

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 6, 2019

C. Barth
Juniper Networks, Inc.
M. Koldychev
S. Sivabalan
Cisco Systems, Inc.
C. Li
Huawei Technologies
March 05, 2019

PCEP extension to support Segment Routing Policy Candidate Paths
draft-barth-pce-segment-routing-policy-cp-02

Abstract

This document introduces a mechanism to specify an Segment Routing (SR) policy, as a collection of SR candidate paths. An SR policy is identified by <headend, color, end-point> tuple. An SR policy can contain one or more candidate paths where each candidate path is identified in PCEP via an PLSP-ID. This document proposes extension to PCEP to support association among candidate paths of a given SR policy. The mechanism proposed in this document is applicable to both MPLS and IPv6 data planes of SR.

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.

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 6, 2019.

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
2. Terminology	4
3. Motivation	4
3.1. Group Candidate Paths belonging to the same SR policy . .	5
3.2. Instantiation of SR policy candidate paths	5
3.3. Avoid computing lower preference candidate paths	5
3.4. Minimal signaling overhead	5
4. Overview	6
5. SR Policy Association Group	7
5.1. SR Policy Association Group Policy Identifiers TLV . . .	8
5.2. SR Policy Association Group Candidate Path Identifiers TLV	9
5.3. SR Policy Association Group Candidate Path Attributes TLV	10
6. Examples	11
6.1. PCC Initiated SR Policy with single candidate-path . . .	11
6.2. PCC Initiated SR Policy with multiple candidate-paths . .	11
6.3. PCE Initiated SR Policy with single candidate-path . . .	12
6.4. PCE Initiated SR Policy with multiple candidate-paths . .	13
7. IANA Considerations	13
7.1. Association Type	13
7.2. PCEP Errors	13
7.3. SRPAG TLVs	14
8. Security Considerations	14
9. Acknowledgement	15
10. References	15
10.1. Normative References	15
10.2. Informative References	16
Appendix A. Contributors	16
Authors' Addresses	17

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.

PCEP Extensions for Establishing Relationships Between Sets of LSPs [I-D.ietf-pce-association-group] introduces a generic mechanism to create a grouping of LSPs which can then be used to define associations between a set of LSPs and a set of attributes (such as configuration parameters or behaviors) and is equally applicable to stateful PCE (active and passive modes) and stateless PCE.

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.

An SR policy contains one or more candidate paths where one or more such paths can be computed via PCE. This document specifies PCEP extensions to signal additional information to map candidate paths to their SR policies. Each candidate path maps to a unique PLSP-ID in PCEP. By associating multiple candidate paths together, a PCE becomes aware of the hierarchical structure of an SR policy. Thus the PCE can take computation and control decisions about the candidate paths, with the additional knowledge that these candidate paths belong to the same SR policy. This is accomplished via the use of the existing PCEP Association object, by defining a new association type specifically for associating SR candidate paths into a single SR policy.

[Editor's Note- Currently it is assumed that each candidate path has only one ERO (SID-List) within the scope of this document. A future

update or another document will deal with a way to allow multiple ERO/SID-Lists for a candidate path within PCEP.]

2. Terminology

The following terminologies 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].

Association parameters: As described in [I-D.ietf-pce-association-group], the combination of the mandatory fields Association type, Association ID and Association Source in the ASSOCIATION object uniquely identify the association group. If the optional TLVs - Global Association Source or Extended Association ID are included, then they MUST be included in combination with mandatory fields to uniquely identify the association group.

Association information: As described in [I-D.ietf-pce-association-group], the ASSOCIATION object could also include other optional TLVs based on the association types, that provides 'information' related to the association type.

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.

3. Motivation

The new Association Type (SR Policy Association) and the new TLVs for the ASSOCIATION object, defined in this document, allow a PCEP peer to exchange additional parameters of SR candidate paths and of their parent SR policy. For the SR policy, the parameters are: color and endpoint. For the candidate path, the parameters are: protocol origin, originator, discriminator and preference. [I-D.ietf-spring-segment-routing-policy] describes the concept of SR Policy and these parameters.

The motivation for signaling these parameters is summarized in the following subsections.

3.1. Group Candidate Paths belonging to the same SR policy

Since each candidate path of an SR policy appears as a different LSP (identified via a PLSP-ID) in PCEP, it is useful to group together all the candidate paths that belong to the same SR policy. Furthermore, it is useful for the PCE to have knowledge of the SR candidate path parameters such as color, protocol origin, discriminator, and preference.

3.2. Instantiation of SR policy candidate paths

A PCE may want to instantiate one or more candidate paths on the PCC, as specified in [RFC8281]. In this scenario, the PCE needs to signal to a PCC <headend, color, end-point, originator, discriminator, preference> tuple using which the PCC can instantiate a candidate path for the SR policy identified. Current PCEP standards (as of the time of this writing) do not provide a way to signal color and preference. Although end-point can be signaled via the PCEP END-POINTS object, this object may not be suitable because the end-point to which the path is computed is not required to be the same IPv4/IPv6 address as the actual endpoint of the SR policy. Thus, a separate way to specify SR policy's end-point is provided in this document.

3.3. Avoid computing lower preference candidate paths

When a PCE knows that a given set of candidate paths all belong to the same SR policy, then path computation MAY be done on only the highest preference candidate-path(s). Path computation for lower preference paths is not necessary if one or two higher preference paths are already computed. Since computing their paths will not affect traffic steering, it MAY be postponed until the higher preference paths become invalid, thus saving computation resources on the PCE.

3.4. Minimal signaling overhead

When an SR policy contains multiple candidate paths computed by a PCE, such candidate paths can be created, updated and deleted independently of each other. This is achieved by making each candidate path correspond to a unique LSP (identified via PLSP-ID). For example, if an SR policy has 4 candidate paths, then if the PCE wants to update one of those candidate paths, only one set of PCUpd and PCRpt messages needs to be exchanged.

