC MUST include PATH-SETUP-TYPE TLV [RFC8408] with PST=PCECC

[I-D.ietf-pce-pcep-extension-for-pce-controller] in the SRP object to clearly identify the PCECC LSP is intended.

4.3.1. Basic PCECC LSP Setup

In order to setup a P2MP LSP based on PCECC mechanism, a PCC MUST delegate the P2MP LSP by sending a PCRpt message with PST set for PCECC and D (Delegate) flag (see [RFC8623]) set in the LSP object.

P2MP-LSP-IDENTIFIER TLV [RFC8623] MUST be included for PCECC LSP, the tuple uniquely identifies the P2MP LSP in the network. As per [I-D.ietf-pce-pcep-extension-for-pce-controller], the LSP object is included in central controller's instructions (label download) to identify the PCECC LSP for this instruction.

When a PCE receives PCRpt message with D flags and PST Type set, it calculates the P2MP tree and assigns labels along the path; and set up the path by sending PCInitiate message to each node along the path of the LSP, similar to

[I-D.ietf-pce-pcep-extension-for-pce-controller]. The new extension required is the instructions on the branch nodes for replications to more than one outgoing interfaces with respective labels. The rest

of the operations remains same as
[I-D.ietf-pce-pcep-extension-for-pce-controller] and [RFC8623].

4.3.2. Central Control Instructions

The new central controller's instructions (CCI) for the label operations in PCEP is done via the PCInitiate message, by defining a new PCEP Objects for CCI operations. Local label range of each PCC is assumed to be known at both the PCC and the PCE.

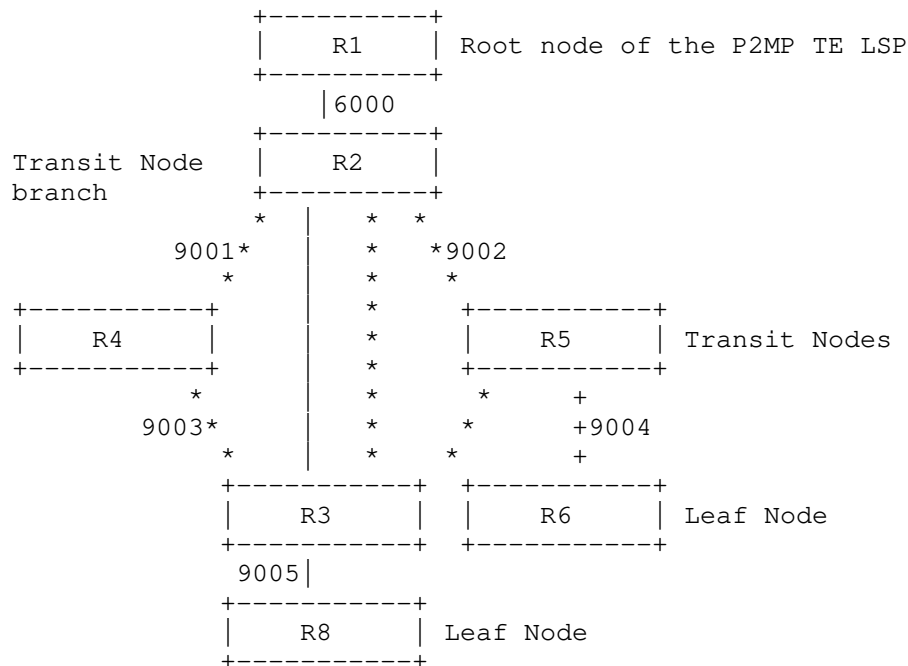
4.3.2.1. Label Download

In order to setup an LSP based on PCECC, the PCE sends a PCInitiate message to each node along the path to download the Label instruction as described in Section 4.3.1.

The CCI object MUST be included, along with the LSP object in the PCInitiate message. The LSP-IDENTIFIER TLV MUST be included in LSP object. The SPEAKER-ENTITY-ID TLV SHOULD be included in LSP object.

As described in [I-D.ietf-pce-pcep-extension-for-pce-controller], if a node (PCC) receives a PCInitiate message which includes a Label to download as part of CCI, that is out of the range set aside for the PCE, it send a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (Label out of range). If a PCC receives a PCInitiate message but failed to download the Label entry, it sends a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (instruction failed).

Consider the example in the [I-D.ietf-teas-pcecc-use-cases] -



PCECC would provision each node along the path and assign incoming and outgoing labels from R1 to {R6, R8} with the path: {R1, 6000}, {6000, R2, {9001,9002}}, {9001, R4, 9003}, {9002, R5, 9004} {9003, R3, 9005}, {9004, R6}, {9005, R8}. The operations on all nodes except R2 are same as [I-D.ietf-pce-pcep-extension-for-pce-controller]. The branch node (R2) needs to be instructed to replicate two copies of the incoming packet, and sent towards R4 and R5 with 9001 and 9002 labels respectively). This done via including 3 instances of CCI objects in the PCEP messages, one for each label in the example, 6000 for incoming and 9001/9002 for outgoing (along with remote nexthop). The message and procedure remains exactly as [I-D.ietf-pce-pcep-extension-for-pce-controller] with only distinction that more than one outgoing CCI MAY be present for the P2MP LSP.

4.3.2.2. Label Cleanup

In order to delete an P2MP LSP based on PCECC, the PCE sends a central controller instructions via a PCInitiate message to each node along the path of the LSP to cleanup the Label forwarding instruction as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. In case of branch nodes all instances of CCIs needs to be present in the PCEP message.

4.3.3. PCE Initiated PCECC LSP

The LSP Instantiation operation is same as defined in [RFC8281] and [RFC8623].

In order to setup a P2MP PCE Initiated LSP based on the PCECC mechanism, a PCE sends PCInitiate message with Path Setup Type set for PCECC (see Section 5.2) to the Ingress PCC (root).

The Ingress PCC MUST also set D (Delegate) flag (see [RFC8231]) and C (Create) flag (see [RFC8281]) in LSP object of PCRpt message. The PCC responds with first PCRpt message with the status as "GOING-UP" and assigned PLSP-ID.

As described in [I-D.ietf-pce-pcep-extension-for-pce-controller], the label forwarding instructions from PCECC are send after the initial PCInitiate and PCRpt exchange. This is done so that the PLSP-ID and other LSP identifiers can be obtained from the ingress and can be included in the label forwarding instruction in the next PCInitiate message. The rest of the PCECC LSP setup operations are same as those described in Section 4.3.1.

4.3.4. PCECC LSP Update

In case of a modification of PCECC P2MP LSP with a new path, the procedure and instructions as described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply.

4.3.5. Re Delegation and Cleanup

In case of a redelgation and cleanup of PCECC P2MP LSP, the procedure and instructions as described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply.

4.3.6. Synchronization of Central Controllers Instructions

The procedure and instructions are as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

4.3.7. PCECC LSP State Report

An Ingress PCC MAY choose to apply any OAM mechanism to check the status of LSP in the Data plane and MAY further send its status in PCRpt message (as per [RFC8623]) to the PCE.

5. PCEP Objects

5.1. OPEN Object

5.1.1. PCECC Capability sub-TLV

The PCECC-CAPABILITY sub-TLV is an optional TLV for use in the OPEN Object for PCECC capability advertisement in PATH-SETUP-TYPE-CAPABILITY TLV as specified in [I-D.ietf-pce-pcep-extension-for-pce-controller].

This document adds a new flag (M-bit) in PCECC-CAPABILITY sub-TLV to indicate the support for P2MP in PCECC. A PCC MUST set M-bit in PCECC-CAPABILITY sub-TLV and set the N (P2MP-CAPABILITY), M (P2MP-LSP-UPDATE-CAPABILITY), and P (P2MP-LSP-INSTITIATION-CAPABILITY) (as per [RFC8623]) in STATEFUL-PCE-CAPABILITY TLV [RFC8231] to support the PCECC P2MP extensions defined in this document. If M-bit is set in PCECC-CAPABILITY sub-TLV and the P2MP bits in STATEFUL-PCE-CAPABILITY TLV are not set in OPEN Object, PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD(P2MP capability was not advertised) and terminate the session.

5.2. PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]; [I-D.ietf-pce-pcep-extension-for-pce-controller] defines a PST value for PCECC, which is also used for P2MP.

5.3. CCI Object

The Central Control Instructions (CCI) Object [I-D.ietf-pce-pcep-extension-for-pce-controller] is used by the PCE to specify the forwarding instructions (Label information in the context of this document) to the PCC, and MAY be carried within PCInitiate or PCRpt message for label download which defined Object Type 1 for MPLS Label, which is also used for P2MP. The address TLVs [I-D.ietf-pce-pcep-extension-for-pce-controller] associates the next-hop information in case of an outgoing label.

If a node (PCC) receives a PCInitiate/PCUpd message with more than one CCI with O-bit set for outgoing label and the node does not support the P2MP branch/replication capability, it MUST respond with PCErr message with Error-Type=2 (Capability not supported).

6. Security Considerations

The security considerations described in [RFC8231], [RFC8281], [RFC8623], and [I-D.ietf-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

7. Manageability Considerations

7.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC P2MP capability as a global configuration.

7.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC P2MP capability.

7.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

7.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

7.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

7.6. Impact On Network Operations

PCEP extensions defined in this document do not put new requirements on network operations.

8. IANA Considerations

8.1. PCECC-CAPABILITY TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY TLV Flag Field registry, as follows:

Bit	Description	Reference
TBD	M((PCECC-P2MP-CAPABILITY))	This document

8.2. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Meaning	
-----	-----	
19	Invalid operation.	
	Error-value = TBD :	P2MP capability was not advertised

9. Acknowledgments

10. References

10.1. 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>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.

- [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>>.
- [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>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8623] Palle, U., Dhody, D., Tanaka, Y., and V. Beeram, "Stateful Path Computation Element (PCE) Protocol Extensions for Usage with Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 8623, DOI 10.17487/RFC8623, June 2019, <<https://www.rfc-editor.org/info/rfc8623>>.
- [I-D.ietf-pce-pcep-extension-for-pce-controller]
Zhao, Q., Li, Z., Negi, M., 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-02 (work in progress), July 2019.

10.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC4857] Fogelstroem, E., Jonsson, A., and C. Perkins, "Mobile IPv4 Regional Registration", RFC 4857, DOI 10.17487/RFC4857, June 2007, <<https://www.rfc-editor.org/info/rfc4857>>.
- [RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <<https://www.rfc-editor.org/info/rfc4875>>.

- [RFC5671] Yasukawa, S. and A. Farrel, Ed., "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) MPLS and GMPLS Traffic Engineering (TE)", RFC 5671, DOI 10.17487/RFC5671, October 2009, <<https://www.rfc-editor.org/info/rfc5671>>.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [RFC8306] Zhao, Q., Dhody, D., Ed., Palleti, R., and D. King, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 8306, DOI 10.17487/RFC8306, November 2017, <<https://www.rfc-editor.org/info/rfc8306>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.
- [I-D.ietf-teas-pcecc-use-cases]
Zhao, Q., Li, Z., Khasanov, B., Dhody, D., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-04 (work in progress), July 2019.

[I-D.ietf-pce-pcep-yang]

Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-12 (work in progress), July 2019.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

Authors' Addresses

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahend.ietf@gmail.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Xuesong Geng
Huawei Technologies
China

EMail: gengxuesong@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: 7 September 2022

Z. Li
S. Peng
X. Geng
Huawei Technologies
M. Negi
RtBrick Inc
6 March 2022

Path Computation Element Communication Protocol (PCEP) Procedures and
Extensions for Using the PCE as a Central Controller (PCECC) of point-
to-multipoint (P2MP) LSPs
draft-dhody-pcep-extension-pce-controller-p2mp-08

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems.

The PCE has been identified as an appropriate technology for the determination of the paths of point-to-multipoint (P2MP) TE Label Switched Paths (LSPs).

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the P2MP LSP can be calculated/set up/initiated and the label-forwarding entries can also be downloaded through a centralized PCE server to each network device along the P2MP path, while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and Path Computation Element Communication Protocol (PCEP) extensions for using the PCE as the central controller for provisioning labels along the path of the static P2MP LSP.

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 7 September 2022.

Copyright Notice

Copyright (c) 2022 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	3
2. Terminology	5
2.1. Requirements Language	5
3. Basic PCECC Mode	5
4. Procedures for Using the PCE as a Central Controller (PCECC) for P2MP	5
4.1. Stateful PCE Model	5
4.2. PCECC Capability Advertisement	6
4.3. LSP Operations	6
4.3.1. PCE-Initiated PCECC LSP	6
4.3.2. PCC-Initiated PCECC LSP	7
4.3.3. Central Control Instructions	7
4.3.3.1. Label Download CCI	7
4.3.3.2. Label Cleanup CCI	9
4.3.4. PCECC LSP Update	9
4.3.5. Re-delegation and Cleanup	9
4.3.6. Synchronization of Central Controllers Instructions	9
4.3.7. PCECC LSP State Report	9
4.3.8. PCC-Based Allocations	9
5. PCEP Messages	10
6. PCEP Objects	10
6.1. OPEN Object	10
6.1.1. PCECC Capability sub-TLV	10
6.2. PATH-SETUP-TYPE TLV	10

6.3. CCI Object	10
7. Security Considerations	11
8. Manageability Considerations	11
8.1. Control of Function and Policy	11
8.2. Information and Data Models	11
8.3. Liveness Detection and Monitoring	12
8.4. Verify Correct Operations	12
8.5. Requirements On Other Protocols	12
8.6. Impact On Network Operations	12
9. IANA Considerations	12
9.1. PCECC-CAPABILITY sub-TLV	12
9.2. PCEP-Error Object	12
10. References	13
10.1. Normative References	13
10.2. Informative References	14
Appendix A. Contributor Addresses	15
Authors' Addresses	16

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload the path computation function from routers in an MPLS traffic-engineered (TE) network. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands. Since then, the role and function of the PCE have grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol.

A PCECC can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/set up/initiated and the label-forwarding entries can also be downloaded through a centralized PCE server to each network device along the path while leveraging the existing PCE technologies as much as possible.

[RFC9050] specify the procedures and PCEP extensions for using the PCE as the central controller for static P2P LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label-forwarding instructions to program and what resources to reserve. The PCE-based controller keeps a view of the network and determines the paths of the end-to-end LSPs, and the controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

[RFC4857] describes how to set up point-to-multipoint (P2MP) Traffic Engineering Label Switched Paths (TE LSPs) for use in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks. The PCE has been identified as a suitable application for the computation of paths for P2MP TE LSPs ([RFC5671]). The extensions of PCEP to request path computation for P2MP TE LSPs are described in [RFC8306]. Further [RFC8623] specify the extensions that are necessary in order for the deployment of stateful PCEs to support P2MP TE LSPs as well as the setup, maintenance and teardown of PCE-initiated P2MP LSPs under the stateful PCE model.

This document extends [RFC9050] to specify the procedures and PCEP extensions for using the PCE as the central controller for static P2MP LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path with an added functionality of a P2MP branch node. As per [RFC4875], a branch node is an LSR that replicates the incoming data on to one or more outgoing interfaces. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for P2MP in PCECC architecture.

2. Terminology

Terminologies used in this document is the same as described in the draft [RFC8283].

2.1. 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.

3. Basic PCECC Mode

Section 3 of [RFC9050] describe the PCECC model of operation.

This document extends the functionality to include support for central control instruction for replication at the branch nodes for the P2MP LSP.

The rest of the processing at the root node is similar to the existing stateful PCE mechanism for P2MP [RFC8623].

4. Procedures for Using the PCE as a Central Controller (PCECC) for P2MP

4.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231] and extended for P2MP [RFC8623]. A PCE as a Central Controller (PCECC) reuses the existing active stateful PCE mechanism as much as possible to control the LSPs.

[RFC9050] extends PCEP messages - PCInitiate, PCRpt, and PCUpd message for the Central Controller's Instructions (CCI) (label-forwarding instructions in the context of this document). This document specify the procedure for additional instruction for branch node needed for P2MP.

4.2. PCECC Capability Advertisement

As per Section 5.4 of [RFC9050], during the PCEP initialization phase, PCEP Speakers (PCE or PCC) advertise their support of and willingness to use PCEP extension for the PCECC using a new Path Setup Type (PST) in PATH-SETUP-TYPE-CAPABILITY TLV and a new PCECC-CAPABILITY sub-TLV.

A new M bit is added in the PCECC-CAPABILITY sub-TLV to indicate support for PCECC-P2MP. A PCC MUST set the M bit in the PCECC-CAPABILITY sub-TLV and include STATEFUL-PCE-CAPABILITY TLV with the P2MP bits set (as per [RFC8623]) in the OPEN object to support the PCECC P2MP extensions defined in this document.

If the M bit is set in PCECC-CAPABILITY sub-TLV and the STATEFUL-PCE-CAPABILITY TLV is not advertised, or is advertised without the N bit set, in the OPEN object, the receiver MUST:

- * send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD2 (P2MP capability was not advertised) and
- * terminate the session.

The rest of the processing is as per [RFC9050].

4.3. LSP Operations

The PCEP messages pertaining to a PCECC includes the PATH-SETUP-TYPE TLV [RFC8408] in the SRP object [RFC8231] with the PST set to '2' to clearly identify the the PCECC LSP is intended as per [RFC9050].

4.3.1. PCE-Initiated PCECC LSP

The LSP instantiation operation is the same as defined in [RFC8281] and [RFC8623].

In order to set up a PCE-Initiated P2MP LSP based on the PCECC mechanism, a PCE sends a PCInitiate message with the PST set to '2' for the PCECC ([RFC9050]) to the ingress PCC (root node).

As described in [RFC9050], the label-forwarding instructions from PCECC are sent after the initial PCInitiate and PCRpt exchange. This is done so that the PCEP-specific identifier for the LSP (PLSP-ID) and other LSP identifiers can be obtained from the ingress and can be included in the label-forwarding instruction in the next set of PCInitiate message along the path.

An P2MP-LSP-IDENTIFIER TLV [RFC8623] MUST be included for the PCECC P2MP LSPs, it uniquely identifies the P2MP LSP in the network. As per [RFC9050], the LSP object is included in the central controller's instructions (label download) to identify the PCECC P2MP LSP for this instruction. The handling of PLSP-ID is as per [RFC9050].

The ingress PCC (root) also sets the D (Delegate) flag (see [RFC8231]) and C (Create) flag (see [RFC8281]) in the LSP object of the PCRpt message. As per [RFC9050], when the PCE receives this PCRpt message with the PLSP-ID, it assigns labels along the path and sets up the path by sending a PCInitiate message to each node along the path of the P2MP Tree as per the PCECC technique. The CC-ID uniquely identifies the central controller instruction within a PCEP session. Each node along the path (PCC) responds with the PCRpt messages to acknowledge the CCI with the PCRpt messages including the CCI and the LSP objects. The only new extension required is the instructions on the branch nodes for replications to more than one outgoing interface with the respective label. The rest of the operations remains the same as [RFC9050] and [RFC8623].

4.3.2. PCC-Initiated PCECC LSP

In order to set up a P2MP LSP based on the PCECC mechanism where the LSP is configured at the PCC, a PCC MUST delegate the P2MP LSP by sending a PCRpt message with the PST set for the PCECC and D (Delegate) flag (see [RFC8623]) set in the LSP object.

When a PCE receives the initial PCRpt message with the D flags and PST Type set to '2', it SHOULD calculate the P2MP tree and assign labels along the P2MP tree in addition to setting up the P2MP LSP by sending PCInitiate message to each node along the path of the P2MP LSP as per [RFC9050]. The only new extension required is the instructions on the branch nodes for replications to more than one outgoing interface with the respective label. The rest of the operations remains the same as [RFC9050] and [RFC8623].

4.3.3. Central Control Instructions

The CCI for the label operations in PCEP are done via the PCInitiate message as described in [RFC9050], by defining a PCEP Objects for CCI operations. The local label range of each PCC is assumed to be known by both the PCC and the PCE.

4.3.3.1. Label Download CCI

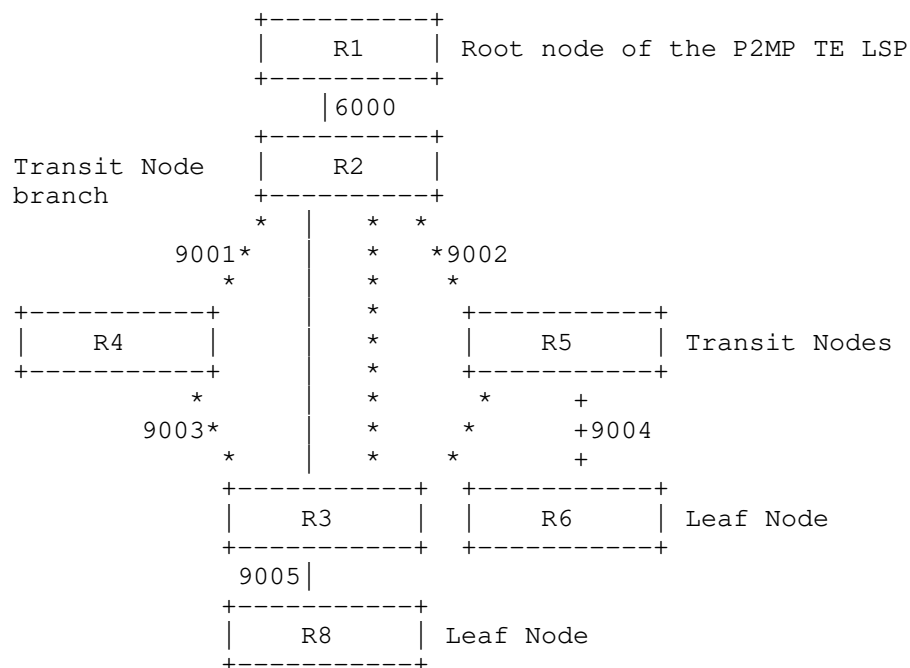
In order to set up an LSP based on the PCECC, the PCE sends a PCInitiate message to each node along the path to download the label instructions, as described in Section 4.3.1 and Section 4.3.2.

The CCI object MUST be included, along with the LSP object in the PCInitiate message. As per [RFC9050], there are at most 2 instances of CCI object in the PCInitiate message. For PCECC-P2MP operations, multiple instances of CCI object for out-labels is allowed. Similarly to acknowledge the central controller instructions, the PCRpt message allows multiple instances of CCI object for PCECC-P2MP operations.

The P2MP-LSP-IDENTIFIERS TLV MUST be included in the LSP object for the PCECC based P2MP LSP. The SPEAKER-ENTITY-ID TLV SHOULD be included in LSP object.

As described in [RFC9050], if a node (PCC) receives a PCInitiate message that includes a label to download (as part of CCI) that is out of the range set aside for the PCE, it send a PCErr message with Error-type=3 (PCECC failure) and Error-value=1 (Label out of range) ([RFC9050]). If a PCC receives a PCInitiate message but fails to download the label entry, it sends a PCErr message with Error-type=3 (PCECC failure) and Error-value=2 (Instruction failed) ([RFC9050]).

Consider the example in the [I-D.ietf-teas-pcecc-use-cases] -



PCECC would provision each node along the path and assign incoming and outgoing labels from R1 to {R6, R8} with the path: {R1, 6000}, {6000, R2, {9001,9002}}, {9001, R4, 9003}, {9002, R5, 9004} {9003, R3, 9005}, {9004, R6}, {9005, R8}. The operations on all nodes except R2 are same as [RFC9050]. The branch node (R2) needs to be instructed to replicate two copies of the incoming packet, and sent towards R4 and R5 with 9001 and 9002 labels respectively). This done via including 3 instances of CCI objects in the PCEP messages, one for each label in the example, 6000 for incoming and 9001/9002 for outgoing (along with remote nexthop). The message and procedure remains exactly as [RFC9050] with only distinction that more than one outgoing CCI MAY be present for the P2MP LSP.

4.3.3.2. Label Cleanup CCI

In order to delete a P2MP LSP based on the PCECC, the PCE sends a Central Controller Instructions via a PCInitiate message to each node along the path of the P2MP tree to clean up the label-forwarding instruction as per [RFC9050]. In case of branch nodes, all instances of CCIs needs to be present in the PCEP message.

4.3.4. PCECC LSP Update

In case of a modification of PCECC P2MP LSP with a new path, the procedure, and instructions as described in [RFC9050] apply.

4.3.5. Re-delegation and Cleanup

In case of a re-delegation and clean up of PCECC P2MP LSP, the procedure, and instructions as described in [RFC9050] apply.

4.3.6. Synchronization of Central Controllers Instructions

The procedure and instructions are as per [RFC9050].

4.3.7. PCECC LSP State Report

An ingress PCC MAY choose to apply any Operations, Administration, and Maintenance (OAM) mechanism to check the status of the LSP in the data plane and MAY further send its status in the PCRpt message (as per [RFC8623]) to the PCE.

4.3.8. PCC-Based Allocations

The PCE can request the PCC to allocate the label using the PCInitiate message. The procedure and instructions are as per Section 5.5.8 of [RFC9050].

5. PCEP Messages

[RFC9050] specify the extension to PCInitiate and PCRpt message for PCECC. For P2P LSP, only two instances of CCI objects can be included. In the case of the P2MP LSP, multiple CCI objects are allowed. The message format and other procedures continue to apply.

6. PCEP Objects

6.1. OPEN Object

6.1.1. PCECC Capability sub-TLV

The PCECC-CAPABILITY sub-TLV is an optional TLV for use in the OPEN Object for PCECC capability advertisement in PATH-SETUP-TYPE-CAPABILITY TLV as specified in [RFC9050].

This document adds a new flag (M Bit) in the PCECC-CAPABILITY sub-TLV to indicate the support for P2MP in PCECC.

M (PCECC-P2MP-CAPABILITY - 1 bit - TBD1): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable of PCECC-P2MP capability.

A PCC MUST set the M Bit in the PCECC-CAPABILITY sub-TLV and set the N (P2MP-CAPABILITY), the M (P2MP-LSP-UPDATE-CAPABILITY), and the P (P2MP-LSP-INstantiation-CAPABILITY) bits (as per [RFC8623]) in the STATEFUL-PCE-CAPABILITY TLV [RFC8231] to support the PCECC-P2MP extensions defined in this document. If the M Bit is set in PCECC-CAPABILITY sub-TLV and the P2MP bits (in the STATEFUL-PCE-CAPABILITY TLV) are not set in the OPEN Object, a PCEP speaker SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD2 (P2MP capability was not advertised) and terminate the session.

6.2. PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]; [RFC9050] defines a PST value for PCECC as '2', which is applicable for P2MP LSP as well.

6.3. CCI Object

The CCI object [RFC9050] is used by the PCE to specify the forwarding instructions (label information in the context of this document) to the PCC, and optionally carried within PCInitiate or PCRpt message for label download/report. The CCI Object Type 1 for MPLS Label is defined in [RFC9050], which is used for the P2MP LSPs as well. The address TLVs are defined in [RFC9050], they associate the next-hop

information in case of an outgoing label.

If a node (PCC) receives a PCInitiate message with more than one CCI with O-bit set for the outgoing label and the node does not support the P2MP branch/replication capability, it MUST respond with PCErr message with Error-Type=2 (Capability not supported) (defined in [RFC5440]).

The rest of the processing is same as [RFC9050].

7. Security Considerations

As per [RFC8283], the security considerations for a PCE-based controller are a little different from those for any other PCE system. That is, the operation relies heavily on the use and security of PCEP, so consideration should be given to the security features discussed in [RFC5440] and the additional mechanisms described in [RFC8253]. It further lists the vulnerability of a central controller architecture, such as a central point of failure, denial of service, and a focus for interception and modification of messages sent to individual Network Elements (NEs).

The security considerations described in [RFC8231], [RFC8281], [RFC8623], and [RFC9050] apply to the extensions described in this document.

As per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

8. Manageability Considerations

8.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC-P2MP capability as a global configuration.

8.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC-P2MP capability.

8.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

8.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

8.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

8.6. Impact On Network Operations

PCEP extensions defined in this document do not put new requirements on network operations.

9. IANA Considerations

9.1. PCECC-CAPABILITY sub-TLV

[RFC9050] defines the PCECC-CAPABILITY sub-TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY sub-TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY sub-TLV Flag Field registry, as follows:

Bit	Description	Reference
TBD1	P2MP	This document

Table 1

9.2. PCEP-Error Object

IANA is requested to allocate a new error value within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Meaning
19	Invalid operation.

Error-value = TBD2 : P2MP capability was
not advertised

The Reference is marked as "This document".

10. References

10.1. 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>>.
- [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>>.
- [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>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8623] Palle, U., Dhody, D., Tanaka, Y., and V. Beeram, "Stateful Path Computation Element (PCE) Protocol Extensions for Usage with Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 8623, DOI 10.17487/RFC8623, June 2019, <<https://www.rfc-editor.org/info/rfc8623>>.

- [RFC9050] Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "Path Computation Element Communication Protocol (PCEP) Procedures and Extensions for Using the PCE as a Central Controller (PCECC) of LSPs", RFC 9050, DOI 10.17487/RFC9050, July 2021, <<https://www.rfc-editor.org/info/rfc9050>>.

10.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC4857] Fogelstroem, E., Jonsson, A., and C. Perkins, "Mobile IPv4 Regional Registration", RFC 4857, DOI 10.17487/RFC4857, June 2007, <<https://www.rfc-editor.org/info/rfc4857>>.
- [RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <<https://www.rfc-editor.org/info/rfc4875>>.
- [RFC5671] Yasukawa, S. and A. Farrel, Ed., "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) MPLS and GMPLS Traffic Engineering (TE)", RFC 5671, DOI 10.17487/RFC5671, October 2009, <<https://www.rfc-editor.org/info/rfc5671>>.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.

- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [RFC8306] Zhao, Q., Dhody, D., Ed., Palleti, R., and D. King, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 8306, DOI 10.17487/RFC8306, November 2017, <<https://www.rfc-editor.org/info/rfc8306>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.
- [I-D.ietf-teas-pcecc-use-cases]
Li, Z. (., Dhody, D., Zhao, Q., Ke, K., Khasanov, B., Fang, L., Zhou, C., Zhang, B., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", Work in Progress, Internet-Draft, draft-ietf-teas-pcecc-use-cases-08, 25 October 2021, <<https://datatracker.ietf.org/doc/html/draft-ietf-teas-pcecc-use-cases-08>>.
- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V. P., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-yang-18, 25 January 2022, <<https://datatracker.ietf.org/doc/html/draft-ietf-pce-pcep-yang-18>>.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

Email: dhruv.ietf@gmail.com

Udayasree Palle

Email: udayasreereddy@gmail.com

Authors' Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: pengshuping@huawei.com

Xuesong Geng
Huawei Technologies
China
Email: gengxuesong@huawei.com

Mahendra Singh Negi
RtBrick Inc
N-17L, 18th Cross Rd, HSR Layout
Bangalore 560102
Karnataka
India
Email: mahend.ietf@gmail.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 23, 2020

M. Negi
Z. Li
X. Geng
Huawei Technologies
August 22, 2019

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) for SRv6
draft-dhody-pce-pcep-extension-pce-controller-srv6-02

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

PCE was developed to derive paths for MPLS Label Switched Paths (LSPs), which are supplied to the head end of the LSP using the Path Computation Element Communication Protocol (PCEP). But SDN has a broader applicability than signaled (G)MPLS traffic-engineered (TE) networks, and the PCE may be used to determine paths in a range of use cases. PCEP has been proposed as a control protocol for use in these environments to allow the PCE to be fully enabled as a central controller.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. This document specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers for Segment Routing in IPv6 (SRv6), in addition to computing the SRv6 paths for packet flows and telling the edge routers what instructions to attach to packets as they enter the network.

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 February 23, 2020.

Copyright Notice

Copyright (c) 2019 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	3
1.1. Requirements Language	5
2. Terminology	5
3. PCECC SRv6	5
4. PCEP Requirements	6
5. Procedures for Using the PCE as the Central Controller (PCECC) in SRv6	6
5.1. Stateful PCE Model	6
5.2. New Functions	6
5.3. PCECC Capability Advertisement	7
5.4. PCEP session IP address and TEDB Router ID	7
5.5. SRv6 Path Operations	7
5.5.1. PCECC Segment Routing in IPv6 (SRv6)	7
5.5.1.1. PCECC SRv6 Node/Prefix SID allocation	7
5.5.1.2. PCECC SRv6 Adjacency SID allocation	8
5.5.1.3. Redundant PCEs	9
5.5.1.4. Re Delegation and Cleanup	9
5.5.1.5. Synchronization of SRv6 SID Allocations	9
6. PCEP messages	9
7. PCEP Objects	9
7.1. OPEN Object	9
7.1.1. PCECC Capability sub-TLV	9
7.2. PATH-SETUP-TYPE TLV	10
7.3. CCI Object	10

7.4. FEC Object	11
8. Security Considerations	11
9. Manageability Considerations	11
9.1. Control of Function and Policy	11
9.2. Information and Data Models	11
9.3. Liveness Detection and Monitoring	12
9.4. Verify Correct Operations	12
9.5. Requirements On Other Protocols	12
9.6. Impact On Network Operations	12
10. IANA Considerations	12
10.1. PCECC-CAPABILITY TLV	12
10.2. New Path Setup Type Registry	12
10.3. PCEP-Error Object	13
11. Acknowledgments	13
12. References	13
12.1. Normative References	13
12.2. Informative References	14
Appendix A. Contributor Addresses	17
Authors' Addresses	17

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440].

This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol.

[I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCECC architecture.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment Routing (SR) technology leverages the source routing and tunneling paradigms. A source node can choose a path without relying on hop-by-hop signaling protocols such as LDP or RSVP-TE. Each path is specified as a set of "segments" advertised by link-state routing protocols (IS-IS or OSPF). [RFC8402] provides an introduction to SR architecture. The corresponding IS-IS and OSPF extensions are specified in [I-D.ietf-isis-segment-routing-extensions] and [I-D.ietf-ospf-segment-routing-extensions], respectively. It relies on a series of forwarding instructions being placed in the header of a packet. The list of segment forming the path is called the Segment List and is encoded in the packet header. Segment Routing can be applied to the IPv6 architecture with the Segment Routing Header (SRH) [I-D.ietf-6man-segment-routing-header]. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is indicated by the Destination Address of the packet. Upon completion of a segment, a pointer in the new routing header is incremented and indicates the next segment. The segment routing architecture supports operations that can be used to steer packet flows in a network, thus providing a form of traffic engineering. [I-D.ietf-pce-segment-routing] and [I-D.ietf-pce-segment-routing-ipv6] specify the SR specific PCEP extensions.

PCECC may further use PCEP protocol for SR SID (Segment Identifier) distribution on the SR nodes with some benefits.

[I-D.zhao-pce-pcep-extension-pce-controller-sr] specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR SID distribution in this case), in addition to computing the paths for packet flows in a segment routing network and telling

the edge routers what instructions to attach to packets as they enter the network. This document extends this to include SRv6 SID distribution as well.

1.1. 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.

2. Terminology

Terminologies used in this document is same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

3. PCECC SRv6

[I-D.ietf-pce-segment-routing] specifies extensions to PCEP that allow a stateful PCE to compute, update or initiate SR-TE paths for MPLS dataplane. An ingress node of an SR-TE path appends all outgoing packets with a list of MPLS labels (SIDs). This is encoded in SR-ERO subobject, capable of carrying a label (SID) as well as the identity of the node/adjacency label (SID).

[I-D.ietf-pce-segment-routing-ipv6] extends the procedure to include support for SRv6 paths.

As per [I-D.ietf-6man-segment-routing-header], an SRv6 Segment is a 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment". Further details are in an illustration provided in [I-D.ietf-spring-srv6-network-programming]. The SR is applied to IPV6 forwarding plane using SRH. A SR path can be derived from an IGP Shortest Path Tree (SPT), but SR-TE paths may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool, or a PCE and provisioned on the ingress node.

[I-D.ietf-pce-segment-routing-ipv6] extended SR-ERO subobject capable of carrying an SRv6 SID as well as the identity of the node/adjacency represented by the SID.

As per [RFC8283], PCE as a central controller can allocate and provision the node/prefix/adjacency label (SID) via PCEP. As per [I-D.ietf-teas-pcecc-use-cases] this is also applicable to SRv6 SIDs.

Rest of the processing is similar to existing stateful PCE with SRv6 mechanism.

4. PCEP Requirements

Following key requirements for PCECC-SRv6 should be considered when designing the PCECC based solution:

- o PCEP speaker supporting this draft MUST have the capability to advertise its PCECC-SRv6 capability to its peers.
- o PCEP speaker not supporting this draft MUST be able to reject PCECC-SRv6 related message with a reason code that indicates no support for it.
- o PCEP procedures MUST provide a means to update (or cleanup) the SRv6 SID to the PCC.
- o PCEP procedures SHOULD provide a means to synchronize the SRv6 SID allocations between PCE to PCC in the PCEP messages.

5. Procedures for Using the PCE as the Central Controller (PCECC) in SRv6

5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. PCE as a central controller (PCECC) reuses existing Active stateful PCE mechanism as much as possible to control the LSP.

5.2. New Functions

This document uses the same PCEP messages and its extensions which are described in [I-D.ietf-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr] for PCECC-SRv6 as well.

PCEP messages PCRpt, PCInitiate, PCUpd are also used to send LSP Reports, LSP setup and LSP update respectively. The extended PCInitiate message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is used to download or cleanup central controller's instructions (CCIs) (SRv6 SID in scope of this document). The extended PCRpt message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is also used to report the CCIs (SRv6 SIDs) from PCC to PCE.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify an object called CCI for the encoding of central controller's instructions. [I-D.zhao-pce-pcep-extension-pce-controller-sr] extends the CCI by defining a object-type for segment routing. This document further extends the CCI by defining another object-type for SRv6.

5.3. PCECC Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions. A PCEP Speaker includes the "PCECC Capability" sub-TLV, described in [I-D.ietf-pce-pcep-extension-for-pce-controller].

A S-bit is added in PCECC-CAPABILITY sub-TLV to indicate support for PCECC-SR in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. This document adds another I-bit to indicate support for SR in IPv6. A PCC MUST set I-bit in PCECC-CAPABILITY sub-TLV and include SRv6-PCE-CAPABILITY sub-TLV ([I-D.ietf-pce-segment-routing-ipv6]) in OPEN Object (inside the the PATH-SETUP-TYPE-CAPABILITY TLV) to support the PCECC SRv6 extensions defined in this document. If I-bit is set in PCECC-CAPABILITY sub-TLV and SRv6-PCE-CAPABILITY sub-TLV is not advertised in OPEN Object, PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD (SRv6 capability was not advertised) and terminate the session.

5.4. PCEP session IP address and TEDB Router ID

As described in [I-D.zhao-pce-pcep-extension-pce-controller-sr], it is important to link the session IP address with the Router ID in TEDB for successful PCECC operations.

5.5. SRv6 Path Operations

The PCEP messages pertaining to PCECC-SRv6 MUST include PATH-SETUP-TYPE TLV [RFC8408] with PST=TBD in the SRP object to clearly identify the PCECC-SRv6 setup is intended.

5.5.1. PCECC Segment Routing in IPv6 (SRv6)

Segment Routing (SR) as described in [RFC8402] depends on "segments" that are advertised by Interior Gateway Protocols (IGPs). The SR-node allocates and advertises the SID (node, adj etc) and flood via the IGP. This document proposes a new mechanism where PCE allocates the SRv6 SID centrally and uses PCEP to advertise the SRv6 SID. In some deployments PCE (and PCEP) are better suited than IGP because of centralized nature of PCE and direct TCP based PCEP session to the node.

5.5.1.1. PCECC SRv6 Node/Prefix SID allocation

Each node (PCC) is allocated a node SRv6 SID by the PCECC. The PCECC sends PCInitiate message to update the SID table of each node. The TE router ID is determined from the TEDB or from "IPv4/IPv6 Router-ID" Sub-TLV [I-D.dhodylee-pce-pcep-ls], in the OPEN Object.

On receiving the SRv6 node SID allocation, each node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly. The PCInitiate message in this case MUST have FEC object.

On receiving the SRv6 node SID allocation:

For the local SID, node (PCC) needs to update SID with associated function (END function in this case) in "My Local SID Table" ([I-D.ietf-spring-srv6-network-programming]).

For the non-local SID, node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly.

The PCInitiate message in this case MUST have FEC object.

The forwarding behavior and the end result is similar to IGP based "Node-SID" in SRv6. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node.

PCE relies on the Node/Prefix SRv6 SID cleanup using the same PCInitiate message.

5.5.1.2. PCECC SRv6 Adjacency SID allocation

[I-D.ietf-pce-segment-routing] extends PCEP to allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

For PCECC SR, apart from node-SID, Adj-SID is used where each adjacency is allocated an Adj-SID by the PCECC. The PCECC sends PCInitiate message to update the label map of each Adj to the corresponding nodes in the domain. Each node (PCC) download the SRv6 SID instructions accordingly. Similar to SRv6 Node/Prefix Label allocation, the PCInitiate message in this case uses the FEC object.

The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SRv6.

The Path Setup Type for segment routing MUST be set for PCECC SRv6 = TBD (see Section 7.2). All PCEP procedures and mechanism are similar to [I-D.ietf-pce-segment-routing].

PCE relies on the Adj label cleanup using the same PCInitiate message.

5.5.1.3. Redundant PCEs

[I-D.litkowski-pce-state-sync] describes synchronization mechanism between the stateful PCEs. The SRv6 SIDs allocated by a PCE MUST also be synchronized among PCEs for PCECC SRv6 state synchronization. Note that the SRv6 SIDs are independent to the PCECC-SRv6 paths, and remains intact till any topology change. The redundant PCEs MUST have a common view of all SRv6 SIDs allocated in the domain.

5.5.1.4. Re Delegation and Cleanup

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the action needed for CCIs for the Basic PCECC LSP on this terminated session. Similarly actions should be applied for the SRv6 SID as well.

5.5.1.5. Synchronization of SRv6 SID Allocations

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the synchronization of Central Controller's Instructions (CCI) via LSP state synchronization as described in [RFC8231] and [RFC8232]. Same procedures should be applied for SRv6 SIDs as well.

6. PCEP messages

The PCEP message is as per
[I-D.zhao-pce-pcep-extension-pce-controller-sr].

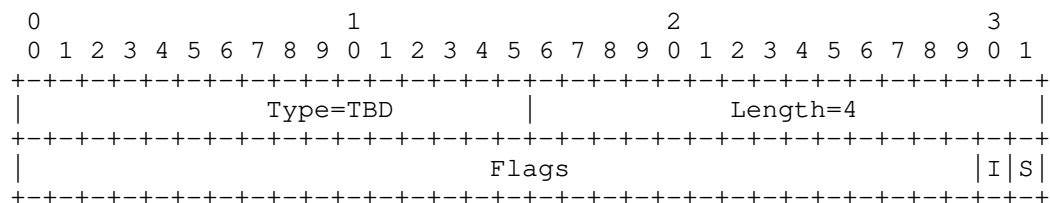
7. PCEP Objects

7.1. OPEN Object

7.1.1. PCECC Capability sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defined the PCECC-CAPABILITY TLV.

A new I-bit is defined in PCECC-CAPABILITY sub-TLV for PCECC-SRv6:



I (PCECC-SRv6-CAPABILITY - 1 bit): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable for PCECC-SRv6 capability and PCE would allocate node and Adj SRv6 SID on this session.

7.2. PATH-SETUP-TYPE TLV

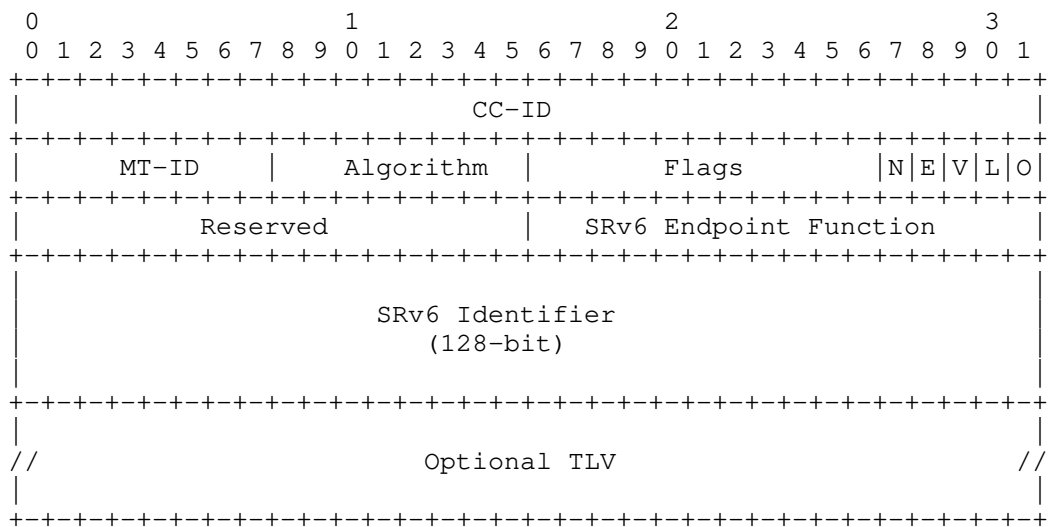
The PATH-SETUP-TYPE TLV is defined in [RFC8408]. PST = TBD is used when Path is setup via PCECC SRv6 mode.

On a PCRpt/PCUpd/PCInitiate message, the PST=TBD indicates that this path was setup via a PCECC-SRv6 based mechanism where either the SIDs were allocated/instructed by PCE via PCECC mechanism.

7.3. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the forwarding instructions is defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]. This document defines another object-type for SRv6 purpose.

CCI Object-Type is TBD for SRv6 as below -



The field CC-ID is as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. The field MT-ID, Algorithm, Flags are defined in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

Reserved: MUST be set to 0 while sending and ignored on receipt.

SRv6 Endpoint Function: 16 bit field representing supported functions associated with SRv6 SIDs.

SRv6 Identifier: 128 bit IPv6 addresses representing SRv6 segment.

[Editor's Note - It might be useful to separate the LOC:FUNC part in the SRv6 SID]

7.4. FEC Object

The FEC Object is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message.

FEC Object (and various Object-Types) are described in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. SRv6 Node SID MUST include the FEC Object-Type 2 for IPv6 Node. SRv6 Adjacency SID MUST include the FEC Object-Type=4 for IPv6 adjacency. Further FEC object types would be added in future revisions.

8. Security Considerations

The security considerations described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

As per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

9. Manageability Considerations

9.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC SR capability as a global configuration.

9.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC SR capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC SR capability.

9.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

9.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

9.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

9.6. Impact On Network Operations

PCEP implementation SHOULD allow a limit to be placed on the rate of PCInitiate/PCUpd messages (as per [RFC8231]) sent by PCE and processed by PCC. It SHOULD also allow sending a notification when a rate threshold is reached.

10. IANA Considerations

10.1. PCECC-CAPABILITY TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY TLV Flag Field registry, as follows:

Bit	Description	Reference
TBD	I((PCECC-SRv6-CAPABILITY))	This document

10.2. New Path Setup Type Registry

IANA is requested to allocate new PST Field in PATH- SETUP-TYPE TLV. The allocation policy for this new registry should be by IETF Consensus. The new registry should contain the following value:

Value	Description	Reference
TBD	Path is setup using PCECC-SRv6 mode	This document

10.3. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Meaning
-----	-----
19	Invalid operation.
	Error-value = TBD : SRv6 capability was not advertised

11. Acknowledgments

12. References

12.1. 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>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.
- [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>>.
- [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>>.

[RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.

[I-D.ietf-pce-segment-routing-ipv6]
Negi, M., Li, C., Sivabalan, S., Kaladharan, P., and Y. Zhu, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-ietf-pce-segment-routing-ipv6-02 (work in progress), April 2019.

[I-D.ietf-pce-pcep-extension-for-pce-controller]
Zhao, Q., Li, Z., Negi, M., 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-02 (work in progress), July 2019.

[I-D.zhao-pce-pcep-extension-pce-controller-sr]
Zhao, Q., Li, Z., Negi, M., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of SR-LSPs", draft-zhao-pce-pcep-extension-pce-controller-sr-05 (work in progress), July 2019.

12.2. Informative References

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.

[RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.

[RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.

- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.
- [RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <<https://www.rfc-editor.org/info/rfc8232>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [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>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.
- [I-D.ietf-teas-pcecc-use-cases]
Zhao, Q., Li, Z., Khasanov, B., Dhody, D., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-04 (work in progress), July 2019.
- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-12 (work in progress), July 2019.

- [I-D.ietf-pce-segment-routing]
Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W.,
and J. Hardwick, "PCEP Extensions for Segment Routing",
draft-ietf-pce-segment-routing-16 (work in progress),
March 2019.
- [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.litkowski-pce-state-sync]
Litkowski, S., Sivabalan, S., Li, C., and H. Zheng, "Inter
Stateful Path Computation Element (PCE) Communication
Procedures.", draft-litkowski-pce-state-sync-06 (work in
progress), July 2019.
- [I-D.dhodylee-pce-pcep-ls]
Dhody, D., Lee, Y., and D. Ceccarelli, "PCEP Extension for
Distribution of Link-State and TE Information.", draft-
dhodylee-pce-pcep-ls-13 (work in progress), February 2019.
- [I-D.ietf-spring-srv6-network-programming]
Filsfils, C., Camarillo, P., Leddy, J.,
daniel.voyer@bell.ca, d., Matsushima, S., and Z. Li, "SRv6
Network Programming", draft-ietf-spring-srv6-network-
programming-01 (work in progress), July 2019.
- [I-D.ietf-6man-segment-routing-header]
Filsfils, C., Dukes, D., Previdi, S., Leddy, J.,
Matsushima, S., and d. daniel.voyer@bell.ca, "IPv6 Segment
Routing Header (SRH)", draft-ietf-6man-segment-routing-
header-22 (work in progress), August 2019.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Authors' Addresses

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahend.ietf@gmail.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Xuesong Geng
Huawei Technologies
China

EMail: gengxuesong@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: 7 September 2022

Z. Li
S. Peng
X. Geng
Huawei Technologies
M. Negi
RtBrick Inc
6 March 2022

Path Computation Element Communication Protocol (PCEP) Procedures and
Extensions for Using the PCE as a Central Controller (PCECC) for SRv6
SID Allocation and Distribution.
draft-dhody-pce-pcep-extension-pce-controller-srv6-08

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems.

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. This document specifies the procedures and PCEP extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers, in addition to computing the paths for packet flows in the for Segment Routing (SR) in IPv6 (SRv6) network and telling the edge routers what instructions to attach to packets as they enter the network. PCECC is further enhanced for SRv6 SID (Segment Identifier) allocation and distribution.

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 7 September 2022.

Copyright Notice

Copyright (c) 2022 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	3
2. Terminology	5
2.1. Requirements Language	5
3. PCECC SRv6	5
4. PCEP Requirements	6
5. Procedures for Using the PCE as a Central Controller (PCECC) in SRv6	6
5.1. Stateful PCE Model	6
5.2. New Functions	6
5.3. PCECC Capability Advertisement	7
5.4. PCEP session IP address and TED Router ID	7
5.5. SRv6 Path Operations	7
5.5.1. PCECC Segment Routing in IPv6 (SRv6)	8
5.5.1.1. PCECC SRv6 Node/Prefix SID allocation	8
5.5.1.2. PCECC SRv6 Adjacency SID allocation	9
5.5.1.3. Redundant PCEs	9
5.5.1.4. Re-Delegation and Cleanup	9
5.5.1.5. Synchronization of SRv6 SID Allocations	9
6. PCEP Messages	9
7. PCEP Objects	10
7.1. OPEN Object	10
7.1.1. PCECC Capability sub-TLV	10
7.2. SRv6 Path Setup	10
7.3. CCI Object	10
7.4. FEC Object	11
8. Security Considerations	12
9. Manageability Considerations	12
9.1. Control of Function and Policy	12
9.2. Information and Data Models	12
9.3. Liveness Detection and Monitoring	13
9.4. Verify Correct Operations	13
9.5. Requirements On Other Protocols	13

9.6. Impact On Network Operations	13
10. IANA Considerations	13
10.1. PCECC-CAPABILITY sub-TLV	13
10.2. PCEP Object	13
10.3. PCEP-Error Object	14
11. References	14
11.1. Normative References	14
11.2. Informative References	15
Appendix A. Contributor Addresses	18
Authors' Addresses	18

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload the path computation function from routers in an MPLS traffic-engineered (TE) network. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands. Since then, the role and function of the PCE have grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and

applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCECC architecture.

[RFC9050] specify the procedures and PCEP extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment Routing (SR) technology leverages the source routing and tunneling paradigms. A source node can choose a path without relying on hop-by-hop signaling protocols such as LDP or RSVP-TE. Each path is specified as a set of "segments" advertised by link-state routing protocols (IS-IS or OSPF). [RFC8402] provides an introduction to SR architecture. The corresponding IS-IS and OSPF extensions are specified in [RFC8667] and [RFC8665], respectively. It relies on a series of forwarding instructions being placed in the header of a packet. The list of segments forming the path is called the Segment List and is encoded in the packet header. Segment Routing can be applied to the IPv6 architecture with the Segment Routing Header (SRH) [RFC8754]. A segment is encoded as an IPv6 address. An ordered list of segments is encoded as an ordered list of IPv6 addresses in the routing header. The active segment is indicated by the Destination Address of the packet. Upon completion of a segment, a pointer in the new routing header is incremented and indicates the next segment. The segment routing architecture supports operations that can be used to steer packet flows in a network, thus providing a form of traffic engineering. [RFC8664] and [I-D.ietf-pce-segment-routing-ipv6] specify the SR specific PCEP extensions.

PCECC may further use PCEP for SR SID (Segment Identifier) allocation and distribution to all the SR nodes with some benefits. The SR nodes continue to rely on IGP for distributed computation (nexthop selection, protection etc) where PCE (and PCEP) does only the allocation and distribution of SRv6 SIDs in the network. Note that the topology at PCE is still learned via existing mechanisms.

[I-D.ietf-pce-pcep-extension-pce-controller-sr] specifies the procedures and PCEP extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR-MPLS SID distribution), in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network. This document extends this to include SRv6 SID distribution as well.

2. Terminology

Terminologies used in this document is the same as described in the document [RFC8283].

2.1. 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.

3. PCECC SRv6

[RFC8664] specifies extensions to PCEP that allow a stateful PCE to compute, update, or initiate SR-TE paths for MPLS dataplane. An ingress node of an SR-TE path appends all outgoing packets with a list of MPLS labels (SIDs). This is encoded in SR-ERO subobject, capable of carrying a label (SID) as well as the identity of the node/adjacency label (SID). [I-D.ietf-pce-segment-routing-ipv6] extends the procedure to include support for SRv6 paths.