4. Overview

As per [I-D.ietf-pce-association-group], LSPs are placed into an association group. In this document, each LSP corresponds to a candidate path of an SR policy, and the association group corresponds to the SR policy itself. Segment-lists within a candidate path are not represented by different LSPs (and identified via PLSP-IDs).

[Editor's Note - The subject of encoding multiple segment lists within a candidate path is left to a future document and is not specified in this document. It is not a good idea to have each segment-list correspond to a different LSP/PLSP-ID, because when the PCC wants to get a path, it must know in advance how many multipaths (i.e., segment-lists) there will be and create that many LSPs/PLSP-IDs. For example, if the PCC supports 32 multipaths, then it must delegate 32 LSPs/PLSP-IDs for every candidate path, which may not be scalable.]

A new Association Type is defined in this document, based on the generic ASSOCIATION object. Association type = TBD1 "SR Policy Association Type" for SR Policy Association Group (SRPAG).

The SRPAG Association is only meant to be used for SR LSPs and with PCEP peers which advertise SR capability.

An Association object of SRPAG group contains TLVs that carry Association Information. The association information can be subdivided into three parts: Policy identifiers, Candidate path identifiers, and Candidate path attributes.

Policy Identifiers uniquely identify the SR policy to which a given LSP belongs, within the context of the head-end. Policy Identifiers MUST be the same for all candidate paths in the same SRPAG. Policy Identifiers MUST NOT change for a given LSP during its lifetime. Policy Identifiers MUST be different for different SRPAG associations. When these rules are not satisfied, the PCE MUST send a PCERR message with Error Code = 26 "Association Error", Error Type = TBD5 "Conflicting SRPAG TLV". Policy Identifiers consist of:

- o Color of SR policy.
- o End-point of SR policy.

Candidate Path Identifiers uniquely identify the SR candidate path within the context of an SR policy. Candidate path Identifiers MUST NOT change for a given LSP during its lifetime. Candidate path Identifiers MUST be different for different LSPs within the same SRPAG. When these rules are not satisfied, the PCE MUST send a PCERR

message with Error Code = 26 "Association Error", Error Type = TBD5 "Conflicting SRPAG TLV". Candidate path Identifiers consist of:

- o Protocol Origin of candidate path.
- o Originator of candidate path.
- o Discriminator of candidate path.

Candidate Path Attributes MUST NOT be used to identify the candidate path. Candidate path attributes carry additional information about the candidate path and MAY change during the lifetime of the LSP. Candidate path Attributes consist of:

- o Preference of candidate path.

As described in [RFC8231], an LSP is uniquely identified in PCEP via PLSP-ID.

A mapping between the Association Parameters (see Section 2) and Policy Identifiers (the Color and End-point) needs to be maintained. The mapping is left up to the implementation. An implementation MAY choose Association Parameters in such a way that every possible Color and End-point maps to a unique value of Association Parameters, which may require the use of Extended Association ID TLV. Alternatively, an implementation MAY implement a lookup table to generate Association Parameters incrementally as new Color and End-point values are created, which may not require the use of Extended Association ID TLV.

As per the processing rules specified in section 5.4 of [I-D.ietf-pce-association-group], if a PCEP speaker does not support the SRPAG association type, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 1 "Association-type is not supported". Please note that the corresponding PCEP session is not reset.

5. SR Policy Association Group

Two ASSOCIATION object types for IPv4 and IPv6 are defined in [I-D.ietf-pce-association-group]. The ASSOCIATION object includes "Association type" indicating the type of the association group. This document adds a new Association type.

Association type = TBD1 "SR Policy Association Type" for SR Policy Association Group (SRPAG).

The operator configured Association Range SHOULD NOT be set for this association type and MUST be ignored.

SRPAG MUST carry additional TLVs to communicate Association Information. This document specifies three new TLVs to carry Association Information: SRPAG-POL-ID-TLV, SRPAG-CPATH-ID-TLV, SRPAG-CPATH-ATTR-TLV. These three TLVs encode the Policy Identifiers, Candidate path Identifiers and Candidate path Attributes, respectively. When any of the mandatory TLVs are missing from the SRPAG association object, the PCE MUST send a PCErr message with Error Code = 26 "Association Error", Error Type = TBD6 "Missing mandatory SRPAG TLV".

A given LSP MUST belong to at most one SRPAG, since a candidate path cannot belong to multiple SR policies. If a PCEP speaker receives a PCEP message with more than one SRPAG for an LSP, then the PCEP speaker MUST send a PCErr message with Error-Type 26 "Association Error" and Error-Value TBD7 "Multiple SRPAG for one LSP". If the message is a PCRpt message, then the PCEP speaker MUST close the PCEP connection. Closing the PCEP connection is necessary because ignoring PCRpt messages may lead to inconsistent LSP DB state between the two PCEP peers.

If the PCEP speaker receives the SRPAG association when the SR capability (as per [I-D.ietf-pce-segment-routing] or [I-D.negi-pce-segment-routing-ipv6]) was not exchanged, the PCEP speaker MUST send a PCErr message with Error-Type 26 "Association Error" and Error-Value TBD8 "Use of SRPAG without SR capability exchange". If the Path Setup Type (PST) of the LSP in SRPAG is not set to SR or SRv6, then the PCEP speaker MUST send a PCErr message with Error-Type 26 "Association Error" and Error-Value TBD9 "non-SR LSP in SRPAG".

5.1. SR Policy Association Group Policy Identifiers TLV

The SRPOLICY-POL-ID TLV is a mandatory TLV for the SRPAG Association. Only one SRPOLICY-POL-ID TLV can be carried and only the first occurrence is processed and any others MUST be ignored.

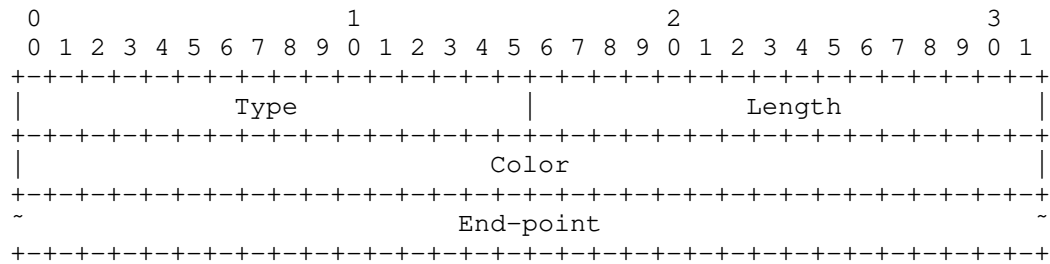


Figure 1: The SRPOLICY-POL-ID TLV format

Type: TBD2 for "SRPOLICY-POL-ID" TLV.

Length: 8 or 20, depending on length of End-point (IPv4 or IPv6)

Color: any unsigned 32-bit number.

End-point: can be either IPv4 or IPv6, depending on whether the policy endpoint has IPv4 or IPv6 address. This value may be different from the one contained in the END-POINTS object, or in the LSP IDENTIFIERS TLV of the LSP object. Endpoint is meant to strictly correspond to the endpoint of the SR policy, as it is defined in [I-D.ietf-spring-segment-routing-policy].

5.2. SR Policy Association Group Candidate Path Identifiers TLV

The SRPOLICY-CPATH-ID TLV is a mandatory TLV for the SRPAG Association. Only one SRPOLICY-CPATH-ID TLV can be carried and only the first occurrence is processed and any others MUST be ignored.

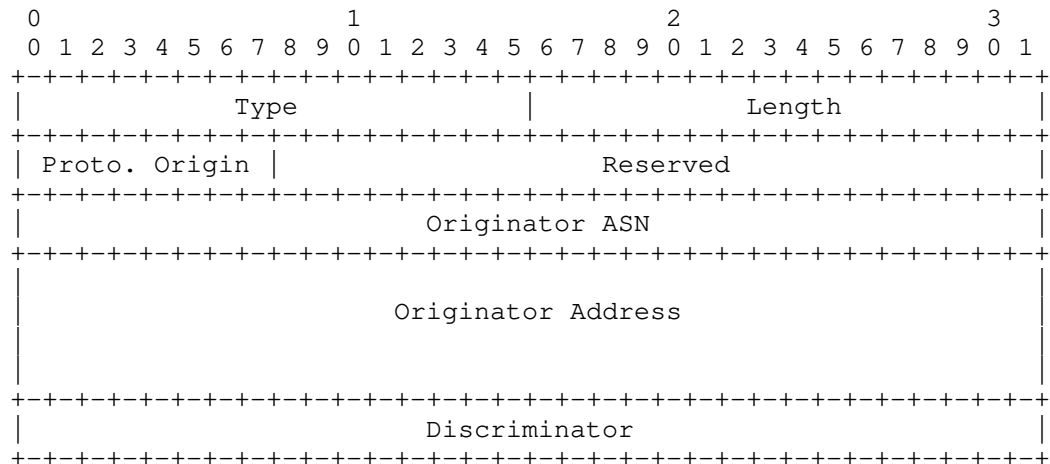


Figure 2: The SRPOLICY-CPATH-ID TLV format

Type: TBD3 for "SRPOLICY-CPATH-ID" TLV.

Length: 28.

Protocol Origin: 8-bit value that encodes the protocol origin, as specified in [I-D.ietf-spring-segment-routing-policy] Section 2.3.

Reserved: MUST be set to zero on transmission and ignored on receipt.

Originator ASN: Represented as 4 byte number, part of the originator identifier, as specified in [I-D.ietf-spring-segment-routing-policy] Section 2.4.

Originator Address: Represented as 128 bit value where IPv4 address are encoded in lowest 32 bits, part of the originator identifier, as specified in [I-D.ietf-spring-segment-routing-policy] Section 2.4.

Discriminator: 32-bit value that encodes the Discriminator of the candidate path.

5.3. SR Policy Association Group Candidate Path Attributes TLV

The SRPOLICY-CPATH-ATTR TLV is an optional TLV for the SRPAG Association. Only one SRPOLICY-CPATH-ATTR TLV can be carried and only the first occurrence is processed and any others MUST be ignored.

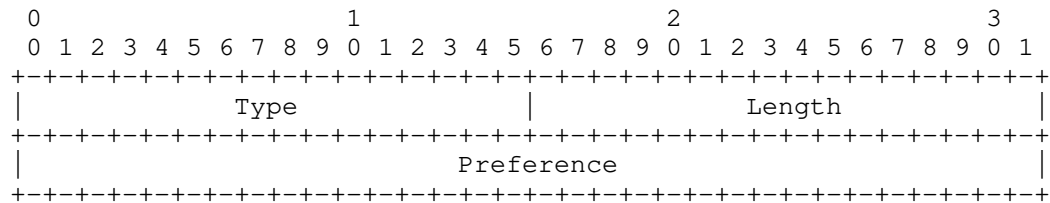


Figure 3: The SRPOLICY-CPATH-ATTR TLV format

Type: TBD4 for "SRPOLICY-CPATH-ATTR" TLV.

Length: 4.

Preference: Numerical preference of the candidate path, as specified in [I-D.ietf-spring-segment-routing-policy] Section 2.7.

If the TLV is missing, a default preference of 100 as specified in [I-D.ietf-spring-segment-routing-policy] is used.

6. Examples

6.1. PCC Initiated SR Policy with single candidate-path

PCReq and PCRep messages are exchanged in the following sequence:

1. PCC sends PCReq message to the PCE, encoding the SRPAG ASSOCIATION object and TLVs in the PCReq message.
2. PCE returns the path in PCRep message, and echoes back the SRPAG object that was used in the computation.

PCRpt and PCUpd messages are exchanged in the following sequence:

1. PCC sends PCRpt message to the PCE, including the LSP object and the SRPAG ASSOCIATION object.
2. PCE computes path, possibly making use of the Association Information from the SRPAG ASSOCIATION object.
3. PCE updates the SR policy candidate path's ERO using PCUpd message.

6.2. PCC Initiated SR Policy with multiple candidate-paths

PCReq and PCRep messages are exchanged using the sequence specified in section 6.1 with individual query for each candidate-path.