As per [RFC8754], an SRv6 Segment is a 128-bit value. "SRv6 SID" or simply "SID" are often used as a shorter reference for "SRv6 Segment". Further details are in an illustration provided in [RFC8986]. The SR is applied to IPV6 data plane using SRH. An SR path can be derived from an IGP Shortest Path Tree (SPT), but SR-TE paths may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool, or a PCE and provisioned on the ingress node. [I-D.ietf-pce-segment-routing-ipv6] specify the SRv6-ERO subobject capable of carrying an SRv6 SID as well as the identity of the node/adjacency represented by the SID.

[RFC8283] examines the motivations and applicability for PCECC and use of PCEP as an SBI. Section 3.1.5. of [RFC8283] highlights the use of PCECC for configuring the forwarding actions on the routers and assume responsibility for managing the identifier space. It simplifies the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. This allows the operator to introduce the advantages of SDN (such as programmability) into the network. Further Section 3.3. of [I-D.ietf-teas-pcecc-use-cases] describes some of the scenarios where the PCECC technique could be useful. Section 4 of [RFC8283] also describe the implications on the protocol when used as an SDN SBI. The operator needs to evaluate the advantages offered by PCECC against the operational and scalability needs of the PCECC.

As per [RFC8283], PCECC can allocate and provision the node/prefix/adjacency label (SID) via PCEP. As per [I-D.ietf-teas-pcecc-use-cases] this is also applicable to SRv6 SIDs.

The rest of the processing is similar to existing stateful PCE for SRv6 [I-D.ietf-pce-segment-routing-ipv6].

4. PCEP Requirements

Following key requirements for PCECC-SRv6 should be considered when designing the PCECC-based solution:

- * A PCEP speaker supporting this document needs to have the capability to advertise its PCECC-SRv6 capability to its peers.
- * PCEP procedures need to allow for PCC-based SRv6 SID allocations.
- * PCEP procedures need to provide a means to update (or clean up) the SRv6 SID to the PCC.
- * PCEP procedures need to provide a means to synchronize the SRv6 SID allocations between the PCE to the PCC in the PCEP messages.

5. Procedures for Using the PCE as a Central Controller (PCECC) in SRv6

5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. A PCE as a Central Controller (PCECC) reuses the existing active stateful PCE mechanism as much as possible to control the LSPs.

5.2. New Functions

This document uses the same PCEP messages and its extensions which are described in [RFC9050] and [I-D.ietf-pce-pcep-extension-pce-controller-sr] for PCECC-SRv6 as well.

The PCEP messages PCRpt, PCInitiate, PCUpd are used to send LSP Reports, LSP setup, and LSP update respectively. The extended PCInitiate message described in [RFC9050] is used to download or clean up CCIs (a new CCI Object-Type=TBD3 for SRv6 SID). The extended PCRpt message described in [RFC9050] is also used to report the CCIs (SRv6 SIDs) from PCC to PCE.

[RFC9050] specify an object called CCI for the encoding of the central controller's instructions.

[I-D.ietf-pce-pcep-extension-pce-controller-sr] defined a CCI object-type for SR-MPLS. This document further defines a new CCI object-type=TBD3 for SRv6.

5.3. PCECC Capability Advertisement

During the PCEP initialization phase, PCEP speakers (PCE or PCC) advertise their support of and willingness to use PCEP extensions for the PCECC. A PCEP speaker includes the PCECC-CAPABILITY sub-TLV in the PATH-SETUP-TYPE-CAPABILITY TLV as per [RFC9050].

A new S bit is added in the PCECC-CAPABILITY sub-TLV to indicate support for PCECC-SR-MPLS in [I-D.ietf-pce-pcep-extension-pce-controller-sr]. This document adds another I bit to indicate support for SR in IPv6. A PCC MUST set the I bit in the PCECC-CAPABILITY sub-TLV and include the SRv6-PCE-CAPABILITY sub-TLV ([I-D.ietf-pce-segment-routing-ipv6]) in the OPEN object (inside the PATH-SETUP-TYPE-CAPABILITY TLV) to support the PCECC SRv6 extensions defined in this document.

If the I bit is set in PCECC-CAPABILITY sub-TLV and the SRv6-PCE-CAPABILITY sub-TLV is not advertised, or is advertised without the I bit set, in the OPEN object, the receiver MUST:

- * send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD4 (SRv6 capability was not advertised) and
- * terminate the session.

The rest of the processing is as per [RFC9050] and [I-D.ietf-pce-pcep-extension-pce-controller-sr].

5.4. PCEP session IP address and TED Router ID

As described in [I-D.ietf-pce-pcep-extension-pce-controller-sr], it is important to link the session IP address with the Router ID in TED for successful PCECC-SRv6 operations.

5.5. SRv6 Path Operations

[RFC8664] specify the PCEP extension to allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks. [I-D.ietf-pce-segment-routing-ipv6] extends it to support SRv6.

The Path Setup Type for SRv6 (PST=TBD) is used on the PCEP session with the Ingress as per [I-D.ietf-pce-segment-routing-ipv6].

5.5.1. PCECC Segment Routing in IPv6 (SRv6)

Segment Routing (SR) as described in [RFC8402] depends on "segments" that are advertised by Interior Gateway Protocols (IGPs). The SR-node allocates and advertises the SID (node, adj, etc) and floods them via the IGP. This document proposes a new mechanism where PCE allocates the SRv6 SID centrally and uses PCEP to distribute them to all nodes. In some deployments, PCE (and PCEP) are better suited than IGP because of the centralized nature of PCE and direct TCP based PCEP sessions to the node. Note that only the SRv6 SID allocation and distribution is done by the PCEP, all other SRv6 operations (nexthop selection, protection, etc) are still done by the node (and the IGPs).

5.5.1.1. PCECC SRv6 Node/Prefix SID allocation

Each node (PCC) is allocated a node SRv6 SID by the PCECC. The PCECC sends the PCInitiate message to update the SRv6 SID table of each node. The TE router ID is determined from the TED or from "IPv4/IPv6 Router-ID" sub-TLV [I-D.dhodylee-pce-pcep-ls], in the OPEN Object.

On receiving the SRv6 node SID allocation, each node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly. The PCInitiate message uses the FEC object [I-D.ietf-pce-pcep-extension-pce-controller-sr].

On receiving the SRv6 node SID allocation:

For the local SID, the node (PCC) needs to update SID with associated function (END function in this case) in "My Local SID Table" ([RFC8986]).

For the non-local SID, the node (PCC) uses the local routing information to determine the next-hop and download the forwarding instructions accordingly.

The forwarding behavior and the end result is similar to IGP based "Node-SID" in SRv6. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node as per [RFC8402].

PCE relies on the Node/Prefix SRv6 SID clean up using the same PCInitiate message as per [RFC8281].

5.5.1.2. PCECC SRv6 Adjacency SID allocation

For PCECC-SRv6, apart from node-SID, Adj-SID is used where each adjacency is allocated an Adj-SID by the PCECC. The PCECC sends PCInitiate message to update the SRv6 SID entry for each adjacency to all nodes in the domain. Each node (PCC) download the SRv6 SID instructions accordingly. Similar to SRv6 Node/Prefix Label allocation, the PCInitiate message in this case uses the FEC object.

The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SRv6 as per [RFC8402].

The handling of adjacencies on the LAN subnetworks is specified in [RFC8402]. PCECC MUST assign Adj-SID for every pair of routers in the LAN. The rest of the protocol mechanism remains the same.

PCE relies on the Adj label clean up using the same PCInitiate message as per [RFC8281].

5.5.1.3. Redundant PCEs

[I-D.ietf-pce-state-sync] describes the synchronization mechanism between the stateful PCEs. The SRv6 SIDs allocated by a PCE MUST also be synchronized among PCEs for PCECC-SRv6 state synchronization. Note that the SRv6 SIDs are independent of the SRv6 paths, and remains intact till any topology change. The redundant PCEs MUST have a common view of all SRv6 SIDs allocated in the domain.

5.5.1.4. Re-Delegation and Cleanup

[RFC9050] describes the action needed for CCIs for the static LSPs on a terminated session. Same holds true for the CCI for SRv6 SID as well.

5.5.1.5. Synchronization of SRv6 SID Allocations

[RFC9050] describes the synchronization of CCIs via the LSP state synchronization as described in [RFC8231] and [RFC8232]. Same procedures are applied for the SRv6 SID CCIs.

6. PCEP Messages

The PCEP messages are as per [I-D.ietf-pce-pcep-extension-pce-controller-sr].

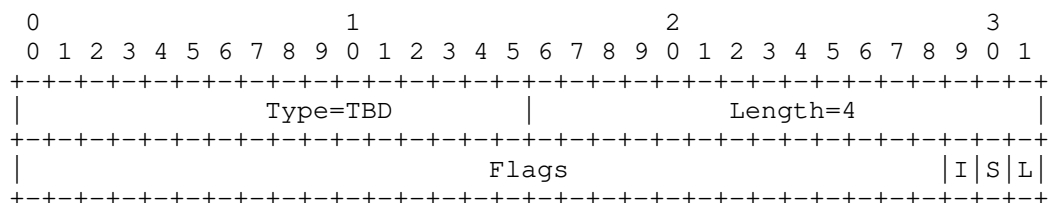
7. PCEP Objects

7.1. OPEN Object

7.1.1. PCECC Capability sub-TLV

[RFC9050] defined the PCECC-CAPABILITY sub-TLV.

A new I-bit is defined in PCECC-CAPABILITY sub-TLV for PCECC-SRv6:



[Editor's Note - The above figure is included for ease of the reader but should be removed before publication.]

I (PCECC-SRv6-CAPABILITY - 1 bit - TBD1): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable of PCECC-SRv6 capability and the PCE allocates the Node and Adj SRv6 SID on this session.

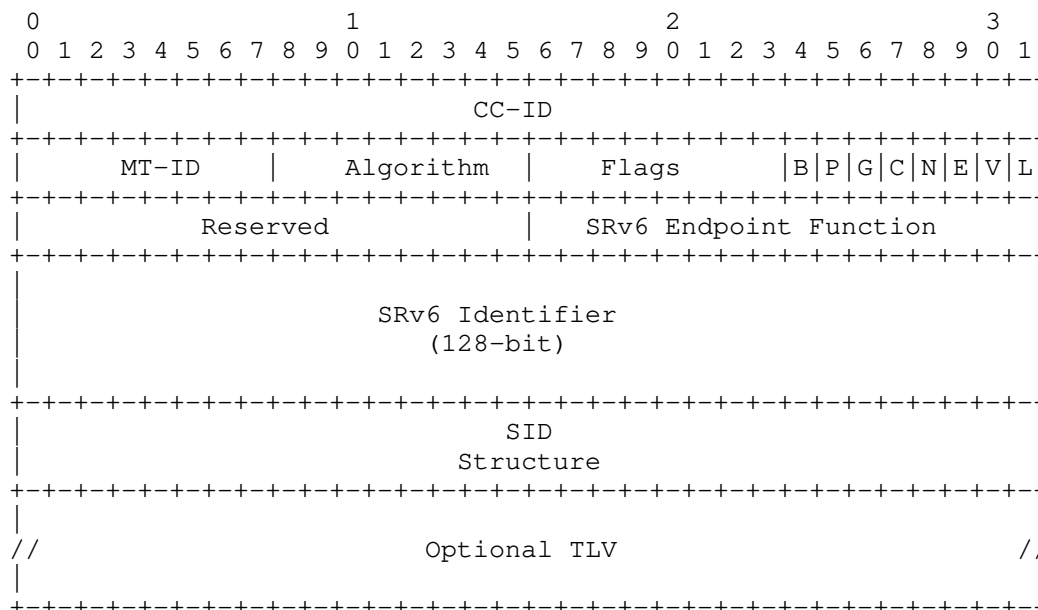
7.2. SRv6 Path Setup

The PATH-SETUP-TYPE TLV is defined in [RFC8408]. A PST value of TBD is used when Path is setup via SRv6 mode as per [I-D.ietf-pce-segment-routing-ipv6]. The procedure for SRv6 path setup as specified in [I-D.ietf-pce-segment-routing-ipv6] remains unchanged.

7.3. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the controller instructions is defined in [RFC9050]. This document defines another object-type for SRv6 purpose.

CCI Object-Type is TBD3 for SRv6 as below -



The field CC-ID is as described in [RFC9050]. The field MT-ID, Algorithm, Flags are defined in [I-D.ietf-pce-pcep-extension-pce-controller-sr].

Reserved: MUST be set to 0 while sending and ignored on receipt.

SRv6 Endpoint Function: 16-bit field representing supported functions associated with SRv6 SIDs.

SRv6 Identifier: 128-bit IPv6 addresses representing SRv6 segment.

SID Structure: 64-bit field formatted as per "SID Structure" in [I-D.ietf-pce-segment-routing-ipv6]. The sum of all four sizes in the SID Structure must be lower or equal to 128 bits. If the sum of all four sizes advertised in the SID Structure is larger than 128 bits, the corresponding SRv6 SID MUST be considered invalid and a PCERR message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Invalid SRv6 SID Structure") is returned.

7.4. FEC Object

The FEC Object is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message.

FEC Object (and various Object-Types) are described in [I-D.ietf-pce-pcep-extension-pce-controller-sr]. SRv6 Node SID MUST include the FEC Object-Type 2 for IPv6 Node. SRv6 Adjacency SID MUST include the FEC Object-Type=4 for IPv6 adjacency. Further FEC object types could be added in future extensions.

8. Security Considerations

As per [RFC8283], the security considerations for a PCE-based controller are a little different from those for any other PCE system. That is, the operation relies heavily on the use and security of PCEP, so consideration should be given to the security features discussed in [RFC5440] and the additional mechanisms described in [RFC8253]. It further lists the vulnerability of a central controller architecture, such as a central point of failure, denial of service, and a focus for interception and modification of messages sent to individual Network Elements (NEs).

The PCECC extension builds on the existing PCEP messages; thus, the security considerations described in [RFC5440], [RFC8231], [RFC8281], [RFC9050], and [I-D.ietf-pce-pcep-extension-pce-controller-sr] continue to apply.

As per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on mutually-authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

9. Manageability Considerations

9.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC SRv6 capability as a global configuration.

9.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC SRv6 capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC SRv6 capability.

9.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

9.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

9.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

9.6. Impact On Network Operations

PCEP implementation SHOULD allow a limit to be placed on the rate of PCInitiate/PCUpd messages (as per [RFC8231]) sent by PCE and processed by PCC. It SHOULD also allow sending a notification when a rate threshold is reached.

10. IANA Considerations

10.1. PCECC-CAPABILITY sub-TLV

[RFC9050] defines the PCECC-CAPABILITY sub-TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY sub-TLV's Flag field. IANA is requested to allocate a new bit in the PCECC-CAPABILITY sub-TLV Flag Field registry, as follows:

Bit	Description	Reference
TBD1	SRv6	This document

Table 1

10.2. PCEP Object

IANA is requested to allocate a new code-point for the new CCI object-type in "PCEP Objects" sub-registry as follows:

Object-Class Value	Name	Object-Type	Reference
TBD	CCI		[RFC9050]
		TBD3: SRv6	This document

Table 2

10.3. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Meaning
19	Invalid operation. Error-value = TBD4 :
	SRv6 capability was not advertised

The Reference is marked as "This document".

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>>.
- [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>>.
- [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>>.
- [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>>.

- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <<https://www.rfc-editor.org/info/rfc8664>>.
- [I-D.ietf-pce-segment-routing-ipv6]
Li(Editor), C., Negi, M. S., Sivabalan, S., Koldychev, M., Kaladharan, P., and Y. Zhu, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", Work in Progress, Internet-Draft, draft-ietf-pce-segment-routing-ipv6-11, 10 January 2022, <<https://datatracker.ietf.org/doc/html/draft-ietf-pce-segment-routing-ipv6-11>>.
- [RFC9050] Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "Path Computation Element Communication Protocol (PCEP) Procedures and Extensions for Using the PCE as a Central Controller (PCECC) of LSPs", RFC 9050, DOI 10.17487/RFC9050, July 2021, <<https://www.rfc-editor.org/info/rfc9050>>.
- [I-D.ietf-pce-pcep-extension-pce-controller-sr]
Li, Z., Peng, S., Negi, M. S., Zhao, Q., and C. Zhou, "Path Computation Element Communication Protocol (PCEP) Procedures and Extensions for Using PCE as a Central Controller (PCECC) for Segment Routing (SR) MPLS Segment Identifier (SID) Allocation and Distribution.", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-extension-pce-controller-sr-04, 6 March 2022, <<https://datatracker.ietf.org/doc/html/draft-ietf-pce-pcep-extension-pce-controller-sr-04>>.

11.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.

- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.
- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.
- [RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <<https://www.rfc-editor.org/info/rfc8232>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.

- [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>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.
- [RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <<https://www.rfc-editor.org/info/rfc8665>>.
- [RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <<https://www.rfc-editor.org/info/rfc8667>>.
- [RFC8754] Filsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/info/rfc8754>>.
- [RFC8986] Filsfils, C., Ed., Camarillo, P., Ed., Leddy, J., Voyer, D., Matsushima, S., and Z. Li, "Segment Routing over IPv6 (SRv6) Network Programming", RFC 8986, DOI 10.17487/RFC8986, February 2021, <<https://www.rfc-editor.org/info/rfc8986>>.
- [I-D.ietf-teas-pcecc-use-cases]
Li, Z. (., Dhody, D., Zhao, Q., Ke, K., Khasanov, B., Fang, L., Zhou, C., Zhang, B., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", Work in Progress, Internet-Draft, draft-ietf-teas-pcecc-use-cases-08, 25 October 2021, <<https://datatracker.ietf.org/doc/html/draft-ietf-teas-pcecc-use-cases-08>>.

[I-D.ietf-pce-pcep-yang]

Dhody, D., Hardwick, J., Beeram, V. P., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-yang-18, 25 January 2022, <<https://datatracker.ietf.org/doc/html/draft-ietf-pce-pcep-yang-18>>.

[I-D.ietf-pce-state-sync]

Litkowski, S., Sivabalan, S., Li, C., and H. Zheng, "Inter Stateful Path Computation Element (PCE) Communication Procedures.", Work in Progress, Internet-Draft, draft-ietf-pce-state-sync-01, 20 October 2021, <<https://datatracker.ietf.org/doc/html/draft-ietf-pce-state-sync-01>>.

[I-D.dhodylee-pce-pcep-ls]

Dhody, D., Peng, S., Lee, Y., Ceccarelli, D., Wang, A., Mishra, G., and S. Sivabalan, "PCEP extensions for Distribution of Link-State and TE Information", Work in Progress, Internet-Draft, draft-dhodylee-pce-pcep-ls-23, 5 March 2022, <<https://datatracker.ietf.org/doc/html/draft-dhodylee-pce-pcep-ls-23>>.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Authors' Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China
Email: pengshuping@huawei.com

Xuesong Geng
Huawei Technologies
China
Email: gengxuesong@huawei.com

Mahendra Singh Negi
RtBrick Inc
N-17L, 18th Cross Rd, HSR Layout
Bangalore 560102
Karnataka
India
Email: mahend.ietf@gmail.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 6, 2020

Q. Zhao
Z. Li
M. Negi
Huawei Technologies
C. Zhou
Cisco Systems
November 3, 2019

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) of LSPs
draft-ietf-pce-pcep-extension-for-pce-controller-03

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

PCE was developed to derive paths for MPLS Label Switched Paths (LSPs), which are supplied to the head end of the LSP using the Path Computation Element Communication Protocol (PCEP). But SDN has a broader applicability than signaled (G)MPLS traffic-engineered (TE) networks, and the PCE may be used to determine paths in a range of use cases. PCEP has been proposed as a control protocol for use in these environments to allow the PCE to be fully enabled as a central controller.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the path while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP protocol extensions for using the PCE as the central controller.

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 May 6, 2020.

Copyright Notice

Copyright (c) 2019 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	3
1.1. Requirements Language	5
2. Terminology	5
3. Basic PCECC Mode	5
4. PCEP Requirements	5
5. Procedures for Using the PCE as the Central Controller (PCECC)	6
5.1. Stateful PCE Model	6
5.2. New LSP Functions	6
5.3. PCECC Capability Advertisement	7
5.4. LSP Operations	8
5.4.1. Basic PCECC LSP Setup	8
5.4.2. Central Control Instructions	12
5.4.2.1. Label Download CCI	12
5.4.2.2. Label Cleanup CCI	12
5.4.3. PCE Initiated PCECC LSP	13
5.4.4. PCECC LSP Update	15
5.4.5. Re Delegation and Cleanup	17
5.4.6. Synchronization of Central Controllers Instructions	17
5.4.7. PCECC LSP State Report	17
5.4.8. PCC Based Allocations	18

5.4.9. Binding Label	18
6. PCEP messages	19
6.1. The PCInitiate message	19
6.2. The PCRpt message	21
7. PCEP Objects	21
7.1. OPEN Object	22
7.1.1. PCECC Capability sub-TLV	22
7.2. PATH-SETUP-TYPE TLV	22
7.3. CCI Object	23
7.3.1. Address TLVs	24
8. Implementation Status	25
8.1. Huawei's Proof of Concept based on ONOS	26
9. Security Considerations	26
9.1. Malicious PCE	26
10. Manageability Considerations	27
10.1. Control of Function and Policy	27
10.2. Information and Data Models	27
10.3. Liveness Detection and Monitoring	27
10.4. Verify Correct Operations	27
10.5. Requirements On Other Protocols	27
10.6. Impact On Network Operations	27
11. IANA Considerations	27
11.1. PCEP TLV Type Indicators	27
11.2. PATH-SETUP-TYPE-CAPABILITY Sub-TLV Type Indicators	28
11.3. New Path Setup Type Registry	28
11.4. PCEP Object	28
11.5. CCI Object Flag Field	28
11.6. PCEP-Error Object	29
12. Acknowledgments	29
13. References	30
13.1. Normative References	30
13.2. Informative References	31
Appendix A. Contributor Addresses	34
Authors' Addresses	35

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave

in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCECC architecture.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the path while leveraging the existing PCE technologies as much as possible.

This draft specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label-forwarding instructions to program and what resources to reserve. The PCE-based controller keeps a view of the network and determines the paths of the end-to-end LSPs, and the controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

The extension for PCECC in Segment Routing (SR) is specified in a separate draft [I-D.zhao-pce-pcep-extension-pce-controller-sr].

1.1. 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.

2. Terminology

Terminologies used in this document is same as described in the draft [RFC8283].

3. Basic PCECC Mode

In this mode LSPs are provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label forwarding instructions to program and what resources to reserve. The controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

Note that the PCE-based controller will take responsibility for managing some part of the MPLS label space for each of the routers that it controls, and may take wider responsibility for partitioning the label space for each router and allocating different parts for different uses. This is also described in section 3.1.2. of [RFC8283]. For the purpose of this document, it is assumed that label range to be used by a PCE is known and set on both PCEP peers. A future extension could add this capability to advertise the range via possible PCEP extensions as well (see [I-D.li-pce-controlled-id-space]). The rest of processing is similar to the existing stateful PCE mechanism.

This document also allow a case where the label space is maintained by PCC itself, and the labels are allocated by the PCC, in this case, the PCE should request the allocation from PCC as described in Section 5.4.8.

4. PCEP Requirements

Following key requirements associated PCECC should be considered when designing the PCECC based solution:

1. PCEP speaker supporting this draft needs to have the capability to advertise its PCECC capability to its peers.
2. PCEP speaker needs a means to identify PCECC based LSP in the PCEP messages.

3. PCEP procedures needs to allow for PCC based label allocations.
 4. PCEP procedures needs to provide a means to update (or cleanup) the label- download entry to the PCC.
 5. PCEP procedures needs to provide a means to synchronize the labels between PCE to PCC in PCEP messages.
5. Procedures for Using the PCE as the Central Controller (PCECC)

5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. PCE as a central controller (PCECC) reuses existing Active stateful PCE mechanism as much as possible to control the LSP.

5.2. New LSP Functions

This document defines the following new PCEP messages and extends the existing messages to support PCECC:

(PCInitiate): a PCEP message described in [RFC8281]. PCInitiate message is used to setup PCE-Initiated LSP based on PCECC mechanism. It is also extended for Central Controller's Instructions (CCI) (download or cleanup the Label forwarding instructions in the context of this document) on all nodes along the path.

(PCRpt): a PCEP message described in [RFC8231]. PCRpt message is used to send PCECC LSP Reports. It is also extended to report the set of Central Controller's Instructions (CCI) (label forwarding instructions in the context of this document) received from the PCE. See Section 5.4.6 for more details.

(PCUpd): a PCEP message described in [RFC8231]. PCUpd message is used to send PCECC LSP Update.

The new LSP functions defined in this document are mapped onto the messages as shown in the following table.

Function	Message
PCECC Capability advertisement	Open
Label entry Add	PCInitiate
Label entry Cleanup	PCInitiate
PCECC Initiated LSP	PCInitiate
PCECC LSP Update	PCUpd
PCECC LSP State Report	PCRpt
PCECC LSP Delegation	PCRpt
PCECC Label Report	PCRpt

This document specify a new PCEP object called CCI (see Section 7.3) for the encoding of central controller's instructions. In the scope of this document this is limited to Label forwarding instructions. Future documents can create new CCI object-type for other types of central control instructions. The CC-ID is the unique identifier for the central controller's instructions in PCEP. The PCEP messages are extended in this document to handle the PCECC operations.

5.3. PCECC Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions.

This document defines a new Path Setup Type (PST) [RFC8408] for PCECC, as follows:

- o PST = TBD1: Path is setup via PCECC mode.

A PCEP speaker MUST indicate its support of the function described in this document by sending a PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object with this new PST included in the PST list.

This document also defines the PCECC Capability sub-TLV
Section 7.1.1. PCEP speakers use this sub-TLV to exchange information about their PCECC capability. If a PCEP speaker includes PST=TBD1 in the PST List of the PATH-SETUP-TYPE-CAPABILITY TLV then it MUST also include the PCECC Capability sub-TLV inside the PATH-SETUP-TYPE-CAPABILITY TLV. If the sub-TLV is absent, then the PCEP speaker MUST send a PCErr message with Error-Type 10 (Reception of an invalid object) and Error-Value TBD2 (Missing PCECC Capability sub-TLV) and MUST then close the PCEP session. If a PCEP speaker receives a PATH-SETUP-TYPE-CAPABILITY TLV with a SR-PCE-CAPABILITY sub-TLV, but the PST list does not contain PST=1, then the PCEP speaker MUST ignore the SR-PCE-CAPABILITY sub- TLV.

The presence of the PST and PCECC Capability sub-TLV in PCC's OPEN Object indicates that the PCC is willing to function as a PCECC client. The presence of the PST and PCECC Capability sub-TLV in PCE's OPEN message indicates that the PCE is interested in function as a PCECC server.

The PCEP protocol extensions for PCECC MUST NOT be used if one or both PCEP Speakers have not included the PST or the PCECC Capability sub-TLV in their respective OPEN message. If the PCEP Speakers support the extensions of this draft but did not advertise this capability then a PCErr message with Error-Type=19(Invalid Operation) and Error-Value=TBD3 (Attempted PCECC operations when PCECC capability was not advertised) will be generated and the PCEP session will be terminated.

A PCC or a PCE MUST include both PCECC-CAPABILITY sub-TLV and STATEFUL-PCE-CAPABILITY TLV ([RFC8231]) (with I flag set [RFC8281]) in OPEN Object to support the extensions defined in this document. If PCECC-CAPABILITY sub-TLV is advertised and STATEFUL-PCE-CAPABILITY TLV is not advertised in OPEN Object, it MUST send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD4 (stateful PCE capability was not advertised) and terminate the session. This error is also triggered if PCECC-CAPABILITY sub-TLV is advertised and I flag is not set.

5.4. LSP Operations

The PCEP messages pertaining to PCECC MUST include PATH-SETUP-TYPE TLV [RFC8408] in the SRP object to clearly identify the PCECC LSP is intended.

5.4.1. Basic PCECC LSP Setup

In order to setup a LSP based on PCECC mechanism, a PCC MUST delegate the LSP by sending a PCRpt message with PST set for PCECC (see Section 7.2) and D (Delegate) flag (see [RFC8231]) set in the LSP object.

LSP-IDENTIFIER TLV MUST be included for PCECC LSP, the tuple uniquely identifies the LSP in the network. The LSP object is included in central controller's instructions (label download) to identify the PCECC LSP for this instruction. The PLSP-ID is the original identifier used by the ingress PCC, so the transit LSR could have multiple central controller instructions that have the same PLSP-ID. The PLSP-ID in combination with the source (in LSP-IDENTIFIER TLV) MUST be unique. The PLSP-ID is included for maintainability reasons to ease debugging. As per [RFC8281], the LSP object could include SPEAKER-ENTITY-ID TLV to identify the PCE that initiated these

instructions. Also the CC-ID is unique on the PCEP session as described in Section 7.3.

When a PCE receives PCRpt message with D flags and PST Type set, it calculates the path and assigns labels along the path; and set up the path by sending PCInitiate message to each node along the path of the LSP. The PCC generates a Path Computation State Report (PCRpt) and include the central controller's instruction (CCI) and the identified LSP. The CC-ID is uniquely identify the central controller's instruction within a PCEP session. The PCC further responds with the PCRpt messages including the CCI and LSP objects.

The Ingress node would receive one CCI object with O bit (out-label) set. The transit node(s) would receive two CCI object with the in-label CCI without O bit set and the out-label CCI with O bit set. The egress node would receive one CCI object without O bit set. A node can determine its role based on the setting of the O bit in the CCI object(s).

Once the central controller's instructions (label operations) are completed, the PCE MUST send the PCUpd message to the Ingress PCC. This PCUpd message is as per [RFC8231] SHOULD include the path information as calculated by the PCE.

Note that the PCECC LSPs MUST be delegated to a PCE at all times.

LSP deletion operation for PCECC LSP is same as defined in [RFC8231]. If the PCE receives PCRpt message for LSP deletion then it does Label cleanup operation as described in Section 5.4.2.2 for the corresponding LSP.

The Basic PCECC LSP setup sequence is as shown below.

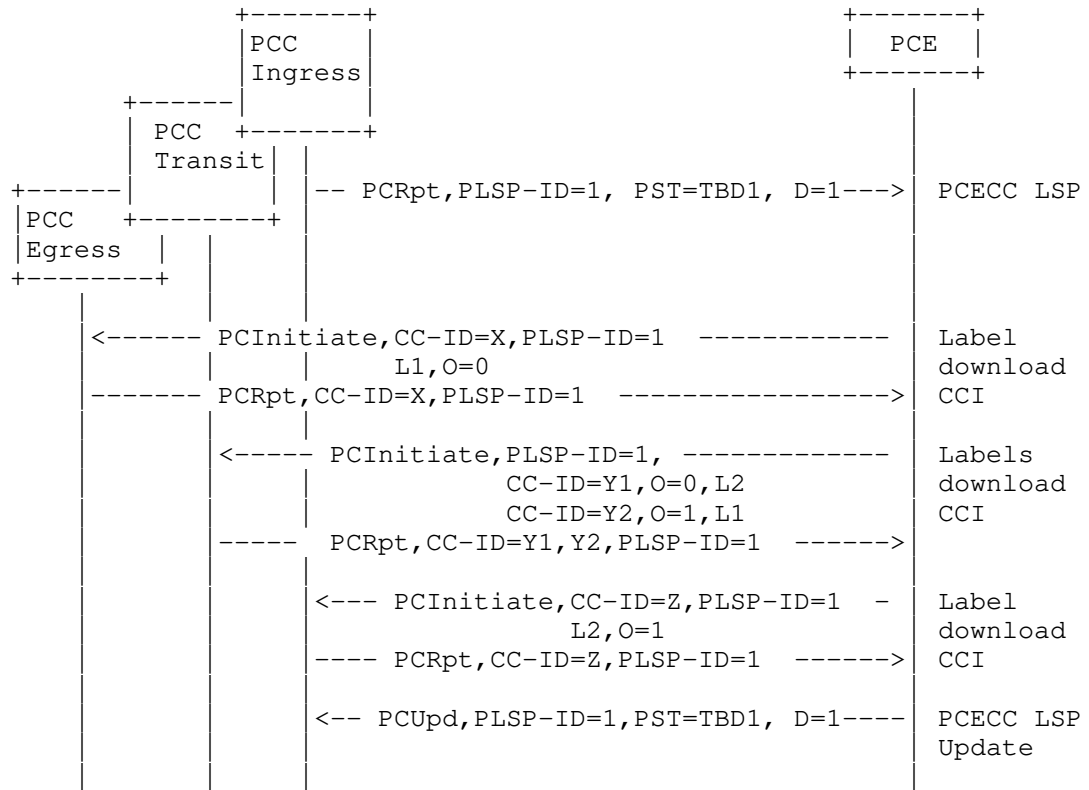
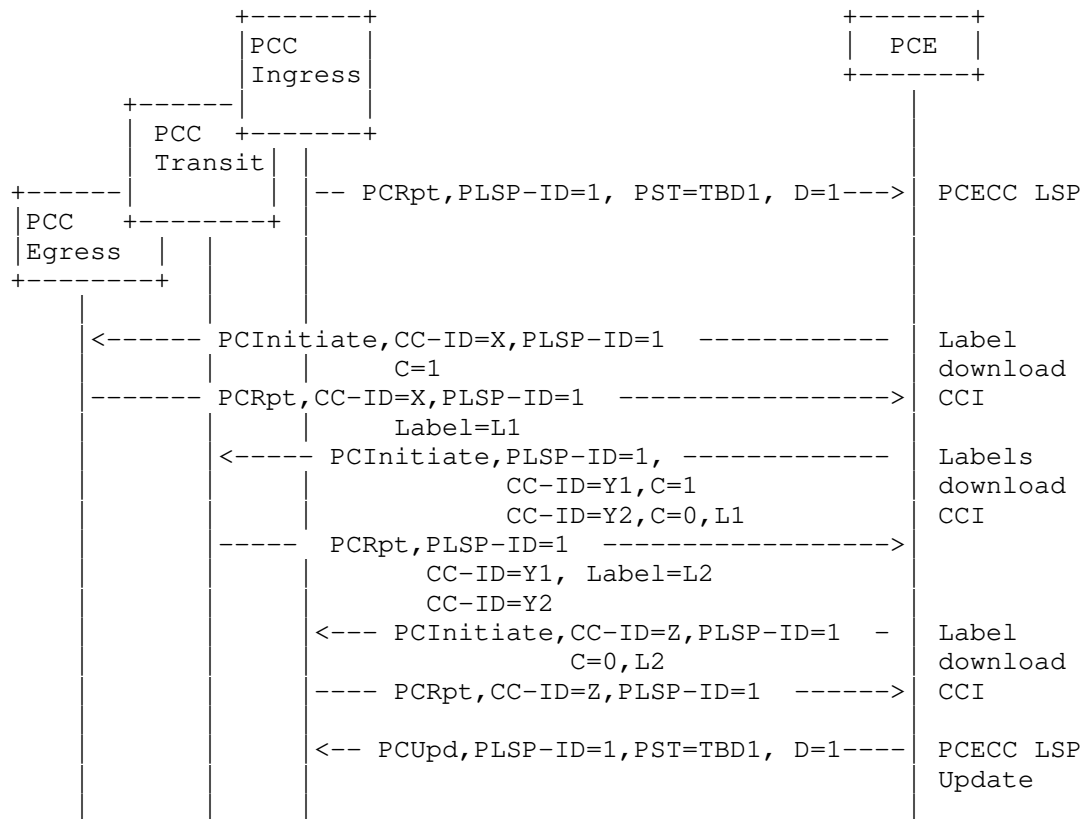


Figure 2: Basic PCECC LSP setup

The PCECC LSP are considered to be 'up' by default (on receipt of PCUpd message from PCE). The Ingress MAY further choose to deploy a data plane check mechanism and report the status back to the PCE via PCRpt message.

In case where the label allocation are made by the PCC itself (see Section 5.4.8), the PCE could request an allocation to be made by the PCC, and where the PCC would send a PCRpt with the allocated label encoded in the CC-ID object as shown below -



- The 0 bit is set as before (and thus not included)

Figure 3: Basic PCECC LSP setup (PCC allocation)

It should be noted that in this example, the request is made to the egress node with C bit set in the CCI object to indicate that the label allocation needs to be done by the egress and it responds with the allocated label to the PCE. The PCE would further inform the transit PCC without setting the C bit in the CCI object for out-label but the C-bit is unset for in-label so the transit node make the label allocation (for the in-label) and report to the PCE. Similarly C bit is unset towards the ingress to complete all the label allocation for the PCECC LSP.

5.4.2. Central Control Instructions

The new central controller's instructions (CCI) for the label operations in PCEP is done via the PCInitiate message, by defining a new PCEP Objects for CCI operations. Local label range of each PCC is assumed to be known at both the PCC and the PCE.

5.4.2.1. Label Download CCI

In order to setup an LSP based on PCECC, the PCE sends a PCInitiate message to each node along the path to download the Label instruction as described in Section 5.4.1.

The CCI object MUST be included, along with the LSP object in the PCInitiate message. The LSP-IDENTIFIER TLV MUST be included in LSP object. The SPEAKER-ENTITY-ID TLV SHOULD be included in LSP object.

If a node (PCC) receives a PCInitiate message which includes a Label to download as part of CCI, that is out of the range set aside for the PCE, it MUST send a PCErr message with Error-type=TBD5 (PCECC failure) and Error-value=TBD6 (Label out of range) and MUST include the SRP object to specify the error is for the corresponding label update via PCInitiate message. If a PCC receives a PCInitiate message but failed to download the Label entry, it MUST send a PCErr message with Error-type=TBD5 (PCECC failure) and Error-value=TBD7 (instruction failed) and MUST include the SRP object to specify the error is for the corresponding label update via PCInitiate message.

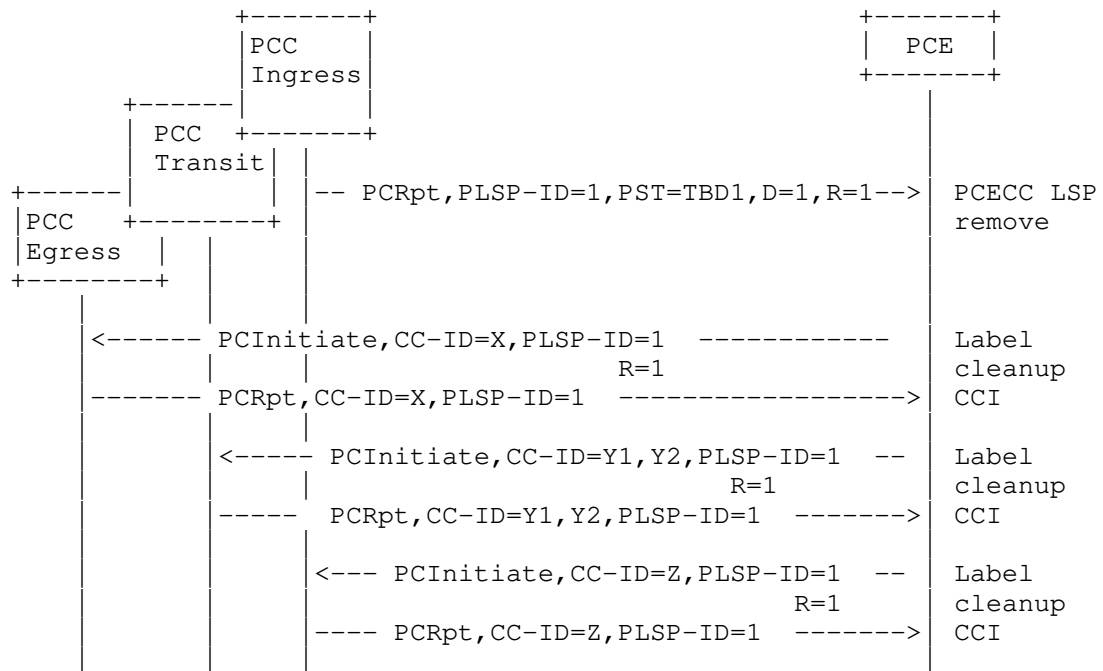
New PCEP object for central control instructions (CCI) is defined in Section 7.3.

5.4.2.2. Label Cleanup CCI

In order to delete an LSP based on PCECC, the PCE sends a central controller instructions via a PCInitiate message to each node along the path of the LSP to cleanup the Label forwarding instruction.

If the PCC receives a PCInitiate message but does not recognize the label in the CCI, the PCC MUST generate a PCErr message with Error-Type 19(Invalid operation) and Error-Value=TBD8, "Unknown Label" and MUST include the SRP object to specify the error is for the corresponding label cleanup (via PCInitiate message).

The R flag in the SRP object defined in [RFC8281] specifies the deletion of Label Entry in the PCInitiate message.



As per [RFC8281], following the removal of the Label forwarding instruction, the PCC MUST send a PCRpt message. The SRP object in the PCRpt MUST include the SRP-ID-number from the PCInitiate message that triggered the removal. The R flag in the SRP object MUST be set.

In case where the label allocation are made by the PCC itself (see Section 5.4.8), the removal procedure remains the same.

5.4.3. PCE Initiated PCECC LSP

The LSP Instantiation operation is same as defined in [RFC8281].

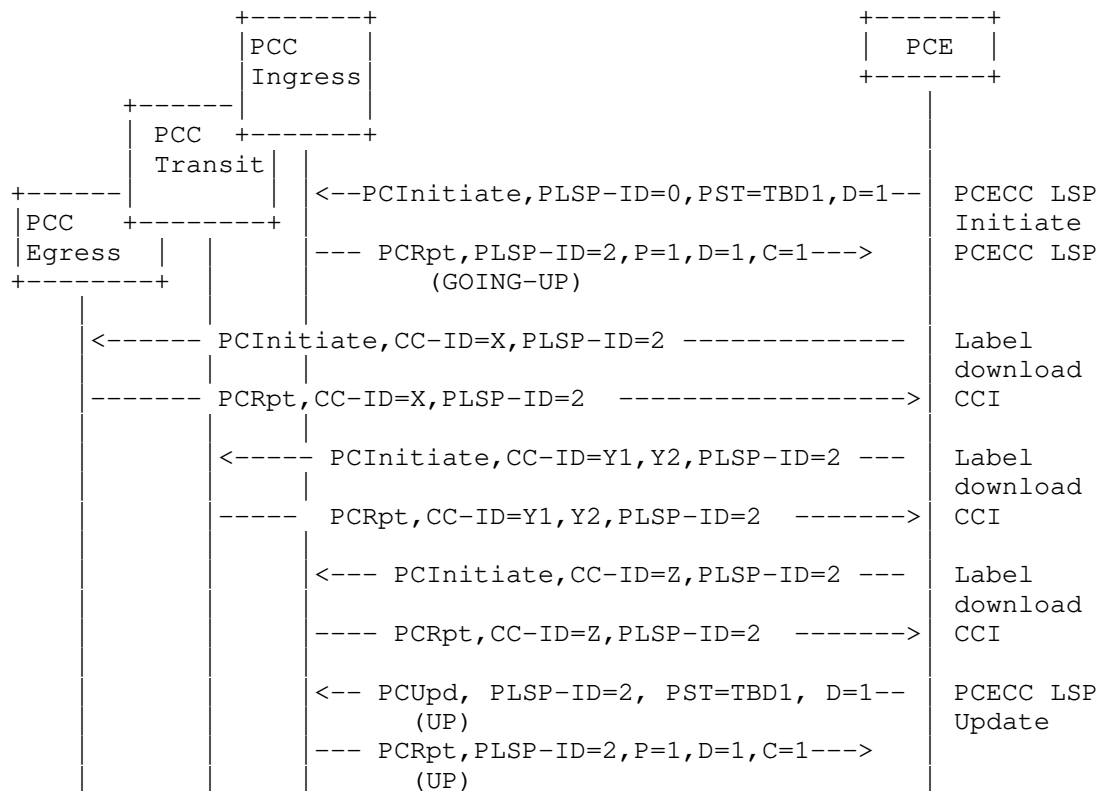
In order to setup a PCE Initiated LSP based on the PCECC mechanism, a PCE sends PCInitiate message with Path Setup Type set for PCECC (see Section 7.2) to the Ingress PCC.

The Ingress PCC MUST also set D (Delegate) flag (see [RFC8231]) and C (Create) flag (see [RFC8281]) in LSP object of PCRpt message. The PCC responds with first PCRpt message with the status as "GOING-UP" and assigned PLSP-ID.

Note that the label forwarding instructions from PCECC are send after the initial PCInitiate and PCRpt exchange. This is done so that the PLSP-ID and other LSP identifiers can be obtained from the ingress and can be included in the label forwarding instruction in the next PCInitiate message. The rest of the PCECC LSP setup operations are same as those described in Section 5.4.1.

The LSP deletion operation for PCE Initiated PCECC LSP is same as defined in [RFC8281]. The PCE should further perform Label entry cleanup operation as described in Section 5.4.2.2 for the corresponding LSP.

The PCE Initiated PCECC LSP setup sequence is shown below -



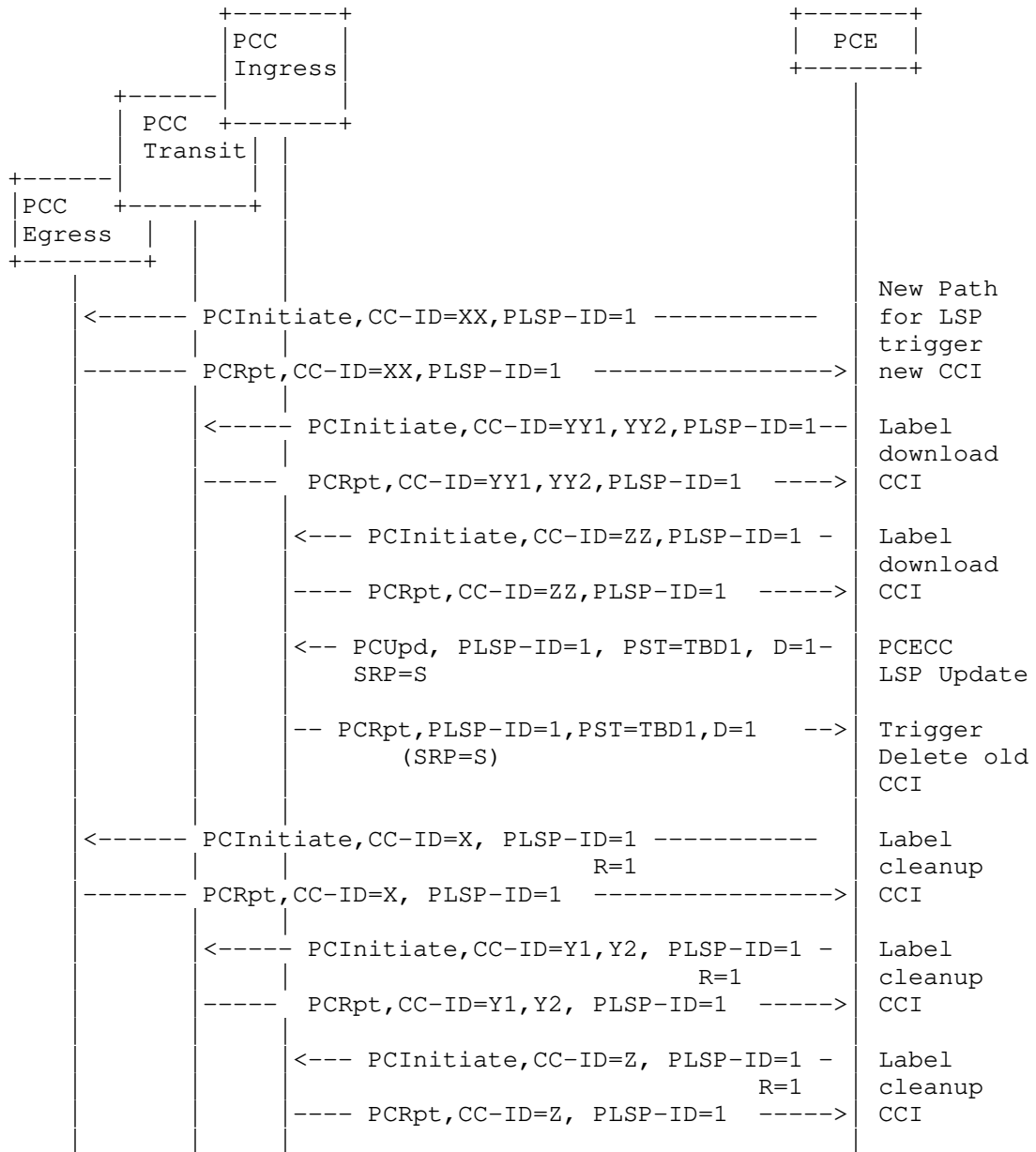
Once the label operations are completed, the PCE SHOULD send the PCUpd message to the Ingress PCC. The PCUpd message is as per [RFC8231].

In case where the label allocation are made by the PCC itself (see Section 5.4.8), the procedure remains similar.

5.4.4. PCECC LSP Update

In case of a modification of PCECC LSP with a new path, a PCE sends a PCUpd message to the Ingress PCC. But to follow the make-before-break procedures, the PCECC first update new instructions based on the updated LSP and then update to ingress to switch traffic, before cleaning up the old instructions. A new CC-ID is used to identify the updated instruction, the existing identifiers in the LSP object identify the existing LSP. Once new instructions are downloaded, the PCE further updates the new path at the ingress which triggers the traffic switch on the updated path. The Ingress PCC acknowledges with a PCRpt message, on receipt of PCRpt message, the PCE does cleanup operation for the old LSP as described in Section 5.4.2.2.

The PCECC LSP Update sequence is shown below -



The modified PCECC LSP are considered to be 'up' by default. The Ingress MAY further choose to deploy a data plane check mechanism and report the status back to the PCE via PCRpt message.

In case where the label allocation are made by the PCC itself (see Section 5.4.8), the procedure remains similar.

5.4.5. Re Delegation and Cleanup

As described in [RFC8281], a new PCE can gain control over the orphaned LSP. In case of PCECC LSP, the new PCE MUST also gain control over the central controllers instructions in the same way by sending a PCInitiate message that includes the SRP, LSP and CCI objects and carries the CC-ID and PLSP-ID identifying the instruction, it wants to take control of.

Further, as described in [RFC8281], the State Timeout Interval timer ensures that a PCE crash does not result in automatic and immediate disruption for the services using PCE-initiated LSPs. Similarly the central controller instructions are not removed immediately upon PCE failure. Instead, they are cleaned up on the expiration of this timer. This allows for network cleanup without manual intervention. The PCC MUST support removal of CCI as one of the behaviors applied on expiration of the State Timeout Interval timer.

5.4.6. Synchronization of Central Controllers Instructions

The purpose of Central Controllers Instructions synchronization (labels in the context of this document) is to make sure that the PCE's view of CCI (Labels) matches with the PCC's Label allocation. This synchronization is performed as part of the LSP state synchronization as described in [RFC8231] and [RFC8233].

As per LSP State Synchronization [RFC8231], a PCC reports the state of its LSPs to the PCE using PCRpt messages and as per [RFC8281], PCE would initiate any missing LSPs and/or remove any LSPs that are not wanted. The same PCEP messages and procedure is also used for the Central Controllers Instructions synchronization. The PCRpt message includes the CCI and the LSP object to report the label forwarding instructions. The PCE would further remove any unwanted instructions or initiate any missing instructions.

5.4.7. PCECC LSP State Report

As mentioned before, an Ingress PCC MAY choose to apply any OAM mechanism to check the status of LSP in the Data plane and MAY further send its status in PCRpt message to the PCE.

5.4.8. PCC Based Allocations

The PCE can request the PCC to allocate the label using the PCInitiate message. The C flag in the CCI object is set to 1 to indicate that the allocation needs to be done by the PCC. The PCC would allocate the Label and would report to the PCE using the PCRpt message.

If the value of the Label is 0 and the C flag is set, it indicates that the PCE is requesting the allocation to be done by the PCC. If the Label is 'n' and the C flag is set in the CCI object, it indicates that the PCE requests a specific value 'n' for the Label. If the allocation is successful, the PCC should report via PCRpt message with the CCI object. Else, it MUST send a PCErr message with Error-Type = TBD5 ("PCECC failure") and Error Value = TBD9 ("Invalid CCI"). If the value of the the Label in the CCI object is valid, but the PCC is unable to allocate it, it MUST send a PCErr message with Error-Type = TBD5 ("PCECC failure") and Error Value = TBD10 ("Unable to allocate the specified CCI").

If the PCC wishes to withdrawn or modify the previously assigned label, it MUST send a PCRpt message without any Label or with the Label containing the new value respectively in the CCI object. The PCE would further trigger the removal of the central controller instruction as per this document.

5.4.9. Binding Label

As per [I-D.sivabalan-pce-binding-label-sid], when a stateful PCE is deployed for setting up TE paths, it may be desirable to report the binding label to the stateful PCE for the purpose of enforcing end-to-end TE. In case of PCECC, the binding label may be allocated by the PCE itself as described in this section. This procedure is thus applicable for all path setup types including PCECC.

A P flag in LSP object is introduced in [I-D.li-pce-sr-path-segment] to indicate the allocation needs to be made by the PCE. This flag is used to indicate that the allocation needs to be made by the PCE. A PCC would set this bit to 1 (and carry the TE-PATH-BINDING TLV [I-D.sivabalan-pce-binding-label-sid] in LSP object) to request for allocation of the binding label by the PCE in the PCReq or PCRpt message. A PCE would also set this bit to 1 to indicate that the binding label is allocated by PCE and encoded in the PCRep, PCUpd or PCInitiate message (the TE-PATH-BINDING TLV is present in LSP object). Further, a PCE would set this bit to 0 to indicate that the allocation is done by the PCC instead.

The ingress PCC could request the binding label to be allocated by the PCE via PCRpt message as per [RFC8231]. The delegate flag (D-flag) MUST also be set for this LSP. The TE-PATH-BINDING TLV MUST be included with no Binding Value. The PCECC would allocate the binding label and further respond to Ingress PCC with PCUpd message as per [RFC8231] and MUST include the TE-PATH-BINDING TLV in a LSP object. The P flag in the LSP object would be set to 1 to indicate that the allocation is made by the PCE.

The PCE could allocate the binding label on its own accord for a PCE-Initiated (or delegated LSP). The allocated binding label needs to be informed to the PCC. The PCE would use the PCInitiate message [RFC8281] or PCUpd message [RFC8231] towards the PCC and MUST include the TE-PATH-BINDING TLV in the LSP object. The P flag in the LSP object would be set to 1 to indicate that the allocation is made by the PCE.