PCRpt and PCUpd messages are exchanged in the following sequence:

1. Step 1: For each candidate path of the SR policy, the PCC generates a different PLSP-ID and symbolic-name and sends multiple PCRpt messages (or one message with multiple LSP objects) to the PCE. Each LSP object is followed by SRPAG ASSOCIATION object with identical Color and Endpoint values.
2. Step 2: PCE takes into account that all the LSPs belong to the same SR policy. PCE prioritizes computation for the highest preference LSP and sends PCUpd message(s) back to the PCC.
3. Step 3: If a new candidate path is added on the PCC by the operator, then a new PLSP-ID and symbolic name is generated for that candidate path and a new PCRpt is sent to the PCE.
4. Step 4: If an existing candidate path is removed from the PCC by the operator, then that PLSP-ID is deleted from the PCE by sending PCRpt with the R-flag in the LSP object set.

6.3. PCE Initiated SR Policy with single candidate-path

A candidate-path is created using the following steps:

1. PCE sends PCInitiate message, as usual containing the SRPAG Association object. PCE needs to generate a symbolic-name for this LSP that will not clash with other symbolic names on the same PCC.
2. PCC uses the color, endpoint and preference from the SRPAG object to create a new candidate path. If no SR policy exists to hold the candidate path, then a new SR policy is created to hold the new candidate-path. The Originator of the candidate path is set to be the address of the PCE that is sending the PCInitiate message.
3. PCC allocates a locally unique PLSP-ID for the newly created candidate path. This PLSP-ID is sent to the PCE in the PCRpt message.

A candidate-path is deleted using the following steps:

1. PCE sends PCInitiate message, setting the R-flag in the LSP object.
2. PCC uses the PLSP-ID from the LSP object to find the candidate path and delete it. If this is the last candidate path under the SR policy, then the containing SR policy is deleted as well.

6.4. PCE Initiated SR Policy with multiple candidate-paths

A candidate-path is created using the following steps:

1. PCE sends a separate PCInitiate message for every candidate path that it wants to create, or it sends multiple LSP objects within a single PCInitiate message. Each candidate-path must get a unique symbolic-name generated on the PCE. SRPAG object is sent for every LSP in the PCInitiate message.
2. PCC creates multiple candidate paths under the same SR policy, identified by Color and Endpoint. PCC generates a unique PLSP-ID for every candidate path.
3. PCC allocates a locally unique PLSP-ID for each newly created candidate path. This PLSP-ID is sent to the PCE in the PCRpt message.

A candidate path is deleted using the following steps:

1. PCE sends PCInitiate message, setting the R-flag in the LSP object.
2. PCC uses the PLSP-ID from the LSP object to find the candidate path and delete it.

7. IANA Considerations

7.1. Association Type

This document defines a new association type: SR Policy Association Group (SRPAG). IANA is requested to make the assignment of a new value for the sub-registry "ASSOCIATION Type Field" (request to be created in [I-D.ietf-pce-association-group]), as follows:

Association Type Value	Association Name	Reference
TBD1	SR Policy Association	This document

7.2. PCEP Errors

This document defines three new Error-Values within the "Association Error" Error-Type. IANA is requested to allocate new error values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry, as follows:

Error Type	Error Value	Meaning	Reference
29	TBD5	Conflicting SRPAG TLV	This document
29	TBD6	Missing mandatory SRPAG TLV	This document
29	TBD7	Multiple SRPAG for one LSP	This document
29	TBD8	Use of SRPAG without SR capability exchange	This document
29	TBD9	non-SR LSP in SRPAG	This document

7.3. SRPAG TLVs

This document defines three new TLVs for carrying additional information about SR policy and SR candidate paths. 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
TBD2	SRPOLICY-POL-ID	This document
TBD3	SRPOLICY-CPATH-ID	This document
TBD4	SRPOLICY-CPATH-ATTR	This document

8. Security Considerations

This document defines one new type for association, which do not add any new security concerns beyond those discussed in [RFC5440], [RFC8231], [I-D.ietf-pce-segment-routing], [I-D.negi-pce-segment-routing-ipv6] and [I-D.ietf-pce-association-group] in itself.

The information carried in the SRPAG Association object, as per this document is related to SR Policy. It often reflects information that can also be derived from the SR Database, but association provides a much easier grouping of related LSPs and messages. The SRPAG association could provides an adversary with the opportunity to

eavesdrop on the relationship between the LSPs. Thus securing the PCEP session using Transport Layer Security (TLS) [RFC8253], as per the recommendations and best current practices in [RFC7525], is RECOMMENDED.

9. Acknowledgement

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>>.
- [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-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-02 (work in progress), October 2018.

[I-D.ietf-pce-association-group]

Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "PCEP Extensions for Establishing Relationships Between Sets of LSPs", draft-ietf-pce-association-group-07 (work in progress), December 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.

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>>.

[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>>.

[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>>.

[I-D.negi-pce-segment-routing-ipv6]

Negi, M., Li, C., Sivabalan, S., and P. Kaladharan, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-negi-pce-segment-routing-ipv6-04 (work in progress), February 2019.

Appendix A. Contributors

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

Email: dhruv.ietf@gmail.com

Authors' Addresses

Colby Barth
Juniper Networks, Inc.

Email: cbarth@juniper.net

Mike Koldychev
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario K2K 3E8
Canada

Email: mkoldych@cisco.com

Siva Sivabalan
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario K2K 3E8
Canada

Email: msiva@cisco.com

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: chengli13@huawei.com

Networking Working Group
Internet-Draft
Intended status: Standards Track
Expires: May 19, 2019

Ran. Chen
Zheng. Zhang
ZTE Corporation
November 15, 2018

PCEP Extensions for BIER
draft-chen-pce-bier-04

Abstract

Bit Index Explicit Replication (BIER)-TE shares architecture and packet formats with BIER as described in [I-D.ietf-bier-architecture]. 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. BIER-TE Path can be derived from a Path Computation Element (PCE).