The PCECC capability MUST be exchanged on the PCEP session, before PCE could allocate binding label. Note that the CCI object is not used for binding allocation; this is done to maintain consistency with the rest of the binding label/SID procedures as per [I-D.sivabalan-pce-binding-label-sid].

6. PCEP messages

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable-length body made of a set of objects that can be either mandatory or optional. An object is said to be mandatory in a PCEP message when the object must be included for the message to be considered valid. For each PCEP message type, a set of rules is defined that specify the set of objects that the message can carry. An implementation MUST form the PCEP messages using the object ordering specified in this document.

LSP-IDENTIFIERS TLV MUST be included in the LSP object for PCECC LSP.

6.1. The PCInitiate message

The PCInitiate message [RFC8281] can be used to download or remove the labels, the message has been extended as shown below -

```

<PCInitiate Message> ::= <Common Header>
                           <PCE-initiated-lsp-list>

```

Where:

<Common Header> is defined in [RFC5440]

```

<PCE-initiated-lsp-list> ::= <PCE-initiated-lsp-request>
                             [<PCE-initiated-lsp-list>]

```

```

<PCE-initiated-lsp-request> ::=
    (<PCE-initiated-lsp-instantiation> |
     <PCE-initiated-lsp-deletion> |
     <PCE-initiated-lsp-central-control>)

```

```

<PCE-initiated-lsp-central-control> ::= <SRP>
                                         <LSP>
                                         <cci-list>

```

```

<cci-list> ::= <CCI>
               [<cci-list>]

```

Where:

<PCE-initiated-lsp-instantiation> and
 <PCE-initiated-lsp-deletion> are as per
 [RFC8281].

The LSP and SRP object is defined in [RFC8231].

When PCInitiate message is used for central controller's instructions (labels), the SRP, LSP and CCI objects MUST be present. The SRP object is defined in [RFC8231] and if the SRP object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=10 (SRP object missing). The LSP object is defined in [RFC8231] and if the LSP object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=8 (LSP object missing). The CCI object is defined in Section 7.3 and if the CCI object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD11 (CCI object missing). More than one CCI object MAY be included in the PCInitiate message for the transit LSR.

To cleanup the SRP object must set the R (remove) bit.

At max two instances of CCI object would be included in case of transit LSR to encode both in-coming and out-going label forwarding instructions. Other instances MUST be ignored.

6.2. The PCRpt message

The PCRpt message can be used to report the labels that were allocated by the PCE, to be used during the state synchronization phase.

```
<PCRpt Message> ::= <Common Header>
                        <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report>[<state-report-list>]
```

```
<state-report> ::= (<lsp-state-report>|
                    <central-control-report>)
```

```
<lsp-state-report> ::= [<SRP>]
                        <LSP>
                        <path>
```

```
<central-control-report> ::= [<SRP>]
                              <LSP>
                              <cci-list>
```

```
<cci-list> ::= <CCI>
               [<cci-list>]
```

Where:

<path> is as per [RFC8231] and the LSP and SRP object are also defined in [RFC8231].

When PCRpt message is used to report the central controller's instructions (labels), the LSP and CCI objects MUST be present. The LSP object is defined in [RFC8231] and if the LSP object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=8 (LSP object missing). The CCI object is defined in Section 7.3 and if the CCI object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD11 (CCI object missing). Two CCI object can be included in the PCRpt message for the transit LSR.

7. PCEP Objects

The PCEP objects defined in this document are compliant with the PCEP object format defined in [RFC5440].

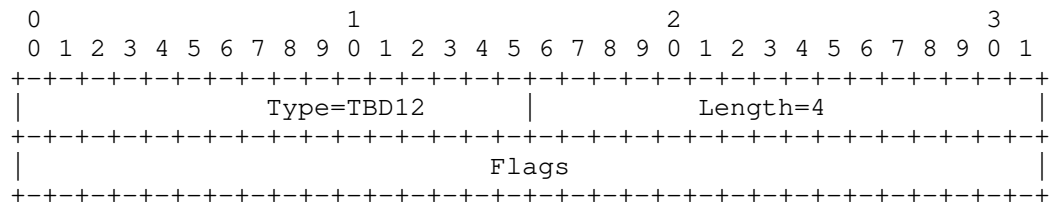
7.1. OPEN Object

This document defines a new optional TLVs for use in the OPEN Object.

7.1.1. PCECC Capability sub-TLV

The PCECC-CAPABILITY sub-TLV is an optional TLV for use in the OPEN Object for PCECC capability advertisement in PATH-SETUP-TYPE-CAPABILITY TLV. Advertisement of the PCECC capability implies support of LSPs that are setup through PCECC as per PCEP extensions defined in this document.

Its format is shown in the following figure:



The type of the TLV is TBD12 and it has a fixed length of 4 octets.

The value comprises a single field - Flags (32 bits).

No flags are assigned right now.

Unassigned bits are considered reserved. They MUST be set to 0 on transmission and MUST be ignored on receipt.

7.2. PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]; this document defines a new PST value:

- o PST = TBD1: Path is setup via PCECC mode.

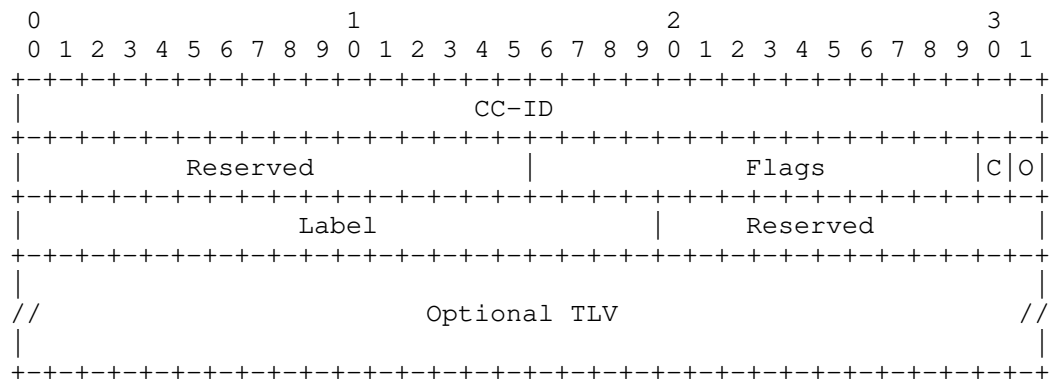
On a PCRpt/PCUpd/PCInitiate message, the PST=TBD1 in PATH-SETUP-TYPE TLV in SRP object indicates that this LSP was setup via a PCECC based mechanism.

7.3. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the forwarding instructions (Label information in the context of this document) to the PCC, and MAY be carried within PCInitiate or PCRpt message for label download.

CCI Object-Class is TBD13.

CCI Object-Type is 1 for the MPLS Label.



The fields in the CCI object are as follows:

CC-ID: A PCEP-specific identifier for the CCI information. A PCE creates an CC-ID for each instruction, the value is unique within the scope of the PCE and is constant for the lifetime of a PCEP session. The values 0 and 0xFFFFFFFF are reserved and MUST NOT be used.

Flags: is used to carry any additional information pertaining to the CCI. Currently, the following flag bit is defined:

- * O bit(Out-label) : If the bit is set, it specifies the label is the OUT label and it is mandatory to encode the next-hop information (via IPV4-ADDRESS TLV or IPV6-ADDRESS TLV or UNNUMBERED-IPV4-ID-ADDRESS TLV in the CCI object). If the bit is not set, it specifies the label is the IN label and it is optional to encode the local interface information (via IPV4-ADDRESS TLV or IPV6-ADDRESS TLV or UNNUMBERED-IPV4-ID-ADDRESS TLV in the CCI object).

- * C-Bit (PCC Allocation): If the bit is set to 1, it indicates that the allocation needs to be done by the PCC for this central controller instruction. A PCE set this bit to request the PCC to make an allocation from its label space. A PCC would set this bit to indicate that it has allocated the CC-ID and report it to the PCE.

Label (20-bit): The Label information.

Reserved (12 bit): Set to zero while sending, ignored on receive.

7.3.1. Address TLVs

This document defines the following TLVs for the CCI object to associate the next-hop information in case of an outgoing label and local interface information in case of an incoming label.

IPV4-ADDRESS TLV:

```

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=TBD14                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     IPv4 address                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

IPV6-ADDRESS TLV:

```

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=TBD15                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     //                                     //
|                                     IPv6 address (16 bytes)                                     |
|                                     //                                     //
+-----+-----+-----+-----+-----+-----+-----+-----+

```

UNNUMBERED-IPV4-ID-ADDRESS TLV:

```

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=TBD16                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Node-ID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```



```

+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Interface ID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

LINKLOCAL-IPV6-ID-ADDRESS TLV:

```

      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=TBD17                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
//                                     Local IPv6 address (16 octets)                                     //
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Local Interface ID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
//                                     Remote IPv6 address (16 octets)                                     //
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Remote Interface ID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

The address TLVs are as follows:

IPV4-ADDRESS TLV: an IPv4 address.

IPV6-ADDRESS TLV: an IPv6 address.

UNNUMBERED-IPV4-ID-ADDRESS TLV: a Node ID / Interface ID tuple.

LINKLOCAL-IPV6-ID-ADDRESS TLV: a pair of (global IPv6 address, interface ID) tuples.

8. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to RFC 7942.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their

features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

8.1. Huawei's Proof of Concept based on ONOS

The PCE function was developed in the ONOS open source platform. This extension was implemented on a private version as a proof of concept for PCECC.

- o Organization: Huawei
- o Implementation: Huawei's PoC based on ONOS
- o Description: PCEP as a southbound plugin was added to ONOS. To support PCECC, an earlier version of this I-D was implemented. Refer <https://wiki.onosproject.org/display/ONOS/PCEP+Protocol>
- o Maturity Level: Prototype
- o Coverage: Partial
- o Contact: satishk@huawei.com

9. Security Considerations

The security considerations described in [RFC8231] and [RFC8281] apply to the extensions described in this document. Additional considerations related to a malicious PCE are introduced.

9.1. Malicious PCE

PCE has complete control over PCC to update the labels and can cause the LSP's to behave inappropriate and cause cause major impact to the network. As a general precaution, it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253], as per the recommendations and best current practices in [RFC7525].

10. Manageability Considerations

10.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC capability as a global configuration.

10.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC capability.

10.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

10.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

10.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

10.6. Impact On Network Operations

PCEP extensions defined in this document do not put new requirements on network operations.

11. IANA Considerations

11.1. PCEP TLV Type Indicators

IANA is requested to allocate the following TLV Type Indicator values within the "PCEP TLV Type Indicators" sub- registry of the PCEP Numbers registry:

Value	Meaning	Reference
TBD14	IPV4-ADDRESS TLV	This document
TBD15	IPV6-ADDRESS TLV	This document
TBD16	UNNUMBERED-IPV4-ID-ADDRESS TLV	This document
TBD17	LINKLOCAL-IPV6-ID-ADDRESS TLV	This document

11.2. PATH-SETUP-TYPE-CAPABILITY Sub-TLV Type Indicators

[I-D.ietf-pce-segment-routing] requested creation of "PATH-SETUP-TYPE-CAPABILITY Sub-TLV Type Indicators" sub-registry. Further IANA is requested to allocate the following code-point:

Value	Meaning	Reference
TBD12	PCECC-CAPABILITY	This document

11.3. New Path Setup Type Registry

[RFC8408] created a sub-registry within the "Path Computation Element Protocol (PCEP) Numbers" registry called "PCEP Path Setup Types". IANA is requested to allocate a new code point within this registry, as follows:

Value	Description	Reference
TBD1	Traffic engineering path is setup using PCECC mode	This document

11.4. PCEP Object

IANA is requested to allocate new code-point in the "PCEP Objects" sub-registry for the CCI object as follows:

Object-Class	Value	Name	Reference
TBD13		CCI Object-Type	This document
	0		Reserved
	1		MPLS Label

11.5. CCI Object Flag Field

IANA is requested to create a new sub-registry to manage the Flag field of the CCI object called "CCI Object Flag Field". New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

- o Bit number (counting from bit 0 as the most significant bit)
- o Capability description
- o Defining RFC

Two bits to be defined for the CCI Object flag field in this document as follows:

Bit	Description	Reference
0-13	Unassigned	This document
14	C Bit - PCC allocation	This document
15	O Bit - Specifies label is out label	This document

11.6. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Meaning	
10	Reception of an invalid object.	
	Error-value = TBD2 :	Missing PCECC Capability sub-TLV
19	Invalid operation.	
	Error-value = TBD3 :	Attempted PCECC operations when PCECC capability was not advertised
	Error-value = TBD4 :	Stateful PCE capability was not advertised
6	Mandatory Object missing.	Unknown Label
	Error-value = TBD8 :	
TBD5	PCECC failure.	CCI object missing
	Error-value = TBD11 :	
	Error-value = TBD6 :	Label out of range.
	Error-value = TBD7 :	Instruction failed.
	Error-value = TBD9 :	Invalid CCI.
	Error-value = TBD10 :	Unable to allocate the specified CCI.

12. Acknowledgments

We would like to thank Robert Tao, Changjing Yan, Tieying Huang and Avantika for their useful comments and suggestions.

13. References

13.1. 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>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [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>>.
- [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>>.
- [RFC8233] Dhody, D., Wu, Q., Manral, V., Ali, Z., and K. Kumaki, "Extensions to the Path Computation Element Communication Protocol (PCEP) to Compute Service-Aware Label Switched Paths (LSPs)", RFC 8233, DOI 10.17487/RFC8233, September 2017, <<https://www.rfc-editor.org/info/rfc8233>>.

- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.

13.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <<https://www.rfc-editor.org/info/rfc7942>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.

- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [I-D.ietf-pce-segment-routing]
Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "PCEP Extensions for Segment Routing", draft-ietf-pce-segment-routing-16 (work in progress), March 2019.
- [I-D.ietf-teas-pcecc-use-cases]
Zhao, Q., Li, Z., Khasanov, B., Dhody, D., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-04 (work in progress), July 2019.
- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-13 (work in progress), October 2019.
- [I-D.zhao-pce-pcep-extension-pce-controller-sr]
Zhao, Q., Li, Z., Negi, M., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of SR-LSPs", draft-zhao-pce-pcep-extension-pce-controller-sr-05 (work in progress), July 2019.
- [I-D.li-pce-controlled-id-space]
Li, C., Chen, M., Dong, J., Li, Z., Wang, A., Cheng, W., and C. Zhou, "PCE Controlled ID Space", draft-li-pce-controlled-id-space-03 (work in progress), June 2019.
- [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.", draft-sivabalan-pce-binding-label-sid-07 (work in progress), July 2019.

[I-D.li-pce-sr-path-segment]

Li, C., Chen, M., Cheng, W., Dong, J., Li, Z., Gandhi, R.,
and Q. Xiong, "Path Computation Element Communication
Protocol (PCEP) Extension for Path Segment in Segment
Routing (SR)", draft-li-pce-sr-path-segment-08 (work in
progress), August 2019.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Satish Karunanithi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: satishk@huawei.com

Adrian Farrel
Juniper Networks, Inc
UK

EMail: adrian@olddog.co.uk

Xuesong Geng
Huawei Technologies
China

Email: gengxuesong@huawei.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

Katherine Zhao
Futurewei Technologies

EMail: katherine.zhao@futurewei.com

Boris Zhang
Telus Ltd.
Toronto
Canada

EMail: boris.zhang@telus.com

Alex Tokar
Cisco Systems

Slovak Republic

EMail: atokar@cisco.com

Authors' Addresses

Quintin Zhao
Huawei Technologies
125 Nagog Technology Park
Acton, MA 01719
USA

EMail: quintin.zhao@huawei.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahend.ietf@gmail.com

Chao Zhou
Cisco Systems

EMail: chao.zhou@cisco.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 5, 2021

Z. Li
S. Peng
Huawei Technologies
M. Negi
RtBrick Inc
Q. Zhao
Etheric Networks
C. Zhou
HPE
March 4, 2021

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) of LSPs
draft-ietf-pce-pcep-extension-for-pce-controller-14

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems.

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/set up/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network device along the path, while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP extensions for using the PCE as the central controller for provisioning labels along the path of the static LSP.

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 September 5, 2021.

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	3
1.1. Requirements Language	5
2. Terminology	5
3. Basic PCECC Mode	5
4. PCEP Requirements	6
5. Procedures for Using the PCE as a Central Controller (PCECC)	6
5.1. Stateful PCE Model	6
5.2. New LSP Functions	6
5.3. New PCEP Object	7
5.4. PCECC Capability Advertisement	7
5.5. LSP Operations	9
5.5.1. PCE-Initiated PCECC LSP	9
5.5.2. PCC-Initiated PCECC LSP	12
5.5.3. Central Controller Instructions	15
5.5.3.1. Label Download CCI	16
5.5.3.2. Label Clean up CCI	16
5.5.4. PCECC LSP Update	17
5.5.5. Re-Delegation and Clean up	20
5.5.6. Synchronization of Central Controllers Instructions	20
5.5.7. PCECC LSP State Report	21
5.5.8. PCC-Based Allocations	21
6. PCEP Messages	21
6.1. The PCInitiate Message	22
6.2. The PCRpt Message	23
7. PCEP Objects	24
7.1. OPEN Object	24
7.1.1. PCECC Capability sub-TLV	25
7.2. PATH-SETUP-TYPE TLV	25
7.3. CCI Object	26

7.3.1. Address TLVs	27
8. Implementation Status	27
8.1. Huawei's Proof of Concept based on ONOS	28
9. Security Considerations	28
9.1. Malicious PCE	29
9.2. Malicious PCC	29
10. Manageability Considerations	29
10.1. Control of Function and Policy	29
10.2. Information and Data Models	30
10.3. Liveness Detection and Monitoring	30
10.4. Verify Correct Operations	30
10.5. Requirements On Other Protocols	30
10.6. Impact On Network Operations	31
11. IANA Considerations	31
11.1. PATH-SETUP-TYPE-CAPABILITY Sub-TLV Type Indicators	31
11.2. PCECC-CAPABILITY sub-TLV's Flag field	31
11.3. Path Setup Type Registry	31
11.4. PCEP Object	32
11.5. CCI Object Flag Field	32
11.6. PCEP-Error Object	32
12. Acknowledgments	33
13. References	33
13.1. Normative References	33
13.2. Informative References	35
Appendix A. Contributor Addresses	38
Authors' Addresses	39

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload the path computation function from routers in an MPLS traffic-engineered network. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way

that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

PCE was developed to derive paths for MPLS Label Switched Paths (LSPs), which are supplied to the head end of the LSP using the Path Computation Element Communication Protocol (PCEP). But SDN has a broader applicability than signaled MPLS and GMPLS traffic-engineered (TE) networks, and the PCE may be used to determine paths in a range of use cases. PCEP has been proposed as a control protocol for use in these environments to allow the PCE to be fully enabled as a central controller.

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCECC architecture.

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network device along the path while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label-forwarding instructions to program and what resources to reserve. The PCE-based controller keeps a view of the network and determines the paths of the end-to-end LSPs, and the controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

While this document is focused on the procedures for the static LSPs (referred to as basic PCECC mode in Section 3), the mechanisms and protocol encodings are specified in such a way that extensions for other use cases are easy to achieve. For example, the extensions for PCECC for Segment Routing (SR) are specified in [I-D.ietf-pce-pcep-extension-pce-controller-sr] and [I-D.dhody-pce-pcep-extension-pce-controller-srv6].

1.1. 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.

2. Terminology

The terminology used in this document is the same as that described in the [RFC8283].

3. Basic PCECC Mode

In this mode, LSPs are provisioned as explicit label instructions at each hop on the end-to-end path. Each router along the path must be told what label forwarding instructions to program and what resources to reserve. The controller uses PCEP to communicate with each router along the path of the end-to-end LSP.

[RFC8283] examines the motivations and applicability for PCECC and use of PCEP as an SBI. Section 3.1.2. of [RFC8283] highlights the use of PCECC for label allocation along the static LSPs and it simplifies the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. This allows the operator to introduce the advantages of SDN (such as programmability) into the network. Further Section 3.3. of [I-D.ietf-teas-pcecc-use-cases] describes some of the scenarios where the PCECC technique could be useful. Section 4 of [RFC8283] also describe the implications on the protocol when used as an SDN SBI. The operator needs to evaluate the advantages offered by PCECC against the operational and scalability needs of the PCECC.

As per Section 3.1.2. of [RFC8283], the PCE-based controller will take responsibility for managing some part of the MPLS label space for each of the routers that it controls, and may take wider responsibility for partitioning the label space for each router and allocating different parts for different uses. The PCC MUST NOT make allocations from the label space set aside for the PCE to avoid

overlap and collisions of label allocations. It is RECOMMENDED that PCE makes allocations (from the label space set aside for the PCE) for all nodes along the path. For the purpose of this document, it is assumed that the exclusive label range to be used by a PCE is known and set on both PCEP peers. A future extension could add the capability to advertise this range via a possible PCEP extension as well (see [I-D.li-pce-controlled-id-space]). The rest of the processing is similar to the existing stateful PCE mechanism.

This document also allows a case where the label space is maintained by the PCC and the labels are allocated by it. In this case, the PCE should request the allocation from PCC as described in Section 5.5.8.

4. PCEP Requirements

The following key requirements should be considered when designing the PCECC-based solution:

1. A PCEP speaker supporting this document needs to have the capability to advertise its PCECC capability to its peers.
2. A PCEP speaker need means to identify PCECC-based LSP in the PCEP messages.
3. PCEP procedures need to allow for PCC-based label allocations.
4. PCEP procedures need to provide a means to update (or clean up) label entries downloaded to the PCC.
5. PCEP procedures need to provide a means to synchronize the labels between the PCE and the PCC via PCEP messages.

5. Procedures for Using the PCE as a Central Controller (PCECC)

5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. PCE as a central controller (PCECC) reuses the existing active stateful PCE mechanism as much as possible to control LSPs.

5.2. New LSP Functions

Several new functions are required in PCEP to support PCECC. This document extends the existing messages to support the new functions required by PCECC:

PCInitiate: a PCEP message described in [RFC8281]. PCInitiate message is used to set up PCE-Initiated LSP based on PCECC

mechanism. It is also extended for Central Controller Instructions (CCI) (download or clean up the Label forwarding instructions in the context of this document) on all nodes along the path as described in Section 6.1.

PCRpt: a PCEP message described in [RFC8231]. PCRpt message is used to send PCECC LSP Reports. It is also extended to report the set of Central Controller Instructions (CCI) (label forwarding instructions in the context of this document) received from the PCE as described in Section 6.2. Section 5.5.6 describes the use of PCRpt message during synchronization.

PCUpd: a PCEP message described in [RFC8231]. PCUpd message is used to send PCECC LSP Updates.

The new functions defined in this document are mapped onto the PCEP messages as shown in Table 1.

Function	Message
PCECC Capability advertisement	Open
Label entry Add	PCInitiate
Label entry Clean up	PCInitiate
PCECC Initiated LSP	PCInitiate
PCECC LSP Update	PCUpd
PCECC LSP State Report	PCRpt
PCECC LSP Delegation	PCRpt
PCECC Label Report	PCRpt

Table 1: Functions mapped to the PCEP messages

5.3. New PCEP Object

This document defines a new PCEP object called CCI (Section 7.3) to specify the central controller instructions. In the scope of this document, this is limited to Label forwarding instructions. Future documents can create new CCI object-types for other types of central controller instructions. The CC-ID is the unique identifier for the central controller instructions in PCEP. The PCEP messages are extended in this document to handle the PCECC operations.

5.4. PCECC Capability Advertisement

During the PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of and willingness to use PCEP extensions for PCECC using these elements in the OPEN message:

- o A new Path Setup Type (PST) (Section 7.2) in the PATH-SETUP-TYPE-CAPABILITY TLV to indicate support for PCEP extensions for PCECC - TBD1 (Path is set up via PCECC mode)
- o A new PCECC-CAPABILITY sub-TLV (Section 7.1.1) with the L bit set to 1 inside the PATH-SETUP-TYPE-CAPABILITY TLV to indicate a willingness to use PCEP extensions for PCECC based central controller instructions for label download
- o The STATEFUL-PCE-CAPABILITY TLV ([RFC8231]) (with the I flag set [RFC8281])

The new Path Setup Type is to be listed in the PATH-SETUP-TYPE-CAPABILITY TLV by all PCEP speakers which support the PCEP extensions for PCECC in this document.

The new PCECC-CAPABILITY sub-TLV is included in PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN object to indicate a willingness to use the PCEP extensions for PCECC during the established PCEP session. Using the L bit in this TLV, the PCE shows the intention to function as a PCECC server, and the PCC shows a willingness to act as a PCECC client for label download instructions (see Section 7.1.1).

If the PCECC-CAPABILITY sub-TLV is advertised and the STATEFUL-PCE-CAPABILITY TLV is not advertised, or is advertised without the I flag set, in the OPEN Object, the receiver MUST:

- o Send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD4 (stateful PCE capability was not advertised)
- o Terminate the session

If a PCEP speaker receives the PATH-SETUP-TYPE-CAPABILITY TLV with the PCECC Path Setup Type but without the PCECC-CAPABILITY sub-TLV, it MUST:

- o Send a PCErr message with Error-Type 10 (Reception of an invalid object) and Error-Value TBD2 (Missing PCECC-CAPABILITY sub-TLV)
- o Terminate the PCEP session

The PCECC-CAPABILITY sub-TLV MUST NOT be used without the corresponding Path Setup Type being listed in the PATH-SETUP-TYPE-CAPABILITY TLV. If it is present without the corresponding Path Setup Type listed in the PATH-SETUP-TYPE-CAPABILITY TLV, it MUST be ignored.

If one or both speakers (PCE and PCC) have not indicated support and willingness to use the PCEP extensions for PCECC, the PCEP extensions for PCECC MUST NOT be used. If a PCECC operation is attempted when both speakers have not agreed in the OPEN messages, the receiver of the message MUST:

- o Send a PCErr message with Error-Type=19 (Invalid Operation) and Error-Value=TBD3 (Attempted PCECC operations when PCECC capability was not advertised)
- o Terminate the PCEP session

A legacy PCEP speaker (that does not recognize the PCECC Capability sub-TLV) will ignore the sub-TLV in accordance with [RFC8408] and [RFC5440]. As per [RFC8408], the legacy PCEP speaker on receipt of an unsupported PST in RP (Request Parameter) /SRP (Stateful PCE Request Parameters) Object will:

- o Send a PCErr message with Error-Type = 21 (Invalid traffic engineering path setup type) and Error-value = 1 (Unsupported path setup type)
- o Terminate the PCEP session

5.5. LSP Operations

The PCEP messages pertaining to a PCECC MUST include PATH-SETUP-TYPE TLV [RFC8408] in the SRP object [RFC8231] with PST set to TBD1 to clearly identify that PCECC LSP is intended.

5.5.1. PCE-Initiated PCECC LSP

The LSP Instantiation operation is defined in [RFC8281]. In order to set up a PCE-Initiated LSP based on the PCECC mechanism, a PCE sends PCInitiate message with PST set to TBD1 for PCECC (see Section 7.2) to the ingress PCC.

The label forwarding instructions (see Section 5.5.3) from PCECC are sent after the initial PCInitiate and PCRpt message exchange with the ingress PCC as per [RFC8281] (see Figure 1). This is done so that the PLSP-ID and other LSP identifiers can be obtained from the ingress and can be included in the label forwarding instruction in the next set of PCInitiate messages along the path as described below.

An LSP-IDENTIFIERS TLV [RFC8231] MUST be included for PCECC LSPs, it uniquely identifies the LSP in the network. Note that the fields in the LSP-IDENTIFIERS TLV are described for the RSVP-signaled LSPs but

are applicable to the PCECC LSP as well. The LSP object is included in the central controller instructions (label download Section 7.3) to identify the PCECC LSP for this instruction. The PLSP-ID is the original identifier used by the ingress PCC, so a transit/egress LSR could have multiple central controller instructions that have the same PLSP-ID. The PLSP-ID in combination with the source (in LSP-IDENTIFIERS TLV) MUST be unique. The PLSP-ID is included for maintainability reasons to ease debugging. As per [RFC8281], the LSP object could also include the SPEAKER-ENTITY-ID TLV to identify the PCE that initiated these instructions. Also, the CC-ID is unique in each PCEP session as described in Section 7.3.

On receipt of PCInitiate message for the PCECC LSP, the PCC responds with a PCRpt message with the status set to "GOING-UP" and carrying the assigned PLSP-ID (see Figure 1). The ingress PCC also sets the D (Delegate) flag (see [RFC8231]) and C (Create) flag (see [RFC8281]) in the LSP object. When the PCE receives this PCRpt message with the PLSP-ID, it assigns labels along the path; and sets up the path by sending a PCInitiate message to each node along the path of the LSP as per the PCECC technique. The CC-ID uniquely identifies the central controller instruction within a PCEP session. Each node along the path (PCC) responds with a PCRpt message to acknowledge the central controller instruction with the PCRpt messages including the central controller instruction (CCI) and the LSP objects.

The ingress node would receive one CCI object with O bit (out-label) set. The transit node(s) would receive two CCI objects with the in-label CCI without an O bit set and the out-label CCI with O bit set. The egress node would receive one CCI object without O bit set (see Figure 1). A node can determine its role based on the setting of the O bit in the CCI object(s) and the LSP-IDENTIFIERS TLV in the LSP object.

The LSP deletion operation for PCE-Initiated PCECC LSP is the same as defined in [RFC8281]. The PCE should further perform Label entry clean up operation as described in Section 5.5.3.2 for the corresponding LSP.

The PCE-Initiated PCECC LSP setup sequence is shown in Figure 1.

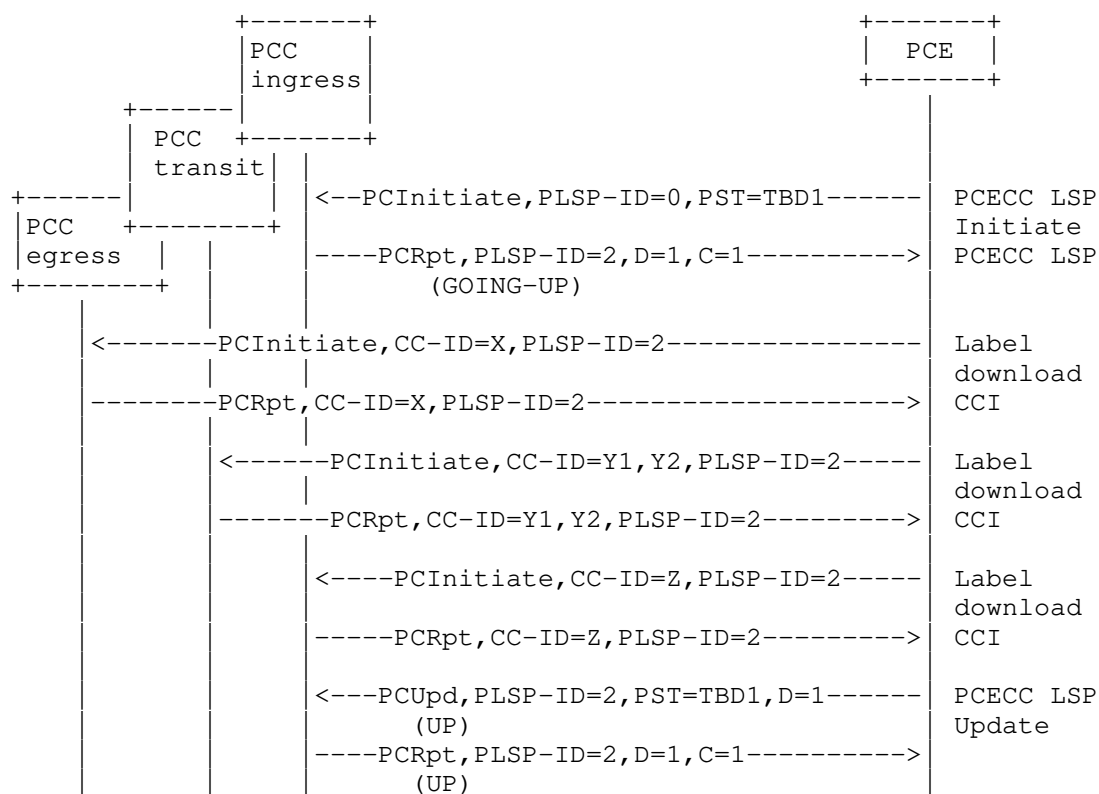


Figure 1: PCE-Initiated PCECC LSP

Once the label operations are completed, the PCE MUST send a PCUpd message to the ingress PCC. The PCUpd message is as per [RFC8231] with D flag set.

The PCECC LSPs are considered to be 'up' by default (on receipt of PCUpd message from PCE). The ingress could further choose to deploy a data plane check mechanism and report the status back to the PCE via a PCRpt message to make sure that the correct label instructions are made along the path of the PCECC LSP (and it is ready to carry traffic). The exact mechanism is out of scope of this document.

In the case where the label allocations are made by the PCC itself (see Section 5.5.8), the PCE could request an allocation to be made by the PCC, and then the PCC would send a PCRpt with the allocated label encoded in the CC-ID object as shown in Figure 2 in the configuration sequence from the egress towards the ingress along the path.

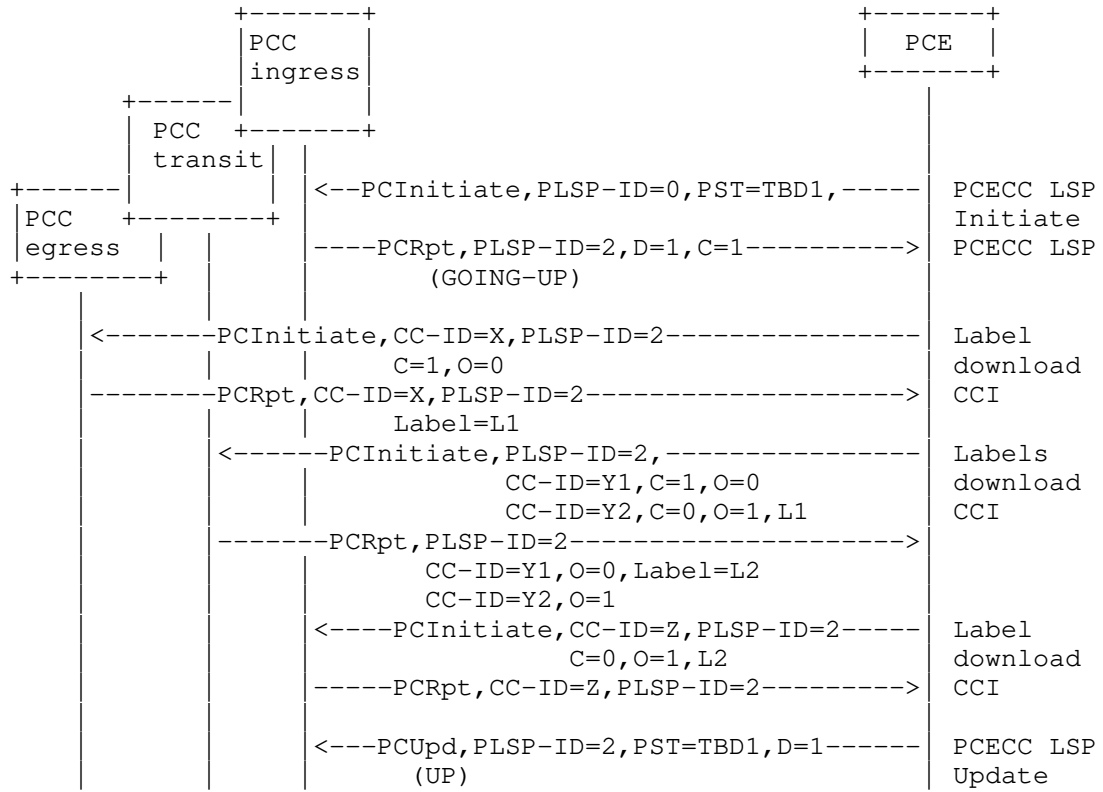


Figure 2: PCE-Initiated PCECC LSP (PCC allocation)

It should be noted that in this example, the request is made to the egress node with the C bit set in the CCI object to indicate that the label allocation needs to be done by the egress and the egress responds with the allocated label to the PCE. The PCE further inform the transit PCC without setting the C bit to 1 in the CCI object for out-label but the C bit is set to 1 for in-label so the transit node make the label allocation (for the in-label) and report to the PCE. Similarly, the C bit is unset towards the ingress to complete all the label allocation for the PCECC LSP.

5.5.2. PCC-Initiated PCECC LSP

In order to set up an LSP based on the PCECC mechanism where the LSP is configured at the PCC, a PCC MUST delegate the LSP by sending a

PCRpt message with PST set for PCECC (see Section 7.2) and D (Delegate) flag (see [RFC8231]) set in the LSP object (see Figure 3).

When a PCE receives the initial PCRpt message with D flag and PST Type set to TBD1, it SHOULD calculate the path and assigns labels along the path; and sets up the path by sending a PCInitiate message to each node along the path of the LSP as per the PCECC technique (see Figure 3). The CC-ID uniquely identifies the central controller instruction within a PCEP session. Each PCC further responds with the PCRpt messages including the central controller instruction (CCI) and the LSP objects.

Once the central controller instructions (label operations) are completed, the PCE MUST send the PCUpd message to the ingress PCC. As per [RFC8231], this PCUpd message should include the path information calculated by the PCE.

Note that the PCECC LSPs MUST be delegated to a PCE at all times.

The LSP deletion operation for PCECC LSPs is the same as defined in [RFC8231]. If the PCE receives a PCRpt message for LSP deletion then it does label clean up operation as described in Section 5.5.3.2 for the corresponding LSP.

The Basic PCECC LSP setup sequence is as shown in Figure 3.

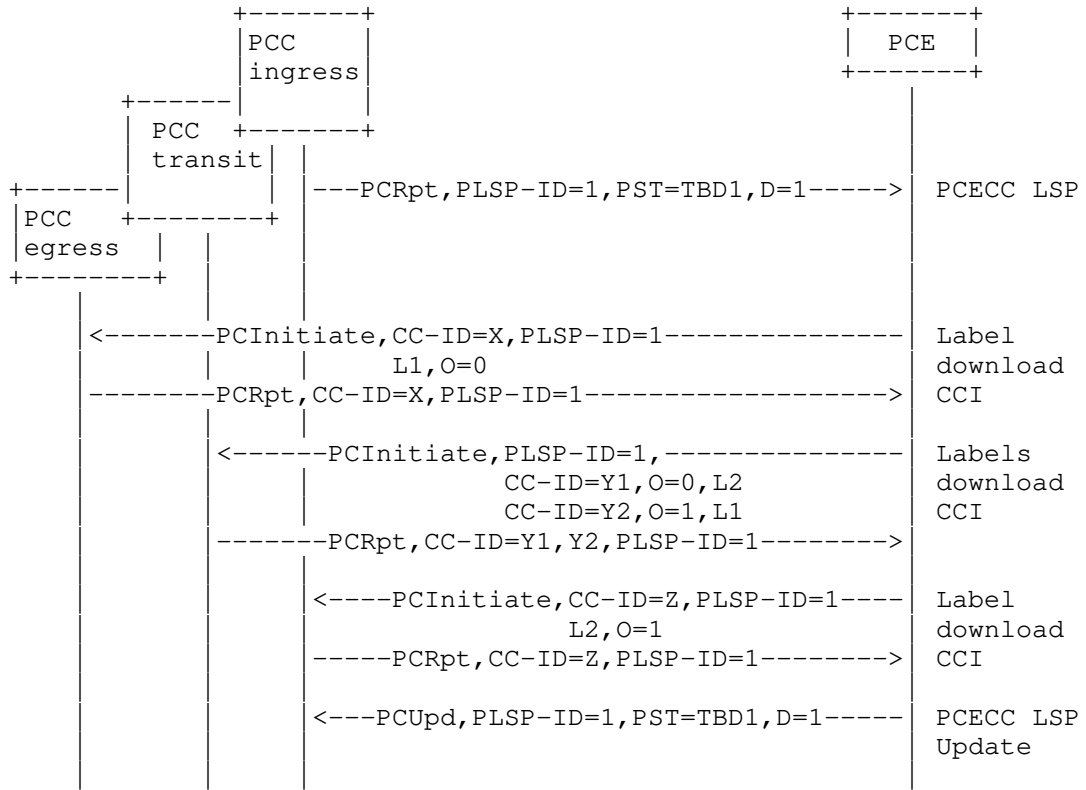
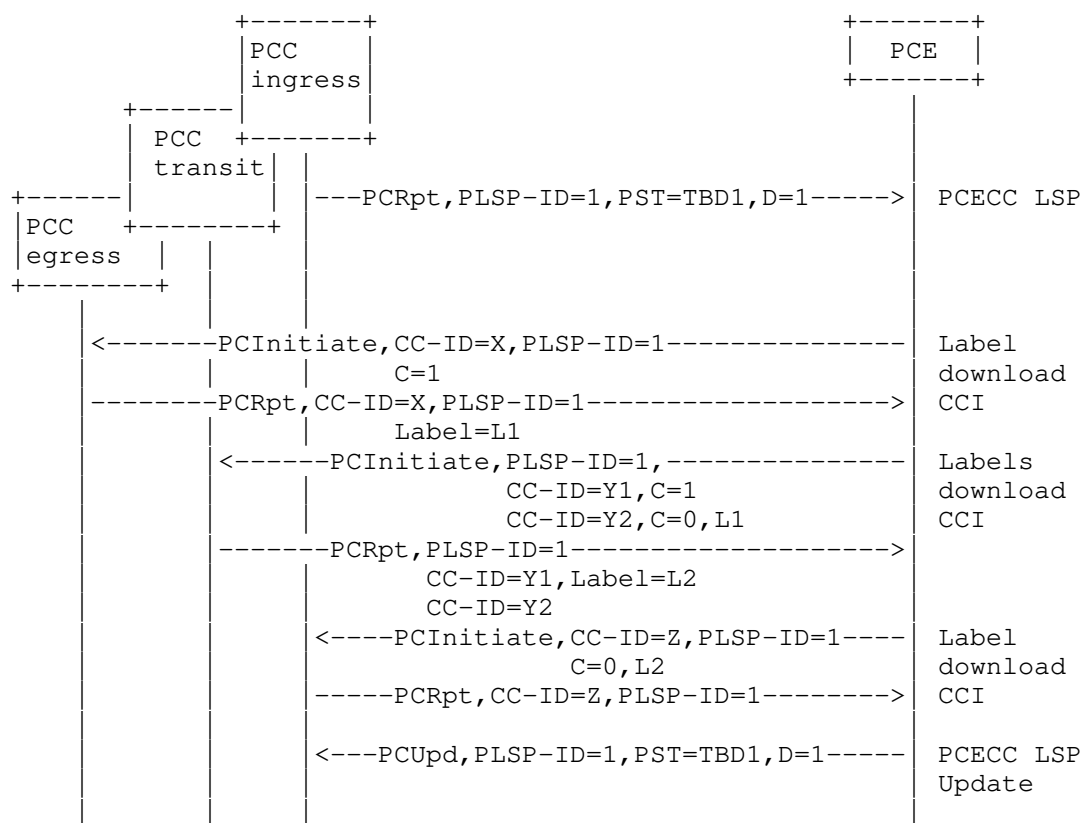


Figure 3: PCC-Initiated PCECC LSP

In the case where the label allocations are made by the PCC itself (see Section 5.5.8), the PCE could request an allocation to be made by the PCC, and then the PCC would send a PCRpt with the allocated label encoded in the CC-ID object as shown in Figure 4.



- The 0 bit is set as before (and thus not included)

Figure 4: PCC-Initiated PCECC LSP (PCC allocation)

In the case where the label allocations are made by the PCC itself (see Section 5.5.8), the procedure remains the same, with just an additional constraint on the configuration sequence.

The rest of the PCC-Initiated PCECC LSP setup operations are the same as those described in Section 5.5.1.

5.5.3. Central Controller Instructions

The new central controller instructions (CCI) for the label operations in PCEP are done via the PCInitiate message (Section 6.1), by defining a new PCEP Object for CCI operations. The local label range of each PCC is assumed to be known by both the PCC and the PCE.

5.5.3.1. Label Download CCI

In order to set up an LSP based on PCECC, the PCE sends a PCInitiate message to each node along the path to download the Label instruction as described in Section 5.5.1 and Section 5.5.2.

The CCI object MUST be included, along with the LSP object in the PCInitiate message. The LSP-IDENTIFIERS TLV MUST be included in the LSP object. The SPEAKER-ENTITY-ID TLV SHOULD be included in the LSP object.

If a node (PCC) receives a PCInitiate message which includes a Label to download, as part of CCI, that is out of the range set aside for the PCE, it MUST send a PCErr message with Error-type=TBD5 (PCECC failure) and Error-value=TBD6 (Label out of range) and MUST include the SRP object to specify the error is for the corresponding label update via PCInitiate message. If a PCC receives a PCInitiate message but fails to download the Label entry, it MUST send a PCErr message with Error-type=TBD5 (PCECC failure) and Error-value=TBD7 (instruction failed) and MUST include the SRP object to specify the error is for the corresponding label update via PCInitiate message.

A new PCEP object for central controller instructions (CCI) is defined in Section 7.3.

5.5.3.2. Label Clean up CCI

In order to delete an LSP based on PCECC, the PCE sends a central controller instructions via a PCInitiate message to each node along the path of the LSP to clean up the Label forwarding instruction.

If the PCC receives a PCInitiate message but does not recognize the label in the CCI, the PCC MUST generate a PCErr message with Error-Type 19(Invalid operation) and Error-Value=TBD8, "Unknown Label" and MUST include the SRP object to specify the error is for the corresponding label clean up (via PCInitiate message).

The R flag in the SRP object defined in [RFC8281] specifies the deletion of Label Entry in the PCInitiate message.

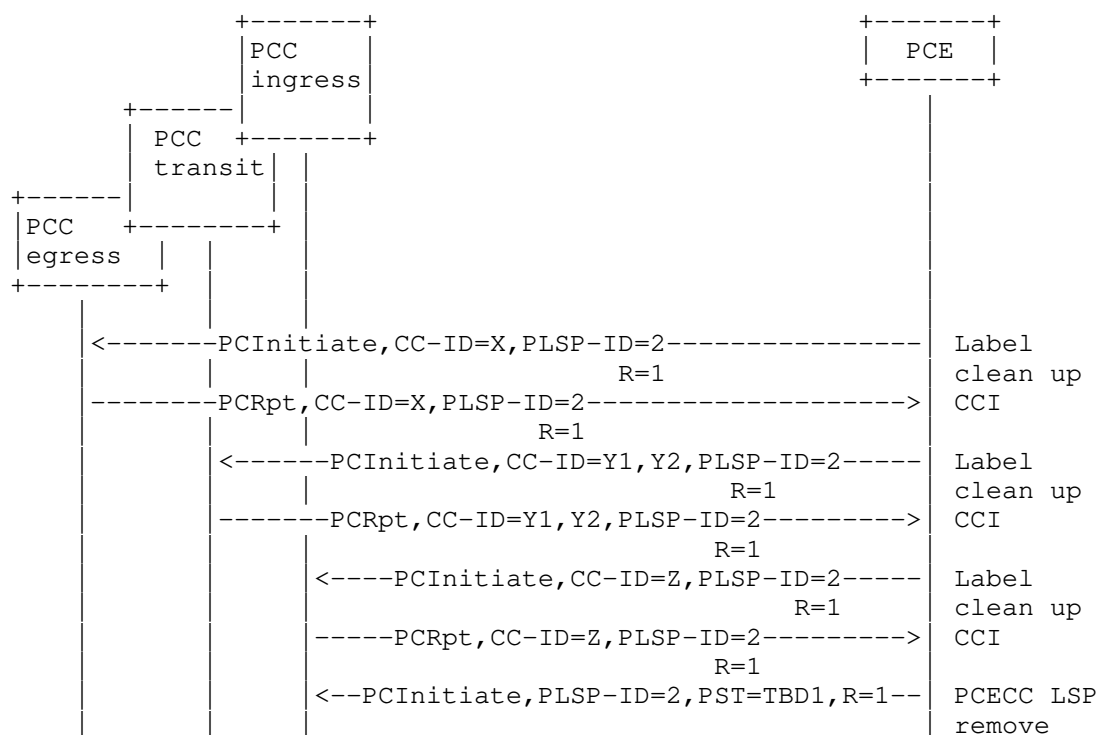


Figure 5: Label Cleanup

As per [RFC8281], following the removal of the Label forwarding instruction, the PCC MUST send a PCRpt message. The SRP object in the PCRpt MUST include the SRP-ID-number from the PCInitiate message that triggered the removal. The R flag in the SRP object MUST be set.

In the case where the label allocation is made by the PCC itself (see Section 5.5.8), the removal procedure remains the same, adding the sequence constraint.

5.5.4. PCECC LSP Update

The update is done as per the make-before-break procedures, i.e. the PCECC first updates new label instructions based on the updated path and then informs the ingress to switch traffic, before cleaning up the former instructions. New CC-IDs are used to identify the updated instructions; the identifiers in the LSP object uniquely identify the existing LSP. Once new instructions are downloaded, the PCE further updates the new path at the ingress which triggers the traffic switch

on the updated path. The ingress PCC acknowledges with a PCRpt message, on receipt of the PCRpt message, the PCE does clean up operation for the former LSP as described in Section 5.5.3.2.

The PCECC LSP Update sequence is shown in Figure 6.

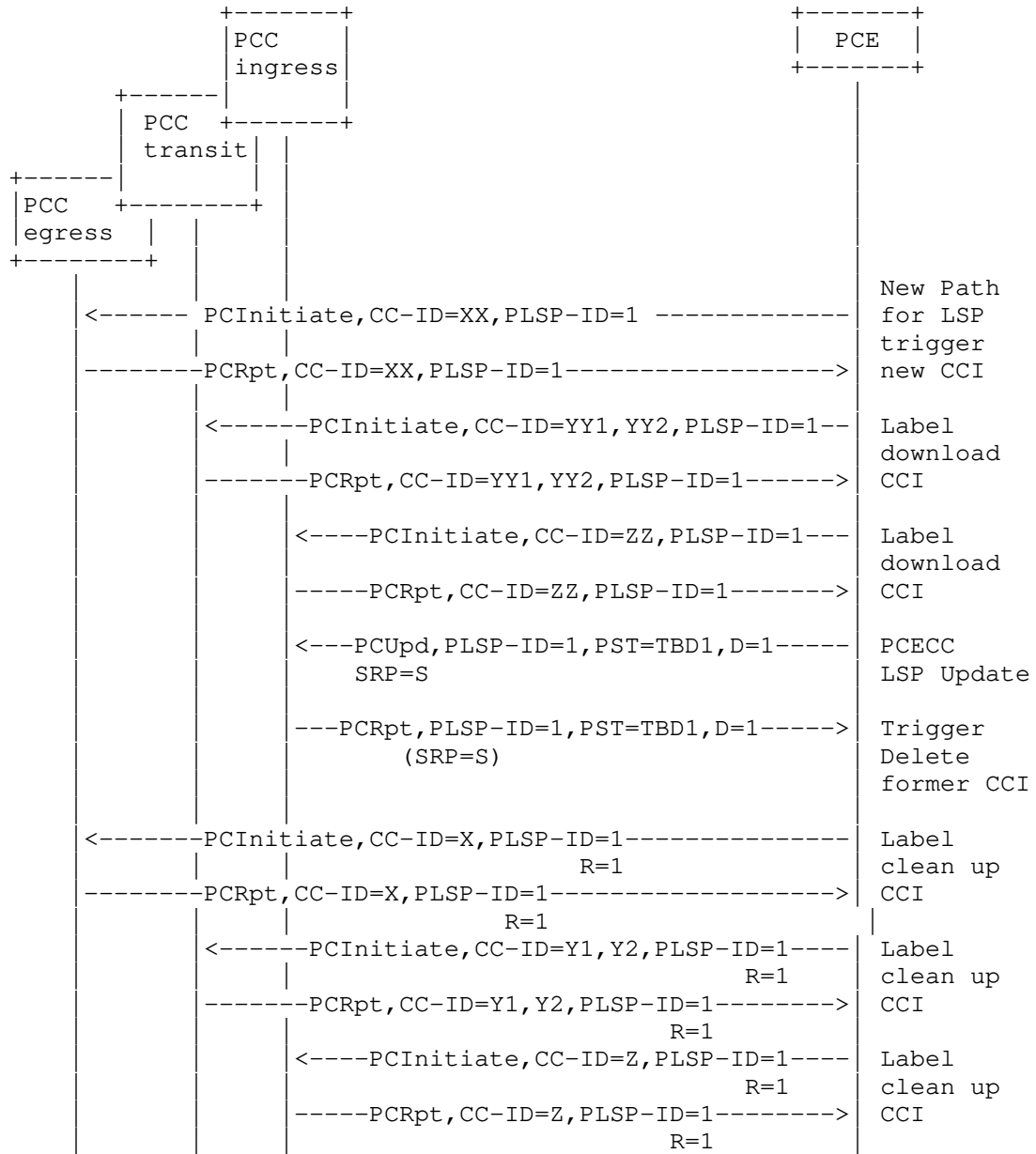


Figure 6: PCECC LSP Update

The modified PCECC LSPs are considered to be 'up' by default. The ingress could further choose to deploy a data plane check mechanism

and report the status back to the PCE via a PCRpt message. The exact mechanism is out of scope of this document.

In the case where the label allocations are made by the PCC itself (see Section 5.5.8), the procedure remains the same.

5.5.5. Re-Delegation and Clean up

As described in [RFC8281], a new PCE can gain control over an orphaned LSP. In the case of a PCECC LSP, the new PCE MUST also gain control over the central controller instructions in the same way by sending a PCInitiate message that includes the SRP, LSP, and CCI objects and carries the CC-ID and PLSP-ID identifying the instruction that it wants to take control of.

Further, as described in [RFC8281], the State Timeout Interval timer ensures that a PCE crash does not result in automatic and immediate disruption for the services using PCE-initiated LSPs. Similarly the central controller instructions are not removed immediately upon PCE failure. Instead, they are cleaned up on the expiration of this timer. This allows for network clean up without manual intervention. The PCC MUST support the removal of CCI as one of the behaviors applied on expiration of the State Timeout Interval timer.

In case of PCC-initiated PCECC LSP, the control over the orphaned LSP at the ingress PCC is taken over by the mechanism specified in [RFC8741] to request delegation. The control over the central controller instructions is described above using [RFC8281].

5.5.6. Synchronization of Central Controllers Instructions

The purpose of Central Controllers Instructions synchronization (labels in the context of this document) is to make sure that the PCE's view of CCI (Labels) matches with the PCC's Label allocation. This synchronization is performed as part of the LSP state synchronization as described in [RFC8231] and [RFC8232].

As per LSP State Synchronization [RFC8231], a PCC reports the state of its LSPs to the PCE using PCRpt messages and as per [RFC8281], PCE would initiate any missing LSPs and/or remove any LSPs that are not wanted. The same PCEP messages and procedures are also used for the Central Controllers Instructions synchronization. The PCRpt message includes the CCI and the LSP object to report the label forwarding instructions. The PCE would further remove any unwanted instructions or initiate any missing instructions.

5.5.7. PCECC LSP State Report

As mentioned before, an ingress PCC MAY choose to apply any OAM mechanism to check the status of LSP in the Data plane and MAY further send its status in a PCRpt message to the PCE.

5.5.8. PCC-Based Allocations

The PCE can request the PCC to allocate the label using the PCInitiate message. The C flag in the CCI object is set to 1 to indicate that the allocation needs to be done by the PCC. The PCC MUST try to allocate the Label and MUST report to the PCE via PCRpt or PCErr message.

If the value of the Label is 0 and the C flag is set to 1, it indicates that the PCE is requesting the allocation to be done by the PCC. If the Label is 'n' and the C flag is set to 1 in the CCI object, it indicates that the PCE requests a specific value 'n' for the Label. If the allocation is successful, the PCC MUST report via the PCRpt message with the CCI object. If the value of the Label in the CCI object is invalid, it MUST send a PCErr message with Error-Type = TBD5 ("PCECC failure") and Error Value = TBD9 ("Invalid CCI"). If it is valid but the PCC is unable to allocate it, it MUST send a PCErr message with Error-Type = TBD5 ("PCECC failure") and Error Value = TBD10 ("Unable to allocate the specified CCI").

If the PCC wishes to withdraw or modify the previously assigned label, it MUST send a PCRpt message without any Label or with the Label containing the new value respectively in the CCI object. The PCE would further trigger the Label cleanup of older label as per Section 5.5.3.2.

6. PCEP Messages

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable-length body made of a set of objects that can be either mandatory or optional. An object is said to be mandatory in a PCEP message when the object must be included for the message to be considered valid. For each PCEP message type, a set of rules is defined that specify the set of objects that the message can carry. An implementation MUST form the PCEP messages using the object ordering specified in this document.

LSP-IDENTIFIERS TLV MUST be included in the LSP object for PCECC LSP.

The message formats in this document are specified using Routing Backus-Naur Form (RBNF) encoding as specified in [RFC5511].

6.1. The PCInitiate Message

The PCInitiate message [RFC8281] can be used to download or remove the labels, this document extends the message as shown below -

```
<PCInitiate Message> ::= <Common Header>
                           <PCE-initiated-lsp-list>
```

Where:

<Common Header> is defined in [RFC5440]

```
<PCE-initiated-lsp-list> ::= <PCE-initiated-lsp-request>
                              [<PCE-initiated-lsp-list>]
```

```
<PCE-initiated-lsp-request> ::=
    (<PCE-initiated-lsp-instantiation>|
     <PCE-initiated-lsp-deletion>|
     <PCE-initiated-lsp-central-control>)
```

```
<PCE-initiated-lsp-central-control> ::= <SRP>
                                          <LSP>
                                          <cci-list>
```

```
<cci-list> ::= <CCI>
               [<cci-list>]
```

Where:

<PCE-initiated-lsp-instantiation> and
<PCE-initiated-lsp-deletion> are as per
[RFC8281].

The LSP and SRP object is defined in [RFC8231].

When PCInitiate message is used for the central controller instructions (labels), the SRP, LSP, and CCI objects MUST be present. The SRP object is defined in [RFC8231] and if the SRP object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=10 (SRP object missing). The LSP object is defined in [RFC8231] and if the LSP object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=8 (LSP object missing). The CCI object is defined in Section 7.3 and if the CCI object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD11 (CCI object missing). More than one CCI object MAY be included in the PCInitiate message for a transit LSR.

To clean up entries, the R (remove) bit MUST be set in the SRP object to be encoded along with the LSP and the CCI object.

The CCI object received at the ingress node MUST have the O bit (out-label) set. The CCI Object received at the egress MUST have the O bit unset. If this is not the case, PCC MUST send a PCErr message with Error-Type = TBD5 ("PCECC failure") and Error Value = TBD9 ("Invalid CCI"). Other instances of the CCI object if present, MUST be ignored.

For the P2P LSP setup via PCECC technique, at the transit LSR two CCI objects are expected for in-coming and outgoing label associated with the LSP object. If any other CCI object is included in the PCInitiate message, it MUST be ignored. If the transit LSR did not receive two CCI object with one of them having the O bit set and another with O bit unset, it MUST send a PCErr message with Error-Type = TBD5 ("PCECC failure") and Error Value = TBD9 ("Invalid CCI").

Note that, on receipt of the PCInitiate message with CCI object, the ingress, egress, or transit role of the PCC is identified via the ingress and egress IP address encoded in the LSP-IDENTIFIERS TLV.

6.2. The PCRpt Message

The PCRpt message can be used to report the labels that were allocated by the PCE, to be used during the state synchronization phase or as an acknowledgment to PCInitiate message.

```
<PCRpt Message> ::= <Common Header>
                        <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report>[<state-report-list>]
```

```
<state-report> ::= (<lsp-state-report>|
                    <central-control-report>)
```

```
<lsp-state-report> ::= [<SRP>]
                        <LSP>
                        <path>
```

```
<central-control-report> ::= [<SRP>]
                              <LSP>
                              <cci-list>
```

```
<cci-list> ::= <CCI>
               [<cci-list>]
```

Where:

<path> is as per [RFC8231] and the LSP and SRP object are also defined in [RFC8231].

When PCRpt message is used to report the central controller instructions (labels), the LSP and CCI objects MUST be present. The LSP object is defined in [RFC8231] and if the LSP object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=8 (LSP object missing). The CCI object is defined in Section 7.3 and if the CCI object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD11 (CCI object missing). Two CCI objects can be included in the PCRpt message for a transit LSR.

7. PCEP Objects

The PCEP objects defined in this document are compliant with the PCEP object format defined in [RFC5440].

7.1. OPEN Object

This document defines a new PST (TBD1) to be included in the PATH-SETUP-TYPE-CAPABILITY TLV in the OPEN Object. Further, a new sub-TLV for PCECC capability exchange is also defined.

7.1.1. PCECC Capability sub-TLV

The PCECC-CAPABILITY sub-TLV is an optional TLV for use in the OPEN Object in the PATH-SETUP-TYPE-CAPABILITY TLV, when the Path Setup Type list includes the PCECC Path Setup Type TBD1. A PCECC-CAPABILITY sub-TLV MUST be ignored if the PST list does not contain PST=TBD1.

Its format is shown in Figure 7.

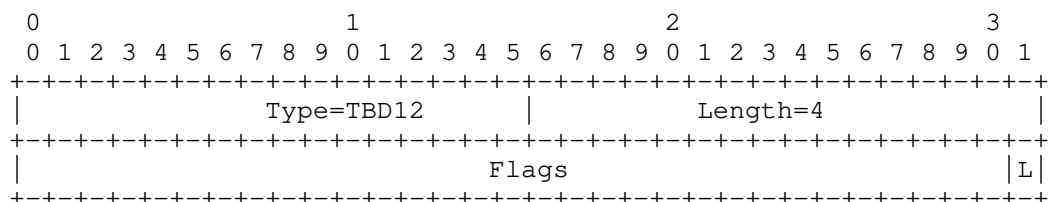


Figure 7: PCECC Capability sub-TLV

The type of the TLV is TBD12 and it has a fixed length of 4 octets.

The value comprises a single field - Flags (32 bits). Currently, the following flag bit is defined:

- o L bit (Label): if set to 1 by a PCEP speaker, the L flag indicates that the PCEP speaker support and is willing to handle the PCECC based central controller instructions for label download. The bit MUST be set to 1 by both a PCC and a PCE for the PCECC label download/report on a PCEP session.
- o Unassigned bits MUST be set to 0 on transmission and MUST be ignored on receipt.

7.2. PATH-SETUP-TYPE TLV

The PATH-SETUP-TYPE TLV is defined in [RFC8408]; this document defines a new PST value:

- o PST = TBD1: Path is set up via PCECC mode.

On a PCRpt/PCUpd/PCInitiate message, the PST=TBD1 in the PATH-SETUP-TYPE TLV in the SRP object MUST be included for a LSP set up via the PCECC-based mechanism.

7.3. CCI Object

The Central Controller Instructions (CCI) Object is used by the PCE to specify the forwarding instructions (Label information in the context of this document) to the PCC, and MAY be carried within PCInitiate or PCRpt message for label download/report.

CCI Object-Class is TBD13.

CCI Object-Type is 1 for the MPLS Label.

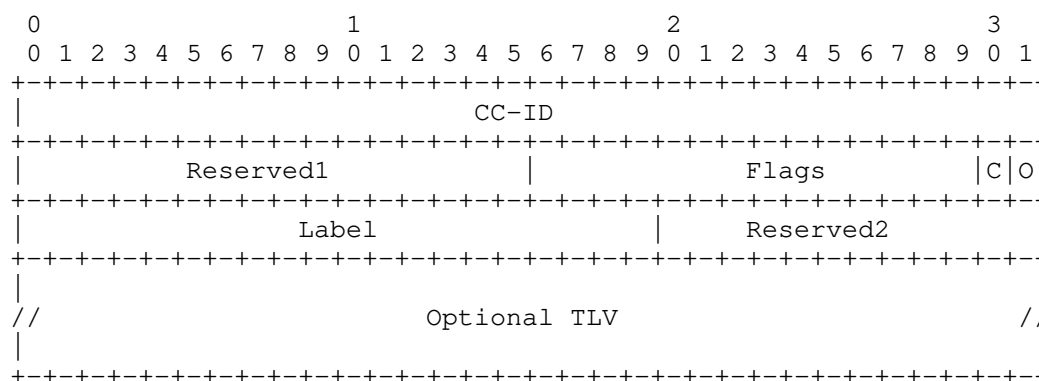


Figure 8: CCI Object

The fields in the CCI object are as follows:

CC-ID: A PCEP-specific identifier for the CCI information. A PCE creates a CC-ID for each instruction, the value is unique within the scope of the PCE and is constant for the lifetime of a PCEP session. The values 0 and 0xFFFFFFFF are reserved and MUST NOT be used. Note that [I-D.gont-numeric-ids-sec-considerations] gives advice on assigning transient numeric identifiers such as the CC-ID so as to minimize security risks.

Reserved1 (16 bit): Set to zero while sending, ignored on receive.

Flags (16 bit): A field used to carry any additional information pertaining to the CCI. Currently, the following flag bits are defined:

- * **O bit (Out-label)** : If the bit is set to 1, it specifies the label is the OUT label and it is mandatory to encode the next-hop information (via Address TLVs Section 7.3.1 in the CCI

object). If the bit is not set, it specifies the label is the IN label and it is optional to encode the local interface information (via Address TLVs in the CCI object).

- * C Bit (PCC Allocation): If the bit is set to 1, it indicates that the label allocation needs to be done by the PCC for this central controller instruction. A PCE sets this bit to request the PCC to make an allocation from its label space. A PCC would set this bit to indicate that it has allocated the label and report it to the PCE.
- * All unassigned bits MUST be set to zero at transmission and ignored at receipt.

Label (20-bit): The Label information.

Reserved2 (12 bit): Set to zero while sending, ignored on receive.

7.3.1. Address TLVs

[RFC8779] defines IPV4-ADDRESS, IPV6-ADDRESS, and UNNUMBERED-ENDPOINT TLVs for the use of Generalized Endpoint. The same TLVs can also be used in the CCI object to associate the next-hop information in the case of an outgoing label and local interface information in the case of an incoming label. The next-hop information encoded in these TLVs needs to be a directly connected IP address/interface information. If the PCC is not able to resolve the next-hop information, it MUST reject the CCI and respond with a PCErr message with Error-Type = TBD5 ("PCECC failure") and Error Value = TBD15 ("Invalid next-hop information").

8. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to RFC 7942.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

8.1. Huawei's Proof of Concept based on ONOS

The PCE function was developed in the ONOS open source platform. This extension was implemented on a private version as a proof of concept for PCECC.

- o Organization: Huawei
- o Implementation: Huawei's PoC based on ONOS
- o Description: PCEP as a southbound plugin was added to ONOS. To support PCECC, an earlier version of this I-D was implemented. Refer <https://wiki.onosproject.org/display/ONOS/PCEP+Protocol>
- o Maturity Level: Prototype
- o Coverage: Partial
- o Contact: satishk@huawei.com

9. Security Considerations

As per [RFC8283], the security considerations for a PCE-based controller is a little different from those for any other PCE system. That is, the operation relies heavily on the use and security of PCEP, so consideration should be given to the security features discussed in [RFC5440] and the additional mechanisms described in [RFC8253]. It further lists the vulnerability of a central controller architecture, such as a central point of failure, denial-of-service, and a focus for interception and modification of messages sent to individual NEs.

In PCECC operations, the PCEP sessions are also required to the internal routers and thus increasing the resources required for the session management at the PCE.

The PCECC extension builds on the existing PCEP messages and thus the security considerations described in [RFC5440], [RFC8231] and [RFC8281] continue to apply. [RFC8253] specify the support of Transport Layer Security (TLS) in PCEP, as it provides support for peer authentication, message encryption, and integrity. It further

provide mechanisms for associating peer identities with different levels of access and/or authoritativeness via an attribute in X.509 certificates or a local policy with a specific accept-list of X.509 certificate. This can be used to check the authority for the PCECC operations. Additional considerations are discussed in following sections.

9.1. Malicious PCE

In this extension, the PCE has complete control over the PCC to download/remove the labels and can cause the LSP's to behave inappropriately and cause a major impact to the network. As a general precaution, it is RECOMMENDED that this PCEP extension be activated on mutually-authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using TLS [RFC8253], as per the recommendations and best current practices in BCP 195 [RFC7525].

Further, an attacker may flood the PCC with PCECC related messages at a rate that exceeds either the PCC's ability to process them or the network's ability to send them, by either spoofing messages or compromising the PCE itself. [RFC8281] provides a mechanism to protect the PCC by imposing a limit. The same can be used for the PCECC operations as well.

As specified in Section 5.5.3.1, a PCC needs to check if the label in the CCI object is in the range set aside for the PCE, otherwise it MUST send a PCErr message with Error-type=TBD5 (PCECC failure) and Error-value=TBD6 (Label out of range).

9.2. Malicious PCC

The PCECC mechanism described in this document requires the PCE to keep labels (CCI) that it downloads and relies on the PCC responding (with either an acknowledgment or an error message) to requests for LSP instantiation. This is an additional attack surface by placing a requirement for the PCE to keep a CCI/label replica for each PCC. It is RECOMMENDED that PCE implementations provide a limit on resources (in this case the CCI) a single PCC can occupy. [RFC8231] provides a notification mechanism when such threshold is reached.

10. Manageability Considerations

10.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow the PCECC capability to be enabled/disabled as part of the global configuration. Section 6.1 of [RFC8664] list various controlling factors regarding path setup type.

They are also applicable to the PCECC path setup types. Further, Section 6.2 of [RFC8664] describe the migration steps when path setup type of an existing LSP is changed.

10.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC capability.

10.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

10.4. Verify Correct Operations

The operator needs the following information to verify that PCEP is operating correctly with respect to the PCECC path setup type.

- o An implementation SHOULD allow the operator to view whether the PCEP speaker sent the PCECC PST capability to its peer.
- o An implementation SHOULD allow the operator to view whether the peer sent the PCECC PST capability.
- o An implementation SHOULD allow the operator to view whether the PCECC PST is enabled on a PCEP session.
- o If one PCEP speaker advertises the PCECC PST capability, but the other does not, then the implementation SHOULD create a log to inform the operator of the capability mismatch.
- o If a PCEP speaker rejects a CCI, then it SHOULD create a log to inform the operator, giving the reason for the decision (local policy, Label issues, etc.).

10.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

10.6. Impact On Network Operations

PCEP extensions defined in this document do not put new requirements on network operations.

11. IANA Considerations

11.1. PATH-SETUP-TYPE-CAPABILITY Sub-TLV Type Indicators

[RFC8408] requested the creation of "PATH-SETUP-TYPE-CAPABILITY Sub-TLV Type Indicators" sub-registry. Further IANA is requested to allocate the following code-point:

Value	Meaning	Reference
TBD12	PCECC-CAPABILITY	This document

11.2. PCECC-CAPABILITY sub-TLV's Flag field

This document defines the PCECC-CAPABILITY sub-TLV and requests that IANA to create a new sub-registry to manage the value of the PCECC-CAPABILITY sub-TLV's 32-bits Flag field. New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

- o Bit number (counting from bit 0 as the most significant bit)
- o Capability description
- o Defining RFC

Currently, there is one allocation in this registry.

Bit	Name	Reference
31	Label	This document
0-30	Unassigned	This document

11.3. Path Setup Type Registry

[RFC8408] created a sub-registry within the "Path Computation Element Protocol (PCEP) Numbers" registry called "PCEP Path Setup Types". IANA is requested to allocate a new code point within this registry, as follows:

Value	Description	Reference
TBD1	Traffic engineering path is set up using PCECC mode	This document

11.4. PCEP Object

IANA is requested to allocate new code-point in the "PCEP Objects" sub-registry for the CCI object as follows:

Object-Class	Value	Name	Reference
TBD13		CCI Object-Type	This document
	0		Reserved
	1		MPLS Label

11.5. CCI Object Flag Field

IANA is requested to create a new sub-registry to manage the Flag field of the CCI object called "CCI Object Flag Field for MPLS Label". New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

- o Bit number (counting from bit 0 as the most significant bit)
- o Capability description
- o Defining RFC

Two bits to be defined for the CCI Object flag field in this document as follows:

Bit	Description	Reference
0-13	Unassigned	This document
14	C Bit - PCC allocation	This document
15	O Bit - Specifies label is out-label	This document

11.6. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Meaning	
-----	-----	
6	Mandatory Object missing.	
	Error-value = TBD11 :	CCI object missing
10	Reception of an invalid object.	

	Error-value = TBD2 :	Missing PCECC Capability sub-TLV
19	Invalid operation.	
	Error-value = TBD3 :	Attempted PCECC operations when PCECC capability was not advertised
	Error-value = TBD4 :	Stateful PCE capability was not advertised
	Error-value = TBD8 :	Unknown Label
TBD5	PCECC failure.	
	Error-value = TBD6 :	Label out of range.
	Error-value = TBD7 :	Instruction failed.
	Error-value = TBD9 :	Invalid CCI.
	Error-value = TBD10 :	Unable to allocate the specified CCI.
	Error-value = TBD15 :	Invalid next-hop information.

12. Acknowledgments

We would like to thank Robert Tao, Changjing Yan, Tieying Huang, Avantika, and Aijun Wang for their useful comments and suggestions.

Thanks to Julien Meuric for shepherding this I-D and providing valuable comments. Thanks to Deborah Brungard for being the responsible AD.

Thanks to Victoria Pritchard for a very detailed RTGDIR review. Thanks to Yaron Sheffer for the SECDIR review. Thanks to Gyan Mishra for the GENART review.

Thanks to Alvaro Retana, Murray Kucherawy, Benjamin Kaduk, Roman Danyliw, Robert Wilton, Eric Vyncke, and Erik Kline for the IESG review.

13. References

13.1. 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>>.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, DOI 10.17487/RFC5511, April 2009, <<https://www.rfc-editor.org/info/rfc5511>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [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>>.
- [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>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.

- [RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <<https://www.rfc-editor.org/info/rfc8664>>.
- [RFC8779] Margaria, C., Ed., Gonzalez de Dios, O., Ed., and F. Zhang, Ed., "Path Computation Element Communication Protocol (PCEP) Extensions for GMPLS", RFC 8779, DOI 10.17487/RFC8779, July 2020, <<https://www.rfc-editor.org/info/rfc8779>>.

13.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.
- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <<https://www.rfc-editor.org/info/rfc7942>>.

- [RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <<https://www.rfc-editor.org/info/rfc8232>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [RFC8741] Raghuram, A., Goddard, A., Karthik, J., Sivabalan, S., and M. Negi, "Ability for a Stateful Path Computation Element (PCE) to Request and Obtain Control of a Label Switched Path (LSP)", RFC 8741, DOI 10.17487/RFC8741, March 2020, <<https://www.rfc-editor.org/info/rfc8741>>.
- [I-D.ietf-teas-pcecc-use-cases]
Li, Z., Khasanov, B., Dhody, D., Zhao, Q., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-06 (work in progress), September 2020.
- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-15 (work in progress), October 2020.
- [I-D.ietf-pce-pcep-extension-pce-controller-sr]
Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for Segment Routing (SR) MPLS Segment Identifier (SID) Allocation and Distribution.", draft-ietf-pce-pcep-extension-pce-controller-sr-00 (work in progress), December 2020.
- [I-D.dhody-pce-pcep-extension-pce-controller-srv6]
Li, Z., Peng, S., Geng, X., and M. Negi, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for SRv6", draft-dhody-pce-pcep-extension-pce-controller-srv6-05 (work in progress), November 2020.

[I-D.li-pce-controlled-id-space]

Li, C., Chen, M., Wang, A., Cheng, W., and C. Zhou, "PCE Controlled ID Space", draft-li-pce-controlled-id-space-07 (work in progress), October 2020.

[I-D.gont-numeric-ids-sec-considerations]

Gont, F. and I. Arce, "Security Considerations for Transient Numeric Identifiers Employed in Network Protocols", draft-gont-numeric-ids-sec-considerations-06 (work in progress), December 2020.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Satish Karunanithi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: satishk@huawei.com

Adrian Farrel
Old Dog Consulting
UK

EMail: adrian@olddog.co.uk

Xuesong Geng
Huawei Technologies
China

Email: gengxuesong@huawei.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

Katherine Zhao
Futurewei Technologies

EMail: katherine.zhao@futurewei.com

Boris Zhang
Telus Ltd.
Toronto
Canada

EMail: boris.zhang@telus.com

Alex Tokar
Cisco Systems

Slovak Republic

EMail: atokar@cisco.com

Authors' Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: pengshuping@huawei.com

Mahendra Singh Negi
RtBrick Inc
N-17L, 18th Cross Rd, HSR Layout
Bangalore, Karnataka 560102
India

EMail: mahend.ietf@gmail.com

Quintin Zhao
Ethereic Networks
1009 S CLAREMONT ST
SAN MATEO, CA 94402
USA

EMail: qzhao@ethericnetworks.com

Chao Zhou
HPE

EMail: chaozhou_us@yahoo.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: May 2, 2020

M. Koldychev
S. Sivabalan
Cisco Systems, Inc.
T. Saad
V. Beeram
Juniper Networks, Inc.
H. Bidgoli
Nokia
B. Yadav
Ciena
October 30, 2019

PCEP Extensions for Signaling Multipath Information
draft-koldychev-pce-multipath-00

Abstract

This document introduces a mechanism to communicate multipath information in PCEP as a set of Explicit Route Objects (EROs). A special object is defined to carry per ERO attributes. This mechanism is applicable to SR-TE and RSVP-TE.

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 <http://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 May 2, 2020.

Copyright Notice

Copyright (c) 2019 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 (<http://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. Terminology	3
2.1. Terms and Abbreviations	3
3. Motivation	3
3.1. Signaling Multiple Segment-Lists of an SR Candidate-Path	4
3.2. Splitting of Requested Bandwidth	4
3.3. Providing Backup ERO for Protection	4
4. Protocol Extensions	4
4.1. Multipath Capability TLV	4
4.2. ERO Attributes Object	5
4.3. Multipath Weight TLV	6
4.4. Multipath Backup TLV	6
5. Operation	7
6. IANA Considerations	8
7. Security Considerations	8
8. Acknowledgement	8
9. References	8
9.1. Normative References	8
9.2. Informative References	9
Authors' Addresses	9

1. Introduction

Path Computation Element (PCE) Communication Protocol (PCEP) [RFC5440] enables the communication between a Path Computation Client (PCC) and a Path Control Element (PCE), or between two PCEs based on the PCE architecture [RFC4655].

PCEP Extensions for the Stateful PCE Model [RFC8231] describes a set of extensions to PCEP to enable active control of Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) tunnels. [RFC8281] describes the setup and teardown of PCE-initiated

LSPs under the active stateful PCE model, without the need for local configuration on the PCC, thus allowing for dynamic centralized control of a network.

PCEP Extensions for Segment Routing [I-D.ietf-pce-segment-routing] specifies extensions to the Path Computation Element Protocol (PCEP)

that allow a stateful PCE to compute and initiate Traffic Engineering (TE) paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

Segment Routing Policy for Traffic Engineering [I-D.ietf-spring-segment-routing-policy] details the concepts of SR Policy and approaches to steering traffic into an SR Policy. In particular, it describes the SR candidate-path as a collection of one or more Segment-Lists. The current PCEP standards only allow for signaling of one Segment-List per Candidate-Path. PCEP extension to support Segment Routing Policy Candidate Paths [I-D.barth-pce-segment-routing-policy-cp] specifically avoids defining how to signal multipath information, and states that this will be defined in another document (this one).

2. Terminology

In this document, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" are to be interpreted as described in BCP 14, RFC 2119 [RFC2119].

2.1. Terms and Abbreviations

The following terms are used in this document:

Endpoint:

The IPv4 or IPv6 endpoint address of the SR policy in question, as described in [I-D.ietf-spring-segment-routing-policy].

PCEP Tunnel:

The object identified by the PLSP-ID, as per [I-D.koldychev-pce-operational].

Tunnel Instance:

The object identified by the LSP-Identifiers TLV, as per [I-D.koldychev-pce-operational].

3. Motivation

This extension is motivated by the use-cases described below.

3.1. Signaling Multiple Segment-Lists of an SR Candidate-Path

The Candidate-Path of an SR Policy corresponds to a PCEP Tunnel, see [I-D.barth-pce-segment-routing-policy-cp]. Each Candidate-Path can contain multiple Segment-Lists and each Segment-List is encoded by one SR-ERO object. However, each Tunnel Instance can contain only a single ERO object, which prevents us from encoding multiple Segment-Lists within the same SR Candidate-Path.

With the help of the protocol extensions defined in this document, this limitation is overcome.

3.2. Splitting of Requested Bandwidth

A PCC may request a path with 100 Gbit of bandwidth, but all links in the network have only 50 Gbit capacity. The PCE can return two paths, that can each carry 50 Gbit. The PCC can then equally or unequally split the incoming 100 Gbit of traffic among the two 50 Gbit paths. Section 4.3 introduces a new TLV that carries the ERO path weight that allows distributing of incoming traffic on to the multiple ERO path(s).

3.3. Providing Backup ERO for Protection

It is desirable for the PCE to compute and signal to the PCC a backup ERO path that is used to protect a primary ERO path. In this case, an indication specify a primary or backup.

When multipath is used, a backup ERO path may protect one or more primary ERO path. For this reason, a primary and backup path identifiers are needed to indicate which backup ERO path(s) protect which primary ERO path(s). Section 4.4 introduces a new TLV that carries the required information.

4. Protocol Extensions

4.1. Multipath Capability TLV

We define the MULTIPATH-CAP TLV that MAY be present in the OPEN object and/or the LSP object. The purpose of this TLV is two-fold:

1. From PCC: it tells how many multipaths the PCC can install in forwarding.
2. From PCE: it tells that the PCE supports this standard and how many multipaths the PCE can compute.

Only the first instance of this TLV can be processed, subsequent instances SHOULD be ignored.

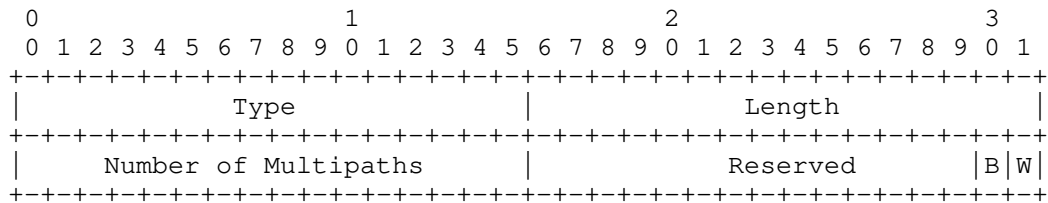


Figure 1: MULTIPATH-CAP TLV format

Type: TBD1 for "MULTIPATH-CAP" TLV.

Length: 4.

Number of Multipaths: the maximum number of multipaths that a PCE can return. The value 0 indicates unlimited number.

B-flag: whether MULTIPATH-BACKUP-TLV is supported.

W-flag: whether MULTIPATH-WEIGHT-TLV is supported.

Reserved: zero on transmit, ignore on receipt.

4.2. ERO Attributes Object

We define the ERO-ATTRIB object that is used to carry per-ERO information and to act as a separator between several ERO objects. The ERO-ATTRIB object always precedes the ERO that it applies to. If multiple ERO objects are present, then each ERO object MUST be preceded by an ERO-ATTRIB object that describes it.

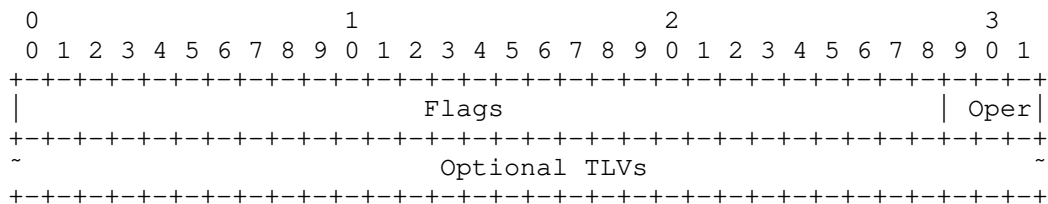


Figure 2: ERO-ATTRIB object format

Flags: to be extended in the future.

Oper: operational state of the ERO, same values as the identically named field in the LSP object.

4.3. Multipath Weight TLV

We define the MULTIPATH-WEIGHT TLV that MAY be present in the ERO_ATTRIBUTES object.

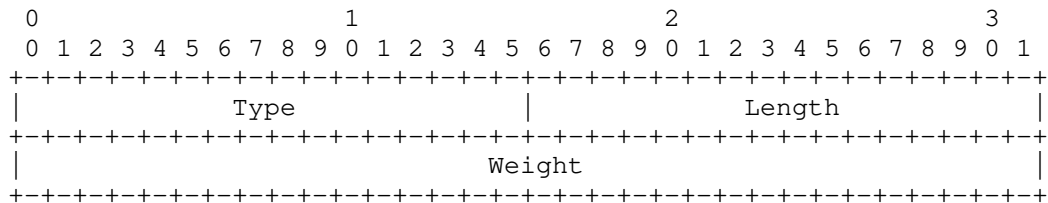


Figure 3: MULTIPATH-WEIGHT TLV format

Type: TBD2 for "MULTIPATH-WEIGHT" TLV.

Length: 4.

Weight: weight of this path within the multipath, if W-ECMP is desired. The fraction of flows a specific ERO carries is derived from the ratio of its weight to the sum of all other multipath ERO weights.

4.4. Multipath Backup TLV

This document introduces a new MULTIPATH-BACKUP TLV that is optional and can be present in the ERO_ATTRIBUTES object.

This TLV is used to indicate the presence of a backup ERO path that is used for protection in case of failure of the primary ERO path. The format of the MULTIPATH-BACKUP TLV is:

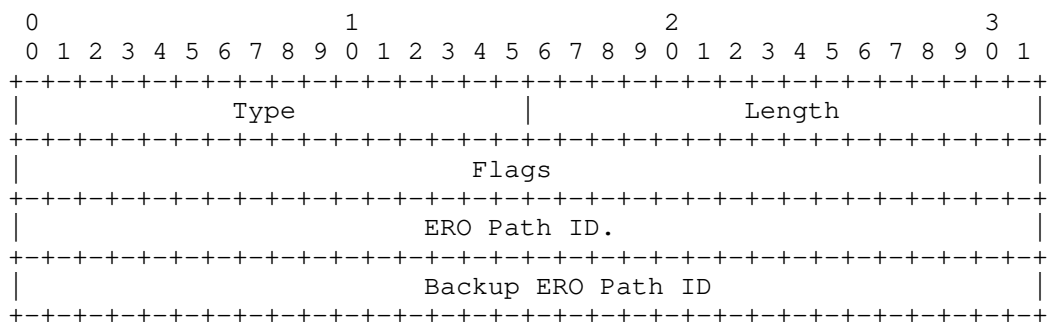


Figure 4: MULTIPATH-BACKUP TLV format

Type: TBD3 for "MULTIPATH-BACKUP" TLV

Length: 8

Flags:

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|P|B|F|           Reserved                                         |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

P: If set, indicates the ERO is for a primary path

B: If set, indicates the ERO is for a backup path

F: If set, indicates this primary ERO is also protected. The backup ERO Path ID indicates the ERO of the backup path.

ERO Path ID: an identifier that identifies a primary path in the set of ERO(s)

Backup ERO Path ID: an identifier that identifies the backup path ERO in the set of ERO(s)

5. Operation

When the PCC wants to indicate to the PCE that it wants to get multipaths instead of a single path, it can do one or both of the following:

1. Send the MULTIPATH-CAP TLV in the OPEN object during session establishment. This applies to all PCEP Tunnels on the PCC, unless overridden by PCEP Tunnel specific information.
2. Send the MULTIPATH-CAP TLV in the LSP object for a particular PCEP Tunnel in the PCRep message. This applies to the specified PCEP Tunnel and overrides the information from the OPEN object.

When PCE computes the path for a PCEP Tunnel, it MUST NOT return more multipaths than the corresponding value of "Number of Multipaths" from the MULTIPATH-CAP TLV. If this TLV is absent (from both OPEN and LSP objects), then the "Number of Multipaths" is assumed to be 1.

If the PCE supports this standard, then it MUST include the MULTIPATH-CAP TLV in the OPEN object. This tells the PCC that it can report multiple ERO objects to this PCE. If the PCE does not include the MULTIPATH-CAP TLV in the OPEN object, then the PCC MUST assume that the PCE does not support this standard and fall back to reporting only a single ERO.

6. IANA Considerations

IANA is requested to make the assignment of a new value for the existing "PCEP TLV Type Indicators" registry as follows:

TLV Type Value	TLV Name	Reference
TBD1	MULTIPATH-CAP	This document
TBD2	MULTIPATH-WEIGHT	This document
TBD3	MULTIPATH-BACKUP	This document

7. Security Considerations

None at this time.

8. Acknowledgement

Thanks to Dhruv Dhody for ideas and discussion.

9. References

9.1. Normative References

- [I-D.barth-pce-segment-routing-policy-cp]
Koldychev, M., Sivabalan, S., Barth, C., and C. Li, "PCEP extension to support Segment Routing Policy Candidate Paths", draft-barth-pce-segment-routing-policy-cp-03 (work in progress), July 2019.
- [I-D.ietf-pce-segment-routing]
Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "PCEP Extensions for Segment Routing", draft-ietf-pce-segment-routing-16 (work in progress), March 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.koldychev-pce-operational]

Koldychev, M., Sivabalan, S., Negi, M., Achaval, D., and H. Kotni, "PCEP Operational Clarification", draft-koldychev-pce-operational-00 (work in progress), July 2019.

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

[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>>.

[RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.

9.2. Informative References

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.

Authors' Addresses

Mike Koldychev
Cisco Systems, Inc.

Email: mkoldych@cisco.com

Siva Sivabalan
Cisco Systems, Inc.

Email: msiva@cisco.com

Tarek Saad
Juniper Networks, Inc.

Email: tsaad@juniper.net

Vishnu Pavan Beeram
Juniper Networks, Inc.

Email: vbeeram@juniper.net

Hooman Bidgoli
Nokia

Email: hooman.bidgoli@nokia.com

Bhupendra Yadav
Ciena

Email: byadav@ciena.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 20, 2021

M. Koldychev
Cisco Systems, Inc.
S. Sivabalan
Ciena Corporation
T. Saad
V. Beeram
Juniper Networks, Inc.
H. Bidgoli
Nokia
B. Yadav
Ciena
S. Peng
Huawei Technologies
February 16, 2021

PCEP Extensions for Signaling Multipath Information
draft-koldychev-pce-multipath-05

Abstract

Current PCEP standards allow only one intended and/or actual path to be present in a PCEP report or update. Applications that require multipath support such as SR Policy require an extension to allow signaling multiple intended and/or actual paths within a single PCEP message. This document introduces such an extension. Encoding of multiple intended and/or actual paths is done by encoding multiple Explicit Route Objects (EROs) and/or multiple Record Route Objects (RROs). A special separator object is defined in this document, to facilitate this. This mechanism is applicable to SR-TE and RSVP-TE and is dataplane agnostic.

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 August 20, 2021.

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	3
2. Terminology	3
2.1. Terms and Abbreviations	4
3. Motivation	4
3.1. Signaling Multiple Segment-Lists of an SR Candidate-Path	4
3.2. Splitting of Requested Bandwidth	4
3.3. Providing Backup path for Protection	4
4. Protocol Extensions	5
4.1. Multipath Capability TLV	5
4.2. Path Attributes Object	6
4.3. Multipath Weight TLV	6
4.4. Multipath Backup TLV	7
4.5. Composite Candidate Path	8
5. Operation	9
5.1. Signaling Multiple Paths for Loadbalancing	10
5.2. Signaling Multiple Paths for Protection	10
6. PCEP Message Extensions	11
7. Examples	11
7.1. SR Policy Candidate-Path with Multiple Segment-Lists . .	11
7.2. Two Primary Paths Protected by One Backup Path	13
7.3. Composite Candidate Path	13
8. IANA Considerations	14
8.1. PCEP Object	14
8.2. PCEP TLV	14
8.3. PCEP-Error Object	14
8.4. Flags in the Multipath Capability TLV	15
8.5. Flags in the Path Attribute Object	15
8.6. Flags in the Multipath Backup TLV	16
9. Security Considerations	16
10. Acknowledgement	16
11. Contributors	16

12. References	16
12.1. Normative References	16
12.2. Informative References	17
Authors' Addresses	18

1. Introduction

Path Computation Element (PCE) Communication Protocol (PCEP) [RFC5440] enables the communication between a Path Computation Client (PCC) and a Path Control Element (PCE), or between two PCEs based on the PCE architecture [RFC4655].

PCEP Extensions for the Stateful PCE Model [RFC8231] describes a set of extensions to PCEP that enable active control of Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) tunnels. [RFC8281] describes the setup and teardown of PCE-initiated LSPs under the active stateful PCE model, without the need for local configuration on the PCC, thus allowing for dynamic centralized control of a network.

PCEP Extensions for Segment Routing [RFC8664] specifies extensions to the Path Computation Element Protocol (PCEP) that allow a stateful PCE to compute and initiate Traffic Engineering (TE) paths, as well as for a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

Segment Routing Policy for Traffic Engineering [I-D.ietf-spring-segment-routing-policy] details the concepts of SR Policy and approaches to steering traffic into an SR Policy. In particular, it describes the SR candidate-path as a collection of one or more Segment-Lists. The current PCEP standards only allow for signaling of one Segment-List per Candidate-Path. PCEP extension to support Segment Routing Policy Candidate Paths [I-D.ietf-pce-segment-routing-policy-cp] specifically avoids defining how to signal multipath information, and states that this will be defined in another document.

This document defines the required extensions that allow the signaling of multipath information via PCEP.

2. Terminology

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.

2.1. Terms and Abbreviations

The following terms are used in this document:

PCEP Tunnel:

The object identified by the PLSP-ID, see [I-D.koldychev-pce-operational] for more details.

3. Motivation

This extension is motivated by the use-cases described below.

3.1. Signaling Multiple Segment-Lists of an SR Candidate-Path

The Candidate-Path of an SR Policy is the unit of report/update in PCEP, see [I-D.ietf-pce-segment-routing-policy-cp]. Each Candidate-Path can contain multiple Segment-Lists and each Segment-List is encoded by one ERO. However, each PCEP LSP can contain only a single ERO (containing multiple SR-ERO subobject), which prevents us from encoding multiple Segment-Lists within the same SR Candidate-Path.

With the help of the protocol extensions defined in this document, this limitation is overcome.

3.2. Splitting of Requested Bandwidth

A PCC may request a path with 80 Gbps of bandwidth, but all links in the network have only 50 Gbps capacity. The PCE can return two paths, that can together carry 80 Gbps. The PCC can then equally or unequally split the incoming 80 Gbps of traffic among the two paths. Section 4.3 introduces a new TLV that carries the path weight that allows for distribution of incoming traffic on to the multiple paths.

3.3. Providing Backup path for Protection

It is desirable for the PCE to compute and signal to the PCC a backup path that is used to protect a primary path within the multipaths in a given LSP.

Note that [RFC8745] specify the Path Protection association among LSPs. The use of [RFC8745] with multipath is out of scope of this document and is for future study.

When multipath is used, a backup path may protect one or more primary paths. For this reason, primary and backup path identifiers are needed to indicate which backup path(s) protect which primary

path(s). Section 4.4 introduces a new TLV that carries the required information.

4. Protocol Extensions

4.1. Multipath Capability TLV

We define the MULTIPATH-CAP TLV that MAY be present in the OPEN object and/or the LSP object. The purpose of this TLV is two-fold:

1. From PCC: it tells how many multipaths per PCEP Tunnel, the PCC can install in forwarding.
2. From PCE: it tells that the PCE supports this standard and how many multipaths per PCEP Tunnel, the PCE can compute.

Only the first instance of this TLV can be processed, subsequent instances SHOULD be ignored.

Section 5 specify the usage of this TLV with Open message (within the OPEN object) and other PCEP messages (within the LSP object).

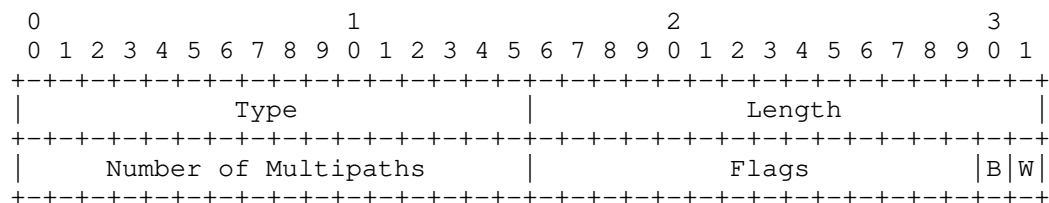


Figure 1: MULTIPATH-CAP TLV format

Type: TBD1 for "MULTIPATH-CAP" TLV.

Length: 4.

Number of Multipaths: the maximum number of multipaths per PCEP Tunnel. The value 0 indicates unlimited number.

Flags: Following bits are defined:

W-flag: whether MULTIPATH-WEIGHT-TLV is supported.

B-flag: whether MULTIPATH-BACKUP-TLV is supported.

Unassigned bits are for future use. They MUST be set to 0 on transmission and MUST be ignored on receipt.

4.2. Path Attributes Object

We define the PATH-ATTRIB object that is used to carry per-path information and to act as a separator between several ERO/RRO objects in the <intended-path>/<actual-path> RBNF element. The PATH-ATTRIB object always precedes the ERO/RRO that it applies to. If multiple ERO/RRO objects are present, then each ERO/RRO object MUST be preceded by an PATH-ATTRIB object that describes it.

The PATH-ATTRIB Object-Class value is TBD2.

The PATH-ATTRIB Object-Type value is 1.

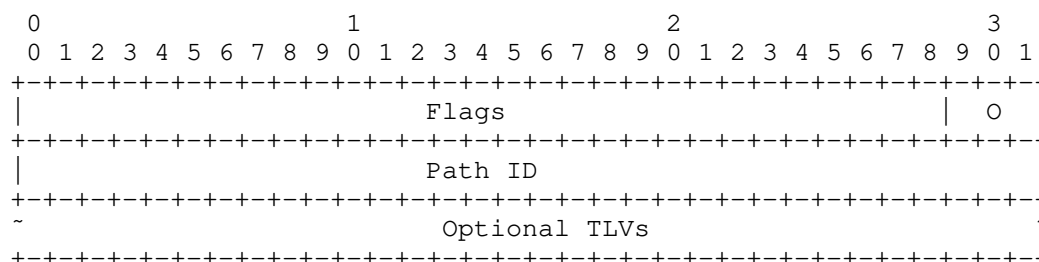


Figure 2: PATH-ATTRIB object format

Flags (32-bits): Following bits are assigned -

0 (Operational - 3 bits): operational state of the path, same values as the identically named field in the LSP object {{RFC8231}}.

Unassigned bits are for future use. They MUST be set to 0 on transmission and MUST be ignored on receipt.

Path ID: 4-octet identifier that identifies a path in the set of multiple paths. It uniquely identifies a path (encoded in the ERO/RRO) within the set of multiple paths under the PCEP LSP. Once a path changes, a new Path ID is assigned.

TLVs that may be included in the PATH-ATTRIB object are described in the following sections. Other optional TLVs could be defined by future documents to be included within the PATH-ATTRIB object body.

4.3. Multipath Weight TLV

We define the MULTIPATH-WEIGHT TLV that MAY be present in the PATH-ATTRIB object.

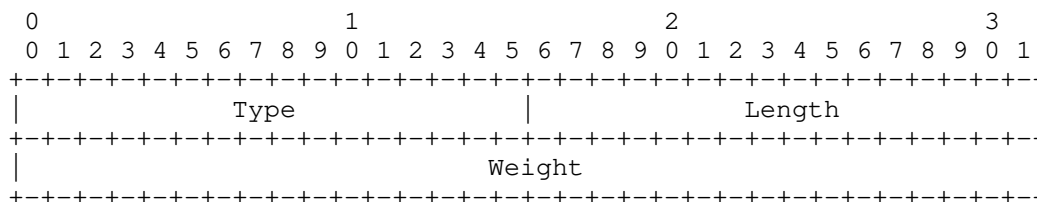


Figure 3: MULTIPATH-WEIGHT TLV format

Type: TBD3 for "MULTIPATH-WEIGHT" TLV.

Length: 4.

Weight: weight of this path within the multipath, if W-ECMP is desired. The fraction of flows a specific ERO/RRO carries is derived from the ratio of its weight to the sum of all other multipath ERO/RRO weights.

When the MULTIPATH-WEIGHT TLV is absent from the PATH-ATTRIB object, or the PATH-ATTRIB object is absent from the <intended-path>/<actual-path>, then the Weight of the corresponding path is taken to be "1".

4.4. Multipath Backup TLV

This document introduces a new MULTIPATH-BACKUP TLV that is optional and can be present in the PATH-ATTRIB object.

This TLV is used to indicate the presence of a backup path that is used for protection in case of failure of the primary path. The format of the MULTIPATH-BACKUP TLV is:

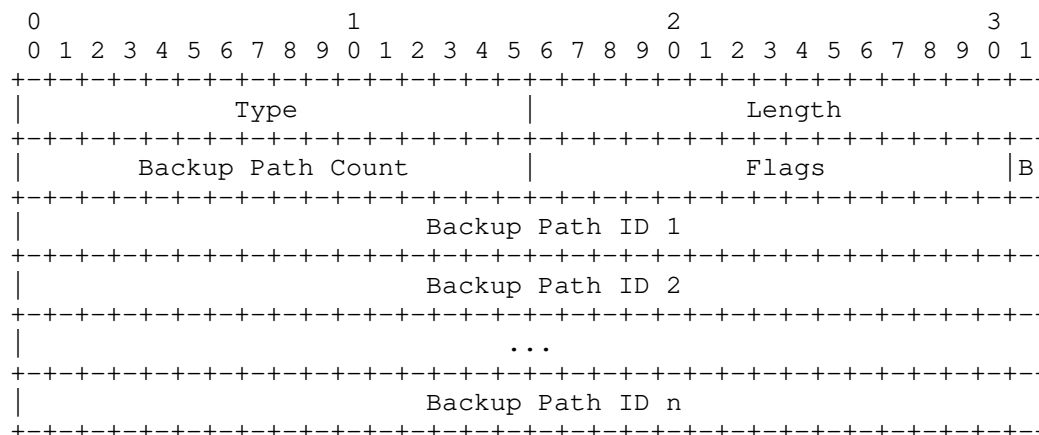


Figure 4: MULTIPATH-BACKUP TLV format

Type: TBD4 for "MULTIPATH-BACKUP" TLV

Length: $4 + (N * 4)$ (where N is the Backup Path Count)

Backup Path Count: Number of backup path(s).

Flags (16 bits): a flag field. Currently a single flag "B bit" is defined.

Unused flags MUST be set to zero while sending and ignored on receipt.

B: If set, indicates a pure backup path. This is a path that only carries rerouted traffic after the protected path fails. If this flag is not set, or if the MULTIPATH-BACKUP TLV is absent, then the path is assumed to be primary that carries normal traffic.

Backup Path ID(s): a series of 4-octet identifier(s) that identify the backup path(s) in the set that protect this primary path.

4.5. Composite Candidate Path

SR Policy Architecture [I-D.ietf-spring-segment-routing-policy] defines the concept of a Composite Candidate Path. Unlike a Non-Composite Candidate Path, which contains Segment Lists, the Composite Candidate Path contains Colors of other policies. The traffic that is steered into a Composite Candidate Path is split among the policies that are identified by the Colors contained in the Composite Candidate Path. The split can be either ECMP or UCMP by adjusting

the weight of each color in the Composite Candidate Path, in the same manner as the weight of each Segment List in the Non-Composite Candidate Path is adjusted.

To signal the Composite Candidate Path, we make use of the COLOR TLV, defined in [I-D.peng-pce-te-constraints]. For a Composite Candidate Path, the COLOR TLV is included in the PATH-ATTRIB Object, thus allowing each Composite Candidate Path to do ECMP/UCMP among SR Policies or Tunnels identified by its constituent Colors. Only one COLOR TLV SHOULD be included into the PATH-ATTRIB object. If multiple COLOR TLVs are contained in the PATH-ATTRIB object, only the first one MUST be processed and the others SHOULD be ignored.

An empty SR-ERO/SR-RRO object MUST be included as per the existing RBNF, i.e., SR-ERO/SR-RRO MUST contain no sub-objects. If the head-end receives a non-empty SR-ERO/SR-RRO, then it MUST send PCErr message with Error-Type 19 ("Invalid Operation") and Error-Value = TBD8 ("Non-empty path").

See Section 7.3 for an example of the encoding.

5. Operation

When the PCC wants to indicate to the PCE that it wants to get multipaths for a PCEP Tunnel, instead of a single path, it can do (1) or both (1) and (2) of the following:

(1) Send the MULTIPATH-CAP TLV in the OPEN object during session establishment. This applies to all PCEP Tunnels on the PCC, unless overridden by PCEP Tunnel specific information.

(2) Additionally send the MULTIPATH-CAP TLV in the LSP object for a particular PCEP Tunnel in the PCRpt or PCReq message. This applies to the specified PCEP Tunnel and overrides the information from the OPEN object.

When PCE computes the path for a PCEP Tunnel, it MUST NOT return more multipaths than the corresponding value of "Number of Multipaths" from the MULTIPATH-CAP TLV. If this TLV is absent (from both OPEN and LSP objects), then the "Number of Multipaths" is assumed to be 1.

If the PCE supports this standard, then it MUST include the MULTIPATH-CAP TLV in the OPEN object. This tells the PCC that it can report multiple ERO/RRO objects per PCEP Tunnel to this PCE. If the PCE does not include the MULTIPATH-CAP TLV in the OPEN object, then the PCC MUST assume that the PCE does not support this standard and fall back to reporting only a single ERO/RRO. The PCE MUST NOT

include MULTIPATH-CAP TLV in the LSP object in any other PCEP message towards the PCC and the PCC MUST ignore it if received.

The Path ID of each ERO/RRO MUST be unique within that LSP. If a PCEP speaker detects that there are two paths with the same Path ID, then the PCEP speaker SHOULD send PCError message with Error-Type = 1 ("Reception of an invalid object") and Error-Value = TBD5 ("Conflicting Path ID").

5.1. Signaling Multiple Paths for Loadbalancing

The PATH-ATTRIB object can be used to signal multiple path(s) and indicate (un)equal loadbalancing amongst the set of multipaths. In this case, the PATH-ATTRIB is populated for each ERO as follows:

1. The PCE assigns a unique Path ID to each ERO path and populates it inside the PATH-ATTRIB object. The Path ID is unique within the context of a PLSP or PCEP Tunnel.
2. The MULTIPATH-WEIGHT TLV MAY be carried inside the PATH-ATTRIB object. A weight is populated to reflect the relative loadshare that is to be carried by the path. If the MULTIPATH-WEIGHT is not carried inside a PATH-ATTRIB object, the default weight 1 MUST be assumed when computing the loadshare.
3. The fraction of flows carried by a specific primary path is derived from the ratio of its weight to the sum of all other multipath weights.

5.2. Signaling Multiple Paths for Protection

The PATH-ATTRIB object can be used to describe a set of backup path(s) protecting a primary path within a PCEP Tunnel. In this case, the PATH-ATTRIB is populated for each ERO as follows:

1. The PCE assigns a unique Path ID to each ERO path and populates it inside the PATH-ATTRIB object. The Path ID is unique within the context of a PLSP or PCEP Tunnel.
2. The MULTIPATH-BACKUP TLV MUST be added inside the PATH-ATTRIB object for each ERO that is protected. The backup path ID(s) are populated in the MULTIPATH-BACKUP TLV to reflect the set of backup path(s) protecting the primary path. The Length field and Backup Path Number in the MULTIPATH-BACKUP are updated according to the number of backup path ID(s) included.
3. The MULTIPATH-BACKUP TLV MAY be added inside the PATH-ATTRIB object for each ERO that is unprotected. In this case,

MULTIPATH-BACKUP does not carry any backup path IDs in the TLV. If the path acts as a pure backup - i.e. the path only carries rerouted traffic after the protected path(s) fail- then the B flag MUST be set.

Note that if a given path has the B-flag set, then there MUST be some other path within the same LSP that uses the given path as a backup. If this condition is violated, then the PCEP speaker SHOULD send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD6 ("No primary path for pure backup").

Note that a given PCC may not support certain backup combinations, such as a backup path that is itself protected by another backup path, etc. If a PCC is not able to implement a requested backup scenario, the PCC SHOULD send a PCErr message with Error-Type = 19 ("Invalid Operation") and Error-Value = TBD7 ("Not supported path backup").