This document specifies extensions to the Path Computation Element Protocol (PCEP) to handle requests and responses for the computation of paths for 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 19, 2019.

Copyright Notice

Copyright (c) 2018 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. BIER PCEP Message Extensions	3
4.1. BIER Capability Advertisement	3
4.1.1. The OPEN Object	3
4.1.1.1. The BIER PCE Capability TLV	3
4.2. Path Computation Request/Reply Message Extensions	4
4.2.1. The RP/SPR Object	4
4.2.2. The New BIER END-POINT Object	5
4.2.3. ERO Object	5
4.2.3.1. BIER-ERO Subobject	6
4.2.4. RRO Object	7
4.2.4.1. RRO Processing	7
5. Security Considerations	7
6. IANA Considerations	7
6.1. PCEP Objects	7
6.2. PCEP-Error Objects and Types	8
6.3. PCEP TLV Type Indicators	8
6.4. New Path Setup Type	8
7. References	8
7.1. Normative references	9
7.2. Informative references	9
Authors' Addresses	9

1. Introduction

Bit Index Explicit Replication (BIER)-TE shares architecture and packet formats with BIER as described in [I-D.ietf-bier-architecture]. 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. BIER-TE Path can be derived from a Path Computation Element (PCE).

This document specifies extensions to the Path Computation Element Protocol (PCEP) to handle requests and responses for the computation of paths for BIER-TE.

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. 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 RRO subobjects, new PCEP error codes and procedures.

4. BIER PCEP Message Extensions

The following section describes the protocol extensions required to support BIER-TE path.

4.1. BIER Capability Advertisement

4.1.1. The OPEN Object

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

4.1.1.1. The BIER PCE Capability TLV

The BIER-PCE-CAPABILITY TLV is an optional TLV associated with the OPEN Object to exchange BIER capability of PCEP speakers. The format of the BIER-PCE-CAPABILITY TLV is shown in the following figure:

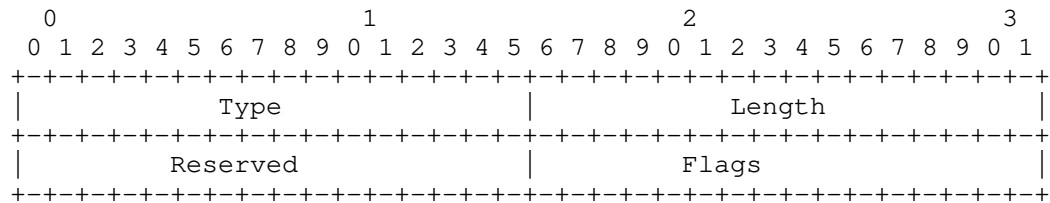


Figure 1

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.

4.1.1.1. Exchanging BIER Capability

This document defines a new optional BIER-PCE-CAPABILITY TLV for use in the OPEN object to negotiate the BIER capability. The inclusion of this TLV in the OPEN message destined to a PCC indicates the PCE's capability to perform BIER-TE path computations, and the inclusion of this TLV in the OPEN message destined to a PCE indicates the PCC's capability to support BIER-TE Path.

A PCE that is able to support the BIER extensions defined in this document SHOULD include the BIER-PCE-CAPABILITY TLV on the OPEN message. If the PCE does not include the BIER-PCE-CAPABILITY TLV in the OPEN message and PCC does include the TLV, it is RECOMMENDED that the PCC indicates a mismatch of capabilities.

4.2. Path Computation Request/Reply Message Extensions

4.2.1. The RP/SPR Object

In order to setup an BIER-TE, a new PATH-SETUP-TYPE TLV[I-D.ietf-pce-lsp-setup-type] MUST be contained in RP or SRP object. This document defines a new Path Setup Type (PST) for BIER as follows:

- o PST = 2: Path is setup using BIER Traffic Engineering technique.

If a PCEP speaker does not recognize the PATH-SETUP-TYPE TLV, it MUST ignore the TLV in accordance with [RFC5440]. If a PCEP speaker recognizes the TLV but does not support the TLV, it MUST send PCErr with Error-Type = 2 (Capability not supported).

4.2.2. The New BIER END-POINT Object

The END-POINTS object 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 define a new END-POINT Object for the BIER path:

The format of the new END-POINTS Object is as follows:

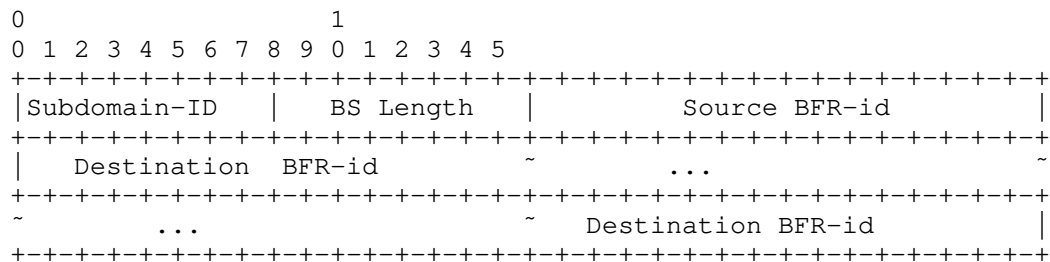


Figure 2

Subdomain-id: Unique value identifying the BIER sub-domain. 1 octet

BS Length: A 1 octet field encoding the supported BitString length.

Source BFR-id: A 2 octet field encoding the source BFR-id.

Destnation BFR-id: A 2 octet field encoding the destnation BFR-id.

4.2.3. 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([I-D.eckert-bier-te-arch]).

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).