6. PCEP Message Extensions

The RBNF of PCReq, PCRep, PCRpt, PCUpd and PCInit messages currently use a combination of <intended-path> and/or <actual-path>. As specified in Section 6.1 of [RFC8231], <intended-path> is represented by the ERO object and <actual-path> is represented by the RRO object:

<intended-path> ::= <ERO>

<actual-path> ::= <RRO>

In this standard, we extend these two elements to allow multiple ERO/RRO objects to be present in the <intended-path>/<actual-path>:

<intended-path> ::= (<ERO> |
 (<PATH-ATTRIB><ERO>)
 [<intended-path>])

<actual-path> ::= (<RRO> |
 (<PATH-ATTRIB><RRO>)
 [<actual-path>])

7. Examples

7.1. SR Policy Candidate-Path with Multiple Segment-Lists

Consider the following sample SR Policy, taken from [I-D.ietf-spring-segment-routing-policy].

```

    SR policy POL1 <headend, color, endpoint>
      Candidate-path CP1 <protocol-origin = 20, originator =
100:1.1.1.1, discriminator = 1>
        Preference 200
        Weight W1, SID-List1 <SID11...SID1i>
        Weight W2, SID-List2 <SID21...SID2j>
      Candidate-path CP2 <protocol-origin = 20, originator =
100:2.2.2.2, discriminator = 2>
        Preference 100
        Weight W3, SID-List3 <SID31...SID3i>
        Weight W4, SID-List4 <SID41...SID4j>

```

As specified in [I-D.ietf-pce-segment-routing-policy-cp], CP1 and CP2 are signaled as separate state-report elements and each has a unique PLSP-ID, assigned by the PCC. Let us assign PLSP-ID 100 to CP1 and PLSP-ID 200 to CP2.

The state-report for CP1 can be encoded as:

```

<state-report> =
  <LSP PLSP_ID=100>
  <ASSOCIATION>
  <END-POINT>
  <PATH-ATTRIB Path_ID=1 <WEIGHT-TLV Weight=W1>>
  <ERO SID-List1>
  <PATH-ATTRIB Path_ID=2 <WEIGHT-TLV Weight=W2>>
  <ERO SID-List2>

```

The state-report for CP2 can be encoded as:

```

<state-report> =
  <LSP PLSP_ID=200>
  <ASSOCIATION>
  <END-POINT>
  <PATH-ATTRIB Path_ID=1 <WEIGHT-TLV Weight=W3>>
  <ERO SID-List3>
  <PATH-ATTRIB Path_ID=2 <WEIGHT-TLV Weight=W4>>
  <ERO SID-List4>

```

The above sample state-report elements only specify the minimum mandatory objects, of course other objects like SRP, LSPA, METRIC, etc., are allowed to be inserted.

Note that the syntax

```

<PATH-ATTRIB Path_ID=1 <WEIGHT-TLV Weight=W1>>

```


, simply means that this is PATH-ATTRIB object with Path ID field set to "1" and with a MULTIPATH-WEIGHT TLV carrying weight of "W1".

7.2. Two Primary Paths Protected by One Backup Path

Suppose there are 3 paths: A, B, C. Where A,B are primary and C is to be used only when A or B fail. Suppose the Path IDs for A, B, C are respectively 1, 2, 3. This would be encoded in a state-report as:

```
<state-report> =
  <LSP>
  <ASSOCIATION>
  <END-POINT>
  <PATH-ATTRIB Path_ID=1 <BACKUP-TLV B=0, Backup_Paths=[3]>>
  <ERO A>
  <PATH-ATTRIB Path_ID=2 <BACKUP-TLV B=0, Backup_Paths=[3]>>
  <ERO B>
  <PATH-ATTRIB Path_ID=3 <BACKUP-TLV B=1, Backup_Paths=[]>>
  <ERO C>
```

Note that the syntax

```
<PATH-ATTRIB Path_ID=1 <BACKUP-TLV B=0, Backup_Paths=[3]>>
```

, simply means that this is PATH-ATTRIB object with Path ID field set to "1" and with a MULTIPATH-BACKUP TLV that has B-flag cleared and contains a single backup path with Backup Path ID of 3.

7.3. Composite Candidate Path

Consider the following Composite Candidate Path, taken from [I-D.ietf-spring-segment-routing-policy].

```
SR policy POL100 <headend = H1, color = 100, endpoint = E1>
  Candidate-path CP1 <protocol-origin = 20, originator =
100:1.1.1.1, discriminator = 1>
    Preference 200
    Weight W1, SR policy <color = 1>
    Weight W2, SR policy <color = 2>
```

This is signaled in PCEP as:

```

<LSP PLSP_ID=100>
<ASSOCIATION>
<END-POINT>
<PATH-ATTRIB Path_ID=1 <WEIGHT-TLV Weight=W1> <COLOR-TLV Color=1>>
<SR-ERO (empty)>
<PATH-ATTRIB Path_ID=2 <WEIGHT-TLV Weight=W2> <COLOR-TLV Color=2>>
<SR-ERO (empty)>

```

8. IANA Considerations

8.1. PCEP Object

IANA is requested to make the assignment of a new value for the existing "PCEP Objects" registry as follows:

Object-Class Value	Name	Object-Type Value	Reference
TBD2	PATH-ATTRIB	1	This document

8.2. PCEP TLV

IANA is requested to make the assignment of a new value for the existing "PCEP TLV Type Indicators" registry as follows:

TLV Type Value	TLV Name	Reference
TBD1	MULTIPATH-CAP	This document
TBD3	MULTIPATH-WEIGHT	This document
TBD4	MULTIPATH-BACKUP	This document

8.3. PCEP-Error Object

IANA is requested to make the assignment of a new value for the existing "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Error-Value	Reference
10	TBD5 - Conflicting Path ID	This document
10	TBD6 - No primary path for pure backup	This document
19	TBD7 - Not supported path backup	This document
19	TBD8 - Non-empty path	This document

8.4. Flags in the Multipath Capability TLV

IANA is requested to create a new sub-registry to manage the Flag field of the MULTIPATH-CAP TLV, called "Flags in MULTIPATH-CAP TLV".

Following bits are defined:

Bit	Description	Reference
0-13	Unassigned	This document
14	B-flag: Backup support	This document
15	W-flag: Weighted ECMP support	This document

8.5. Flags in the Path Attribute Object

IANA is requested to create a new sub-registry to manage the Flag field of the PATH-ATTRIBUTE object, called "Flags in PATH-ATTRIBUTE Object".

Following bits are defined:

Bit	Description	Reference
0-12	Unassigned	This document
13-15	O-flag: Operational state	This document

8.6. Flags in the Multipath Backup TLV

IANA is requested to create a new sub-registry to manage the Flag field of the MULTIPATH-BACKUP TLV, called "Flags in MULTIPATH-BACKUP TLV".

Following bits are defined:

Bit	Description	Reference
0-14	Unassigned	This document
15	B-flag: Pure backup	This document

9. Security Considerations

None at this time.

10. Acknowledgement

Thanks to Dhruv Dhody for ideas and discussion.

11. Contributors

Andrew Stone
Nokia

Email: andrew.stone@nokia.com

12. References

12.1. Normative References

[I-D.ietf-pce-segment-routing-policy-cp]
Koldychev, M., Sivabalan, S., Barth, C., Peng, S., and H. Bidgoli, "PCEP extension to support Segment Routing Policy Candidate Paths", draft-ietf-pce-segment-routing-policy-cp-02 (work in progress), January 2021.

[I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-09 (work in progress), November 2020.

- [I-D.koldychev-pce-operational]
Koldychev, M., Sivabalan, S., Negi, M., Achaval, D., and H. Kotni, "PCEP Operational Clarification", draft-koldychev-pce-operational-02 (work in progress), August 2020.
- [I-D.peng-pce-te-constraints]
Peng, S., Xiong, Q., and F. Qin, "PCE TE Constraints for Network Slicing", draft-peng-pce-te-constraints-04 (work in progress), August 2020.
- [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>>.
- [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>>.
- [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>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <<https://www.rfc-editor.org/info/rfc8664>>.

12.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC8745] Ananthakrishnan, H., Sivabalan, S., Barth, C., Minei, I., and M. Negi, "Path Computation Element Communication Protocol (PCEP) Extensions for Associating Working and Protection Label Switched Paths (LSPs) with Stateful PCE", RFC 8745, DOI 10.17487/RFC8745, March 2020, <<https://www.rfc-editor.org/info/rfc8745>>.

Authors' Addresses

Mike Koldychev
Cisco Systems, Inc.

Email: mkoldych@cisco.com

Siva Sivabalan
Ciena Corporation

Email: ssivabal@ciena.com

Tarek Saad
Juniper Networks, Inc.

Email: tsaad@juniper.net

Vishnu Pavan Beeram
Juniper Networks, Inc.

Email: vbeeram@juniper.net

Hooman Bidgoli
Nokia

Email: hooman.bidgoli@nokia.com

Bhupendra Yadav
Ciena

Email: byadav@ciena.com

Shuping Peng
Huawei Technologies

Email: pengshuping@huawei.com

PCE
Internet-Draft
Intended status: Standards Track
Expires: March 14, 2020

S. Peng
Q. Xiong
ZTE Corporation
September 11, 2019

PCEP Extension for SR-MPLS Entropy Label Position
draft-peng-pce-entropy-label-position-01

Abstract

This document proposes a set of extensions for PCEP to configure the entropy label position for SR-MPLS networks.

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 March 14, 2020.

Copyright Notice

Copyright (c) 2019 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. Conventions used in this document	3
2.1. Terminology	3
2.2. Requirements Language	3
3. PCEP Extensions	4
3.1. The OPEN Object	4
3.2. The LSP Object	4
3.2.1. The LSP-EXTENDED-FLAG TLV	5
3.3. The ERO Object	5
4. Operations	6
5. Security Considerations	6
6. Acknowledgements	6
7. IANA Considerations	6
7.1. New SR PCE Capability Flag Registry	6
7.2. New LSP Flag Registry	6
7.3. New SR-ERO Flag Registry	7
8. Normative References	7
Authors' Addresses	8

1. Introduction

[RFC5440] describes the Path Computation Element Protocol (PCEP) which is used between a Path Computation Element (PCE) and a Path Computation Client (PCC) (or other PCE) to enable computation of Multi-protocol Label Switching (MPLS) for Traffic Engineering Label Switched Path (TE LSP). PCEP Extensions for the Stateful PCE Model [RFC8231] describes a set of extensions to PCEP to enable active control of MPLS-TE and Generalized MPLS (GMPLS) tunnels. [RFC8281] describes the setup and teardown of PCE-initiated LSPs under the active stateful PCE model, without the need for local configuration on the PCC, thus allowing for dynamic centralized control of a network.

Segment Routing (SR) leverages the source routing paradigm. Segment Routing can be instantiated on MPLS data plane which is referred to as SR-MPLS [I-D.ietf-spring-segment-routing-mpls]. SR-MPLS leverages the MPLS label stack to construct the SR path. PCEP Extensions for Segment Routing [I-D.ietf-pce-segment-routing] specifies extensions to the PCEP that allow a stateful PCE to compute and initiate TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

Entropy label (EL) [RFC6790] is a technique used in the MPLS data plane to provide entropy for load-balancing. Entropy Label Indicator (ELI) can be immediately preceding an EL in the MPLS label stack. The idea behind the EL is that the ingress router computes a hash

based on several fields from a given packet and places the result in an additional label, named "entropy label". Then, this entropy label can be used as part of the hash keys used by an LSR. Using the entropy label as part of the hash keys reduces the need for deep packet inspection in the LSR while keeping a good level of entropy in the load-balancing. When the entropy label is used, the keys used in the hashing functions are still a local configuration matter and an LSR may use solely the entropy label or a combination of multiple fields from the incoming packet.

[I-D.ietf-mpls-spring-entropy-label] proposes to use entropy labels for SR-MPLS networks. The Entropy Readable Label Depth (ERLD) is defined as the number of labels which means that the router will perform load-balancing using the ELI/EL. An appropriate algorithm would consider the following goals:

- o a limited number of <ELI, EL> pairs should be inserted deeper in the label-stack.
- o the inserted position should be within the ERLD of most transit nodes.
- o a minimum number of <ELI, EL> to satisfy the above criteria.

In some cases, It is required for the controller (e.g. PCE) to perform the TE path computation as well as the Entropy Label Position (ELP), because the controller has the ERLD information of all nodes, especially for inter-domain scenarios. This document proposes a set of extensions for PCEP to configure the ELP information for SR-MPLS networks.

2. Conventions used in this document

2.1. Terminology

The terminology is defined as [RFC5440], [RFC6790], [I-D.ietf-pce-segment-routing] and [I-D.ietf-mpls-spring-entropy-label].

2.2. 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.

3. PCEP Extensions

3.1. The OPEN Object

As defined in [I-D.ietf-pce-segment-routing], PCEP speakers use SR PCE Capability sub-TLV to exchange information about their SR capability when PST=1 in the PST List of the PATH-SETUP-TYPE-CAPABILITY TLV carried in Open object. This document defined a new flag (E-flag) for SR PCE Capability sub-TLV as shown in Figure 1.

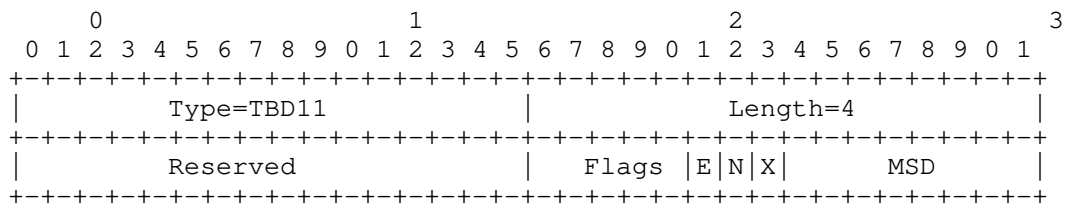


Figure 1: E-flag in SR-PCE-CAPABILITY sub-TLV

E (ELP Configuration is supported) : A PCC or PCE sets this flag bit to 1 carried in Open message to indicate that it supports the SR path with ELP configuration.

3.2. The LSP Object

The LSP Object is defined in Section 7.3 of [RFC8231]. This document defined a new flag (E-flag) for the LSP Object as Figure 2 shown:

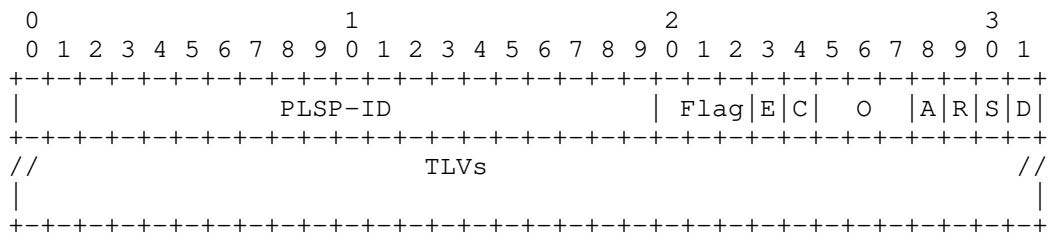


Figure 2: E-flag in LSP Object

E (Request for ELP Configuration) : If the bit is set to 1, it indicates that the PCC requests PCE to compute the SR path with ELP information. A PCE would also set this bit to 1 to indicate that the

ELP information is included by PCE and encoded in the PCRep, PCUpd or PCInitiate message.

3.2.1. The LSP-EXTENDED-FLAG TLV

As defined in [RFC8231], the length of LSP Object Flag field is 12 bits and it defined the value from bit 5 to bit 11. The bits from 1 to 3 are defined in [RFC8623], the bit value 4 is used in [RFC8281]. So all bits of the flag has been used and this document proposes to define a new LSP-EXTENDED-FLAG TLV for LSP object to extend the length of the flag as the Figure 3 shown.

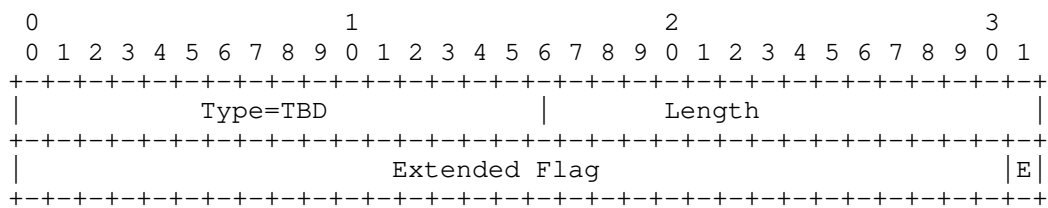


Figure 3: LSP-EXTENDED-FLAG TLV Format

The bit E has the same definition with section 3.2 and the other bits of the Extended flag can be used for other drafts in the future.

3.3. The ERO Object

SR-ERO subobject is used for SR-TE path which consists of one or more SIDs as defined in [I-D.ietf-pce-segment-routing]. This document defines a new flag (E-flag) for the SR-ERO subobject as Figure 3 shown:

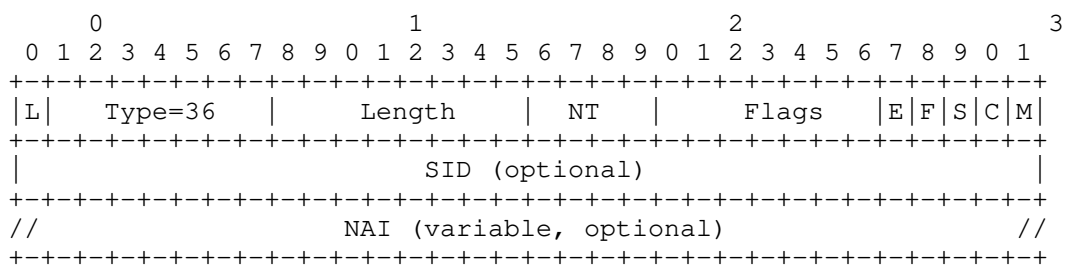


Figure 4: E-flag in SR-ERO subobject

E (ELP Configuration) : If this flag is set, it means that the position after this SR-ERO subobject is the position to insert <ELI, EL>, otherwise it cannot insert <ELI, EL> after this segment.

4. Operations

The SR path is initiated by PCE or PCC with PCReq, PCInitiated or PCUpd messages and the E bit is set to 1 in LSP object to request the ELP configuration. The SR-TE path being received by PCC with SR-ERO segment list, for example, <S1, S2, S3, S4, S5, S6>, especially S3 and S6 with E-flag set. It indicates that two <ELI, EL> pairs MUST be inserted into the label stack of the SR-TE forwarding entry, respectively after the label for S3 and label for S6. With EL information, the label stack for SR-MPLS would be <label1, label2, label3, ELI, EL, label4, label5, label6, ELI, EL>.

5. Security Considerations

TBA

6. Acknowledgements

TBA

7. IANA Considerations

7.1. New SR PCE Capability Flag Registry

SR PCE Capability TLV is defined in [I-D.ietf-pce-segment-routing], and the registry to manage the Flag field of the SR PCE Capability TLV is requested in [I-D.ietf-pce-segment-routing]. IANA is requested to make allocations from the registry, as follows:

Value	Name	Reference
TBD11	ELP Configuration is supported (E)	[this document]

Table 1

7.2. New LSP Flag Registry

[RFC8231] defines the LSP object; per that RFC, IANA created a registry to manage the value of the LSP object's Flag field. IANA is requested to make allocations from the registry, as follows:

Value	Name	Reference
TBD	Request for ELP Configuration (E)	[this document]
TBD	LSP-EXTENDED-FLAG TLV	[this document]

Table 2

7.3. New SR-ERO Flag Registry

SR-ERO subobject is defined in [I-D.ietf-pce-segment-routing], and the registry to manage the Flag field of SR-ERO is requested in [I-D.ietf-pce-segment-routing]. IANA is requested to make allocations from the registry, as follows:

Value	Name	Reference
36	ELP Configuration (E)	[this document]

Table 3

8. Normative References

- [I-D.ietf-mpls-spring-entropy-label]
Kini, S., Kompella, K., Sivabalan, S., Litkowski, S., Shakir, R., and J. Tantsura, "Entropy label for SPRING tunnels", draft-ietf-mpls-spring-entropy-label-12 (work in progress), July 2018.
- [I-D.ietf-pce-segment-routing]
Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "PCEP Extensions for Segment Routing", draft-ietf-pce-segment-routing-16 (work in progress), March 2019.
- [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.
- [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>>.
- [RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <<https://www.rfc-editor.org/info/rfc6790>>.
- [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>>.
- [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>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8623] Palle, U., Dhody, D., Tanaka, Y., and V. Beeram, "Stateful Path Computation Element (PCE) Protocol Extensions for Usage with Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 8623, DOI 10.17487/RFC8623, June 2019, <<https://www.rfc-editor.org/info/rfc8623>>.

Authors' Addresses

Shaofu Peng
ZTE Corporation
No.50 Software Avenue
Nanjing, Jiangsu 210012
China

Email: peng.shaofu@zte.com.cn

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan, Hubei 430223
China

Email: xiong.quan@zte.com.cn

PCE
Internet-Draft
Intended status: Standards Track
Expires: 3 September 2022

Q. Xiong
S. Peng
ZTE Corporation
F. Qin
China Mobile
March 2022

PCEP Extension for SR-MPLS Entropy Label Position
draft-peng-pce-entropy-label-position-07

Abstract

This document proposes a set of extensions for PCEP to configure the entropy label position for SR-MPLS networks.

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 2 September 2022.

Copyright Notice

Copyright (c) 2022 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	2
2. Conventions used in this document	3
2.1. Terminology	3
2.2. Requirements Language	3
3. Entropy Labels in SR-MPLS Scenario with PCE	3
4. PCEP Extensions	5
4.1. The OPEN Object	5
4.2. The LSP-EXTENDED-FLAG TLV	5
4.3. The SR-ERO Object	6
5. Operations	7
6. Security Considerations	7
7. Acknowledgements	7
8. IANA Considerations	7
8.1. New SR PCE Capability Flag Registry	7
8.2. New LSP-EXTENDED-FLAG Flag Registry	7
8.3. New SR-ERO Flag Registry	8
9. Normative References	8
Authors' Addresses	10

1. Introduction

[RFC5440] describes the Path Computation Element Protocol (PCEP) which is used between a Path Computation Element (PCE) and a Path Computation Client (PCC) (or other PCE) to enable computation of Multi-protocol Label Switching (MPLS) for Traffic Engineering Label Switched Path (TE LSP). PCEP Extensions for the Stateful PCE Model [RFC8231] describes a set of extensions to PCEP to enable active control of MPLS-TE and Generalized MPLS (GMPLS) tunnels. [RFC8281] describes the setup and teardown of PCE-initiated LSPs under the active stateful PCE model, without the need for local configuration on the PCC, thus allowing for dynamic centralized control of a network.

Segment Routing (SR) leverages the source routing paradigm. Segment Routing can be instantiated on MPLS data plane which is referred to as SR-MPLS [RFC8660]. SR-MPLS leverages the MPLS label stack to construct the SR path. PCEP Extensions for Segment Routing [RFC8664] specifies extensions to the PCEP that allow a stateful PCE to compute and initiate TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

Entropy label (EL) [RFC6790] is a technique used in the MPLS data plane to improve load-balancing. Entropy Label Indicator (ELI) can be immediately preceding an EL in the MPLS label stack. The idea behind the EL is that the ingress router computes a hash based on several fields from a given packet and places the result in an

additional label, named "entropy label". Then, this entropy label can be used as part of the hash keys used by an LSR. Using the entropy label as part of the hash keys reduces the need for deep packet inspection in the LSR while keeping a good level of entropy in the load-balancing. When the entropy label is used, the keys used in the hashing functions are still a local configuration matter and an LSR may use solely the entropy label or a combination of multiple fields from the incoming packet.

[RFC8662] proposes to use entropy labels for SR-MPLS networks and multiple <ELI, EL> pairs SHOULD be inserted in the SR-MPLS label stack. The ingress node may decide the number and place of the ELI/ELs which need to be inserted into the label stack. The extensions for Border Gateway Protocol (BGP) to indicate the entropy label position in the SR-MPLS label stack has been proposed in [I-D.zhou-idr-bgp-srmpls-elp].

In some cases, the the controller(e.g. PCE) could be used to perform the TE path computation as well as the Entropy Label Position (ELP) which is useful for inter-domain scenarios. This document proposes a set of extensions for PCEP to configure the ELP information for SR-MPLS networks.

2. Conventions used in this document

2.1. Terminology

The terminology is defined as [RFC5440], [RFC6790], [RFC8664] and [RFC8662].

2.2. 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.

3. Entropy Labels in SR-MPLS Scenario with PCE

[RFC8662] proposes to use entropy labels for SR-MPLS networks. The Entropy Readable Label Depth (ERLD) is defined as the number of labels which means that the router will perform load-balancing using the ELI/EL. An appropriate algorithm should consider the following criteria:

- * a limited number of <ELI, EL> pairs SHOULD be inserted in the SR-MPLS label stack;

- * the inserted positions SHOULD be within the ERLD of a maximize number of transit LSRs;
- * a minimum number of <ELI, EL> pairs SHOULD be inserted while satisfying the above criteria.

As described in [RFC8662] section 7, the ERLD value is important for inserting ELI/EL and the ingress node need to evaluate the minimum ERLD value along the node segment path. But it will add complexity in the ELI/EL insertion process. Moreover, the ingress node cannot find the minimum ERLD along the path and does not support the computation of the minimum ERLD especilly in inter-domain scenarios. As the Figure 1 shown, in SR-MPLS inter-domain scenario, the ingress node of the first domain could not get the ERLD information of other nodes of other domains.

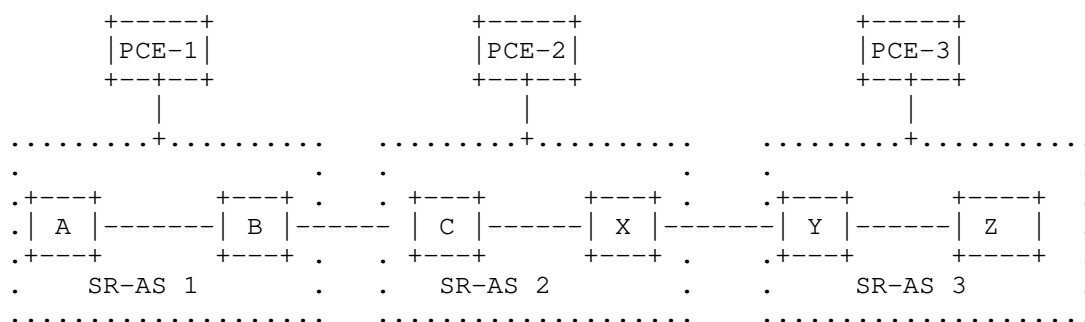


Figure 1: Figure 1: Entropy Labels in SR-MPLS Inter-Domain Scenario

The PCEs could get the information of all nodes such as Maximum SID Depth (MSD) and ERLD through Interior Gateway Protocol (IGP) and can compute the minimum ERLD along the end-to-end path. For example, the ERLD value can be collected via IS-IS [I-D.ietf-isis-mppls-elc], OSPF [I-D.ietf-ospf-mppls-elc]. [RFC8476] and [RFC8491] provide examples of advertisement of the MSD. Moreover, the PCEs also can compute the Entropy Label Position (ELP) including the number and the places of the ELI/ELs. Then the ingress nodes MAY be required to support the capabilities of inserting multiple ELI/ELs and need to advertise the capabilities to the PCEs.

This document proposes the extensions for PCE to perform the computation of the end-to-end path as well as the positions of entropy labels in SR-MPLS networks. The ingress nodes can directly insert the ELI/ELs based on the positions.

4. PCEP Extensions

4.1. The OPEN Object

As defined in [RFC8664], PCEP speakers use SR PCE Capability sub-TLV to exchange information about their SR capability when PST=1 in the PST List of the PATH-SETUP-TYPE-CAPABILITY TLV carried in Open object. This document defined a new flag (E-flag) for SR PCE Capability sub-TLV as shown in Figure 2.

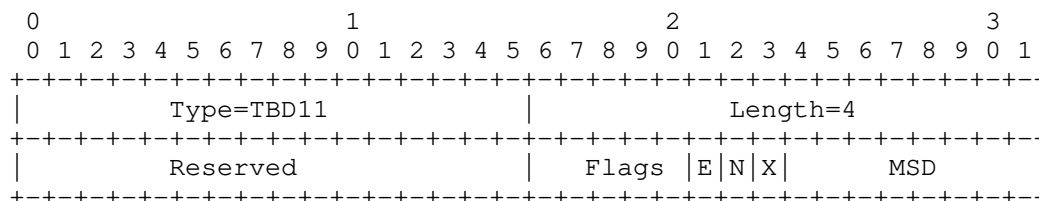


Figure 2: Figure 2: E-flag in SR-PCE-CAPABILITY sub-TLV

E (Entropy Label Configuration is supported) : A PCE sets this flag bit to 1 carried in Open message to indicate that it supports the computation of SR path with ELP information. A PCC sets this flag to 1 to indicate that it supports the capability of inserting multiple ELI/EL pairs and supports the results of SR path with ELP from PCE.

4.2. The LSP-EXTENDED-FLAG TLV

The LSP Object is defined in Section 7.3 of [RFC8231]. This document defiend a new flag (E-flag) for the LSP-EXTENDED-FLAG TLV carried in LSP Object as defined in [I-D.ietf-pce-lsp-extended-flags]. The format is shown as Figure 3:

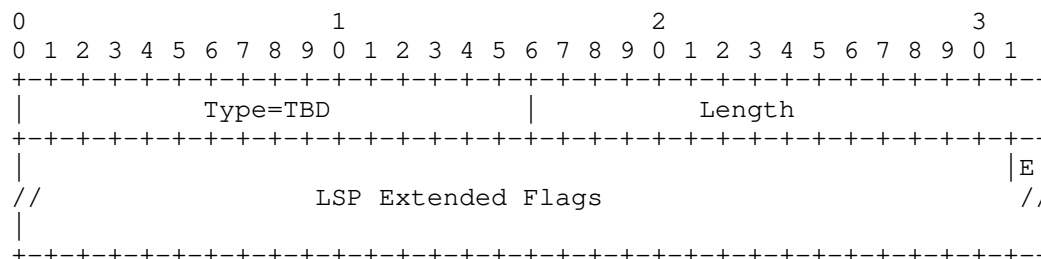


Figure 3: Figure 3: E-flag in LSP-EXTENDED-FLAG TLV

E (Request for ELP Configuration) : If the bit is set to 1, it indicates that the PCC requests PCE to compute the SR path with ELP information. A PCE would also set this bit to 1 to indicate that the ELP information is included by PCE and encoded in the PCRep, PCUpd or PCInitiate message.

4.3. The SR-ERO Object

SR-ERO subobject is used for SR-TE path which consists of one or more SIDs as defined in [RFC8664]. This document defines a new flag (E-flag) for the SR-ERO subobject as Figure 4 shown:

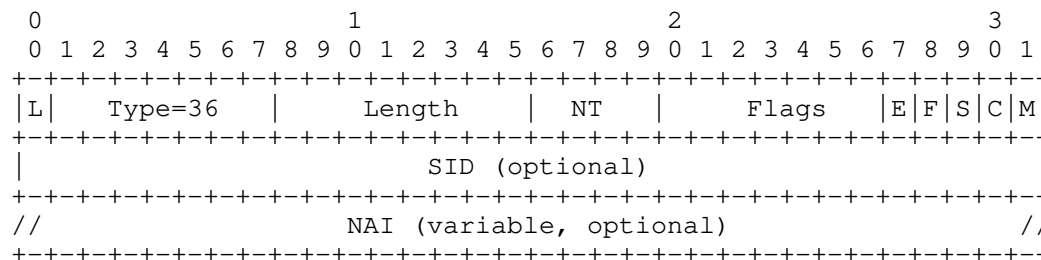


Figure 4: Figure 4: E-flag in SR-ERO subobject

E (ELP Configuration) : If this flag is set, it means that the position after this SR-ERO subobject is the position to insert <ELI, EL>, otherwise it cannot insert <ELI, EL> after this segment.

5. Operations

The SR path is initiated by PCE or PCC with PCReq, PCInitiated or PCUpd messages and the E bit is set to 1 in LSP object to request the ELP configuration. The SR-TE path being received by PCC with SR-ERO segment list, for example, <S1, S2, S3, S4, S5, S6>, especially S3 and S6 with E-flag set. It indicates that two <ELI, EL> pairs MUST be inserted into the label stack of the SR-TE forwarding entry, respectively after the label for S3 and label for S6. With EL information, the label stack for SR-MPLS would be <label1, label2, label3, ELI, EL, label4, label5, label6, ELI, EL>.

6. Security Considerations

Procedures and protocol extensions defined in this document do not introduce any new security considerations beyond those already listed in [RFC8662] and [RFC8664].

7. Acknowledgements

The authors would like to thank Stephane Litkowski, Dhruv Dhody, Tarek Saad, Zhenbin Li and Jeff Tantsura for their review, suggestions and comments to this document.

8. IANA Considerations

8.1. New SR PCE Capability Flag Registry

SR PCE Capability TLV is defined in [RFC8664], and the registry to manage the Flag field of the SR PCE Capability TLV is requested in [RFC8664]. IANA is requested to make allocations from the registry, as follows:

Value	Name	Reference
TBD11	Entropy Label Configuration is supported (E)	[this document]

Table 1

8.2. New LSP-EXTENDED-FLAG Flag Registry

[I-D.ietf-pce-lsp-extended-flags] defines the LSP-EXTENDED-FLAG TLV. IANA is requested to make allocations from the Flag field registry, as follows:

Value	Name	Reference
TBD	Request for ELP Configuration (E)	[this document]

Table 2

8.3. New SR-ERO Flag Registry

SR-ERO subobject is defined in [RFC8664], and the registry to manage the Flag field of SR-ERO is requested in [RFC8664]. IANA is requested to make allocations from the registry, as follows:

Value	Name	Reference
36	ELP Configuration (E)	[this document]

Table 3

9. Normative References

[I-D.ietf-isis-mpls-elc]

Xu, X., Kini, S., Psenak, P., Filsfils, C., Litkowski, S., and M. Bocci, "Signaling Entropy Label Capability and Entropy Readable Label Depth Using IS-IS", Work in Progress, Internet-Draft, draft-ietf-isis-mpls-elc-13, 28 May 2020, <<https://www.ietf.org/archive/id/draft-ietf-isis-mpls-elc-13.txt>>.

[I-D.ietf-ospf-mpls-elc]

Xu, X., Kini, S., Psenak, P., Filsfils, C., Litkowski, S., and M. Bocci, "Signaling Entropy Label Capability and Entropy Readable Label Depth Using OSPF", Work in Progress, Internet-Draft, draft-ietf-ospf-mpls-elc-15, 1 June 2020, <<https://www.ietf.org/archive/id/draft-ietf-ospf-mpls-elc-15.txt>>.

[I-D.ietf-pce-lsp-extended-flags]

Xiong, Q., "LSP Object Flag Extension of Stateful PCE", Work in Progress, Internet-Draft, draft-ietf-pce-lsp-extended-flags-01, 18 October 2021, <<https://www.ietf.org/archive/id/draft-ietf-pce-lsp-extended-flags-01.txt>>.

- [I-D.zhou-idr-bgp-srmppls-elp]
Liu, Y. and S. Peng, "BGP Extension for SR-MPLS Entropy Label Position", Work in Progress, Internet-Draft, draft-zhou-idr-bgp-srmppls-elp-04, 1 March 2022, <<https://www.ietf.org/archive/id/draft-zhou-idr-bgp-srmppls-elp-04.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>>.
- [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>>.
- [RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <<https://www.rfc-editor.org/info/rfc6790>>.
- [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>>.
- [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>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8476] Tantsura, J., Chunduri, U., Aldrin, S., and P. Psenak, "Signaling Maximum SID Depth (MSD) Using OSPF", RFC 8476, DOI 10.17487/RFC8476, December 2018, <<https://www.rfc-editor.org/info/rfc8476>>.
- [RFC8491] Tantsura, J., Chunduri, U., Aldrin, S., and L. Ginsberg, "Signaling Maximum SID Depth (MSD) Using IS-IS", RFC 8491, DOI 10.17487/RFC8491, November 2018, <<https://www.rfc-editor.org/info/rfc8491>>.

- [RFC8623] Palle, U., Dhody, D., Tanaka, Y., and V. Beeram, "Stateful Path Computation Element (PCE) Protocol Extensions for Usage with Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 8623, DOI 10.17487/RFC8623, June 2019, <<https://www.rfc-editor.org/info/rfc8623>>.
- [RFC8660] Bashandy, A., Ed., Filsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/info/rfc8660>>.
- [RFC8662] Kini, S., Kompella, K., Sivabalan, S., Litkowski, S., Shakir, R., and J. Tantsura, "Entropy Label for Source Packet Routing in Networking (SPRING) Tunnels", RFC 8662, DOI 10.17487/RFC8662, December 2019, <<https://www.rfc-editor.org/info/rfc8662>>.
- [RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <<https://www.rfc-editor.org/info/rfc8664>>.

Authors' Addresses

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan
Hubei, 430223
China
Email: xiong.quan@zte.com.cn

Shaofu Peng
ZTE Corporation
No.50 Software Avenue
Nanjing
Jiangsu, 210012
China
Email: peng.shaofu@zte.com.cn

Fengwei Qin
China Mobile
Beijing
China

Email: qinfengwei@chinamobile.com

PCE
Internet-Draft
Intended status: Standards Track
Expires: May 6, 2020

S. Peng
Q. Xiong
ZTE Corporation
F. Qin
China Mobile
November 3, 2019

PCEP Extension for TE Constraints
draft-peng-pce-te-constraints-01

Abstract

This document proposes a set of constraints for PCEP to configure PCE to use specific virtual network topology or application attributes during path computation. A simple COLOR parameter is also introduced to simplify network operations.

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 May 6, 2020.

Copyright Notice

Copyright (c) 2019 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. Conventions used in this document	3
2.1. Terminology	3
2.2. Requirements Language	3
3. PCEP Extensions for Constraints	3
3.1. Source Protocol Object	3
3.2. Multi-topology Object	4
3.3. The AII Object	5
3.4. Application Specific Object	6
3.5. The Color Object	7
4. Security Considerations	8
5. Acknowledgements	8
6. IANA Considerations	8
7. Normative References	9
Authors' Addresses	10

1. Introduction

[RFC5440] describes the Path Computation Element Protocol (PCEP) which is used between a Path Computation Element (PCE) and a Path Computation Client (PCC) (or other PCE) to enable computation of Multi-protocol Label Switching (MPLS) for Traffic Engineering Label Switched Path (TE LSP). PCEP Extensions for the Stateful PCE Model [RFC8231] describes a set of extensions to PCEP to enable active control of MPLS-TE and Generalized MPLS (GMPLS) tunnels. [RFC8281] describes the setup and teardown of PCE-initiated LSPs under the active stateful PCE model, without the need for local configuration on the PCC, thus allowing for dynamic centralized control of a network.

As depicted in [RFC4655], a PCE MUST be able to compute the path of a TE LSP by operating on the TED and considering bandwidth and other constraints applicable to the TE LSP service request. The constraint parameters are provided such as metric, bandwidth, delay, affinity, etc. However these parameters can't meet the virtual network service requirements. A PCE always perform path computation based on the network topology information collected through BGP-LS [RFC7752]. BGP-LS can get multiple link-state data from multiple IGP instance, or multiple virtual topologies from a single IGP instance. It is necessary to restrict the PCE to a small topology scope during path computation for some special purpose. BGP-LS can also get application specific TE attributes for a link, it is also necessary

to restrict PCE to use TE attributes of specific application during path computation.

This document will extend PCEP to support some new constraint parameters during path computation, e.g, IGP instance, virtual network, specific application, as well as a simple COLOR parameter.

2. Conventions used in this document

2.1. Terminology

The terminology is defined as [RFC5440] and [RFC7752].

2.2. 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.

3. PCEP Extensions for Constraints

3.1. Source Protocol Object

The Source Protocol object is optional and can be used for several purposes.

In a PCReq message, a PCC MAY insert one Source Protocol object to indicate the source protocol that MUST be considered by the PCE. The PCE will perform path computation based on the sub-topology identified by the specific source protocol. The absence of the Source Protocol object MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of the source protocols.

In a PCRep/PCInit/PCUpd message, the Source Protocol object MAY be inserted so as to provide the source protocol information for the computed path.

Only one Source Protocol Object could be inserted in the above messages, otherwise the first one MUST be considered and others MUST be ignored.

Source Protocol Object-Class is TBA.

Source Protocol Object-Type is 1.

The format of the Source Protocol object is shown as Figure 1:

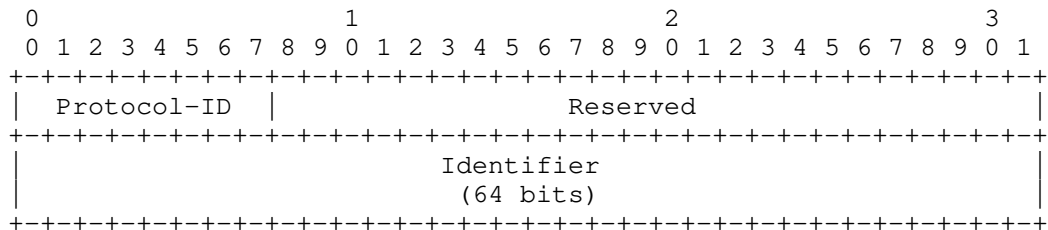


Figure 1: Source Protocol Object

The Source Protocol object body has a fixed length of 12 bytes.

Protocol-ID (8 bits): defined in [RFC7752] section 3.2.

Reserved (24 bits): This field MUST be set to zero on transmission and MUST be ignored on receipt.

Identifier (64 bits): defined in [RFC7752] section 3.2.

3.2. Multi-topology Object

The Multi-topology object is optional and can be used for several purposes.

In a PCReq message, a PCC MAY insert one Multi-topology object to indicate the sub-topology of an IGP instance that MUST be considered by the PCE. The PCE will perform path computation based on the sub-topology identified by the specific Multi-Topology ID within a source protocol. The absence of the Multi-topology object MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of the multi-topologies.

In a PCRep/PCInit/PCUpd message, the Multi-topology object MAY be inserted so as to provide the Multi-topology information for the computed path.

Only one Multi-topology Object could be inserted in the above messages, otherwise the first one MUST be considered and others MUST be ignored. It MUST be inserted with a Source Protocol Object, if not it MUST be ignored.

Multi-topology Object-Class is TBA.

Multi-topology Object-Type is 1.

The format of the Multi-topology object is shown as Figure 2:

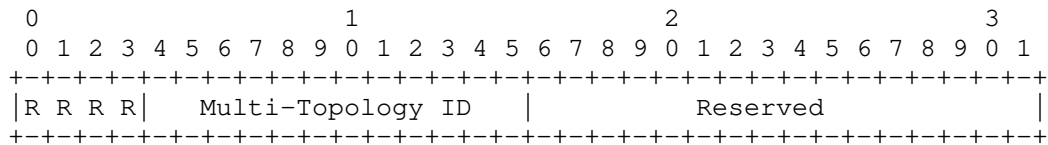


Figure 2: Multi-topology Object

The Multi-topology object body has a fixed length of 4 bytes.

Multi-Topology ID (16 bits): Semantics of the IS-IS MT-ID are defined in Section 7.2 of [RFC5120]. Semantics of the OSPF MT-ID are defined in Section 3.7 of [RFC4915]. If the value is derived from OSPF, then the upper 9 bits MUST be set to 0. Bits R are reserved and SHOULD be set to 0 when originated and ignored on receipt.

Reserved (16 bits): This field **MUST** be set to zero on transmission and **MUST** be ignored on receipt.

3.3. The AII Object

The AII object is optional and can be used for several purposes.

In a PCReq message, a PCC MAY insert one AII object to indicate the global virtual network that MUST be considered by the PCE. The PCE will perform path computation based on the intra or inter-domain sub-topology identified by the specific AII, which is independent of routing protocols such as IGP/BGP. The absence of the AII object MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of the virtual network, i.e., a default AII (0) will be applied.

In a PCRep/PCInit/PCUpd message, the AII object MAY be inserted so as to provide the network slicing information for the computed path.

Only one AII Object could be inserted in the above messages, otherwise the first one MUST be considered and others MUST be ignored.

API Object-Class is TBA.

API Object-Type is 1.

The format of the AII object is shown as Figure 3:

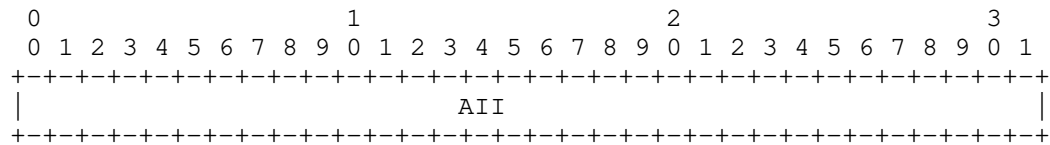


Figure 3: AII Object

The AII object body has a fixed length of 4 bytes.

AII (32 bits): Administrative Instance Identifier defined in [I-D.peng-lsr-network-slicing].

3.4. Application Specific Object

The Application Specific object is optional and can be used for several purposes.

In a PCReq message, a PCC MAY insert one Application Specific object to indicate the application that MUST be considered by the PCE. The PCE will perform path computation using the specific application attributes. The absence of the Application Specific object MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of the Application Specific attributes.

In a PCRep/PCInit/PCUpd message, the Application Specific object MAY be inserted so as to provide the Application Specific information for the computed path.

Only one Application Specific Object could be inserted in the above messages, otherwise the first one MUST be considered and others MUST be ignored.

Application Specific Object-Class is TBA.

Application Specific Object-Type is 1.

The format of the Application Specific object is shown as Figure 4:

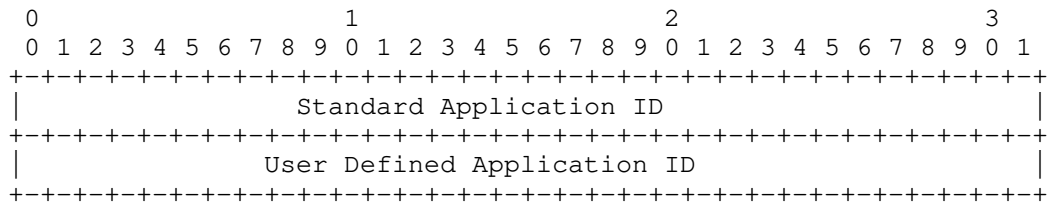


Figure 4: Application Specific Object

The Application Specific object body has a fixed length of 8 bytes.

Standard Application ID : Represents a bit-position value for a single STANDARD application that is defined in the IANA "IGP Parameters" registries under the "Link Attribute Applications" registry [I-D.ietf-isis-te-app].

User Defined Application ID : Represents a single user defined application that is implementation specific.

3.5. The Color Object

The Color object is optional and can be used for several purposes.

In a PCReq message, a PCC MAY insert one Color object to indicate the traffic engineering purpose that is recognized by the both PCE and PCC with no conflict meaning. The PCE will perform path computation based on the color template defined in local and extract the detailed constraints from the color template. Note the same color template is also defined in PCC side. At this time, any other traditional constraints (i.e, metric, bandwidth, dealy, etc) that is directly contained in the message MUST be ignored. The absence of the Color object MUST be interpreted by the PCE as a path computation request for which traditional constraints that are contained in message need be applied.

In a PCRep/PCInit/PCUpd message, the Color object MAY be inserted so as to provide the TE purpose information for the computed path, the PCC recognize the color value that match a local color-template.

Only one Color Object could be inserted in the above messages, otherwise the first one MUST be considered and others MUST be ignored.

Color Object-Class is TBA.

Color Object-Type is 1.

The format of the Color object is shown as Figure 5:

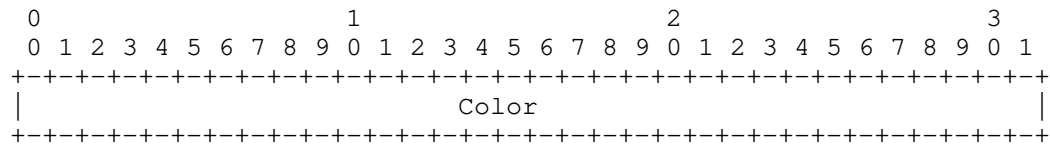


Figure 5: Color Object

The Color object body has a fixed length of 4 bytes.

Color (32 bits): Represent a TE purpose, 0 is invalid value. It is consistent with the meaning of Color Extended Community that is defined in [I-D.ietf-idr-tunnel-encaps], and color of SR policy that is also defined in [I-D.ietf-spring-segment-routing-policy].

Note that Color Object defined in this document is used to represent a TE purpose, it can be suitable for any TE instance such as RSVP-TE, SR-TE, SR-policy. [I-D.barth-pce-segment-routing-policy-cp] has already been using SR policy KEY (that also includes a color information) as an association group KEY to associate many candidate paths, however it is only for association purpose but not constraint purpose for path computation.

A color template can be defined to use any constraints such as traditional metric, bandwidth, delay, affinity parameters, but also any sub-topology parameters above defined in this document. Both PCE and PCC MUST have the same understanding for a same color value.

4. Security Considerations

TBA

5. Acknowledgements

TBA

6. IANA Considerations

IANA is requested to make allocations from the registry, as follows:

Value	Object	Reference
TBA1	Source Protocol Object	[this document]
TBA2	Multi-topology Object	[this document]
TBA3	AII Object	[this document]
TBA4	Application Specific Object	[this document]
TBA5	Color Object	[this document]

Table 1

7. Normative References

- [I-D.barth-pce-segment-routing-policy-cp]
Koldychev, M., Sivabalan, S., Barth, C., Li, C., and H. Bidgoli, "PCEP extension to support Segment Routing Policy Candidate Paths", draft-barth-pce-segment-routing-policy-cp-04 (work in progress), October 2019.
- [I-D.ietf-idr-tunnel-encaps]
Patel, K., Velde, G., and S. Ramachandra, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-14 (work in progress), September 2019.
- [I-D.ietf-isis-te-app]
Ginsberg, L., Psenak, P., Previdi, S., Henderickx, W., and J. Drake, "IS-IS TE Attributes per application", draft-ietf-isis-te-app-09 (work in progress), October 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.peng-lsr-network-slicing]
Peng, S., Chen, R., and G. Mirsky, "Packet Network Slicing using Segment Routing", draft-peng-lsr-network-slicing-00 (work in progress), February 2019.
- [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>>.

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [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>>.
- [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>>.
- [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <<https://www.rfc-editor.org/info/rfc7752>>.
- [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>>.
- [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>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.

Authors' Addresses

Shaofu Peng
ZTE Corporation
No.50 Software Avenue
Nanjing, Jiangsu 210012
China

Email: peng.shaofu@zte.com.cn

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan, Hubei 430223
China

Email: xiong.quan@zte.com.cn

Fengwei Qin
China Mobile
Beijing
China

Email: qinfengwei@chinamobile.com

PCE
Internet-Draft
Intended status: Standards Track
Expires: January 12, 2022

S. Peng
Q. Xiong
ZTE Corporation
F. Qin
China Mobile
M. Koldychev
Cisco Systems
S. Sivabalan
Ciena Corporation
July 11, 2021

PCE TE Constraints
draft-peng-pce-te-constraints-06

Abstract

This document proposes a set of extensions for PCEP to support the TE constraints during path computation, e.g, IGP instance, virtual network, Slice-id, specific application, color template and FA-id etc.

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. Conventions used in this document	3
2.1. Terminology	3
2.2. Requirements Language	3
3. PCEP Extensions for TE Constraints	3
3.1. Source Protocol TLV	3
3.2. Multi-topology TLV	4
3.3. Slice-id TLV	5
3.4. Application Specific TLV	6
3.5. Color TLV	7
3.6. FA-id TLV	9
4. Security Considerations	10
5. Acknowledgements	10
6. IANA Considerations	10
7. Normative References	11
Authors' Addresses	13

1. Introduction

[RFC5440] describes the Path Computation Element Protocol (PCEP) which is used between a Path Computation Element (PCE) and a Path Computation Client (PCC) (or other PCE) to enable computation of Multi-protocol Label Switching (MPLS) for Traffic Engineering Label Switched Path (TE LSP). PCEP Extensions for the Stateful PCE Model [RFC8231] describes a set of extensions to PCEP to enable active control of MPLS-TE and Generalized MPLS (GMPLS) tunnels. As depicted in [RFC4655], a PCE MUST be able to compute the path of a TE LSP by operating on the TED and considering bandwidth and other constraints applicable to the TE LSP service request. The constraint parameters are provided such as metric, bandwidth, delay, affinity, etc. However these parameters can't meet the network slicing requirements.

A PCE always perform path computation based on the network topology information collected through BGP-LS [RFC7752]. BGP-LS can get multiple link-state data from multiple IGP instance, or multiple virtual topologies from a single IGP instance. It is necessary to restrict the PCE to a small topology scope during path computation for some special purpose. BGP-LS can also get application specific TE attributes for a link, it is also necessary to restrict PCE to use

TE attributes of specific application. The PCE MUST take the identifier of slicing into consideration during path computation.

This document proposes a set of extensions for PCEP to support the TE constraints during path computation, e.g, IGP instance, virtual network, Slice-id, specific application, color template and FA-id etc.

2. Conventions used in this document

2.1. Terminology

The terminology is defined as [RFC5440] and [RFC7752].

2.2. 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.

3. PCEP Extensions for TE Constraints

As defined in [RFC5440], the LSPA object is used to specify the LSP attributes to be taken into account by the PCE during path computation such as TE constraints. This document proposes several new TLVs for the LSPA object to carry TE constraints in Network Slicing.

3.1. Source Protocol TLV

The Source Protocol TLV is optional and is defined to carry the source protocol constraint.

In a PCReq/PCRpT message, a PCC MAY insert one or more Source Protocol TLVs to indicate the source protocol that MUST be considered by the PCE. If more than one Source Protocol TLVs are carried, the PCE may perform path computation based on the sub-topology identified by the one of the source protocols. The absence of the Source Protocol TLV MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of the source protocols.

In a PCRep/PCInit/PCUpd message, the Source Protocol TLV MAY be carried so as to provide the source protocol information for the computed path.

The format of the Source Protocol TLV is shown as Figure 1:

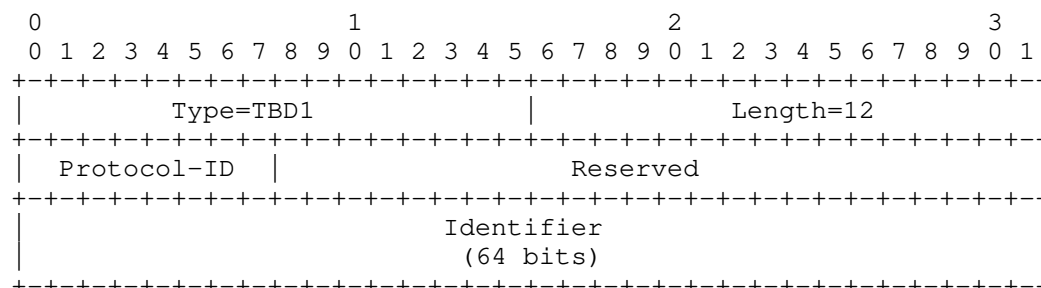


Figure 1: Source Protocol TLV

The code point for the TLV type is TBD1. The TLV length is 12 octets.

Protocol-ID (8 bits): defined in [RFC7752] section 3.2.

Reserved (24 bits): This field MUST be set to zero on transmission and MUST be ignored on receipt.

Identifier (64 bits): defined in [RFC7752] section 3.2.

3.2. Multi-topology TLV

The Multi-topology TLV is optional and is defined to carry the multi-topology protocol constraint.

In a PCReq message, a PCC MAY insert one Multi-topology TLV to indicate the sub-topology of an IGP instance that MUST be considered by the PCE. The PCE will perform path computation based on the sub-topology identified by the specific Multi-Topology ID within a source protocol. The absence of the Multi-topology TLV MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of the multi-topologies.

In a PCRep/PCInit/PCUpd message, the Multi-topology TLV MAY be carried so as to provide the Multi-topology information for the computed path.

The Multi-topology TLV MUST be carried after a Source Protocol TLV, if not it MUST be ignored.

The format of the Multi-topology TLV is shown as Figure 2:

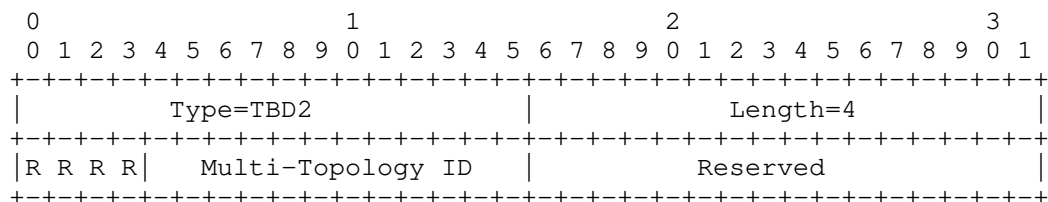


Figure 2: Multi-topology TLV

The code point for the TLV type is TBD2. The TLV length is 4 octets.

Multi-Topology ID (12 bits): Semantics of the IS-IS MT-ID are defined in Section 7.2 of [RFC5120]. Semantics of the OSPF MT-ID are defined in Section 3.7 of [RFC4915]. If the value is derived from OSPF, then the upper 9 bits MUST be set to 0. Bits R are reserved and SHOULD be set to 0 when originated and ignored on receipt.

Reserved (16 bits): This field MUST be set to zero on transmission and MUST be ignored on receipt.

3.3. Slice-id TLV

PCEP message needs to carry Slice ID to let the scope of path calculation to be limited in a specific slice.

There are many control plane technologies to realize slicing. Some control plane technologies may directly maintain resources per slice granularity in the link-state database, only for the case with small slice scalability. [I-D.bestbar-teas-ns-packet] proposes a more scalable slicing scheme. The resource information in link-state database is identified by SA-ID to distinguish the logical topologies corresponding to different slice-aggregate. Within the controller, a slice-aggregate includes one or more slices mapped to it. If the number of slices is small, the resources per slice granularity can be maintained directly in the link-state database. In this case, different slice may be mapped to different slice-aggregate. If the number of slices is large, it is not recommended to maintain the slice granularity resources in the link-state database, but the aggregated SA-ID granularity.

In any case, the slice service (such as VPN service) perceives the Slice ID (not others), so it is natural for the service to include a Slice ID constraint in its TE purpose definition. For example, VPN routes may have Color attribute (refer to [I-D.ietf-idr-tunnel-encaps] and [I-D.ietf-spring-segment-routing-policy]). Color represents a

specific TE purpose, which can contain a Slice ID. Thus it is natural carry Slice ID in PCEP message.

When the controller receives the path computation request with a Slice ID constraint, it can use the resources identified by specific Slice in TED, or firstly look up the Slice ID to SA-ID mapping entry and then use the resources of specific SA-ID in TED, to calculate the path.

In a PCReq message, a PCC MAY insert one Slice-id TLV to indicate the slice based virtual network that MUST be considered by the PCE. The PCE will perform path computation based on the intra-domain or inter-domain sub-topology identified by the specific Slice-id, which is independent of routing protocols such as IGP/BGP. The absence of the Slice-id TLV MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of slice, i.e, a default Slice-id (0) will be applied.

In a PCRep/PCInit/PCUpd message, the Slice-id TLV MAY be carried so as to provide the network slicing information for the computed path. The headend may put the Slice-id to an encapsulated data packet.

The format of the Slice-id TLV is shown as Figure 3:

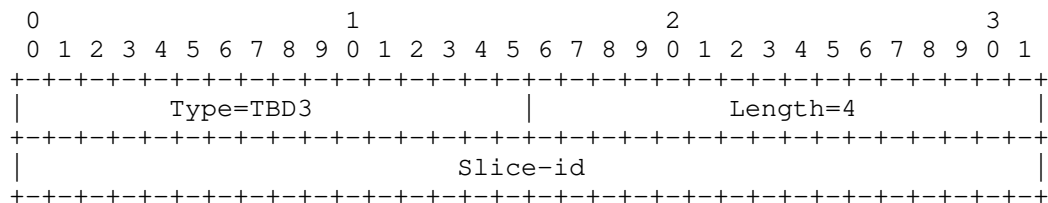


Figure 3: Slice-id TLV

The code point for the TLV type is TBD3. The TLV length is 4 octets.

Slice-id (32 bits): indicate the Slice-id information. The Slice-id is also termed as AII defined in [I-D.peng-lsr-network-slicing] to represent an IETF Network Slice that is defined in [I-D.ietf-teas-ietf-network-slice-definition].

3.4. Application Specific TLV

The Application Specific TLV is optional and is defined to carry the application specific constraints.

In a PCReq message, a PCC MAY insert one Application Specific TLV to indicate the application that MUST be considered by the PCE. The PCE will perform path computation using the specific application attributes. The absence of the Application Specific TLV MUST be interpreted by the PCE as a path computation request for which no constraints need be applied to any of the Application Specific attributes.

In a PCRep/PCInit/PCUpd message, the Application Specific TLV MAY be inserted so as to provide the Application Specific information for the computed path.

The format of the Application Specific TLV is shown as Figure 4:

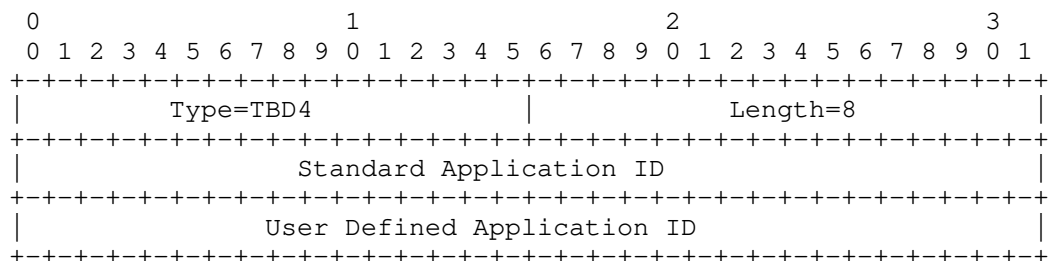


Figure 4: Application Specific TLV

The code point for the TLV type is TBD4. The TLV length is 8 octets.

Standard Application ID: Represents a bit-position value for a single STANDARD application that is defined in the IANA "IGP Parameters" registries under the "Link Attribute Applications" registry [RFC8919].

User Defined Application ID: Represents a single user defined application which is a specific implementation.

3.5. Color TLV

The Color TLV is optional and is defined to carry the color constraints.

In a PCReq message, a PCC MAY insert one Color TLV to indicate the traffic engineering purpose that is recognized by both PCE and PCC with no conflict meaning. The PCE will perform path computation based on the color template. The same color template may be also defined at PCC and the existing constraints (i.e, metric, bandwidth, delay, etc) carried in the message MUST be ignored. The absence of

the Color TLV MUST be interpreted by the PCE as a path computation request for which traditional constraints that are contained in message need be applied.

In a PCRep/PCInit/PCUpd message, the Color TLV MAY be inserted so as to provide the TE purpose information for the computed path, the PCC recognize the color value that match a local color-template. For example, the COLOR TLV can be used to identify the Color of each Candidate Path in the Composite Candidate Path as described in [I-D.ietf-pce-multipath]

The format of the Color TLV is shown as Figure 5:

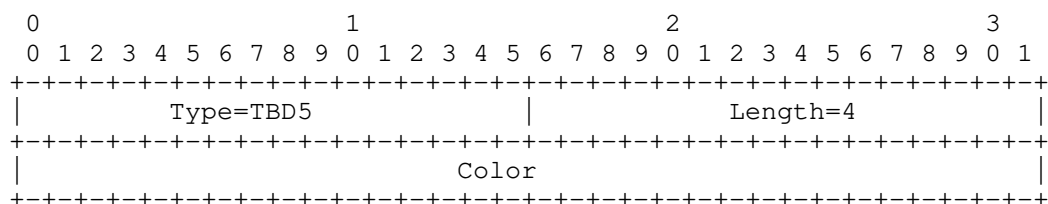


Figure 5: Color TLV

The code point for the TLV type is TBD5. The TLV length is 4 octets.

Color (32 bits): indicate a TE template, 0 is invalid value. It is consistent with the Color Extended Community defined in [I-D.ietf-idr-tunnel-encaps], and color of SR policy defined in [I-D.ietf-spring-segment-routing-policy].

Note that Color TLV defined in this document is used to represent a TE template, it can be suitable for any TE instance such as RSVP-TE, SR-TE, SR-policy. [I-D.ietf-pce-segment-routing-policy-cp] has proposed the SR policy KEY (that also includes a color information) as an association group KEY to associate many candidate paths, however it is only for association purpose but not constraint purpose for path computation.

A color template can be defined to contain existing constraints such as metric, bandwidth, delay, affinity parameters, and the sub-topology constraints above defined in this document.

3.6. FA-id TLV

FA-id defined in [I-D.ietf-lsr-flex-algo] is a short mapping of SR policy color to optimize segment stack depth for the IGP area partial of the entire SR policy. The overlay service that want to be carried over a particular SR-FA path must firstly let the SR policy supplier know that requirement. There are two possible ways to map a color to an FA-id. One is explicit mapping configuration within color template, the other is dynamically replacing a long segment list to short FA segment by headend or controller once the constraints contained in the color-template equal to that contained in FAD.

In addition to the above mapping behavior, it is also possible to merge the constraints contained in the color-template and constraints contained in FAD. The merging behavior can be used to compute SR-TE path within a Flex-algo plane.

In a PCReq message, a PCC MAY insert one FA-id TLV to indicate the above explicit FA-id mapping or merging. For mapping case, the PCE will perform path computation based on the FA-id mapping. In detailed, The PCE will check if there are connectivity within the corresponding Flex-algo plane to the destination. If yes, the path computation result will be represented as segment list with a single prefix-SID@FA for intra-domain case, or several prefix-SID@FA for inter-domain case.

For merging case, the PCE will perform path computation based on the total constraints combined with the ones contained in FAD identified by FA-id and other ones contained in PCReq message. The later constraints can get from color template or directly represent by a color. In this case the computed path will be limited in the specific Flex-algo plane determined by link resource Including/Excluding rules of FAD, and at the same time the path will also meet other constraints for the TE purpose within the Flex-algo plane. The PCE can optimize the strictly path to a loosely path when a part of the strictly path is consistent with the algorithm based path, i.e, some consecutive adjacency SIDs can be replaced with a single algorithm based Prefix-SID.

In a PCRep/PCInit/PCUpd message, the FA-id TLV MAY be inserted so as to provide the FA plane information for the computed path.

In general, the FA-id TLV is only meaningful for the domain (ingress domain) that headend node belongs to. For inter-domain case, operator SHOULD ensure the FA-id configuration of different domain are same for an E2E slice, when he want to explicitly indicate FA-id in PCEP message, otherwise the PCE has to choose different FA-id for

other domain as long as the contents of FAD is consistent with the one of ingress domain.

The format of the FA-id TLV is shown as Figure 6:

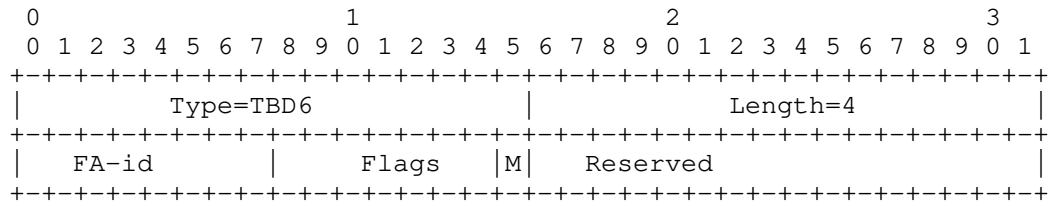


Figure 6: FA-id TLV

The code point for the TLV type is TBD6. The TLV length is 4 octets.

FA-id (8 bits): indicate an explicit FA-id mapping information.

Flags (8 bits): Currently only one flag, Flag-M, is defined.

Flag-M: Indicate mapping behavior when unset, and merging behavior when set.

4. Security Considerations

TBA

5. Acknowledgements

TBA

6. IANA Considerations

IANA is requested to make allocations from the registry, as follows:

Type	TLV	Reference
TBD1	Source Protocol TLV	[this document]
TBD2	Multi-topology TLV	[this document]
TBD3	Slice-id TLV	[this document]
TBD4	Application Specific TLV	[this document]
TBD5	Color TLV	[this document]
TBD6	FA-id TLV	[this document]

Table 1

7. Normative References

[I-D.bestbar-teas-ns-packet]

Saad, T., Beeram, V. P., Wen, B., Ceccarelli, D., Halpern, J., Peng, S., Chen, R., Liu, X., and L. M. Contreras, "Realizing Network Slices in IP/MPLS Networks", draft-bestbar-teas-ns-packet-02 (work in progress), February 2021.

[I-D.ietf-idr-tunnel-encaps]

Patel, K., Velde, G. V. D., Sangli, S. R., and J. Scudder, "The BGP Tunnel Encapsulation Attribute", draft-ietf-idr-tunnel-encaps-22 (work in progress), January 2021.

[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-15 (work in progress), April 2021.

[I-D.ietf-pce-multipath]

Koldychev, M., Sivabalan, S., Saad, T., Beeram, V. P., Bidgoli, H., Yadav, B., and S. Peng, "PCEP Extensions for Signaling Multipath Information", draft-ietf-pce-multipath-00 (work in progress), May 2021.

[I-D.ietf-pce-segment-routing-policy-cp]

Koldychev, M., Sivabalan, S., Barth, C., Peng, S., and H. Bidgoli, "PCEP extension to support Segment Routing Policy Candidate Paths", draft-ietf-pce-segment-routing-policy-cp-04 (work in progress), March 2021.

- [I-D.ietf-spring-segment-routing-policy]
Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", draft-ietf-spring-segment-routing-policy-11 (work in progress), April 2021.
- [I-D.ietf-teas-ietf-network-slice-definition]
Rokui, R., Homma, S., Makhiyani, K., Contreras, L. M., and J. Tantsura, "Definition of IETF Network Slices", draft-ietf-teas-ietf-network-slice-definition-01 (work in progress), February 2021.
- [I-D.peng-lsr-network-slicing]
Peng, S., Chen, R., and G. Mirsky, "Packet Network Slicing using Segment Routing", draft-peng-lsr-network-slicing-00 (work in progress), February 2019.
- [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>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [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-IS)", RFC 5120, DOI 10.17487/RFC5120, February 2008, <<https://www.rfc-editor.org/info/rfc5120>>.
- [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>>.
- [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <<https://www.rfc-editor.org/info/rfc7752>>.

- [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>>.
- [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>>.
- [RFC8919] Ginsberg, L., Psenak, P., Previdi, S., Henderickx, W., and J. Drake, "IS-IS Application-Specific Link Attributes", RFC 8919, DOI 10.17487/RFC8919, October 2020, <<https://www.rfc-editor.org/info/rfc8919>>.

Authors' Addresses

Shaofu Peng
ZTE Corporation
No.50 Software Avenue
Nanjing, Jiangsu 210012
China

Email: peng.shaofu@zte.com.cn

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan, Hubei 430223
China

Email: xiong.quan@zte.com.cn

Fengwei Qin
China Mobile
Beijing
China

Email: qinfengwei@chinamobile.com

Mike Koldychev
Cisco Systems
Canada

Email: mkoldych@cisco.com

Siva Sivabalan
Ciena Corporation
Canada

Email: ssivabal@ciena.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 27, 2020

A. Stone
M. Aissaoui
Nokia
October 25, 2019

Path Protection Enforcement in PCEP
draft-stone-pce-path-protection-enforcement-00

Abstract

This document aims to clarify existing usage of the local protection desired bit signalled in Path Computation Element Protocol (PCEP). This document also introduces a new flag for signalling protection strictness in PCEP.

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 April 27, 2020.

Copyright Notice

Copyright (c) 2019 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
1.1. Requirements Language	3
1.2. Terminology	3
2. Protection Enforcement Flag (E-Flag)	4
3. Security Considerations	5
4. IANA Considerations	6
4.1. LSP Attributes Protection Enforcement Flag	6
5. Normative References	6
Authors' Addresses	7

1. Introduction

Path Computation Element (PCE) Communication Protocol (PCEP) [RFC5440] enables the communication between a Path Computation Client (PCC) and a Path Control Element (PCE), or between two PCEs based on the PCE architecture [RFC4655] .

PCEP [RFC5440] utilizes flags, values and concepts previously defined in RSVP-TE Extensions [RFC3209] and Fast Reroute Extensions to RSVP-TE [RFC4090] . One such concept in PCEP is the "Local Protection Desired" (L-flag in the LSPA Object in RFC5440), which was originally defined in the SESSION-ATTRIBUTE Object in RFC3209. In RSVP, this flag signals to downstream routers that local protection is desired, which indicates to transit routers that they may use a local repair mechanism. The headend router calculating the path does not know whether a downstream router will or will not protect a hop during it's calculation. Therefore, a local protection desired does not require the transit router to satisfy protection in order to establish the RSVP signalled path. This flag is signalled in PCEP as an attribute of the LSP via the LSP Attributes object.

PCEP Extensions for Segment Routing (draft-ietf-pce-segment-routing) extends support in PCEP for Segment Routed LSPs (SR-LSPs) as defined in the Segment Routing Architecture [RFC8402] . As per the Segment Routing Architecture, Adjacency Segment Identifiers(Adj-SID) may be eligible for protection (using IPFRR or MPLS-FRR). The protection eligibility is advertised into IGP (draft-ietf-ospf-segment-routing-extensions and draft-ietf-isis-segment-routing-extensions) as the B-Flag part of the Adjacency SID sub-tlv and can be discovered by a PCE via BGP-LS [RFC7752] using the BGP-LS Segment Routing Extensions (draft-ietf-idr-bgp-ls-segment-routing-ext). An Adjacency SID may or may not have protection eligibility and for a given adjacency between two routers there may be multiple Adjacency SIDs, some of which are protected and some which are not.

A Segment Routed path calculated by PCE may contain various types of segments, as defined in [RFC8402] such as Adjacency, Node or Binding. The protection eligibility for Adjacency SIDs can be discovered by PCE, so therefore the PCE can take the protection eligibility into consideration as a path constraint. If a path is calculated to include other segment identifiers which are not applicable to having their protection state advertised, as they may only be locally significant for each router processing the SID such as Node SIDs, it may not be possible for PCE to include the protection constraint as part of the path calculation.

It is desirable for an operator to define the enforcement, or strictness of the protection requirement when it can be applied.

As defined in [RFC5440] the mechanism to signal protection enforcement in PCEP is with the previously mentioned L-flag defined in the LSPA Object. The name of the flag uses the term "Desired", which by definition means "strongly wished for or intended" and is rooted in the RSVP use case. For RSVP, this is not within control of the PCE. However, [RFC5440] does state "When set, this means that the computed path must include links protected with Fast Reroute as defined in [RFC4090]." Implementations of [RFC5440] have either interpreted the L-Flag as PROTECTION MANDATORY or PROTECTION PREFERRED, leading to operational differences. The boolean bit flag is unable to distinguish between the the different options of PROTECTION MANDATORY, UNPROTECTED MANDATORY, PROTECTION PREFERRED and UNPROTECTED PREFERRED.

1.1. 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].

1.2. Terminology

This document uses the following terminology:

PROTECTION MANDATORY: path MUST have protection eligibility on all links.

UNPROTECTED MANDATORY: path MUST NOT have protection eligibility on all links.

PROTECTION PREFERRED: path SHOULD have protection eligibility on all links but MAY contain links which do not have protection eligibility.

UNPROTECTED PREFERRED: path SHOULD NOT have protection eligibility on all links but MAY contain links which have protection eligibility.

PCC: Path Computation Client. Any client application requesting a path computation to be performed by a Path Computation Element.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

PCEP: Path Computation Element Protocol.

2. Protection Enforcement Flag (E-Flag)

Section 7.11 in Path Computation Element Protocol [RFC5440] describes the encoding of the Local Protection Desired (L-Flag). A new flag is proposed in this document in the LSP Attributes Object which extends the L-Flag to identify the protection enforcement.

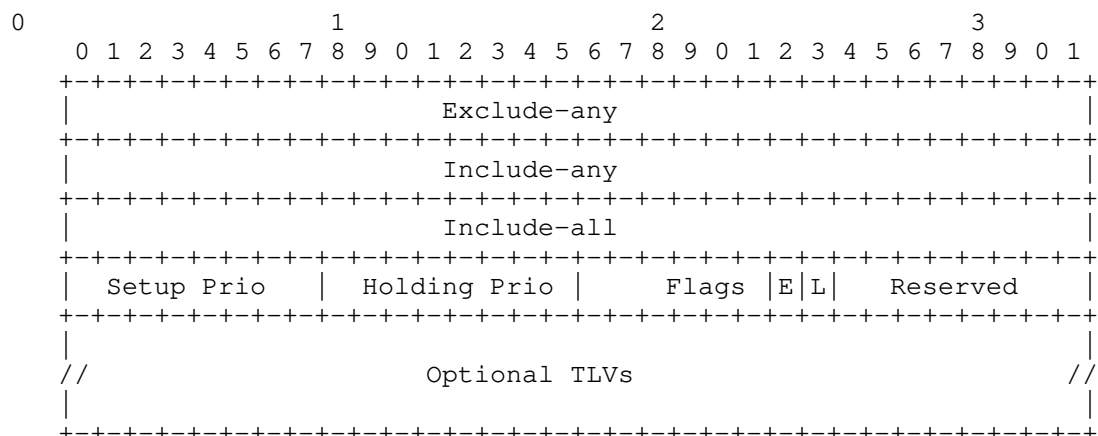
The flag bit is to be allocated by IANA following IETF Consensus.

This draft version proposes using bit 6.

Codespace of the Flag field (LSPA Object)

Bit	Description	Reference
7	Local Protection Desired	RFC5440
6	Local Protection Enforcement	This document

The format of the LSPA Object as defined in [RFC5440] is:



Flags (8 bits)

- o L flag: As defined in [RFC5440] and further updated by this document. When set, protection is desired. When not set, protection is not desired. The enforcement of the protection is identified via the E-Flag.
- o E flag (Protection Enforcement): When set, the value of the L-Flag MUST be treated as a MUST constraint where applicable, when protection state of a SID is known. When E flag is not set, the value of the L-Flag MUST be treated as a MAY constraint.

When L-flag is set and E-flag is set then PCE MUST consider the protection eligibility as PROTECTION MANDATORY constraint.

When L-flag is set and E-flag is not set then PCE MUST consider the protection eligibility as PROTECTION PREFERRED constraint.

When L-flag is not set and E-flag is not set then PCE SHOULD consider the protection eligibility as UNPROTECTED PREFERRED but MAY consider protection eligibility as UNPROTECTED MANDATORY constraint.

When L-flag is not set and E-flag is set then PCE MUST consider the protection eligibility as UNPROTECTED MANDATORY constraint.

For a PCC which does not yet support this draft, the E-flag bit is always set to zero as per [RFC5440] . Therefore, a PCE communicating with a PCC which does not support this draft would treat the L-Flag set as being PROTECTION PREFERRED.

The protection constraint can only be applied to resource selection in which the protection state is known to PCE. A PCE calculating a path that includes resources which does not support the protection state being known to PCE (such as Node SID), then the protection state MAY ignore the protection enforcement constraint.

UNPROTECTED PREFERRED and PROTECTED PREFERRED may seem similar but they indicate the preference of selection if PCE has an option of either protected or unprotected available for a link. When presented with either option, PCE SHOULD select the SID which has a protection state matching the state of the L-Flag.

3. Security Considerations

This document clarifies the behaviour of an existing flag and introduces a new flag to provide further control of that existing behaviour. The introduction of this new flag and behaviour

clarification does not create any new sensitive information. No additional security measure is required.

Securing the PCEP session using Transport Layer Security (TLS) [RFC8253] , as per the recommendations and best current practices in [RFC7525], is RECOMMENDED.

4. IANA Considerations

4.1. LSP Attributes Protection Enforcement Flag

This document defines a new LSP Attribute Flag; IANA is requested to make the following bit allocation from the "LSPA Object" sub registry of the PCEP Numbers registry, as follows:

Value	Name	Reference
6	PROTECTION-ENFORCEMENT	This document

5. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<https://www.rfc-editor.org/info/rfc3209>>.
- [RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, DOI 10.17487/RFC4090, May 2005, <<https://www.rfc-editor.org/info/rfc4090>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [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>>.

- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.
- [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <<https://www.rfc-editor.org/info/rfc7752>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [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

Andrew Stone
Nokia

Email: andrew.stone@nokia.com

Mustapha Aissaoui
Nokia

Email: mustapha.aissaoui@nokia.com

PCE Working Group
Internet Draft
Category: Standards track

Xian Zhang
Haomian Zheng
Huawei Technologies
Oscar Gonzales de Dios
Victor Lopez
Telefonica I+D
Yunbin Xu
CAICT

Expires: April 24, 2020

October 24, 2019

Extensions to the Path Computation Element Protocol (PCEP) to Support
Resource Sharing-based Path Computation

draft-zhang-pce-resource-sharing-11

Abstract

Resource sharing in a network means two or more Label Switched Paths (LSPs) use common pieces of resource along their paths. This can help save network resources and is useful in scenarios such as LSP recovery or when two LSPs do not need to be active at the same time. A Path Computation Element (PCE) is responsible for path computation with such requirement.

Existing extensions to the Path Computation Element Protocol (PCEP) allow one path computation request for an LSP to be associated with other (existing) LSPs through the use of the PCEP Association Object.

This document extends PCEP in order to support resource-sharing-based path computation as another use of the Association Object to enable better efficiency in the computation and in the resultant paths and network resource usage.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 24, 2020.

Copyright Notice

Copyright (c) 2019 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 (<http://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 and Motivation	3
1.1. Requirements Language	4
2. Motivation	5
2.1. Single Domain Use Case	5
2.2. Multiple Layers/Domains Use Case	6
2.3. Bulk Path Computation Use Case	8
3. Extensions to PCEP	9
3.1. Association Group and Type	9
3.2. Resource Sharing TLV	10
3.3. Processing Rules	11
4. Implementation Status	12
5. Manageability Considerations	12
5.1. Control of Function and Policy	12
5.2. Information and Data Models	12
5.3. Liveness Detection and Monitoring	13
5.4. Verify Correct Operations	13
5.5. Requirements on Other Protocols	13

5.6. Impact on Network Operations	13
6. Security Considerations	13
7. IANA Considerations	14
7.1. Association Object Type Indicators	14
7.2. PCEP TLV Definitions	14
8. References	15
8.1. Normative References	15
8.2. Informative References	15
9. Acknowledgements	16
10. Contributor's Address	16
11. Authors' Addresses	17

1. Introduction and Motivation

A Path Computation Element (PCE) is a way to provide path computation function, and it is especially useful in the scenarios where complex constraints and/or a demanding amount of computation resource are required [RFC4655]. The development of PCE standardization has evolved from stateless to stateful. A stateful PCE has access to the LSP database information of the networks it serves as a computation engine [RFC8231]. Unless specified, this document assumes a PCE mentioned is a stateful PCE.

Resource sharing denotes that two or more Label Switched Paths (LSPs) share common pieces of resource, (such as a common time slot of a link in an Optical Transport Network (OTN)). This is usually useful in the scenario where only one of the LSPs is active and the benefit is to save network resources. A simple example of this is dynamically calculating a recovery LSP for an existing LSP undergoing a link failure. Note that resource sharing can be worked out using a stateless PCE, but the mechanism may be complex and is out the scope of this document.

This document considers the requirement that a new LSP may request for resource sharing with one or multiple existing LSPs. Furthermore, if there is resource sharing between a new LSP and existing an LSP, the two LSPs cannot be used to carry traffic simultaneously, the new LSP will take over the traffic from the existing LSP.

In a single domain, this is a common requirement in the recovery cases especially in order to increase traffic resilience against failure while reducing the amount of network resource used for recovery purposes [RFC4428].

The current protocol supporting the communication between a PCE and a Path Computation Client (PCC), i.e. PCE Protocol (PCEP), allows

for re-optimization of an existing LSP [RFC5440]. This is achieved by setting the R bit in the Request Parameter (RP) object, together with some additional information if applicable, in the Path Computation Request (PCReq) message sent from a PCC to the PCE. To support this type of resource sharing, a PCC needs to ask a PCE to compute a new path with the constraints of sharing resource with one or multiple existing LSPs. It is worth noting the "resource sharing" in this draft not only means one LSP re-using the same links of another LSP, but also the same slice of bandwidth in the network. This may occur when an LSP is required for re-routing, or online re-optimization. Current PCEP specifications do not provide such function. More specifically, this document describes the resource sharing issue during the procedure when a new LSP is required to replace an existing LSP for use together with Make-before-break (MBB) described in [RFC3209].

As mentioned in [RFC8231], the PLSP-ID provides a unique identifier for an LSP during a PCEP session between PCC and PCE. Such identification is helpful in supporting the resource sharing requirement for stateful PCEs because it greatly simplifies the operation of a PCC. Instead of the PCC determining all the resources to be shared, the PCC can request that the PCE share the resources of a specific LSP: the stateful PCE is able to determine those resource itself.

Resource sharing can also be required in an inter-layer PCEP session. This is similar to the previous requirement. However, it is more complex and therefore deserves a more detailed explanation here.

In a multi-layer network, LSPs in a lower layer are used to carry higher-layer LSPs across the lower-layer network [RFC5623]. Therefore, the resource sharing constraints in the higher layer might actually relate to resource sharing in the lower layer. Thus, it is useful to consider how this can be achieved and whether additional extensions are needed using the models defined in [RFC5623].

In the next sections, use cases are provided to show what information needs to be exchanged to fulfill these requirements. This memo then provides extensions to PCEP to enable this function.

1.1. 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.

2. Motivation

2.1. Single Domain Use Case

There are two potential cases that request resource to be shared: restoration and re-optimization. Figure 1 shows a single domain network with a stateful PCE, and is used as an example for the resource sharing application.

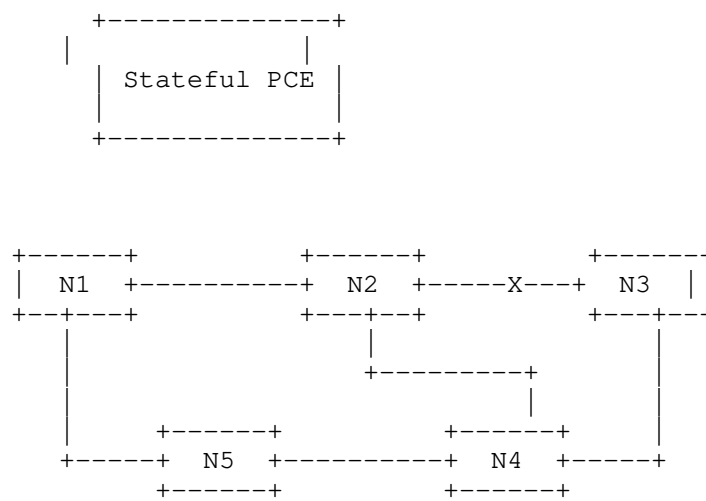


Figure 1: A Single Domain Example

LSP0 (existing): N1-N2-N3

LSP1 (restoration): N1-N2-N4-N3

LSP2 (re-optimization): N1-N5-N4-N3

For the failure restoration, we can assume a working LSP (LSP0) exists in the network. When there is failure on the link N2-N3, it is desired to set up a restoration path for this working LSP. Suppose N1 serves as the PCC and sends a request to the stateful PCE for such an LSP. Before sending the request, N1 may need to check what policy should be applied for the restoration. For example, it might value resource sharing and prefer to share as much resource with the working LSP as possible and specify this policy in the PCReq message. Given such policy, a probable outcome from the path

computation would be LSP1, which shares the link 'N1-N2' with the existing LSP.

Re-optimization does not usually result from a specific failure in the network, but takes place on a stable network when more optimal paths may have become available. Thus switching from the existing LSP to the new LSP happens with live traffic. An example can be found in Figure 1 without failure on the link N2-N3. Instead, an online re-optimization is needed for the working LSP (LSP0) from the stateful PCE. In such cases, the best choice is to set up a backup LSP for the working LSP with totally separate routing (for example, LSP2), and move the traffic to that backup LSP. After that, the working LSP can be torn down, which will not result in any interruption during the optimization procedure. This can actually be implemented with existing PCEP mechanisms. However, if there is no such separate path, existing PCEP mechanisms will return an error. A secondary option for this case is to set up an LSP and complete re-optimization with resource sharing, even if some interruption is introduced.

In the example from Figure 1 it is assumed that the restored LSP or re-optimized LSP have the same source and destination nodes. But in some applications there is no restriction for this assumption, i.e., after an LSP is failed, it can be restored as a new LSP with different source/destination.

In the use cases above it is also assumed that the characteristics of the restored LSP or re-optimized LSP are unchanged. However, it is possible to have parameter changes during the resource sharing computation. For example, the bandwidth of the request LSP may be different from the existing LSP, while resource sharing is still preferred by the PCC. The PCE should consider the sharing request together with the policy and available resources in the network. Details can be found in Section 3.3.

Conversely to resource sharing, it may also be required to apply a disjoint constraint for the path computation. [ietf-pce-association-diversity] discusses the solution under such a scenario, which is a companion work to this document.

2.2. Multiple Layers/Domains Use Case

As Discussed in Section 3 of [RFC5623], there are three models for inter-layer path computation. They are single PCE computation, multiple PCE with inter-PCE communication, and multiple PCE without inter-PCE communication. For the single PCE computation, the process would be similar to that of the use case in Section 2.1.

An inter-layer path computation example is shown in Figure 2. Assume an LSP (LSP1: H2-H3) has been established already, visible as H2-H3 from the view of the higher-layer PCE, and as H2-L1-L2-H3 from the global view (or from the view of the lower-layer PCE). A new request is received by H2 to establish a new LSP (LSP2: from H2 to H5), given the constraint that it can share resources with LSP1. This requirement is possible if only one of the LSPs needs to be active and resource sharing is the target.

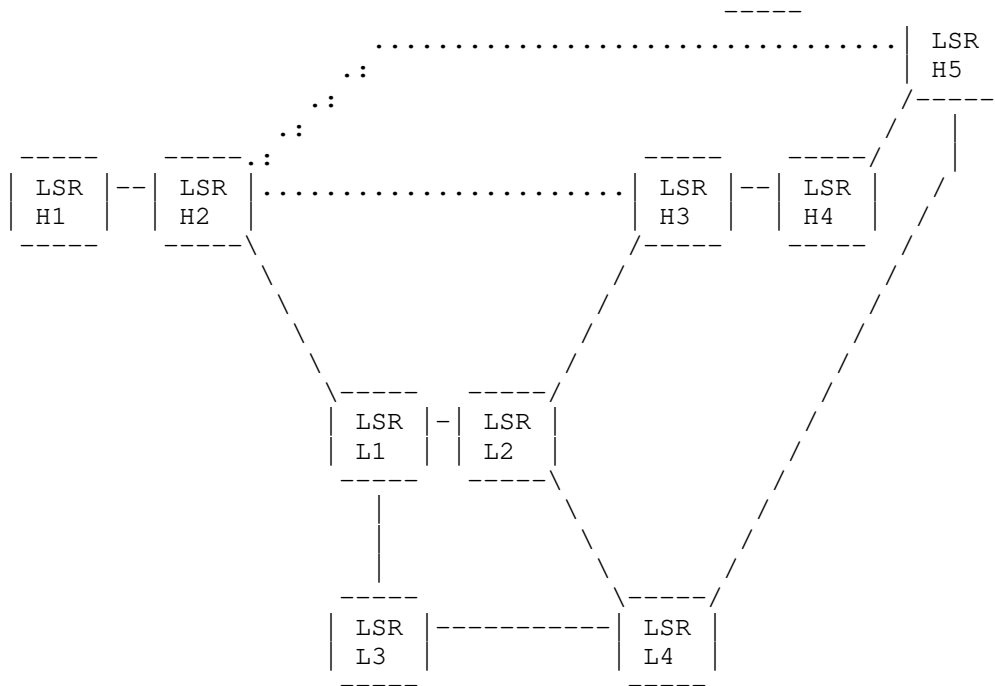


Figure 2: A Two-layer Network Example

If the model of multiple PCEs with inter-PCE communication is employed, the path computation request sent by H2 to higher-layer PCE will be forwarded to lower-layer PCE since there is no resource readily available in the higher layer. So it leaves the lower-layer PCE to compute a path in the lower layer in order to support the higher layer request. In this case, the lower-layer PCE is required to compute a path between H2 and H5 under the constraint that it can share the resource with that of LSP1. At this moment the lower-layer PCE has knowledge of the explicit route of LSP1 (H2-L1-L2-H3), and therefore can map the lower layer LSP with the higher-layer one. So when the lower-layer PCE computes the path for LSP2, it can consider

the resource used by LSP1 as available with higher priority. For example, the lower-layer PCE may choose H2-L1-L2-L4-H5 as the computation result. On the other hand, if the path computation policy is to have a separate path with LSP1, the lower-layer PCE may choose H2-L1-L3-L4-H5.

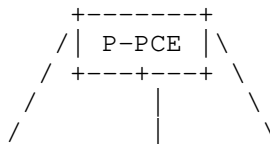
During this procedure the higher-layer PCE can only use information about LSP1 (such as its five-tuple LSP information). An issue to solve is how the lower-layer PCE can resolve this information to the actual resource usage in its own layer, i.e. the lower layer. This could be solved by the edge LSR (L1) reporting this higher-lower LSP correlation to the lower-layer PCE as part of the LSP information during the LSP state synchronization process. If needed, it can be updated later when there is a change in this information. Alternatively, the lower-layer PCE can get this information from other sources, such as a network management system, where this information should be stored.

If the model of multiple PCEs without inter-PCE communication is employed, the path computation request in the lower layer will be initiated by the border LSR node, i.e., L1. The process would be similar to that of the previous scenario. A point worth noting is that the border LSR node may be able to resolve the higher layer LSP information itself, such as by mapping it to the corresponding LSP in the lower layer, in this way the lower-layer PCE does not need to perform this function. Otherwise, the mapping method mentioned above can still be used.

2.3. Bulk Path Computation Use Case

There is a potential need for resource sharing during bulk path computation, especially the processing of the "sticky resources" in [RFC7399]. It would be useful to specify the resources that can be shared among different paths, i.e., the bandwidth information.

Considering the H-PCE architecture in [ietf-pce-stateful-hpce], when the parent PCE asks for a single path across a few domains, such a request may become a bulk path computation to a certain child PCE. Figure 3 shows an example of 3 domains. The parent PCE will select one of these path for establishment.



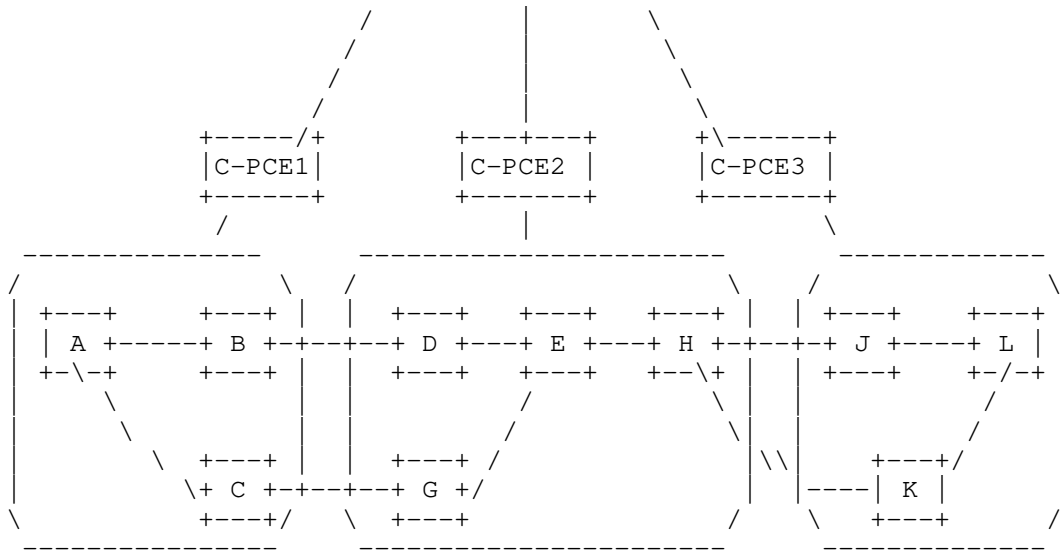


Figure 3: Bulk Request example with Hierarchical PCEs

A 3-domain example is shown in Figure 3, with the hierarchical PCE architecture. In this example nodes A/B/C belong to domain 1, nodes D/E/G/H belong to domain 2, and nodes J/K/L belong to domain 3. Inter-domain links are B-D/C-G between domains 1 and 2, and H-J/H-K between domains 2 and 3. Given a path computation request from A to L, a bulk request from P-PCE would be helpful to understand whether it is possible to have different combinations on the inter-domain links. However, the resources on some specific links become 'sticky' and have to be indicated as 'sharing allowed' to avoid unnecessary resource competition. For example, both the route A-B-D-E-H-J-L and A-C-G-E-H-K-L are qualified, but these routes are competing for the resource on the link E-H and cannot be established simultaneously, so there must be one route failed to be reported to P-PCE. Given the indication of allowing resource sharing on the link E-H, both of these routes can be reported for P-PCE's decision, and there will not be any competition as the P-PCE understands that only one path needs to be set up.

3. Extensions to PCEP

3.1. Association Group and Type

According to the definition in [ietf-pce-association-group], the association group is used to associate multiple LSPs into one group for further path computation considerations, such as disjointness and resource sharing. An association ID will be used to identify the

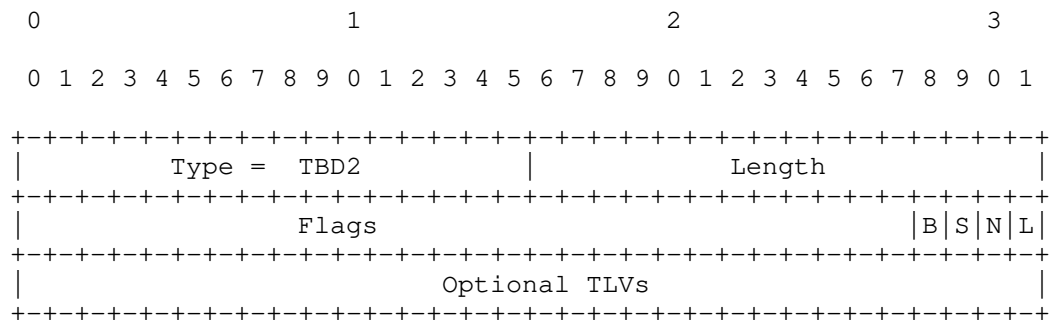
resource sharing group. An association type that described disjointness has been defined in [ietf-pce-association-diversity]. In this document, a new association type is defined as follows:

Association type = TBD1 ("Sharing Association Type").

A sharing group should have multiple LSPs. The number of LSPs and the criteria for how LSPs share among each other are dependent on local policy.

3.2. Resource Sharing TLV

The PCEP Resource Sharing group MAY carry the following TLV. It MAY be carried within a PCReq message from the network element (or other PCCs) so as to indicate the desired resource sharing requirements to be applied by the stateful PCE during path computation.



The following flags are defined:

* L (Link share) bit: when set, this flag indicates that the PCE should prioritize the links shared by existing LSPs within the sharing group for path computation.

* N (Node share) bit: when set, this flag indicates that the PCE should prioritize the nodes shared by existing LSPs within the sharing group for path computation.

* S (SRLG share) bit: when set, this flag indicates that the PCE should set the SRLG (Shared Risk Link Group) of the computed LSP to the same as existing LSPs within the sharing group for path computation.

* B (Bandwidth share) bit: when set, this flag indicates that the PCE should prioritize the bandwidth to be shared by LSPs within the sharing group for bulk path computation.

It is worth noting that there can be multiple flags set which may conflict with each. In this scenario, the result for path computation will be dependent on the policy of PCE.

Optional TLVs may be needed to indicate the LSPs with which the resource is shared. If multiple LSPs are required, the PCE may need to consider different sharing policies, which is implementation dependent and may result in a different computing result. The selection policy among multiple computation result is out of the scope of this document.

3.3. Processing Rules

To request a path allowing resource sharing with one or multiple existing LSPs, a PCC includes a Resource Sharing TLV in the Association Group Object in any kind of path computation request message, such as the PCReq, PCUpd, or PCInitiate messages specified in [RFC8231] and [RFC8281].

On receipt of a PCEP message with a Resource Sharing TLV, a stateful PCE MUST proceed as follows:

- If the Resource Sharing TLV is unknown/unsupported, the PCE will follow procedures defined in [RFC5440]. That is, the PCE sends a PCErr message with error type 26 (Association Error) and error value 6 (Association Information Mismatch), and the related path computation request is discarded.
- If the Resource Sharing TLV is extracted correctly, the PCE MUST apply the requested resource sharing requirement.

The procedure of setting flags follows the rules defined in Section 3.1. The flags in the Resource Sharing TLV may be locally configured on the requesting nodes via external entities, such as a network management system or the entity that imposes the resource sharing requirement.

It is worth noting that the Resource Sharing TLV can be used together with other path indication objects like the IRO/XRO, with different objectives. The first difference is, the use of the Resource Sharing TLV is to set up an alternative path, instead a new path. It is also dependent on the knowledge held by the PCC, e.g., if the PCC has full knowledge of the path information and has a strong preference on the route, it may send the request message with an IRO to specify the route. On the other hand, if the PCC does not know how the path should go but just wants to set up a new LSP to replace the old one, it may use the Resource Sharing TLV instead of an IRO. The second difference is that the Resource Sharing TLV is a loose requirement. For example, if the constraint specified in an IRO/XRO in an A-Z path computation request cannot be satisfied, the reply message from PCE to PCC would be unsuccessful. However it is still possible to have a path from the A-Z. If the target node/link/SRLG/Bandwidth is set in the Resource Sharing TLV rather than an IRO, the PCE may feedback a path from A-Z that does not share the target specified in the Resource Sharing TLV.

4. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to [RFC7942].

Currently the authors are not aware of any implementations.

5. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440] and [RFC8231] apply to the PCEP protocol extensions defined in this document. In addition, requirements and considerations listed in this section apply.

5.1. Control of Function and Policy

A PCE or PCC implementation MUST allow operator-configured associations and SHOULD allow setting of the resource sharing TLV (Section 3.4) as described in this document.

5.2. Information and Data Models

An implementation SHOULD allow the operator to view the resource sharing configured or created dynamically. Further implementation SHOULD allow to view resource sharing associations reported by each peer, and the current set of LSPs in the association. The PCEP YANG module [ietf-pce-pcep-yang] includes association groups information.

5.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

5.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

5.5. Requirements on Other Protocols

Mechanisms defined in this document do not imply any new requirements on other protocols. The configuration on local policy may be accomplished by other protocols, such as Netconf.

5.6. Impact on Network Operations

Mechanisms defined in [RFC5440] and [RFC8231] also apply to PCEP extensions defined in this document.

6. Security Considerations

Security of PCEP is discussed in [RFC5440] and [RFC6952]. The extensions in this document do not change the fundamentals of security for PCEP.

However, the introduction of the Resource Sharing TLV in the Association Group Object provides a vector that may be used to probe for information from a network. For example, a PCC that wants to discover the path of an LSP with which it is not involved can issue a request message with a Resource Sharing TLV and may be able to get back quite a lot of information about the path of the LSP through issuing multiple such requests for different endpoints and analyzing the received results. To protect against this, a PCE SHOULD be configured with access and authorization controls such that only authorized PCCs (for example, those within the network) can make computation requests, only specifically authorized PCCs can make requests for resource sharing, and such requests relating to specific LSPs are further limited to a select few PCCs. How such access controls and authorization is managed is outside the scope of this document, but it will at the least include Access Control Lists.

Furthermore, a PCC must be aware that setting up an LSP that shares resources with another LSP may be a way of attacking the other LSP, for example by depriving it of the resources it needs to operate correctly. Thus it is important that, both in PCEP and the associated signaling protocols, only authorized resource sharing is allowed.

7. IANA Considerations

7.1. Association Object Type Indicators

IANA maintains a registry called the "Path Computation Element Protocol (PCEP) Numbers" registry with a subregistry called the "Association Type Field" subregistry. IANA is requested to make an assignment from that subregistry as follows:

Object Class	Name	Object Type	Reference

TBD1	Sharing-group	Association Type	[this document]

7.2. PCEP TLV Definitions

This document defines the following TLVs to support the resource sharing scenario:

Value	Name	Reference

TBD2	Resource-sharing TLV	[this document]

IANA is requested to allocate the following bit numbers in the flag spaces of Resource-sharing TLV:

Bit	Flag name	Reference
31	Link Share	[this document]
30	Node Share	[this document]
29	SRLG Share	[this document]
28	Bandwidth Share	[this document]

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997. <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001. <<https://www.rfc-editor.org/info/rfc3209>>.
- [RFC5440] Vasseur, J.-P., and Le Roux, J.L., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009. <<https://www.rfc-editor.org/info/rfc5440>>.
- [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>>.
- [RFC8231] Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE", RFC8231, June 2017. <<https://www.rfc-editor.org/info/rfc8231>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", RFC 8281, October 2017. <<https://www.rfc-editor.org/info/rfc8281>>.
- [ietf-pce-association-group] Minei, I., Crabbe E., Sivabalan S., Ananthakrishnan H., Dhody D., Tanaka Y., "PCEP Extensions for Establishing Relationships Between Sets of LSPs", work in progress.
- [ietf-pce-association-diversity] Litkowski, S., Sivabalan, S., Barth, C., Dhody, D., "Path Computation Element communication Protocol extension for signaling LSP diversity constraint", work in progress.

8.2. Informative References

- [RFC4428] Papadimitriou, D., Mannie., E., "Analysis of Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery Mechanisms (including Protection and Restoration)", RFC4428, March 2006. <<https://www.rfc-editor.org/info/rfc4428>>.

- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006. <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC5623] Oki., E., Takeda, T., Le Roux, JL., Farrel, A., "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC5623, September 2009. <<https://www.rfc-editor.org/info/rfc5623>>.
- [RFC6952] Jethanandani, M., Patel, K., Zheng, L., "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC6952, May 2013. <<https://www.rfc-editor.org/info/rfc6952>>.
- [RFC7399] Farrel, A., King, D., "Unanswered Questions in the Path Computation Element Architecture", RFC7399, October 2014. <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7942] Sheffer, Y., Farrel, A., "Improving Awareness of Running Code: The Implementation Status Section", RFC7942, July 2016. <<https://www.rfc-editor.org/info/rfc7942>>.
- [ietf-pce-stateful-hpce] Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., King, D., Gonzalez de Dios, O., "Hierarchical Stateful Path Computation Element (PCE)", work in progress.
- [ietf-pce-pcep-yang] Dhody, D., Hardwick, J., Beeram, V., Tantsura, J., "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", work in progress.

9. Acknowledgements

The authors would like to thank Adrian Farrel for his review and valuable comments.

10. Contributor's Address

Dhruv Dhody
Huawei Technologies
Email: dhruv.dhody@huawei.com

Igor Bryskin
Huawei Technologies
Email: Igor.Bryskin@huawei.com

11. Authors' Addresses

Xian Zhang
Huawei Technologies
Email: zhang.xian@huawei.com

Haomian Zheng
Huawei Technologies
Email: zhenghaomian@huawei.com

Oscar Gonzalez de Dios
Telefonica I+D/gCTIO
Distrito Telefonica
E-28050 Madrid, Spain
EMail: oscar.gonzalezdedios@telefonica.com

Victor Lopez
Telefonica I+D/gCTIO
Distrito Telefonica
E-28050 Madrid, Spain
EMail: victor.lopezalvarez@telefonica.com

Yunbin Xu
CAICT
xuyunbin@caict.ac.cn

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 27, 2022

X. Zhang
H. Zheng
Huawei Technologies
O. Gonzales de Dios
Telefonica
V. Lopez
Nokia
Y. Xu
CAICT
October 24, 2021

Extensions to the Path Computation Element Protocol (PCEP) to Support
Resource Sharing-based Path Computation
draft-zhang-pce-resource-sharing-15

Abstract

Resource sharing in a network means two or more Label Switched Paths (LSPs) use common pieces of resource along their paths. This can help save network resources and is useful in scenarios such as LSP recovery or when two LSPs do not need to be active at the same time. A Path Computation Element (PCE) is responsible for path computation with such requirement.

Existing extensions to the Path Computation Element Protocol (PCEP) allow one path computation request for an LSP to be associated with other (existing) LSPs through the use of the PCEP Association Object.

This document extends PCEP in order to support resource-sharing-based path computation as another use of the Association Object to enable better efficiency in the computation and in the resultant paths and network resource usage.