4.2.3.1. BIER-ERO Subobject

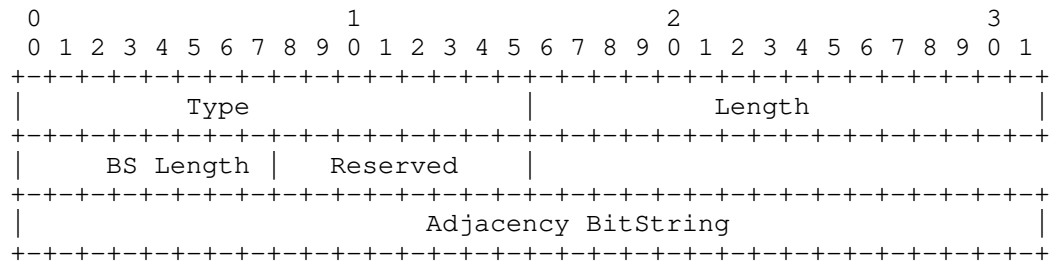


Figure 3

Type: TBD

Length: 4 bytes

BS Length: A 1 octet field encoding the supported BitString length.

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

Adjacency BitString: A 4 octet field encoding the Adjacency BitString where every BitPosition of the BitString indicates one or more adjacencies.

4.2.3.1.1. BIER-ERO Processing

If a PCC finds a non-recognize the BIER-ERO subobject, the PCC MUST respond with a PCErr message with Error-Type=3 ("Unknown Object") and Error-Value=2 ("Unrecognized object Type") or Error-Type=4 ("Not supported object") and Error-Value=2 ("Not supported object Type") as described in [RFC5440] .

If a PCC receives an BIER-ERO subobject in which either BitStringLength or Adjacency BitString 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") and Error-Value = TBD ("BitStringLength is absent ") and Error-Value = TBD ("Adjacency BitString is absent ")

If a PCC detects that all subobjects of BIER-ERO are not identical, it MUST send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Non-identical BIER-ERO subobjects").

If a PCC receives an BIER-ERO subobject in which BitStringLength values are not chosen from: 64, 128, 256, 512, 1024, 2048, and 4096, as it described in ([I-D.ietf-bier-architecture]). The PCC MUST send a PCErr message with Error-Type = 10 ("Reception of an invalid object") and Error-Value = TBD ("Invalid BitStringLength").

4.2.4. RRO Object

A PCC can record BIER-ERO explicit paths and report the paths to a PCE via RRO. An RRO object contains one or more subobjects called "BIER-RRO subobjects" whose formats are the same as that of BIER-ERO subobject.

4.2.4.1. RRO Processing

Processing rules of BIER-RRO subobject are identical to those of BIER-ERO subobject defined in section 4.2.3.1 in this document.

5. Security Considerations

TBD.

6. IANA Considerations

6.1. PCEP Objects

As discussed in Section 4.2.2, a new END-POINTS Object-Type is defined. IANA has made the following Object-Type allocations from the "PCEP Objects" sub-registry:

Object	Object-Class Value
BIER END-POINT Object	TBD

As discussed in Section 4.2.3 and 4.2.4, a new sub-object type for the PCEP explicit route object (ERO), and a new sub-object type for the PCEP record route object (RRO) are defined.

IANA has made the following sub-objects allocation from the RSVP Parameters registry:

Object	Sub-Object	Sub-Object Type

EXPLICIT_ROUTE	BIER-ERO (PCEP-specific)	TBD
ROUTE_RECORD	BIER-RRO (PCEP-specific)	TBD

6.2. PCEP-Error Objects and Types

As described in Section 4.2.3.1.1, a number of new PCEP-ERROR Object Error Values have been defined.

Error-Type	Meaning	Reference

10	Reception of an invalid object.	RFC5
540	Error-value = TBD: BitStringLength is absent	This document
	Error-value = TBD: BitString is absent	This document
	Error-value = TBD: Invalid BitStringLength	This document

6.3. PCEP TLV Type Indicators

IANA is requested to allocate a new code point in the PCEP TLV Type Indicators registry, as follows:

Value	Meaning	Reference

TBD	BIER-PCE-CAPABILITY TLV	This document

6.4. New Path Setup Type

IANA is requested to allocate a new code point in the PCEP PATH_SETUP_TYPE TLV PST field registry, as follows:

Value	Description	Reference

2	Path is setup using BIER Traffic Engineering technique	This document

7. References

7.1. Normative references

- [I-D.ietf-bier-architecture]
Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", draft-ietf-bier-architecture-08 (work in progress), September 2017.
- [I-D.ietf-pce-lsp-setup-type]
Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying path setup type in PCEP messages", draft-ietf-pce-lsp-setup-type-10 (work in progress), May 2018.
- [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>>.

7.2. Informative references

- [I-D.eckert-bier-te-arch]
Eckert, T., Cauchie, G., Braun, W., and M. Menth, "Traffic Engineering for Bit Index Explicit Replication BIER-TE", draft-eckert-bier-te-arch-06 (work in progress), November 2017.

Authors' Addresses

Ran Chen
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing, Jiangsu Province 210012
China

Phone: +86 025 88014636
Email: chen.ran@zte.com.cn

Zheng Zhang
ZTE Corporation
No.50 Software Avenue, Yuhuatai District
Nanjing, Jiangsu Province 210012
China

Email: zhang.zheng@zte.com.cn

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:

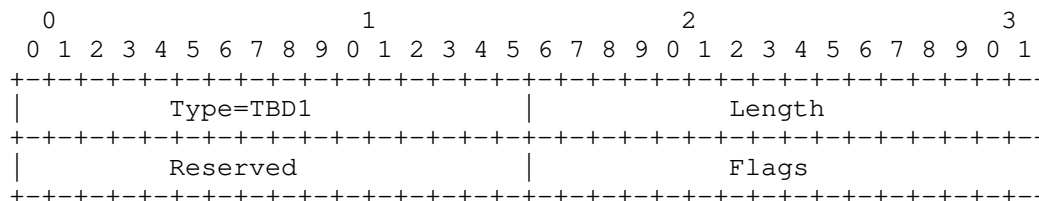


Figure 1 BIER-TE-PCE-CAPABILITY sub-TLV format

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

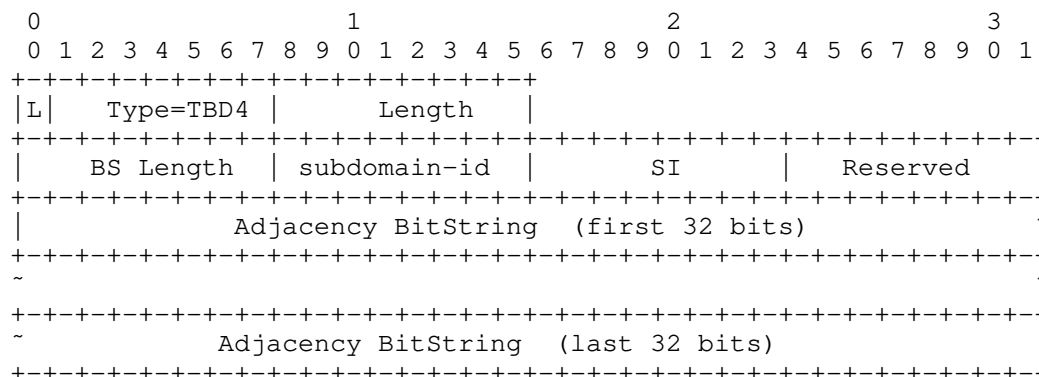


Figure 3

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:

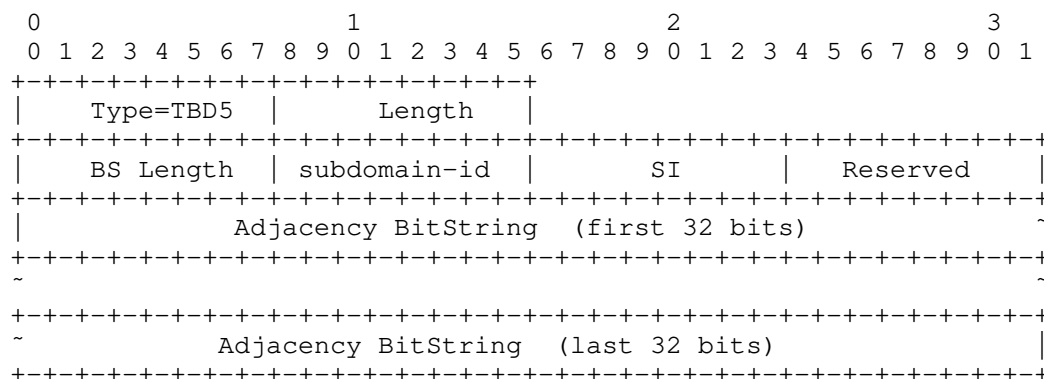


Figure 4

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

Path Computation Element Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 5, 2019

O. Dugeon
J. Meuric
Orange Labs
Y. Lee
Huawei Technologies
D. Ceccarelli
Ericsson
March 04, 2019

PCEP Extension for Stateful Inter-Domain Tunnels
draft-dugeon-pce-stateful-interdomain-02

Abstract

This document proposes to combine a Backward Recursive or Hierarchical method in Stateful PCE with PCInitiate message to setup independent paths per domain, and combine these different paths together in order to operate them as end-to-end inter-domain paths without the need of signaling session between AS border routers. A new Stitching Label is defined, new Path Setup Types and a new Association Type are considered for that purpose.

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 <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 September 5, 2019.

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	3
1.1. General assumptions	5
1.2. Terminology	6
2. Stitching Label	7
2.1. Definition	8
2.2. Inter-domain LSP-TYPE	8
3. Backward Recursive PCInitiate procedure	9
3.1. Mode of operation	9
3.2. Example	12
3.3. Inter-domain LSP setup procedure completion failure	13
4. Hierarchical PCInitiate procedure	14
4.1. Mode of operation	14
4.2. Inter-domain LSP setup procedure completion failure	16
4.3. Example for Stateful H-PCE Stching procedure	17
5. Inter-domain LSP Management	21
5.1. Identification of inter-domain tunnels	21
5.2. Inter-domain association group	21
5.3. Inter-domain LSP management	22
5.4. Modification of inter-domain LSP	23
5.5. Removal of inter-domain LSP	23
6. Applicability	24
6.1. RSVP-TE	24
6.2. Segment Routing	24
6.3. Mixing technology	25
7. IANA Considerations	26
7.1. Path Setup Type values	26
7.2. Association Type value	27
7.3. PCEP Error values	27
8. Security Considerations	27
9. Acknowledgements	27
10. Disclaimer	28

11. References	28
11.1. Normative References	28
11.2. Informative References	29
Authors' Addresses	30

1. Introduction

The Path Computation Element (PCE) working group (WG) has produced a set of RFCs to standardize the behavior of the Path Computation Element as a tool to help MPLS-TE, GMPLS LSP tunnels and Segment Routing paths placement. This also includes the ability to compute inter-domain LSPs or Segment Routing paths following a distributed or hierarchical approach. To complement the original stateless mode, a stateful mode has been added. In particular, the new PCInitiate message allows a PCE to directly ask a PCC to setup an MPLS-TE, GMPLS LSP tunnel or a Segment Routing path. However, once computed, the inter-domain LSPs or Segment Routing path are hard to setup in the underlying network. Especially, in operational network, RSVP-TE signaling is not enabled between AS border routers. But, such RSVP-TE signaling is mandatory to setup contiguous LSP tunnels or to stitch or nest independent LSP tunnels to form the end-to-end inter-domain paths.

Looking to the different RFCs that describe the PCE architecture and in particular PCE based architecture [RFC4655], PCE protocol [RFC5440], BRPC [RFC5441] and H-PCE [RFC6805], the Path Computation Element (PCE) is able to compute inter-domain paths in complement to intra-domain computation. Such inter-domain paths could then serve as the Explicit Route Object input for the RSVP-TE signaling to setup the tunnels within the underlying network. Three kinds of inter-domain paths could be established:

- o Contiguous tunnel ([RFC3209] and [RFC3473]): The RSVP-TE signaling crosses the boundary between two domains, e.g. between two AS Border Routers (ASBRs) as if they were two routers of the same domain. This kind of tunnel is not recommended mostly for security and scalability purpose. In addition, the initiating domain imposes huge constraints on subsequent domains, because they undergo the tunnel request without being able to control it.
- o Stitching tunnel ([RFC5150]): Each domain establishes in its own network the corresponding part of the inter-domain path independently. Then, a second end-to-end RSVP-TE Path message is sent by the initiating domain to stitch the different tunnel parts to form the inter-domain path. In fact, this second RSVP-TE Path message is used by border nodes to exchange the label that must be used by the previous domain to send the traffic in order that the

MPLS packets follow the next LSP tunnel in the following domain. These labels are conveyed in the RSVP-TE Resv message.