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 April 27, 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	3
1.1. Requirements Language	4
2. Motivation	4
2.1. Single Domain Use Case	4
2.2. Multiple Layers/Domains Use Case	6
2.3. Bulk Path Computation Use Case	8
3. Extensions to PCEP	10
3.1. Association Group and Type	10
3.2. Resource Sharing TLV	10
3.3. Processing Rules	11
4. Implementation Status	12
5. Manageability Considerations	12
5.1. Control of Function and Policy	12
5.2. Information and Data Models	12
5.3. Liveness Detection and Monitoring	13
5.4. Verify Correct Operations	13
5.5. Requirements on Other Protocols	13
5.6. Impact on Network Operations	13
6. Security Considerations	13
7. IANA Considerations	14
7.1. Association Object Type Indicators	14
7.2. PCEP TLV Definitions	14
8. References	15
8.1. Normative References	15
8.2. Informational References	15
Authors' Addresses	16

1. Introduction

A Path Computation Element (PCE) is a way to provide path computation function, and it is especially useful in the scenarios where complex constraints and/or a demanding amount of computation resource are required [RFC4655]. The development of PCE standardization has evolved from stateless to stateful. A stateful PCE has access to the LSP database information of the networks it serves as a computation engine [RFC8231]. Unless specified, this document assumes a PCE mentioned is a stateful PCE..

Resource sharing denotes that two or more Label Switched Paths (LSPs) share common pieces of resource, (such as a common time slot of a link in an Optical Transport Network (OTN)). This is usually useful in the scenario where only one of the LSPs is active and the benefit is to save network resources. A simple example of this is dynamically calculating a recovery LSP for an existing LSP undergoing a link failure. Note that resource sharing can be worked out using a stateless PCE, but the mechanism may be complex and is out the scope of this document.

This document considers the requirement that a new LSP may request for resource sharing with one or multiple existing LSPs. Furthermore, if there is resource sharing between a new LSP and existing an LSP, the two LSPs cannot be used to carry traffic simultaneously, the new LSP will take over the traffic from the existing LSP.

In a single domain, this is a common requirement in the recovery cases especially in order to increase traffic resilience against failure while reducing the amount of network resource used for recovery purposes [RFC4428]

The current protocol supporting the communication between a PCE and a Path Computation Client (PCC), i.e. PCE Protocol (PCEP), allows for re-optimization of an existing LSP [RFC5440]. This is achieved by setting the R bit in the Request Parameter (RP) object, together with some additional information if applicable, in the Path Computation Request (PCReq) message sent from a PCC to the PCE. To support this type of resource sharing, a PCC needs to ask a PCE to compute a new path with the constraints of sharing resource with one or multiple existing LSPs. It is worth noting the "resource sharing" in this draft not only means one LSP re-using the same links of another LSP, but also the same slice of bandwidth in the network. This may occur when an LSP is required for re-routing, or online re-optimization. Current PCEP specifications do not provide such function. More specifically, this document describes the resource sharing issue during the procedure when a new LSP is required to replace an

existing LSP for use together with Make-before-break (MBB) described in [RFC3209].

As mentioned in [RFC8231], the PLSP-ID provides a unique identifier for an LSP during a PCEP session between PCC and PCE. Such identification is helpful in supporting the resource sharing requirement for stateful PCEs because it greatly simplifies the operation of a PCC. Instead of the PCC determining all the resources to be shared, the PCC can request that the PCE share the resources of a specific LSP: the stateful PCE is able to determine those resource itself.

Resource sharing can also be required in an inter-layer PCEP session. This is similar to the previous requirement. However, it is more complex and therefore deserves a more detailed explanation here.

In a multi-layer network, LSPs in a lower layer are used to carry higher-layer LSPs across the lower-layer network [RFC5623]. Therefore, the resource sharing constraints in the higher layer might actually relate to resource sharing in the lower layer. Thus, it is useful to consider how this can be achieved and whether additional extensions are needed using the models defined in [RFC5623].

In the next sections, use cases are provided to show what information needs to be exchanged to fulfill these requirements. This memo then provides extensions to PCEP to enable this function.

1.1. 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.

2. Motivation

2.1. Single Domain Use Case

There are two potential cases that request resource to be shared: restoration and re-optimization. Figure 1 shows a single domain network with a stateful PCE, and is used as an example for the resource sharing application.

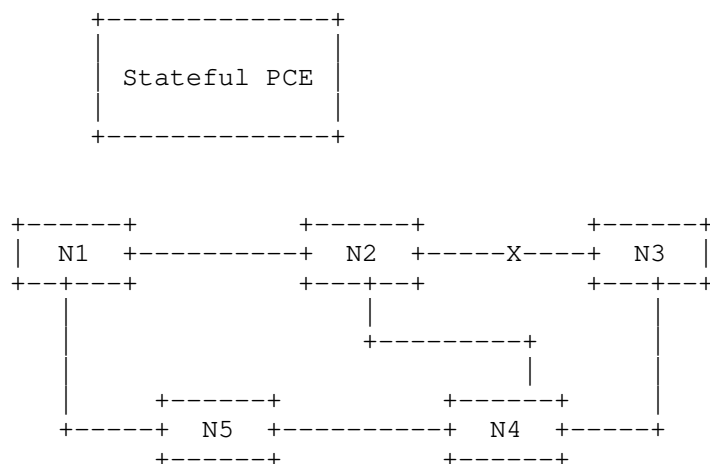


Figure 1: A Single Domain Example

LSP0 (existing): N1-N2-N3.

LSP1 (restoration): N1-N2-N4-N3.

LSP2 (re-optimization): N1-N5-N4-N3.

For the failure restoration, we can assume a working LSP (LSP0) exists in the network. When there is failure on the link N2-N3, it is desired to set up a restoration path for this working LSP. Suppose N1 serves as the PCC and sends a request to the stateful PCE for such an LSP. Besides the head-end and tail-end node of the working LSP, N1 may also need to check what policy should be applied for the restoration. For example, it may evaluate resource sharing and prefer to share as much resource with the working LSP as possible and specify this policy as a special object in the PCReq message. Given such policy, a probable outcome from the path computation would be LSP1, which shares the link 'N1-N2' with the existing LSP. The LSP1 will be set up by PCC via either PCInitiate or RSVP.

Re-optimization does not usually result from a specific failure in the network, but takes place on a stable network when more optimal paths may have become available. Thus switching from the existing LSP to the new LSP happens with live traffic. An example can be found in Figure 1 without failure on the link N2-N3. Instead, an online re-optimization is needed for the working LSP (LSP0) from the stateful PCE. In such cases, the best choice is to set up a backup

LSP for the working LSP with totally separate routing (for example, LSP2), and move the traffic to that backup LSP. After that, the working LSP can be torn down, which will not result in any interruption during the optimization procedure. This can actually be implemented with existing PCEP mechanisms. However, if there is no such separate path, existing PCEP mechanisms will return an error. A secondary option for this case is to set up an LSP and complete re-optimization with resource sharing, even if some interruption is introduced.

In the example from Figure 1 it is assumed that the restored LSP or re-optimized LSP have the same source and destination nodes. But in some applications there is no restriction for this assumption, i.e., after an LSP is failed, it can be restored as a new LSP with different source/destination.

In the use cases above it is also assumed that the characteristics of the restored LSP or re-optimized LSP are unchanged. However, it is possible to have parameter changes during the resource sharing computation. For example, the bandwidth of the request LSP may be different from the existing LSP, while resource sharing is still preferred by the PCC. The PCE should consider the sharing request together with the policy and available resources in the network. Details can be found in Section 3.3.

Conversely to resource sharing, it may also be required to apply a disjoint constraint for the path computation. [RFC8800] discusses the solution under such a scenario, which is a companion work to this document.

2.2. Multiple Layers/Domains Use Case

As Discussed in Section 3 of [RFC5623], there are three models for inter-layer path computation. They are single PCE computation, multiple PCE with inter-PCE communication, and multiple PCE without inter-PCE communication. For the single PCE computation, the process would be similar to that of the use case in Section 2.1.

An inter-layer path computation example is shown in Figure 2. Assume an LSP (LSP1: H2-H3) has been established already, visible as H2-H3 from the view of the higher-layer PCE, and as H2-L1-L2-H3 from the global view (or from the view of the lower-layer PCE). A new request is received by H2 to establish a new LSP (LSP2: from H2 to H5), given the constraint that it can share resources with LSP1. This requirement is possible if only one of the LSPs needs to be active and resource sharing is the target.

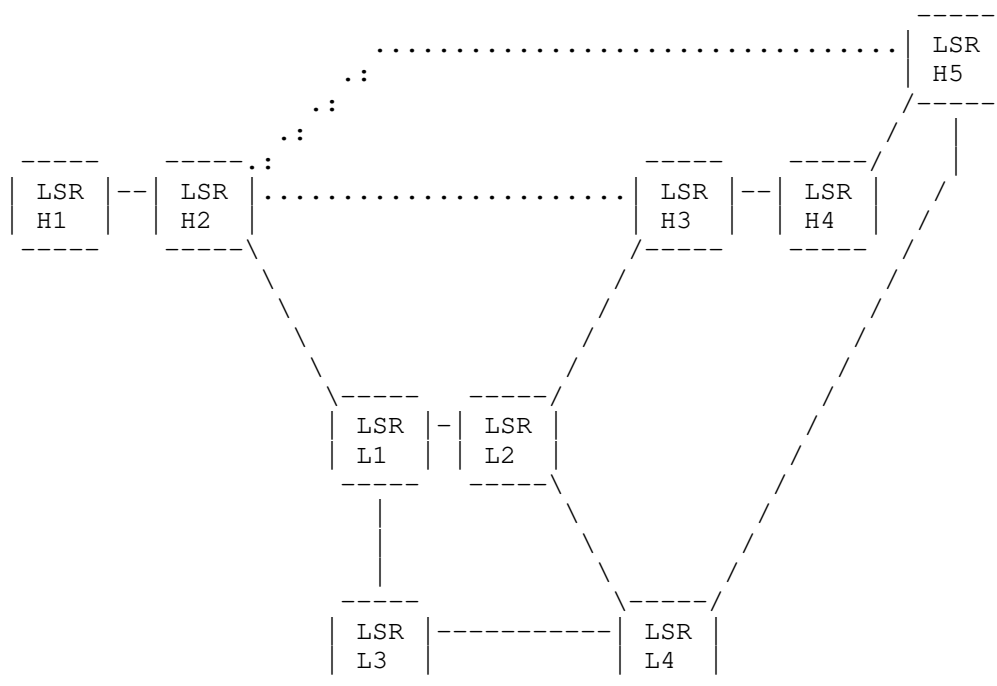


Figure 2: A Two-layer Network Example

If the model of multiple PCEs with inter-PCE communication is employed, the path computation request sent by H2 to higher-layer PCE will be forwarded to lower-layer PCE since there is no resource readily available in the higher layer. So it leaves the lower-layer PCE to compute a path in the lower layer in order to support the higher layer request. In this case, the lower-layer PCE is required to compute a path between H2 and H5 under the constraint that it can share the resource with that of LSP1. At this moment the lower-layer PCE has knowledge of the mapping relationship between the higher-layer link H2-H3 and the lower layer link L1-L2, and therefore can convert the resource to be shared from higher layer to lower layer. So when the lower-layer PCE computes the path for LSP2, it can consider the resource used by L1-L2 as available with higher priority. For example, the lower-layer PCE may choose H2-L1-L2-L4-H5 as the computation result. On the other hand, if the path computation policy is to have a separate path with LSP1, the lower-layer PCE may choose H2-L1-L3-L4-H5.

During this procedure the higher-layer PCE can only use information about LSP1 (such as its five-tuple LSP information). An issue to solve is how the lower-layer PCE can resolve this information to the actual resource usage in its own layer, i.e. the lower layer. This could be solved by the edge LSR (L1) reporting this higher-lower LSP correlation to the lower-layer PCE as part of the LSP information during the LSP state synchronization process. If needed, it can be updated later when there is a change in this information. Alternatively, the lower-layer PCE can get this information from other sources, such as a network management system, where this information should be stored.

If the model of multiple PCEs without inter-PCE communication is employed, the path computation request in the lower layer will be initiated by the border LSR node, i.e., L1. The process would be similar to that of the previous scenario. A point worth noting is that the border LSR node may be able to resolve the higher layer LSP information itself, such as by mapping it to the corresponding LSP in the lower layer, in this way the lower-layer PCE does not need to perform this function. Otherwise, the mapping method mentioned above can still be used.

2.3. Bulk Path Computation Use Case

There is a potential need for resource sharing during bulk path computation, especially the processing of the "sticky resources" in [RFC7399]. It would be useful to specify the resources that can be shared among different paths, i.e., the bandwidth information.

Considering the H-PCE architecture in [RFC8751], when the parent PCE asks for a single path across a few domains, such a request may become a bulk path computation to a certain child PCE. Figure 3 shows an example of 3 domains. The parent PCE will select one of these path for establishment.

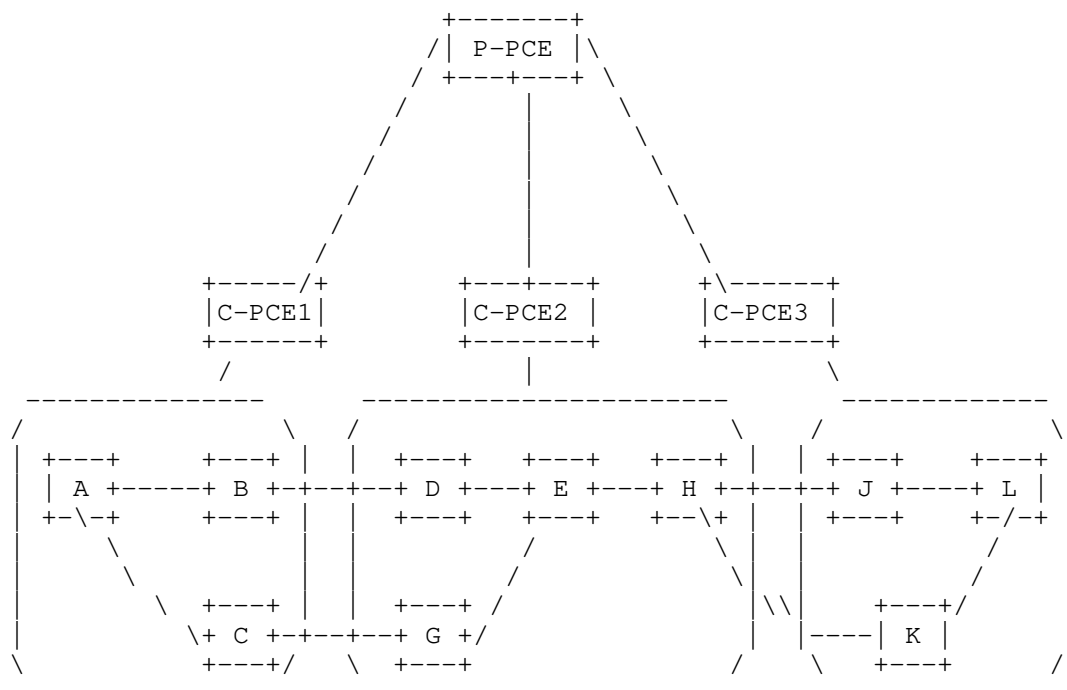


Figure 3: Bulk Request example with Hierarchical PCEs

A 3-domain example is shown in Figure 3, with the hierarchical PCE architecture. In this example nodes A/B/C belong to domain 1, nodes D/E/G/H belong to domain 2, and nodes J/K/L belong to domain 3. Inter-domain links are B-D/C-G between domains 1 and 2, and H-J/H-K between domains 2 and 3. Given a path computation request from A to L, a bulk request from P-PCE would be helpful to understand whether it is possible to have different combinations on the inter-domain links. However, the resources on some specific links become 'sticky' and have to be indicated as 'sharing allowed' to avoid unnecessary resource competition. For example, both the route A-B-D-E-H-J-L and A-C-G-E-H-K-L are qualified, but these routes are competing for the resource on the link E-H and cannot be established simultaneously, so there must be one route failed to be reported to P-PCE. Given the indication of allowing resource sharing on the link E-H, both of these routes can be reported for P-PCE's decision, and there will not be any competition as the P-PCE understands that only one path needs to be set up.

3. Extensions to PCEP

3.1. Association Group and Type

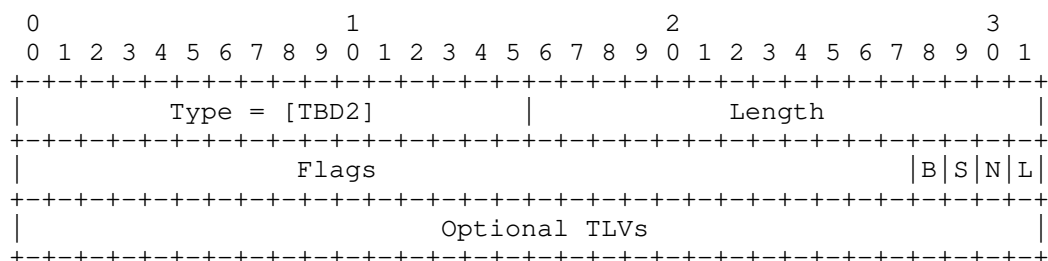
According to the definition in [RFC8697], the association group is used to associate multiple LSPs into one group for further path computation considerations, such as disjointness and resource sharing, in the messages when requesting path computation. An association ID will be used to identify the resource sharing group. An association type that described disjointness has been defined in [RFC8800]. In this document, a new association type is defined as follows:

- o Association type = TBD1 ("Sharing Association Type").

A sharing group should have multiple LSPs. The number of LSPs and the criteria for how LSPs share among each other are dependent on local policy.

3.2. Resource Sharing TLV

The PCEP Resource Sharing group MAY carry the following TLV. It MAY be carried within a PCReq message from the network element (or other PCCs) so as to indicate the desired resource sharing requirements to be applied by the stateful PCE during path computation.



The following flags are defined:

- o L (Link share) bit: when set, this flag indicates that the PCE should prioritize the links shared by existing LSPs within the sharing group for path computation. The existing LSP identifier and its available link identifiers can be contained in the optional TLVs.

- o N (Node share) bit: when set, this flag indicates that the PCE should prioritize the nodes shared by existing LSPs within the sharing group for path computation. The existing LSP identifier and its available node identifiers can be contained in the optional TLVs.
- o S (SRLG share) bit: when set, this flag indicates that the PCE should set the SRLG (Shared Risk Link Group) of the computed LSP to the same as existing LSPs within the sharing group for path computation. The existing LSP identifier and SRLG information can be contained in the optional TLVs.
- o B (Bandwidth share) bit: when set, this flag indicates that the PCE should prioritize the bandwidth to be shared by LSPs within the sharing group for bulk path computation. The LSP identifiers can be contained in the optional TLVs.

It is worth noting that there can be multiple flags set which may conflict with each other. In this scenario, the result for path computation may not be unique, and is dependent on the implementation. The selection among multiple computation results is out of the scope of this document.

3.3. Processing Rules

To request a path allowing resource sharing with one or multiple existing LSPs, a PCC includes a Resource Sharing TLV in the Association Group Object in any kind of path computation request message, such as the PCReq, PCUpd, or PCInitiate messages specified in [RFC8231] and [RFC8281].

On receipt of a PCEP message with a Resource Sharing TLV, a stateful PCE MUST proceed as follows:

- o If the Resource Sharing TLV is unknown/unsupported, the PCE will follow procedures defined in [RFC5440]. That is, the PCE sends a PCErr message with error type 26 (Association Error) and error value 6 (Association Information Mismatch), and the related path computation request is discarded.
- o If the Resource Sharing TLV is extracted correctly, the PCE MUST apply the requested resource sharing requirement, i.e., try to share as much resource as possible with the LSP specified in Resource Sharing TLV.

The procedure of setting flags follows the rules defined in Section 3.1. The flags in the Resource Sharing TLV may be locally configured on the requesting nodes via external entities, such as a

network management system or the entity that imposes the resource sharing requirement.

It is worth noting that the Resource Sharing TLV can be used together with other path indication objects like the IRO/XRO, with different objectives. The first difference is, the use of the Resource Sharing TLV is to set up an alternative path, instead a new path. It is also dependent on the knowledge held by the PCC, e.g., if the PCC has full knowledge of the path information and has a strong preference on the route, it may send the request message with an IRO to specify the route. On the other hand, if the PCC does not know how the path should go but just wants to set up a new LSP to replace the old one, it may use the Resource Sharing TLV instead of an IRO. The second difference is that the Resource Sharing TLV is a loose requirement. For example, if the constraint specified in an IRO/XRO in an A-Z path computation request cannot be satisfied, the reply message from PCE to PCC would be unsuccessful. However it is still possible to have a path from the A-Z. If the target node/link/SRLG/Bandwidth is set in the Resource Sharing TLV rather than an IRO, the PCE may feedback a path from A-Z that does not share the target specified in the Resource Sharing TLV.

4. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to [RFC7942].

Currently the authors are not aware of any implementations.

5. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440] and [RFC8231] apply to the PCEP protocol extensions defined in this document. In addition, requirements and considerations listed in this section apply.

5.1. Control of Function and Policy

A PCE or PCC implementation MUST allow operator-configured associations and SHOULD allow setting of the resource sharing TLV (Section 3.2) as described in this document.

5.2. Information and Data Models

An implementation SHOULD allow the operator to view the resource sharing configured or created dynamically. Further implementation SHOULD allow to view resource sharing associations reported by each peer, and the current set of LSPs in the association. The PCEP YANG

module [I-D.ietf-pce-pcep-yang] includes association groups information.

5.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

5.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

5.5. Requirements on Other Protocols

Mechanisms defined in this document do not imply any new requirements on other protocols. The configuration on local policy may be accomplished by other protocols, such as Netconf.

5.6. Impact on Network Operations

Mechanisms defined in [RFC5440] and [RFC8231] also apply to PCEP extensions defined in this document.

6. Security Considerations

Security of PCEP is discussed in [RFC5440] and [RFC6952]. The extensions in this document do not change the fundamentals of security for PCEP.

However, the introduction of the Resource Sharing TLV in the Association Group Object provides a vector that may be used to probe for information from a network. For example, a PCC that wants to discover the path of an LSP with which it is not involved can issue a request message with a Resource Sharing TLV and may be able to get back quite a lot of information about the path of the LSP through issuing multiple such requests for different endpoints and analyzing the received results. To protect against this, a PCE SHOULD be configured with access and authorization controls such that only authorized PCCs (for example, those within the network) can make computation requests, only specifically authorized PCCs can make requests for resource sharing, and such requests relating to specific LSPs are further limited to a select few PCCs. How such access controls and authorization is managed is outside the scope of this document, but it will at the least include Access Control Lists.

Furthermore, a PCC must be aware that setting up an LSP that shares resources with another LSP may be a way of attacking the other LSP, for example by depriving it of the resources it needs to operate correctly. Thus it is important that, both in PCEP and the associated signaling protocols, only authorized resource sharing is allowed.

7. IANA Considerations

7.1. Association Object Type Indicators

IANA maintains a registry called the "Path Computation Element Protocol (PCEP) Numbers" registry with a subregistry called the "Association Type Field" subregistry. IANA is requested to make an assignment from that subregistry as follows:

Object Class	Name	Object Type	Reference
TBD1	Sharing-group	Association Type	[this document]

7.2. PCEP TLV Definitions

This document defines the following TLVs to support the resource sharing scenario:

Value	Name	Reference
TBD2	Resource-sharing TLV	[this document]

IANA is requested to allocate the following bit numbers in the flag spaces of Resource-sharing TLV:

Bit	Flag name	Reference
31	Link Share	[this document]
30	Node Share	[this document]
29	SRLG Share	[this document]
28	Bandwidth Share	[this document]

8. References

8.1. 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>>.
- [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>>.
- [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>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8697] Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "Path Computation Element Communication Protocol (PCEP) Extensions for Establishing Relationships between Sets of Label Switched Paths (LSPs)", RFC 8697, DOI 10.17487/RFC8697, January 2020, <<https://www.rfc-editor.org/info/rfc8697>>.
- [RFC8800] Litkowski, S., Sivabalan, S., Barth, C., and M. Negi, "Path Computation Element Communication Protocol (PCEP) Extension for Label Switched Path (LSP) Diversity Constraint Signaling", RFC 8800, DOI 10.17487/RFC8800, July 2020, <<https://www.rfc-editor.org/info/rfc8800>>.

8.2. Informational References

- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V. P., and J. Tantsura,
"A YANG Data Model for Path Computation Element
Communications Protocol (PCEP)", draft-ietf-pce-pcep-
yang-16 (work in progress), February 2021.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V.,
and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP
Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001,
<<https://www.rfc-editor.org/info/rfc3209>>.
- [RFC4428] Papadimitriou, D., Ed. and E. Mannie, Ed., "Analysis of
Generalized Multi-Protocol Label Switching (GMPLS)-based
Recovery Mechanisms (including Protection and
Restoration)", RFC 4428, DOI 10.17487/RFC4428, March 2006,
<<https://www.rfc-editor.org/info/rfc4428>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
Element (PCE)-Based Architecture", RFC 4655,
DOI 10.17487/RFC4655, August 2006,
<<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC5623] Oki, E., Takeda, T., Le Roux, JL., and A. Farrel,
"Framework for PCE-Based Inter-Layer MPLS and GMPLS
Traffic Engineering", RFC 5623, DOI 10.17487/RFC5623,
September 2009, <<https://www.rfc-editor.org/info/rfc5623>>.
- [RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of
BGP, LDP, PCEP, and MSDP Issues According to the Keying
and Authentication for Routing Protocols (KARP) Design
Guide", RFC 6952, DOI 10.17487/RFC6952, May 2013,
<<https://www.rfc-editor.org/info/rfc6952>>.
- [RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running
Code: The Implementation Status Section", BCP 205,
RFC 7942, DOI 10.17487/RFC7942, July 2016,
<<https://www.rfc-editor.org/info/rfc7942>>.

Authors' Addresses

Xian Zhang
Huawei Technologies
China

Email: zhang.xian@huawei.com

Haomian Zheng
Huawei Technologies
H1, Xiliu Beipo Village, Songshan Lake,
Dongguan, Guangdong 523808
China

Email: zhenghaomian@huawei.com

Oscar Gonzales de Dios
Telefonica
Spain

Email: oscar.gonzalezdedios@telefonica.com

Victor Lopez
Nokia
Spain

Email: victor.lopez@nokia.com

Yunbin Xu
CAICT
China

Email: xuyunbin@caict.ac.cn

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 9, 2020

Q. Zhao
Z. Li
M. Negi
Huawei Technologies
C. Zhou
Cisco Systems
July 8, 2019

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) of SR-LSPs
draft-zhao-pce-pcep-extension-pce-controller-sr-05

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

PCE was developed to derive paths for MPLS Label Switched Paths (LSPs), which are supplied to the head end of the LSP using the Path Computation Element Communication Protocol (PCEP). But SDN has a broader applicability than signaled (G)MPLS traffic-engineered (TE) networks, and the PCE may be used to determine paths in a range of use cases. PCEP has been proposed as a control protocol for use in these environments to allow the PCE to be fully enabled as a central controller.

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/setup/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network devices along the path while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers, in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network.

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 9, 2020.

Copyright Notice

Copyright (c) 2019 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	3
1.1. Requirements Language	5
2. Terminology	5
3. PCECC SR	5
4. PCEP Requirements	6
5. Procedures for Using the PCE as the Central Controller (PCECC) in Segment Routing	6
5.1. Stateful PCE Model	6
5.2. New LSP Functions	6
5.3. PCECC Capability Advertisement	7
5.4. PCEP session IP address and TEDB Router ID	7
5.5. LSP Operations	8
5.5.1. PCECC Segment Routing (SR)	8
5.5.1.1. PCECC SR Node/Prefix SID allocation	8

5.5.1.2.	PCECC SR Adjacency Label allocation	10
5.5.1.3.	Redundant PCEs	12
5.5.1.4.	Re Delegation and Cleanup	12
5.5.1.5.	Synchronization of Label Allocations	13
5.5.1.6.	PCC Based Allocations	13
5.5.1.7.	Binding SID	13
6.	PCEP messages	14
6.1.	Central Control Instructions	14
6.1.1.	The PCInitiate message	14
6.1.2.	The PCRpt message	15
7.	PCEP Objects	16
7.1.	OPEN Object	16
7.1.1.	PCECC Capability sub-TLV	16
7.2.	PATH-SETUP-TYPE TLV	17
7.3.	CCI Object	17
7.4.	FEC Object	19
8.	Implementation Status	21
8.1.	Huawei's Proof of Concept based on ONOS	21
9.	Security Considerations	22
10.	Manageability Considerations	22
10.1.	Control of Function and Policy	22
10.2.	Information and Data Models	22
10.3.	Liveness Detection and Monitoring	22
10.4.	Verify Correct Operations	23
10.5.	Requirements On Other Protocols	23
10.6.	Impact On Network Operations	23
11.	IANA Considerations	23
11.1.	PCECC-CAPABILITY sub-TLV	23
11.2.	New Path Setup Type Registry	23
11.3.	PCEP Object	24
11.4.	PCEP-Error Object	24
12.	Acknowledgments	24
13.	References	24
13.1.	Normative References	24
13.2.	Informative References	26
Appendix A.	Contributor Addresses	30
Authors'	Addresses	31

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCECC architecture.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment Routing (SR) technology leverages the source routing and tunneling paradigms. A source node can choose a path without relying on hop-by-hop signaling protocols such as LDP or RSVP-TE. Each path is specified as a set of "segments" advertised by link-state routing protocols (IS-IS or OSPF). [RFC8402] provides an introduction to SR architecture. The corresponding IS-IS and OSPF extensions are specified in [I-D.ietf-isis-segment-routing-extensions] and [I-D.ietf-ospf-segment-routing-extensions], respectively. It relies on a series of forwarding instructions being placed in the header of a packet. The segment routing architecture supports operations that can be used to steer packet flows in a network, thus providing a form

of traffic engineering. [I-D.ietf-pce-segment-routing] specify the SR specific PCEP extensions.

PCECC may further use PCEP protocol for SR SID (Segment Identifier) distribution on the SR nodes with some benefits.

This document specifies the procedures and PCEP protocol extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR SID distribution in this case), in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network.

1.1. 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.

2. Terminology

Terminologies used in this document is same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

3. PCECC SR

[I-D.ietf-pce-segment-routing] specifies extensions to PCEP that allow a stateful PCE to compute, update or initiate SR-TE paths. An ingress node of an SR-TE path appends all outgoing packets with a list of MPLS labels (SIDs). This is encoded in SR-ERO subobject, capable of carrying a label (SID) as well as the identity of the node/adjacency label (SID).

The notion of segment and SID is defined in [RFC8402], which fits the MPLS architecture [RFC3031] as the label which is managed by a local allocation process of LSR (similarly to other MPLS signaling protocols) [I-D.ietf-spring-segment-routing-mpls]. The SR information such as node/adjacency label (SID) is flooded via IGP as specified in [I-D.ietf-isis-segment-routing-extensions] and [I-D.ietf-ospf-segment-routing-extensions].

As per [RFC8283], PCE as a central controller can allocate and provision the node/prefix/adjacency label (SID) via PCEP.

Rest of the processing is similar to existing stateful PCE with SR mechanism.

For the purpose of this document, it is assumed that label range to be used by a PCE is set on both PCEP peers. Further, a global label range is assumed to be set on all PCEP peers in the SR domain. This document also allow a case where the label space is maintained by PCC itself, and the labels are allocated by the PCC, in this case, the PCE should request the allocation from PCC as described in Section 5.5.1.6.

4. PCEP Requirements

Following key requirements for PCECC-SR should be considered when designing the PCECC based solution:

- o PCEP speaker supporting this draft MUST have the capability to advertise its PCECC-SR capability to its peers.
- o PCEP speaker not supporting this draft MUST be able to reject PCECC-SR related message with a reason code that indicates no support for PCECC.
- o PCEP procedures MUST provide a means to update (or cleanup) the label- map entry to the PCC.
- o PCEP procedures SHOULD provide a means to synchronize the SR labels allocations between PCE to PCC in the PCEP messages.
- o PCEP procedures MAY allow for PCC based label allocations.

5. Procedures for Using the PCE as the Central Controller (PCECC) in Segment Routing

5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. PCE as a central controller (PCECC) reuses existing Active stateful PCE mechanism as much as possible to control the LSP.

5.2. New LSP Functions

This document uses the same PCEP messages and its extensions which are described in [I-D.ietf-pce-pcep-extension-for-pce-controller] for PCECC-SR as well.

PCEP messages PCRpt, PCInitiate, PCUpd are also used to send LSP Reports, LSP setup and LSP update respectively. The extended PCInitiate message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is used to download or cleanup central controller's instructions (CCIs) (SR SID in scope

of this document). The extended PCRpt message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is also used to report the CCIs (SR SIDs) from PCC to PCE.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify an object called CCI for the encoding of central controller's instructions. This document extends the CCI by defining a new object-type for segment routing. The PCEP messages are extended in this document to handle the PCECC operations for SR.

5.3. PCECC Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions. A PCEP Speaker includes the "PCECC Capability" sub-TLV, described in [I-D.ietf-pce-pcep-extension-for-pce-controller].

A new S-bit is added in PCECC-CAPABILITY sub-TLV to indicate support for PCECC-SR. A PCC MUST set S-bit in PCECC-CAPABILITY sub-TLV and include SR-PCE-CAPABILITY sub-TLV ([I-D.ietf-pce-segment-routing]) in OPEN Object (inside the the PATH-SETUP-TYPE-CAPABILITY TLV) to support the PCECC SR extensions defined in this document. If S-bit is set in PCECC-CAPABILITY sub-TLV and SR-PCE-CAPABILITY sub-TLV is not advertised in OPEN Object, PCE SHOULD send a PCERR message with Error-Type=19 (Invalid Operation) and Error-value=TBD(SR capability was not advertised) and terminate the session.

5.4. PCEP session IP address and TEDB Router ID

PCE may construct its TEDB by participating in the IGP ([RFC3630] and [RFC5305] for MPLS-TE; [RFC4203] and [RFC5307] for GMPLS). An alternative is offered by BGP-LS [RFC7752] and [I-D.dhodylee-pce-pcep-ls].

PCEP [RFC5440] speaker MAY use any IP address while creating a TCP session. It is important to link the session IP address with the Router ID in TEDB for successful PCECC operations.

During PCEP Initialization Phase, PCC SHOULD advertise the TE mapping information. Thus a PCC includes the "Node Attributes TLV" [I-D.dhodylee-pce-pcep-ls] with "IPv4/IPv6 Router-ID of Local Node", in the OPEN Object for this purpose. [RFC7752] describes the usage as auxiliary Router-IDs that the IGP might be using, e.g., for TE purposes. If there are more than one auxiliary Router-ID of a given type, then multiple TLVs are used to encode them.

If "IPv4/IPv6 Router-ID" TLV is not present, the TCP session IP address is directly used for the mapping purpose.

5.5. LSP Operations

The PCEP messages pertaining to PCECC-SR MUST include PATH-SETUP-TYPE TLV [RFC8408] with PST=TBD in the SRP object to clearly identify the PCECC-SR LSP is intended.

5.5.1. PCECC Segment Routing (SR)

Segment Routing (SR) as described in [RFC8402] depends on "segments" that are advertised by Interior Gateway Protocols (IGPs). The SR-node allocates and advertises the SID (node, adj etc) and flood via the IGP. This document proposes a new mechanism where PCE allocates the SID (label/index/SID) centrally and uses PCEP to advertise the SID. In some deployments PCE (and PCEP) are better suited than IGP because of centralized nature of PCE and direct TCP based PCEP session to the node.

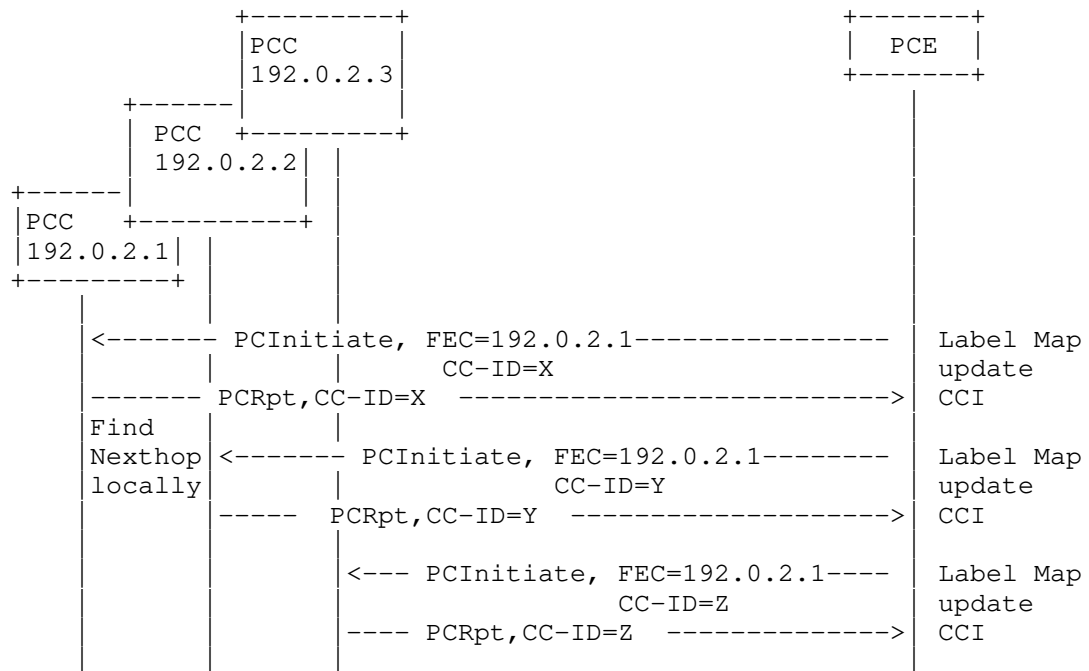
5.5.1.1. PCECC SR Node/Prefix SID allocation

Each node (PCC) is allocated a node-SID by the PCECC. The PCECC sends PCInitiate message to update the label map of each node to all the nodes in the domain. The TE router ID is determined from the TEDB or from "IPv4/IPv6 Router-ID" Sub-TLV [I-D.dhodylee-pce-pcep-ls], in the OPEN Object Section 5.4.

It is RECOMMENDED that PCEP session with PCECC SR capability to use a different session IP address during TCP session establishment than the node Router ID in TEDB, to make sure that the PCEP session does not get impacted by the SR Node/Prefix Label maps (Section 5.4).

If a node (PCC) receives a PCInitiate message with a CCI encoding a SID, out of the range set aside for the SRGB, it MUST send a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (SID out of range) and MUST include the SRP object to specify the error is for the corresponding label update via PCInitiate message.

On receiving the label map, each node (PCC) uses the local information to determine the next-hop and download the label forwarding instructions accordingly. The PCInitiate message in this case MUST NOT have LSP object but uses the new FEC object defined in this document.

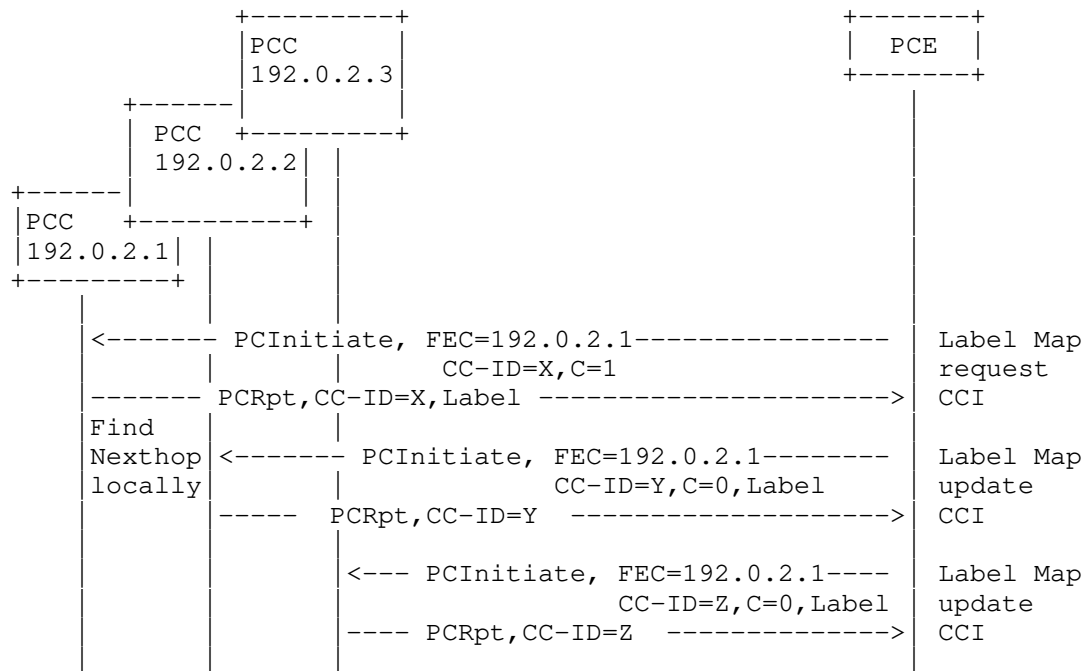


The forwarding behavior and the end result is similar to IGP based "Node-SID" in SR. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node.

PCE relies on the Node/Prefix Label cleanup using the same PCInitiate message.

The above example Figure 1 depict FEC and PCEP speakers that uses IPv4 address. Similarly IPv6 address (such as 2001:DB8::1) can be used during PCEP session establishment as well in FEC object as described in this specification.

In case where the label allocation are made by the PCC itself (see Section 5.5.1.6), the PCE could still request an allocation to be made by the PCC, and where the PCC would send a PCRpt with the allocated label encoded in the CC-ID object as shown below -

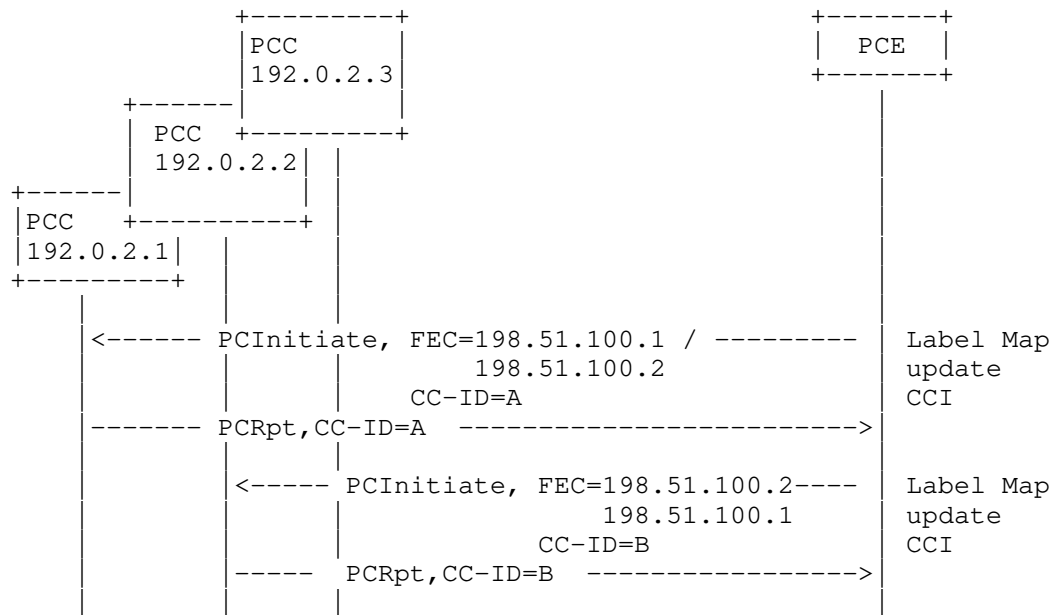


It should be noted that in this example, the request is made to the node 192.0.2.1 with C bit set in the CCI object to indicate that the allocation needs to be done by this PCC and it responds with the allocated label/SID to the PCE. The PCE would further inform the other PCCs in the network about the allocation without setting the C bit.

5.5.1.2. PCECC SR Adjacency Label allocation

[I-D.ietf-pce-segment-routing] extends PCEP to allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

For PCECC SR, apart from node-SID, Adj-SID is used where each adjacency is allocated an Adj-SID by the PCECC. The PCECC sends PCInitiate message to update the label map of each Adj to the corresponding nodes in the domain. Each node (PCC) download the label forwarding instructions accordingly. Similar to SR Node/Prefix Label allocation, the PCInitiate message in this case MUST NOT have LSP object but uses the new FEC object defined in this document.



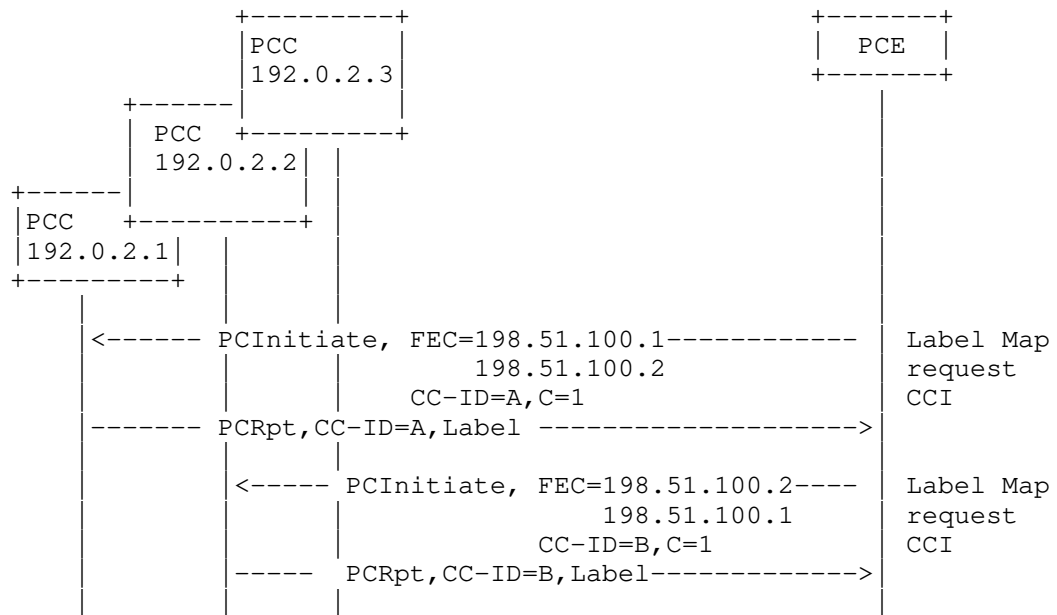
The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SR.

The Path Setup Type for segment routing MUST be set for PCECC SR = TBD (see Section 7.2). All PCEP procedures and mechanism are similar to [I-D.ietf-pce-segment-routing].

PCE relies on the Adj label cleanup using the same PCInitiate message.

The above example Figure 3 depict FEC and PCEP speakers that uses IPv4 address. Similarly IPv6 address (such as 2001:DB8::1, 2001:DB8::2) can be used during PCEP session establishment as well in FEC object as described in this specification.

In case where the label allocation are made by the PCC itself (see Section 5.5.1.6), the PCE could still request an allocation to be made by the PCC, and where the PCC would send a PCRpt with the allocated label encoded in the CC-ID object as shown below -



In this example the request is made to the node 192.0.2.1 with C bit set in the CCI object to indicate that the allocation needs to be done by this PCC for the adjacency (198.51.100.1 - 198.51.100.2) and it responds with the allocated label/SID to the PCE. Similarly, another request is made to the node 192.0.2.2 with C bit set in the CCI object to indicate that the allocation needs to be done by this PCC for the adjacency (198.51.100.2 - 198.51.100.1).

5.5.1.3. Redundant PCEs

[I-D.litkowski-pce-state-sync] describes synchronization mechanism between the stateful PCEs. The SR SIDs allocated by a PCE MUST also be synchronized among PCEs for PCECC SR state synchronization. Note that the SR SIDs are independent to the PCECC-SR LSP, and remains intact till any topology change. The redundant PCEs MUST have a common view of all SR SIDs allocated in the domain.

5.5.1.4. Re Delegation and Cleanup

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the action needed for CCIs for the Basic PCECC LSP on this terminated session. Similarly actions should be applied for the SR SID as well.

5.5.1.5. Synchronization of Label Allocations

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the synchronization of Central Controller's Instructions (CCI) via LSP state synchronization as described in [RFC8231] and [RFC8232]. Same procedures should be applied for SR SIDs as well.

5.5.1.6. PCC Based Allocations

The PCE can request the PCC to allocate the label/SID using the PCInitiate message. The C flag in the CCI object is set to 1 to indicate that the allocation needs to be done by the PCC. The PCC would allocate the SID/Label/Index and would report to the PCE using the PCRpt message.

If the value of the SID/Label/Index is 0 and the C flag is set, it indicates that the PCE is requesting the allocation to be done by the PCC. If the SID/Label/Index is 'n' and the C flag is set in the CCI object, it indicates that the PCE requests a specific value 'n' for the SID/Label/Index. If the allocation is successful, the PCC should report via PCRpt message with the CCI object. Else, it MUST send a PCErr message with Error-Type = TBD ("PCECC failure") and Error Value = TBD ("Invalid CCI"). If the value of the the SID/Label/Index in the CCI object is valid, but the PCC is unable to allocate it, it MUST send a PCErr message with Error-Type = TBD ("PCECC failure") and Error Value = TBD ("Unable to allocate the specified CCI").

If the PCC wishes to withdrawn or modify the previously assigned label/SID, it MUST send a PCRpt message without any SID/Label/Index or with the SID/Label/Index containing the new value respectively in the CCI object. The PCE would further trigger the removal of the central controller instruction as per this document.

5.5.1.7. Binding SID

A PCE as a central controller can allocate and provision the node/prefix/adjacency label (SID) via PCEP. One such SID is binding SID as described in [I-D.sivabalan-pce-binding-label-sid], the PCECC mechanism can also be used to allocate the binding SID as described in this section.

A procedure for binding label/SID allocation is described in [I-D.ietf-pce-pcep-extension-for-pce-controller] and is applicable for all path setup types (including SR paths).

6. PCEP messages

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable-length body made of a set of objects that can be either mandatory or optional. An object is said to be mandatory in a PCEP message when the object must be included for the message to be considered valid. For each PCEP message type, a set of rules is defined that specify the set of objects that the message can carry. An implementation **MUST** form the PCEP messages using the object ordering specified in this document.

6.1. Central Control Instructions

6.1.1. The PCInitiate message

The PCInitiate Message defined in [RFC8281] and extended in [I-D.ietf-pce-pcep-extension-for-pce-controller] is further extended to support SR based central control instructions.

The format of the extended PCInitiate message is as follows:

```
<PCInitiate Message> ::= <Common Header>
                           <PCE-initiated-lsp-list>
```

Where:

<Common Header> is defined in [RFC5440]

```
<PCE-initiated-lsp-list> ::= <PCE-initiated-lsp-request>
                              [<PCE-initiated-lsp-list>]
```

```
<PCE-initiated-lsp-request> ::=
    (<PCE-initiated-lsp-instantiation> |
     <PCE-initiated-lsp-deletion> |
     <PCE-initiated-lsp-central-control>)
```

```
<PCE-initiated-lsp-central-control> ::= <SRP>
                                          (<LSP>
                                           <cci-list>) |
                                          (<FEC>
                                           <CCI>)
```

```
<cci-list> ::= <CCI>
                [<cci-list>]
```

Where:

<PCE-initiated-lsp-instantiation> and
<PCE-initiated-lsp-deletion> are as per
[RFC8281].

The LSP and SRP object is defined in [RFC8231].

When PCInitiate message is used to distribute SR SIDs, the SRP, FEC and CCI objects MUST be present. The error handling for missing SRP or CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. If the FEC object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD (FEC object missing).

To cleanup the SRP object must set the R (remove) bit.

6.1.2. The PCRpt message

The PCRpt message can be used to report the SR instructions received from the central controller (PCE) during the state synchronization phase.

The format of the PCRpt message is as follows:

```
<PCRpt Message> ::= <Common Header>
                        <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report>[<state-report-list>]
```

```
<state-report> ::= (<lsp-state-report>|
                    <central-control-report>)
```

```
<lsp-state-report> ::= [<SRP>]
                        <LSP>
                        <path>
```

```
<central-control-report> ::= [<SRP>]
                              (<LSP>
                               <cci-list>)|
                              (<FEC>
                               <CCI>)
```

```
<cci-list> ::= <CCI>
               [<cci-list>]
```

Where:

<path> is as per [RFC8231] and the LSP and SRP object are also defined in [RFC8231].

When PCRpt message is used to report the label map allocations, the FEC and CCI objects MUST be present. The error handling for CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. If the FEC object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD (FEC object missing).

7. PCEP Objects

7.1. OPEN Object

7.1.1. PCECC Capability sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defined the PCECC-CAPABILITY TLV.

A new S-bit is defined in PCECC-CAPABILITY sub-TLV for PCECC-SR:



S (PCECC-SR-CAPABILITY - 1 bit): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable for PCECC-SR capability and PCE would allocate node and Adj label on this session.

7.2. PATH-SETUP-TYPE TLV

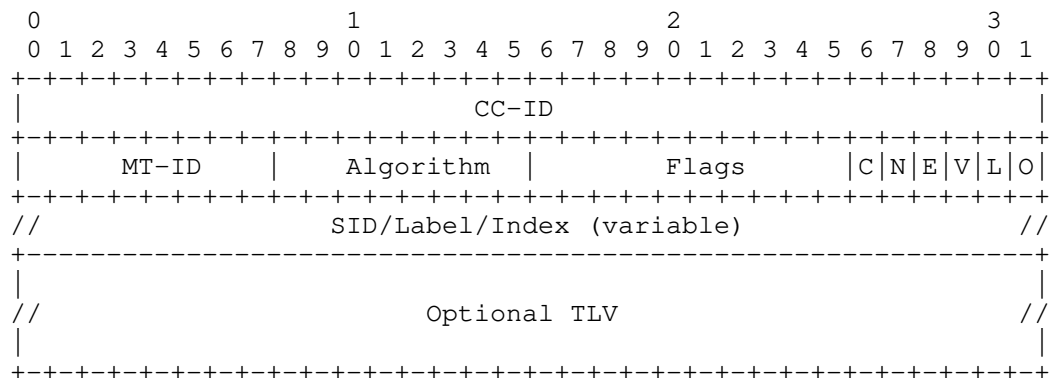
The PATH-SETUP-TYPE TLV is defined in [RFC8408]. PST = TBD is used when Path is setup via PCECC SR mode.

On a PCRpt/PCUpd/PCInitiate message, the PST=TBD indicates that this LSP was setup via a PCECC-SR based mechanism where either the SIDs were allocated/instructed by PCE via PCECC mechanism.

7.3. CCI Object

The Central Control Instructions (CCI) Object is used by the PCE to specify the forwarding instructions is defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]. This document defines another object-type for SR purpose.

CCI Object-Type is TBD for SR as below -



The field CC-ID is as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. Following new fields are defined for CCI Object-Type TBD -

MT-ID: Multi-Topology ID (as defined in [RFC4915]).

Algorithm: Single octet identifying the algorithm the SID is associated with. See [I-D.ietf-ospf-segment-routing-extensions].

Flags: is used to carry any additional information pertaining to the CCI. The O bit was defined in [I-D.ietf-pce-pcep-extension-for-pce-controller], this document further defines following bits-

- * L-Bit (Local/Global): If set, then the value/index carried by the CCI object has local significance. If not set, then the value/index carried by this object has global significance.
- * V-Bit (Value/Index): If set, then the CCI carries an absolute value. If not set, then the CCI carries an index.
- * E-Bit (Explicit-Null): If set, any upstream neighbor of the node that advertised the SID MUST replace the SID with the Explicit-NULL label (0 for IPv4) before forwarding the packet.
- * N-Bit (No-PHP): If set, then the penultimate hop MUST NOT pop the SID before delivering packets to the node that advertised the SID.
- * C-Bit (PCC Allocation): If the bit is set to 1, it indicates that the allocation needs to be done by the PCC for this central controller instruction. A PCE set this bit to request the PCC to make an allocation from its SR label/ID space. A PCC would set this bit to indicate that it has allocated the CC-ID and report it to the PCE.

SID/Label/Index: According to the V and L flags, it contains either:

A 32-bit index defining the offset in the SID/Label space advertised by this router.

A 24-bit label where the 20 rightmost bits are used for encoding the label value.

7.4. FEC Object

The FEC Object is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message.

FEC Object-Class is TBD.

FEC Object-Type is 1 'IPv4 Node ID'.

```

      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 Node ID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

FEC Object-Type is 2 'IPv6 Node ID'.

```

      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 Node ID (16 bytes)                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

FEC Object-Type is 3 'IPv4 Adjacency'.

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Local IPv4 address                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Remote IPv4 address                                    |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

FEC Object-Type is 4 'IPv6 Adjacency'.

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Local IPv6 address (16 bytes)                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Remote IPv6 address (16 bytes)                                    |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

```

|
+-----+
FEC Object-Type is 5 'Unnumbered Adjacency with IPv4 NodeIDs'.

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
+-----+
|                               Local Node-ID                               |
+-----+
|                               Local Interface ID                          |
+-----+
|                               Remote Node-ID                              |
+-----+
|                               Remote Interface ID                         |
+-----+

```

FEC Object-Type is 6 'Linklocal IPv6 Adjacency'.

```

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
+-----+
//                               Local IPv6 address (16 octets)                               //
+-----+
|                               Local Interface ID                          |
+-----+
//                               Remote IPv6 address (16 octets)                               //
+-----+
|                               Remote Interface ID                         |
+-----+

```

The FEC objects are as follows:

IPv4 Node ID: where IPv4 Node ID is specified as an IPv4 address of the Node. FEC Object-type is 1, and the Object-Length is 4 in this case.

IPv6 Node ID: where IPv6 Node ID is specified as an IPv6 address of the Node. FEC Object-type is 2, and the Object-Length is 16 in this case.

IPv4 Adjacency: where Local and Remote IPv4 address is specified as pair of IPv4 address of the adjacency. FEC Object-type is 3, and the Object-Length is 8 in this case.

IPv6 Adjacency: where Local and Remote IPv6 address is specified as pair of IPv6 address of the adjacency. FEC Object-type is 4, and the Object-Length is 32 in this case.

Unnumbered Adjacency with IPv4 NodeID: where a pair of Node ID / Interface ID tuples is used. FEC Object-type is 5, and the Object-Length is 16 in this case.

Linklocal IPv6 Adjacency: where a pair of (global IPv6 address, interface ID) tuples is used. FEC object-type is 6, and the Object-Length is 40 in this case.

8. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to RFC 7942.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

8.1. Huawei's Proof of Concept based on ONOS

The PCE function was developed in the ONOS open source platform. This extension was implemented on a private version as a proof of concept for PCECC.

- o Organization: Huawei
- o Implementation: Huawei's PoC based on ONOS

- o Description: PCEP as a southbound plugin was added to ONOS. To support PCECC-SR, an earlier version of this I-D was implemented. Refer <https://wiki.onosproject.org/display/ONOS/PCEP+Protocol>
- o Maturity Level: Prototype
- o Coverage: Partial
- o Contact: satishk@huawei.com

9. Security Considerations

The security considerations described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

As per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

10. Manageability Considerations

10.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC SR capability as a global configuration.

10.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC SR capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC SR capability.

10.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

10.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

10.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

10.6. Impact On Network Operations

PCEP implementation SHOULD allow a limit to be placed on the rate of PCLabelUpd messages sent by PCE and processed by PCC. It SHOULD also allow sending a notification when a rate threshold is reached.

11. IANA Considerations

11.1. PCECC-CAPABILITY sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY sub-TLV and requests that IANA creates a registry to manage the value of the PCECC-CAPABILITY sub-TLV's Flag field. New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

- o Bit number (counting from bit 0 as the most significant bit)
- o Capability description
- o Defining RFC

IANA is requested to allocate a new bit in the PCECC-CAPABILITY sub-TLV Flag Field registry, as follows:

Bit	Description	Reference
31	S((PCECC-SR-CAPABILITY))	This document

11.2. New Path Setup Type Registry

[RFC8408] created a sub-registry within the "Path Computation Element Protocol (PCEP) Numbers" registry called "PCEP Path Setup Types". IANA is requested to allocate a new code point within this registry, as follows:

Value	Description	Reference
TBD	Traffic engineering path is setup using PCECC-SR mode	This document

11.3. PCEP Object

IANA is requested to allocate new code-point for the new FEC object in "PCEP Objects" sub-registry as follows:

Object-Class	Value	Name	Reference
TBD		FEC	This document
		Object-Type : 1	IPv4 Node ID
		Object-Type : 2	IPv6 Node ID
		Object-Type : 3	IPv4 Adjacency
		Object-Type : 4	IPv6 Adjacency
		Object-Type : 5	Unnumbered Adjacency with IPv4 NodeID
		Object-Type : 6	Linklocal IPv6 Adjacency

11.4. PCEP-Error Object

IANA is requested to allocate new error types and error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type	Meaning	
6	Mandatory Object missing.	
19	Error-value = TBD : Invalid operation.	FEC object missing
	Error-value = TBD :	SR capability was not advertised

12. Acknowledgments

We would like to thank Robert Tao, Changjing Yan, Tieying Huang and Avantika for their useful comments and suggestions.

13. References

13.1. 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>>.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <<https://www.rfc-editor.org/info/rfc3630>>.
- [RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <<https://www.rfc-editor.org/info/rfc4203>>.
- [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>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <<https://www.rfc-editor.org/info/rfc5305>>.
- [RFC5307] Kompella, K., Ed. and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, DOI 10.17487/RFC5307, October 2008, <<https://www.rfc-editor.org/info/rfc5307>>.
- [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>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.

- [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <<https://www.rfc-editor.org/info/rfc7752>>.
- [RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <<https://www.rfc-editor.org/info/rfc7942>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [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>>.
- [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>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.

13.2. Informative References

- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <<https://www.rfc-editor.org/info/rfc3031>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <<https://www.rfc-editor.org/info/rfc8232>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [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>>.
- [I-D.ietf-teas-pcecc-use-cases]
Zhao, Q., Li, Z., Khasanov, B., Dhody, D., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-04 (work in progress), July 2019.

[I-D.ietf-pce-pcep-yang]

Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-12 (work in progress), July 2019.

[I-D.ietf-pce-pcep-extension-for-pce-controller]

Zhao, Q., Li, Z., Negi, M., 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-02 (work in progress), July 2019.

[I-D.ietf-pce-segment-routing]

Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "PCEP Extensions for Segment Routing", draft-ietf-pce-segment-routing-16 (work in progress), March 2019.

[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.litkowski-pce-state-sync]

Litkowski, S., Sivabalan, S., Li, C., and H. Zheng, "Inter Stateful Path Computation Element (PCE) Communication Procedures.", draft-litkowski-pce-state-sync-06 (work in progress), July 2019.

[I-D.dhodylee-pce-pcep-ls]

Dhody, D., Lee, Y., and D. Ceccarelli, "PCEP Extension for Distribution of Link-State and TE Information.", draft-dhodylee-pce-pcep-ls-13 (work in progress), February 2019.

[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.

[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.", draft-sivabalan-pce-binding-
label-sid-07 (work in progress), July 2019.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Satish Karunanithi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: satishk@huawei.com

Adrian Farrel
Juniper Networks, Inc
UK

EMail: adrian@olddog.co.uk

Xuesong Geng
Huawei Technologies
China

Email: gengxuesong@huawei.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

Katherine Zhao
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA

EMail: katherine.zhao@huawei.com

Boris Zhang
Telus Ltd.
Toronto
Canada

EMail: boris.zhang@telus.com

Alex Tokar
Cisco Systems
Slovak Republic

EMail: atokar@cisco.com

Authors' Addresses

Quintin Zhao
Huawei Technologies
125 Nagog Technology Park
Acton, MA 01719
USA

EMail: quintinzhao@gmail.com

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Mahendra Singh Negi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: mahendrasingh@huawei.com

Chao Zhou
Cisco Systems

EMail: choa.zhou@cisco.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 29, 2021

Z. Li
S. Peng
Huawei Technologies
M. Negi
RtBrick Inc
Q. Zhao
Etheric Networks
C. Zhou
HPE
November 25, 2020

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) for Segment Routing (SR) MPLS Segment Identifier
(SID) Allocation and Distribution.
draft-zhao-pce-pcep-extension-pce-controller-sr-09

Abstract

The Path Computation Element (PCE) is a core component of Software-Defined Networking (SDN) systems. It can compute optimal paths for traffic across a network and can also update the paths to reflect changes in the network or traffic demands.

PCE was developed to derive paths for MPLS Label Switched Paths (LSPs), which are supplied to the head end of the LSP using the Path Computation Element Communication Protocol (PCEP). But SDN has a broader applicability than signaled (G)MPLS traffic-engineered (TE) networks, and the PCE may be used to determine paths in a range of use cases. PCEP has been proposed as a control protocol for use in these environments to allow the PCE to be fully enabled as a central controller.

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by blending it with elements of SDN and without necessarily completely replacing it. Thus, the LSP can be calculated/set up/initiated and the label forwarding entries can also be downloaded through a centralized PCE server to each network device along the path while leveraging the existing PCE technologies as much as possible.

This document specifies the procedures and PCEP extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers, in addition to computing the paths for packet flows in a segment routing (SR) network and telling the edge routers what instructions to attach to packets as they enter the network. PCECC is further enhanced for SR SID (Segment Identifier) allocation and distribution.

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 May 29, 2021.

Copyright Notice

Copyright (c) 2020 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	3
1.1. Requirements Language	5
2. Terminology	5
3. PCECC SR	5
4. PCEP Requirements	6
5. Procedures for Using the PCE as a Central Controller (PCECC) in Segment Routing	6
5.1. Stateful PCE Model	6
5.2. New LSP Functions	6
5.3. PCECC Capability Advertisement	7
5.4. PCEP session IP address and TED Router ID	7
5.5. LSP Operations	8
5.5.1. PCECC Segment Routing (SR)	8
5.5.1.1. PCECC SR Node/Prefix SID allocation	8

5.5.1.2.	PCECC SR Adjacency Label allocation	10
5.5.1.3.	Redundant PCEs	12
5.5.1.4.	Re Delegation and Clean up	12
5.5.1.5.	Synchronization of Label Allocations	13
5.5.1.6.	PCC-Based Allocations	13
5.5.1.7.	Binding SID	13
6.	PCEP Messages	14
6.1.	Central Control Instructions	14
6.1.1.	The PCInitiate Message	14
6.1.2.	The PCRpt message	15
7.	PCEP Objects	16
7.1.	OPEN Object	16
7.1.1.	PCECC Capability sub-TLV	16
7.2.	SR-TE Path Setup	17
7.3.	CCI Object	17
7.4.	FEC Object	19
8.	Implementation Status	21
8.1.	Huawei's Proof of Concept based on ONOS	22
9.	Security Considerations	22
10.	Manageability Considerations	22
10.1.	Control of Function and Policy	22
10.2.	Information and Data Models	23
10.3.	Liveness Detection and Monitoring	23
10.4.	Verify Correct Operations	23
10.5.	Requirements On Other Protocols	23
10.6.	Impact On Network Operations	23
11.	IANA Considerations	23
11.1.	PCECC-CAPABILITY sub-TLV	23
11.2.	PCEP Object	24
11.3.	PCEP-Error Object	24
11.4.	CCI Object Flag Field for SR	24
12.	Acknowledgments	25
13.	References	25
13.1.	Normative References	25
13.2.	Informative References	27
Appendix A.	Contributor Addresses	30
Authors'	Addresses	31

1. Introduction

The Path Computation Element (PCE) [RFC4655] was developed to offload the path computation function from routers in an MPLS traffic-engineered network. Since then, the role and function of the PCE has grown to cover a number of other uses (such as GMPLS [RFC7025]) and to allow delegated control [RFC8231] and PCE-initiated use of network resources [RFC8281].

According to [RFC7399], Software-Defined Networking (SDN) refers to a separation between the control elements and the forwarding components so that software running in a centralized system, called a controller, can act to program the devices in the network to behave in specific ways. A required element in an SDN architecture is a component that plans how the network resources will be used and how the devices will be programmed. It is possible to view this component as performing specific computations to place traffic flows within the network given knowledge of the availability of network resources, how other forwarding devices are programmed, and the way that other flows are routed. This is the function and purpose of a PCE, and the way that a PCE integrates into a wider network control system (including an SDN system) is presented in [RFC7491].

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281].

[RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol. [I-D.ietf-teas-pcecc-use-cases] describes the use cases for the PCE-based Central Controller (PCECC) architecture.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP extensions for using the PCE as the central controller for static LSPs, where LSPs can be provisioned as explicit label instructions at each hop on the end-to-end path.

Segment Routing (SR) technology leverages the source routing and tunneling paradigms. A source node can choose a path without relying on hop-by-hop signaling protocols such as LDP or RSVP-TE. Each path is specified as a set of "segments" advertised by link-state routing protocols (IS-IS or OSPF). [RFC8402] provides an introduction to SR architecture. The corresponding IS-IS and OSPF extensions are specified in [RFC8667] and [RFC8665], respectively. It relies on a series of forwarding instructions being placed in the header of a packet. The segment routing architecture supports operations that can be used to steer packet flows in a network, thus providing a form of traffic engineering. [RFC8664] specify the SR specific PCEP extensions.

PCECC may further use PCEP for SR SID (Segment Identifier) allocation and distribution on the SR nodes with some benefits.

This document specifies the procedures and PCEP extensions when a PCE-based controller is also responsible for configuring the forwarding actions on the routers (SR SID allocation and distribution in this case), in addition to computing the paths for packet flows in a segment routing network and telling the edge routers what instructions to attach to packets as they enter the network.

Only SR using MPLS dataplane (SR-MPLS) is in the scope of this document. Refer [I-D.dhody-pce-pcep-extension-pce-controller-srv6] for use of PCECC technique for SR in IPv6 (SRv6) dataplane.

1.1. 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.

2. Terminology

Terminologies used in this document is the same as described in the draft [RFC8283] and [I-D.ietf-teas-pcecc-use-cases].

3. PCECC SR

[RFC8664] specifies extensions to PCEP that allow a stateful PCE to compute, update, or initiate SR-TE paths. An ingress node of an SR-TE path appends all outgoing packets with a list of MPLS labels (SIDs). This is encoded in SR-ERO subobject, capable of carrying a label (SID) as well as the identity of the node/adjacency label (SID).

The notion of segment and SID is defined in [RFC8402], which fits the MPLS architecture [RFC3031] as the label which is managed by a local allocation process of LSR (similarly to other MPLS signaling protocols) [RFC8660]. The SR information such as node/adjacency label (SID) is flooded via IGP as specified in [RFC8667] and [RFC8665].

As per [RFC8283], PCE as a central controller can allocate and provision the node/prefix/adjacency label (SID) via PCEP.

The rest of the processing is similar to existing stateful PCE with SR mechanism.

For the purpose of this document, it is assumed that the label range to be used by a PCE is set on both PCEP peers. Further, a global label range is assumed to be set on all PCEP peers in the SR domain. This document also allows a case where the label space is maintained by PCC itself, and the labels are allocated by the PCC, in this case, the PCE should request the allocation from PCC as described in Section 5.5.1.6.

4. PCEP Requirements

Following key requirements for PCECC-SR should be considered when designing the PCECC-based solution:

- o A PCEP speaker supporting this draft needs to have the capability to advertise its PCECC-SR capability to its peers.
- o PCEP procedures need to allow for PCC-based label/SID allocations.
- o PCEP procedures need means to update (or clean up) the label-map entry to the PCC.
- o PCEP procedures need to provide a mean to synchronize the SR labels allocations between the PCE to the PCC via PCEP messages.

5. Procedures for Using the PCE as a Central Controller (PCECC) in Segment Routing

5.1. Stateful PCE Model

Active stateful PCE is described in [RFC8231]. PCE as a Central Controller (PCECC) reuses the existing active stateful PCE mechanism as much as possible to control the LSPs.

5.2. New LSP Functions

Several new functions are required in PCEP to support PCECC as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. This document reuses the existing messages to support PCECC-SR.

The PCEP messages PCRpt, PCInitiate, PCUpd are used to send LSP Reports, LSP setup, and LSP update respectively. The extended PCInitiate message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is used to download or clean up central controller's instructions (CCIs) (SR SID in the scope of this document). The extended PCRpt message described in [I-D.ietf-pce-pcep-extension-for-pce-controller] is also used to report the CCIs (SR SIDs) from PCC to PCE.

[I-D.ietf-pce-pcep-extension-for-pce-controller] specify an object called CCI for the encoding of the central controller's instructions. This document extends the CCI by defining a new object-type for segment routing. The PCEP messages are extended in this document to handle the PCECC operations for SR.

5.3. PCECC Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of PCECC extensions. A PCEP Speaker includes the "PCECC Capability" sub-TLV, described in [I-D.ietf-pce-pcep-extension-for-pce-controller].

A new S-bit is added in the PCECC-CAPABILITY sub-TLV to indicate support for PCECC-SR. A PCC MUST set the S-bit in the PCECC-CAPABILITY sub-TLV and include the SR-PCE-CAPABILITY sub-TLV ([RFC8664]) in the OPEN Object (inside the PATH-SETUP-TYPE-CAPABILITY TLV) to support the PCECC SR extensions defined in this document. If the S-bit is set in the PCECC-CAPABILITY sub-TLV and the SR-PCE-CAPABILITY sub-TLV is not advertised in the OPEN Object, PCE SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBd4 (SR capability was not advertised) and terminate the session.

The rest of the processing is as per [I-D.ietf-pce-pcep-extension-for-pce-controller].

5.4. PCEP session IP address and TED Router ID

A PCE may construct its Traffic Engineering Database (TED) by participating in the IGP ([RFC3630] and [RFC5305] for MPLS-TE; [RFC4203] and [RFC5307] for GMPLS). An alternative is offered by BGP-LS [RFC7752] and [I-D.dhodylee-pce-pcep-ls].

A PCEP [RFC5440] speaker could use any local IP address while creating a TCP session. It is important to link the session IP address with the Router ID in TED for successful PCECC operations.

During PCEP Initialization Phase, the PCC SHOULD advertise the TE mapping information by including the "Node Attributes TLV" [I-D.dhodylee-pce-pcep-ls] with "IPv4/IPv6 Router-ID of Local Node", in the OPEN Object for this purpose. [RFC7752] describes the usage as auxiliary Router-IDs that the IGP might be using, e.g., for TE purposes. If there are more than one auxiliary Router-ID of a given type, then multiple TLVs are used to encode them.

If "IPv4/IPv6 Router-ID" TLV is not present, the TCP session IP address is directly used for mapping purpose.

5.5. LSP Operations

[RFC8664] specify the PCEP extension to allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request a path subject to certain constraint(s) and optimization criteria in SR networks.

The Path Setup Type for segment routing (PST=1) is used on the PCEP session with the Ingress as per [RFC8664].

5.5.1. PCECC Segment Routing (SR)

Segment Routing (SR) as described in [RFC8402] depends on "segments" that are advertised by Interior Gateway Protocols (IGPs). The SR-node allocates and advertises the SID (node, adj, etc) and flood them via the IGP. This document proposes a new mechanism where PCE allocates the SID (label/index/SID) centrally and uses PCEP to advertise them. In some deployments, PCE (and PCEP) are better suited than IGP because of the centralized nature of PCE and direct TCP based PCEP sessions to the node.

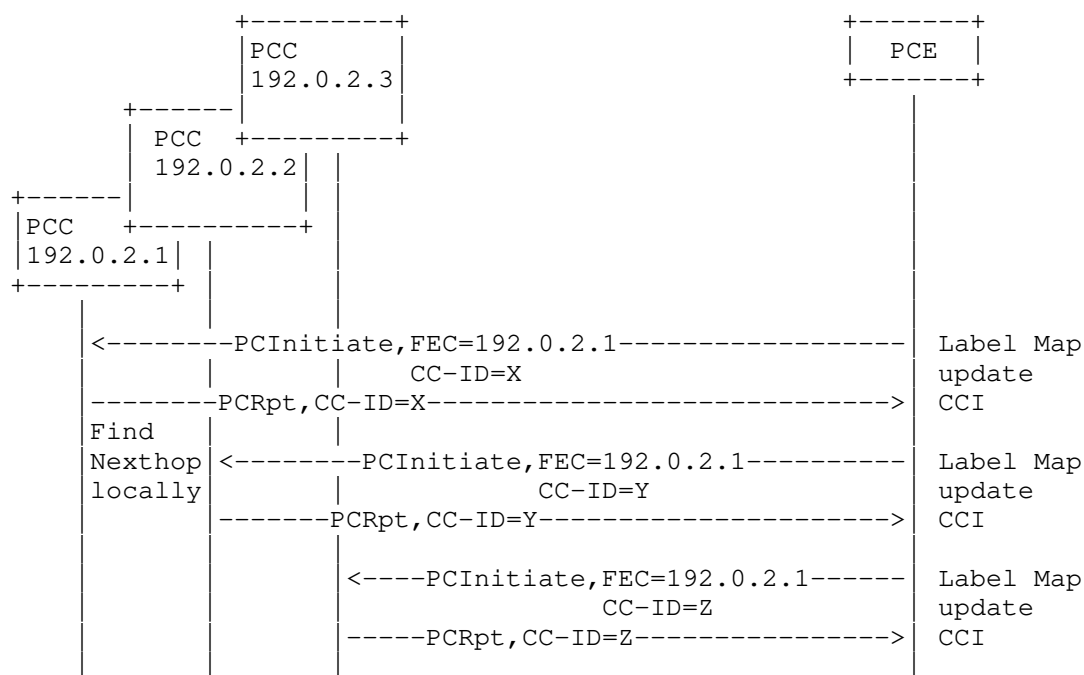
5.5.1.1. PCECC SR Node/Prefix SID allocation

Each node (PCC) is allocated a node-SID by the PCECC. The PCECC sends PCInitiate message to update the label map of each node to all the nodes in the domain. The TE router ID is determined from the TED or from "IPv4/IPv6 Router-ID" Sub-TLV [I-D.dhodylee-pce-pcep-ls], in the OPEN Object Section 5.4.

It is RECOMMENDED that PCEP session with PCECC-SR capability to use a different session IP address during TCP session establishment than the node Router ID in TEDB, to make sure that the PCEP session does not get impacted by the SR Node/Prefix Label maps (Section 5.4).

If a node (PCC) receives a PCInitiate message with a CCI encoding a SID, out of the range set aside for the SR Global Block (SRGB), it MUST send a PCErr message with Error-type=TBD (PCECC failure) and Error-value=TBD (Label out of range) (defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]) and MUST include the SRP object to specify the error is for the corresponding central control instruction via the PCInitiate message.

On receiving the label map, each node (PCC) uses the local routing information to determine the next-hop and download the label forwarding instructions accordingly. The PCInitiate message in this case does not use the LSP object but uses a new FEC object defined in this document.

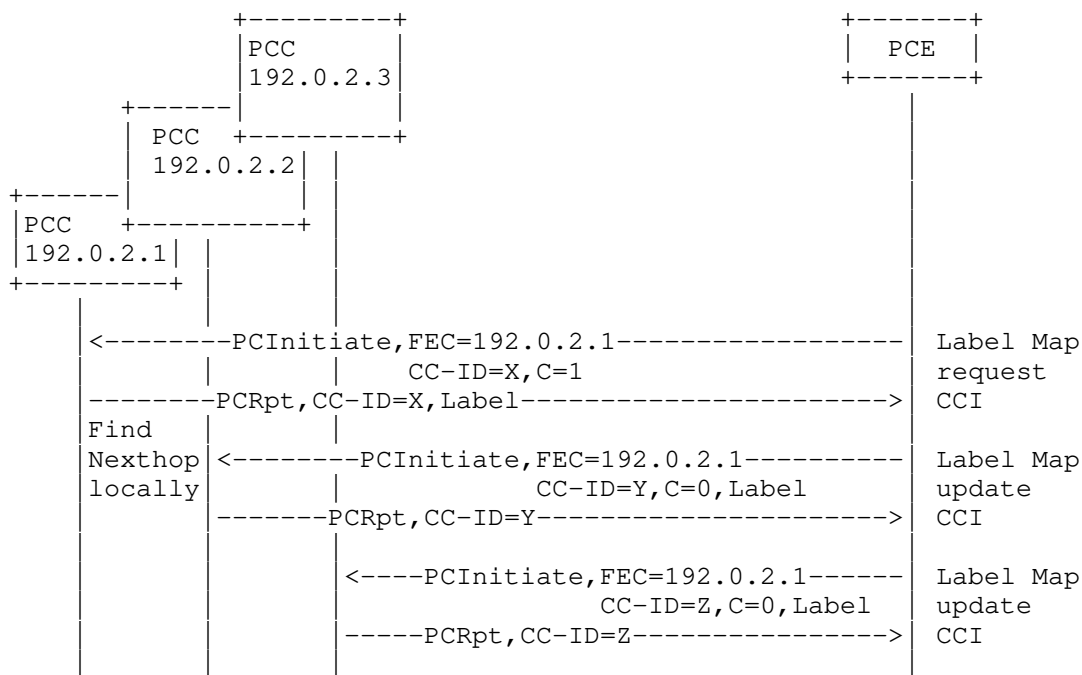


The forwarding behavior and the end result is similar to IGP based "Node-SID" in SR. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node as per [RFC8402].

PCE relies on the Node/Prefix Label clean up using the same PCInitiate message as per [RFC8281].

The above example Figure 1 depicts the FEC and PCEP speakers that uses IPv4 address. Similarly an IPv6 address (such as 2001:DB8::1) can be used during PCEP session establishment in the FEC object as described in this specification.

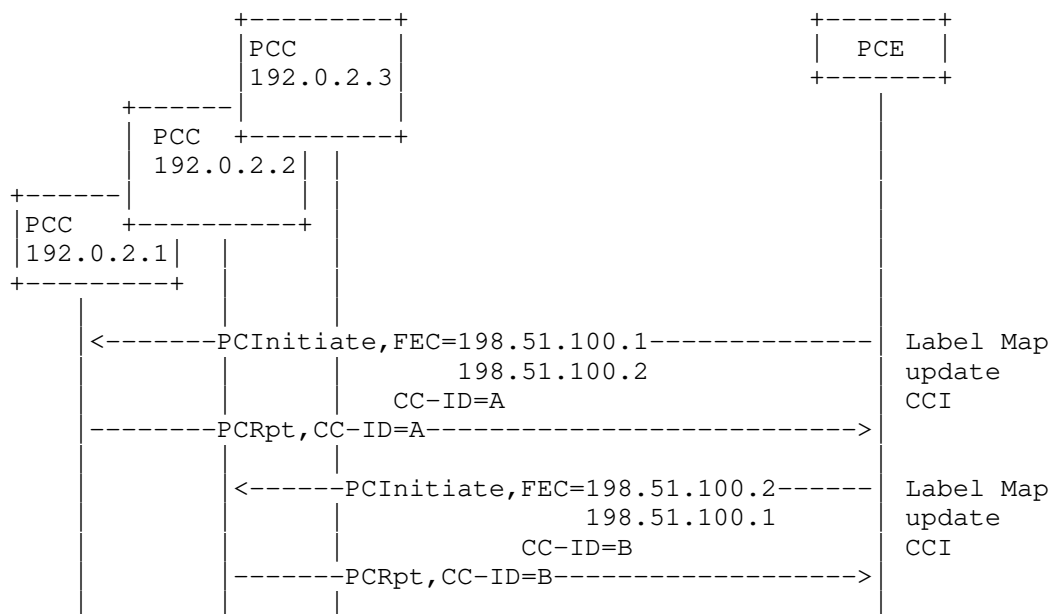
In the case where the label/SID allocation is made by the PCC itself (see Section 5.5.1.6), the PCE could request an allocation to be made by the PCC, and where the PCC would send a PCRpt with the allocated label/SID encoded in the CC-ID object as shown in Figure 2.



It should be noted that in this example, the request is made to the node 192.0.2.1 with C bit set in the CCI object to indicate that the allocation needs to be done by this PCC and it responds with the allocated label/SID to the PCE. The PCE would further inform the other PCCs in the network about the label-map allocation without setting the C bit.

5.5.1.2. PCECC SR Adjacency Label allocation

For PCECC-SR, apart from node-SID, Adj-SID is used where each adjacency is allocated an Adj-SID by the PCECC. The PCECC sends the PCInitiate message to update the label map of each adjacency to the corresponding nodes in the domain. Each node (PCC) download the label forwarding instructions accordingly. Similar to SR Node/Prefix Label allocation, the PCInitiate message in this case does not use the LSP object but uses the new FEC object defined in this document.



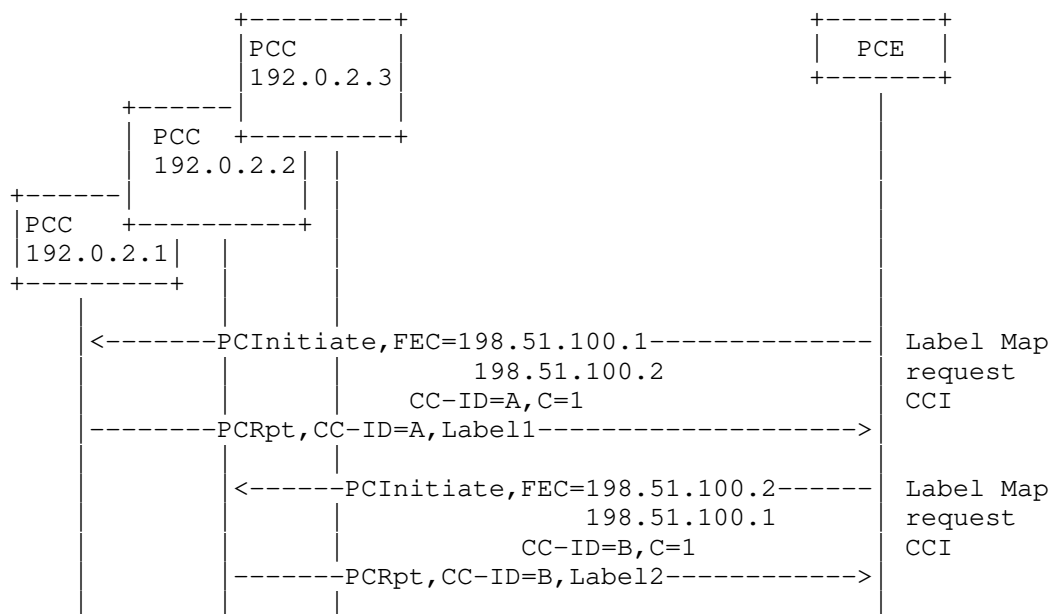
The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SR.

PCE relies on the Adj label clean up using the same PCInitiate message as per [RFC8281].

The above example Figure 3 depicts FEC object and PCEP speakers that uses an IPv4 address. Similarly an IPv6 address (such as 2001:DB8::1, 2001:DB8::2) can be used during the PCEP session establishment in the FEC object as described in this specification.

The handling of adjacencies on the LAN subnetworks is specified in [RFC8402]. PCECC MUST assign Adj-SID for every pair of routers in the LAN. The rest of the protocol mechanism remains the same.

In the case where the label/SID map allocation is made by the PCC itself (see Section 5.5.1.6), the PCE could request an allocation to be made by the PCC, and where the PCC would send a PCRpt with the allocated label/SID encoded in the CC-ID object as shown in Figure 4.



In this example, the request is made to the node 192.0.2.1 with the C bit set in the CCI object to indicate that the allocation needs to be done by this PCC for the adjacency (198.51.100.1 - 198.51.100.2) and it responds with the allocated label/SID to the PCE. Similarly, another request is made to the node 192.0.2.2 with the C bit set in the CCI object to indicate that the allocation needs to be done by this PCC for the adjacency (198.51.100.2 - 198.51.100.1).

5.5.1.3. Redundant PCEs

[I-D.litkowski-pce-state-sync] describes the synchronization mechanism between the stateful PCEs. The SR SIDs allocated by a PCE MUST also be synchronized among PCEs for PCECC SR state synchronization. Note that the SR SIDs are independent of the SR-TE LSPs, and remains intact till any topology change. The redundant PCEs MUST have a common view of all SR SIDs allocated in the domain.

5.5.1.4. Re Delegation and Clean up

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the action needed for CCIs for the static LSPs on a terminated session. Same holds true for the CCI for SR SID as well.

5.5.1.5. Synchronization of Label Allocations

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes the synchronization of Central Controller's Instructions (CCI) via LSP state synchronization as described in [RFC8231] and [RFC8232]. Same procedures are applied for the CCI for SR SID as well.

5.5.1.6. PCC-Based Allocations

The PCE can request the PCC to allocate the label/SID using the PCInitiate message. The C flag in the CCI object is set to 1 to indicate that the allocation needs to be done by the PCC. The PCC would allocate the SID/Label/Index and would report to the PCE using the PCRpt message.

If the value of the SID/Label/Index is 0 and the C flag is set to 1, it indicates that the PCE is requesting the allocation to be done by the PCC. If the SID/Label/Index is 'n' and the C flag is set to 1 in the CCI object, it indicates that the PCE requests a specific value 'n' for the SID/Label/Index. If the allocation is successful, the PCC should report via PCRpt message with the CCI object. Else, it MUST send a PCErr message with Error-Type = TBD ("PCECC failure") and Error Value = TBD ("Invalid CCI") (defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]). If the value of the SID/Label/Index in the CCI object is valid, but the PCC is unable to allocate it, it MUST send a PCErr message with Error-Type = TBD ("PCECC failure") and Error Value = TBD ("Unable to allocate the specified CCI") (defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]).

If the PCC wishes to withdraw or modify the previously assigned label/SID, it MUST send a PCRpt message without any SID/Label/Index or with the SID/Label/Index containing the new value respectively in the CCI object. The PCE would further trigger the removal of the central controller instruction as per this document.

5.5.1.7. Binding SID

A PCECC can allocate and provision the node/prefix/adjacency label (SID) via PCEP. Another SID called binding SID is described in [I-D.ietf-pce-binding-label-sid], the PCECC mechanism can also be used to allocate the binding SID.

A procedure for binding label/SID allocation is described in [I-D.ietf-pce-pcep-extension-for-pce-controller] and is applicable for all path setup types (including SR paths).

6. PCEP Messages

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable-length body made of a set of objects that can be either mandatory or optional. An object is said to be mandatory in a PCEP message when the object must be included for the message to be considered valid. For each PCEP message type, a set of rules is defined that specify the set of objects that the message can carry. An implementation **MUST** form the PCEP messages using the object ordering specified in this document.

6.1. Central Control Instructions

6.1.1. The PCInitiate Message

The PCInitiate message defined in [RFC8281] and extended in [I-D.ietf-pce-pcep-extension-for-pce-controller] is further extended to support SR based central control instructions.

The format of the extended PCInitiate message is as follows:

```
<PCInitiate Message> ::= <Common Header>
                           <PCE-initiated-lsp-list>
```

Where:

<Common Header> is defined in [RFC5440]

```
<PCE-initiated-lsp-list> ::= <PCE-initiated-lsp-request>
                              [<PCE-initiated-lsp-list>]
```

```
<PCE-initiated-lsp-request> ::=
    (<PCE-initiated-lsp-instantiation> |
     <PCE-initiated-lsp-deletion> |
     <PCE-initiated-lsp-central-control>)
```

```
<PCE-initiated-lsp-central-control> ::= <SRP>
                                          (<LSP>
                                           <cci-list>) |
                                          (<FEC>
                                           <CCI>)
```

```
<cci-list> ::= <CCI>
                [<cci-list>]
```

Where:

<PCE-initiated-lsp-instantiation> and
<PCE-initiated-lsp-deletion> are as per
[RFC8281].

The LSP and SRP object is defined in [RFC8231].

When the PCInitiate message is used to distribute SR SIDs, the SRP, the FEC and the CCI objects MUST be present. The error handling for missing SRP or CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. If the FEC object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD5 (FEC object missing).

To clean up, the R (remove) bit in the SRP object and the corresponding FEC and the CCI object are included.

6.1.2. The PCRpt message

The PCRpt message can be used to report the SR central controller instructions received from the PCECC during the state synchronization phase or as an acknowledgment to the PCInitiate message.

The format of the PCRpt message is as follows:

```
<PCRpt Message> ::= <Common Header>
                        <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report> [<state-report-list>]
```

```
<state-report> ::= (<lsp-state-report> |
                    <central-control-report>)
```

```
<lsp-state-report> ::= [<SRP>]
                        <LSP>
                        <path>
```

```
<central-control-report> ::= [<SRP>]
                              (<LSP>
                               <cci-list>) |
                              (<FEC>
                               <CCI>)
```

```
<cci-list> ::= <CCI>
                [<cci-list>]
```

Where:

<path> is as per [RFC8231] and the LSP and SRP object are also defined in [RFC8231].

When PCRpt message is used to report the label map allocations, the FEC and CCI objects MUST be present. The error handling for the missing CCI object is as per [I-D.ietf-pce-pcep-extension-for-pce-controller]. If the FEC object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD5 (FEC object missing).

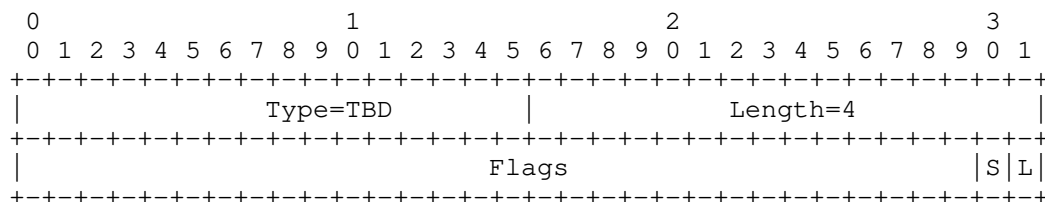
7. PCEP Objects

7.1. OPEN Object

7.1.1. PCECC Capability sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defined the PCECC-CAPABILITY sub-TLV.

A new S-bit is defined in PCECC-CAPABILITY sub-TLV for PCECC-SR:



[Editor's Note - The above figure is included for ease of the reader but should be removed before publication.]

S (PCECC-SR-CAPABILITY - 1 bit - TBD1): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable of PCECC-SR capability and the PCE allocates the Node and Adj label/SID on this session.

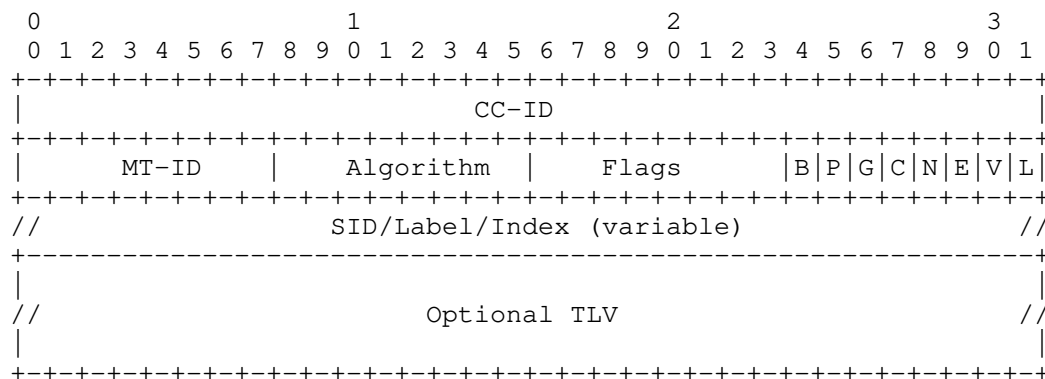
7.2. SR-TE Path Setup

The PATH-SETUP-TYPE TLV is defined in [RFC8408]. A PST value of 1 is used when Path is setup via SR mode as per [RFC8664]. The procedure for SR-TE path setup as specified in [RFC8664] remains unchanged.

7.3. CCI Object

The Central Control Instructions (CCI) Object used by the PCE to specify the controller instructions is defined in [I-D.ietf-pce-pcep-extension-for-pce-controller]. This document defines another object-type for SR-MPLS purpose.

CCI Object-Type is TBD6 for SR-MPLS as below -



The field CC-ID is as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. Following new fields are defined for CCI Object-Type TBD6 -

MT-ID: Multi-Topology ID (as defined in [RFC4915]).

Algorithm: Single octet identifying the algorithm the SID is associated with. See [RFC8665].

Flags: is used to carry any additional information pertaining to the CCI. The following bits are defined -

- * L-Bit (Local/Global): If set, then the value/index carried by the CCI object has local significance. If not set, then the value/index carried by this object has global significance.
- * V-Bit (Value/Index): If set, then the CCI carries an absolute value. If not set, then the CCI carries an index.
- * E-Bit (Explicit-Null): If set, any upstream neighbor of the node that advertised the SID MUST replace the SID with the Explicit-NULL label (0 for IPv4) before forwarding the packet.
- * N-Bit (No-PHP): If set, then the penultimate hop MUST NOT pop the SID before delivering packets to the node that advertised the SID.
- * C-Bit (PCC Allocation): If the bit is set to 1, it indicates that the allocation needs to be done by the PCC for this central controller instruction. A PCE set this bit to request the PCC to make an allocation from its SR label/ID space. A PCC would set this bit to indicate that it has allocated the CC-ID and report it to the PCE.
- * Following bits are applicable when the SID represents an Adj-SID only, it MUST be ignored for others -
 - + G-Bit (Group): When set, the G-Flag indicates that the Adj-SID refers to a group of adjacencies (and therefore MAY be assigned to other adjacencies as well).
 - + P-Bit (Persistent): When set, the P-Flag indicates that the Adj-SID is persistently allocated, i.e., the Adj-SID value remains consistent across router restart and/or interface flap.
 - + B-Bit (Backup): If set, the Adj-SID refers to an adjacency that is eligible for protection (e.g., using IP Fast Reroute

or MPLS-FRR (MPLS-Fast Reroute) as described in Section 2.1 of [RFC8402].

- + All unassigned bits MUST be set to zero at transmission and ignored at receipt.

SID/Label/Index: According to the V and L flags, it contains either:

A 32-bit index defining the offset in the SID/Label space advertised by this router.

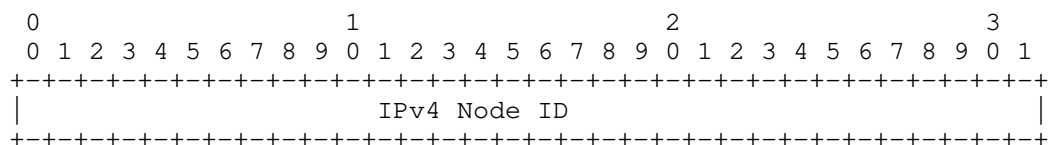
A 24-bit label where the 20 rightmost bits are used for encoding the label value.

7.4. FEC Object

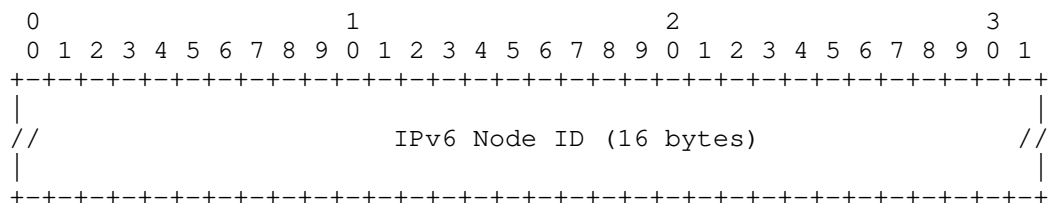
The FEC Object is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message.

FEC Object-Class is TBD3.

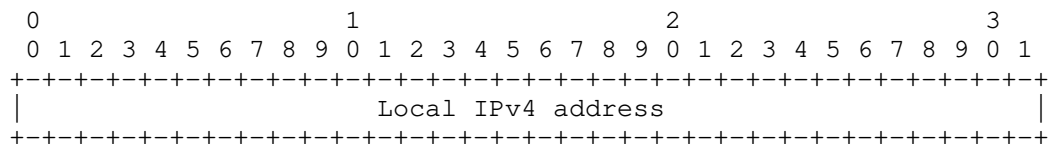
FEC Object-Type is 1 'IPv4 Node ID'.



FEC Object-Type is 2 'IPv6 Node ID'.



FEC Object-Type is 3 'IPv4 Adjacency'.




```

|----- Remote IPv4 address -----|
+-----+

```

FEC Object-Type is 4 'IPv6 Adjacency'.

```

      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
+-----+
|
//               Local IPv6 address (16 bytes)               //
|
+-----+
|
//               Remote IPv6 address (16 bytes)               //
|
+-----+

```

FEC Object-Type is 5 'Unnumbered Adjacency with IPv4 NodeIDs'.

```

      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
+-----+
|               Local Node-ID               |
+-----+
|               Local Interface ID           |
+-----+
|               Remote Node-ID              |
+-----+
|               Remote Interface ID         |
+-----+

```

FEC Object-Type is 6 'Linklocal IPv6 Adjacency'.

```

      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
+-----+
//               Local IPv6 address (16 octets)               //
+-----+
|               Local Interface ID           |
+-----+
//               Remote IPv6 address (16 octets)               //
+-----+
|               Remote Interface ID         |
+-----+

```

The FEC objects are as follows:

IPv4 Node ID: where IPv4 Node ID is specified as an IPv4 address of the Node. FEC Object-type is 1, and the Object-Length is 4 in this case.

IPv6 Node ID: where IPv6 Node ID is specified as an IPv6 address of the Node. FEC Object-type is 2, and the Object-Length is 16 in this case.

IPv4 Adjacency: where Local and Remote IPv4 address is specified as pair of IPv4 addresses of the adjacency. FEC Object-type is 3, and the Object-Length is 8 in this case.

IPv6 Adjacency: where Local and Remote IPv6 address is specified as pair of IPv6 addresses of the adjacency. FEC Object-type is 4, and the Object-Length is 32 in this case.

Unnumbered Adjacency with IPv4 NodeID: where a pair of Node ID / Interface ID tuple is used. FEC Object-type is 5, and the Object-Length is 16 in this case.

Linklocal IPv6 Adjacency: where a pair of (global IPv6 address, interface ID) tuple is used. FEC object-type is 6, and the Object-Length is 40 in this case.

8. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to RFC 7942.]

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature.

It is up to the individual working groups to use this information as they see fit".

8.1. Huawei's Proof of Concept based on ONOS

The PCE function was developed in the ONOS open source platform. This extension was implemented on a private version as a proof of concept for PCECC.

- o Organization: Huawei
- o Implementation: Huawei's PoC based on ONOS
- o Description: PCEP as a southbound plugin was added to ONOS. To support PCECC-SR, an earlier version of this I-D was implemented. Refer <https://wiki.onosproject.org/display/ONOS/PCEP+Protocol>
- o Maturity Level: Prototype
- o Coverage: Partial
- o Contact: satishk@huawei.com

9. Security Considerations

The security considerations described in [I-D.ietf-pce-pcep-extension-for-pce-controller] apply to the extensions described in this document.

As per [RFC8231], it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525] (unless explicitly set aside in [RFC8253]).

10. Manageability Considerations

10.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC SR capability as a global configuration. The implementation SHOULD also allow setting the local IP address used by the PCEP session.

10.2. Information and Data Models

[RFC7420] describes the PCEP MIB, this MIB can be extended to get the PCECC SR capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC SR capability.

10.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

10.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440] and [RFC8231].

10.5. Requirements On Other Protocols

PCEP extensions defined in this document do not put new requirements on other protocols.

10.6. Impact On Network Operations

PCEP implementation SHOULD allow a limit to be placed on the rate of PCLabelUpd messages sent by PCE and processed by PCC. It SHOULD also allow sending a notification when a rate threshold is reached.

11. IANA Considerations

11.1. PCECC-CAPABILITY sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY sub-TLV and requests that IANA to create a new sub-registry to manage the value of the PCECC-CAPABILITY sub-TLV's Flag field.

IANA is requested to allocate a new bit in the PCECC-CAPABILITY sub-TLV Flag Field sub-registry, as follows:

Bit	Description	Reference
TBD1	SR	This document

11.2. PCEP Object

IANA is requested to allocate new code-points for the new FEC object and a new Object-Type for CCI object in "PCEP Objects" sub-registry as follows:

Object-Class Value	Name	Object-Type	Reference
TBD3	FEC	1: IPv4 Node ID	This document
		2: IPv6 Node ID	This document
		3: IPv4 Adjacency	This document
		4: IPv6 Adjacency	This document
		5: Unnumbered Adjacency with IPv4 NodeID	This document
		6: Linklocal IPv6 Adjacency	This document
TBD	CCI	TBD6: SR-MPLS	This document

11.3. PCEP-Error Object

IANA is requested to allocate a new error-value within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry for the following errors:

Error-Type -----	Meaning -----	
6	Mandatory Object missing.	
19	Error-value = TBD5 :	FEC object missing
	Invalid operation.	
	Error-value = TBD4 :	SR capability was not advertised

11.4. CCI Object Flag Field for SR

IANA is requested to create a new sub-registry to manage the Flag field of the CCI Object-Type=TBD6 for SR called "CCI Object Flag Field for SR". New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

- o Bit number (counting from bit 0 as the most significant bit)
- o Capability description
- o Defining RFC

Following bits are defined for the CCI Object flag field for SR in this document as follows:

Bit	Description	Reference
0-7	Unassigned	This document
8	B-Bit - Backup	This document
9	P-Bit - Persistent	This document
10	G-Bit - Group	This document
11	C-Bit - PCC Allocation	This document
12	N-Bit - No-PHP	This document
13	E-Bit - Explicit-Null	This document
14	V-Bit - Value/Index	This document
15	L-Bit - Local/Global	This document

12. Acknowledgments

We would like to thank Robert Tao, Changjing Yan, Tieying Huang, Avantika, and Aijun Wang for their useful comments and suggestions.

13. References

13.1. 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>>.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <<https://www.rfc-editor.org/info/rfc3630>>.
- [RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <<https://www.rfc-editor.org/info/rfc4203>>.
- [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>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <<https://www.rfc-editor.org/info/rfc5305>>.
- [RFC5307] Kompella, K., Ed. and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, DOI 10.17487/RFC5307, October 2008, <<https://www.rfc-editor.org/info/rfc5307>>.

- [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>>.
- [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <<https://www.rfc-editor.org/info/rfc7752>>.
- [RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <<https://www.rfc-editor.org/info/rfc7942>>.
- [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>>.
- [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>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.
- [RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <<https://www.rfc-editor.org/info/rfc8664>>.

[I-D.ietf-pce-pcep-extension-for-pce-controller]
Li, Z., Peng, S., Negi, M., 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-08 (work in progress), November 2020.

13.2. Informative References

- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <<https://www.rfc-editor.org/info/rfc3031>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.
- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.

- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <<https://www.rfc-editor.org/info/rfc8232>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [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>>.
- [RFC8660] Bashandy, A., Ed., Filtsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/info/rfc8660>>.
- [RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filtsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <<https://www.rfc-editor.org/info/rfc8665>>.
- [RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filtsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <<https://www.rfc-editor.org/info/rfc8667>>.
- [I-D.ietf-teas-pcecc-use-cases]
Li, Z., Khasanov, B., Dhody, D., Zhao, Q., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-06 (work in progress), September 2020.

[I-D.ietf-pce-pcep-yang]

Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-15 (work in progress), October 2020.

[I-D.ietf-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.", draft-ietf-pce-binding-label-sid-05 (work in progress), October 2020.

[I-D.litkowski-pce-state-sync]

Litkowski, S., Sivabalan, S., Li, C., and H. Zheng, "Inter Stateful Path Computation Element (PCE) Communication Procedures.", draft-litkowski-pce-state-sync-09 (work in progress), November 2020.

[I-D.dhodylee-pce-pcep-ls]

Dhody, D., Peng, S., Lee, Y., Ceccarelli, D., Wang, A., and G. Mishra, "PCEP extensions for Distribution of Link-State and TE Information", draft-dhodylee-pce-pcep-ls-19 (work in progress), November 2020.

[I-D.dhody-pce-pcep-extension-pce-controller-srv6]

Li, Z., Peng, S., Geng, X., and M. Negi, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for SRv6", draft-dhody-pce-pcep-extension-pce-controller-srv6-05 (work in progress), November 2020.

Appendix A. Contributor Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com

Satish Karunanithi
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: satishk@huawei.com

Adrian Farrel
Juniper Networks, Inc
UK

EMail: adrian@olddog.co.uk

Xuesong Geng
Huawei Technologies
China

Email: gengxuesong@huawei.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

Katherine Zhao
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA

EMail: katherine.zhao@huawei.com

Boris Zhang
Telus Ltd.
Toronto
Canada

EMail: boris.zhang@telus.com

Alex Tokar
Cisco Systems
Slovak Republic

EMail: atokar@cisco.com

Authors' Addresses

Zhenbin Li
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: pengshuping@huawei.com

Mahendra Singh Negi
RtBrick Inc
N-17L, 18th Cross Rd, HSR Layout
Bangalore, Karnataka 560102
India

EMail: mahend.ietf@gmail.com

Quintin Zhao
Etheric Networks
1009 S CLAREMONT ST
SAN MATEO, CA 94402
USA

EMail: qzhao@ethericnetworks.com

Chao Zhou
HPE

EMail: chaozhou_us@yahoo.com