- o Nesting tunnel ([RFC4206]): This is similar to the stitching mode but, this time, with the possibility to setup tunnel hierarchy. For example, an LSP tunnel between two edge domains crossing a transit domain could be carried over a tunnel of a higher level in the transit domain. Again, a second end-to-end RSVP-TE Path message is sent from the source to the destination. Labels that must be used by the ASBRs of transit domains to identify flows to be nested are carried by the RSVP-TE Resv message.

In all case, RSVP-TE signaling must be exchanged between the different domains. However, from an operational point of view, looking to different networks under the responsibility of different administrative entities, only BGP sessions are setup and configured between ASBRs. Technologically speaking, this is possible and many RFCs describe how to use RSVP-TE for inter-domain. But, due to security, scalability, management and contract constraints, RSVP-TE is not exposed at the network boundary. To circumvent some of the security issues, RSVP-TE can be carried inside an IPsec tunnel between ASBRs, but, this does not eliminate the scalability aspect nor the constraints imposed by setting up inter-domain paths.

The purpose of this memo is to take the benefit of PCE Stateful [RFC8231] and PCE Initiated [RFC8281] modes to stitch or nest inter-domain paths directly using PCEP between domains' PCEs instead of using RSVP-TE signaling at the inter-domain border nodes, while keeping each operator free to independently setup their respective part of the inter-domain paths. PCInitiate message is used in a Backward Recursive way like the PCReq message in BRPC [RFC5441], to recursively setup the end-to-end tunnel. PCRep message is used to automatically stitch or nest the different local LSP tunnels. And, PCRep in conjunction of PCUpd messages are used to maintain, modify and remove inter-domain paths. This method is also applicable to Segment Routing to build inter-domain segment paths.

H-PCE [RFC6805] describes a Hierarchical PCE architecture which can be used for computing end-to-end paths for inter-domain MPLS Traffic Engineering (TE) and GMPLS Label Switched Paths (LSPs). Within this architecture, the Parent PCE (P-PCE) is used to compute a multi-domain path based on the domain connectivity information. A Child PCE (C-PCE) may be responsible for a single domain or multiple domains, it is used to compute the intra-domain path based on its domain topology information.

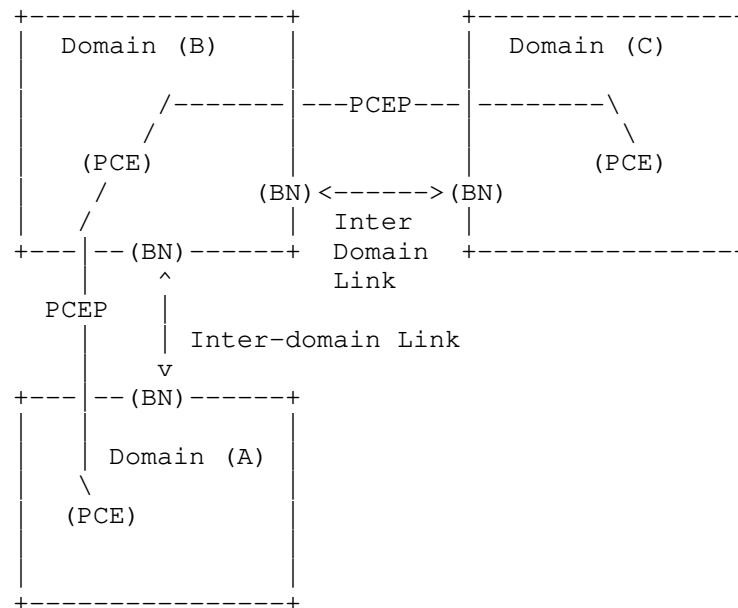
Stateful H-PCE [I-D.ietf-pce-stateful-hpce] presents general considerations for stateful PCE(s) in hierarchical PCE architecture.

In particular, the behavior changes and additions to the existing stateful PCE mechanisms (including PCE-initiated LSP setup and active PCE usage) in the context of networks using the H-PCE architecture. Section 3.3.1 [I-D.ietf-pce-stateful-hpce] describes the per domain stitched LSP mode, where the individual per domain LSP are stitched together. PCInitiate message is also used to stitch the end-to-end tunnel. See section 4 for details.

1.1. General assumptions

In the rest of this document, we used the same references as per BRPC [RFC5441] and make the following set of assumptions (see figure below):

- o Domain refers to an IGP area or an Autonomous System (AS).
- o Inter-domain path is used to refer to a path that cross two or more different domains as defined previously,
- o At least, one PCE is deployed in each domain. These PCE are all stateful active capable and could request to enforce LSP tunnels in their respective domain by means of PCInitiate messages.
- o LSRs, including border nodes, are PCC enable and support stateful active mode. PCEP sessions is established between these routers and their domains' PCE.
- o Each PCE establishes a PCEP session with its respective neighbor domain's PCE. The way a PCE discover its neighboring PCE is out of scope of this draft. These information could be fulfill administratively or automatically discovered through, for example per draft 'BGP Extensions for Path Computation Element (PCE) Discovery' [I-D.dong-pce-discovery-proto-bgp].
- o PCEs are able to compute and end-o-end path as per BRPC procedure [RFC5441] or as per H-PCE procedure (stateless [RFC6805] or stateful [I-D.ietf-pce-stateful-hpce]).
- o Tunnels refer to LSPs setup by mean of RSVP-TE or Segment Path in a Segment Routing network.



Example of the representation of 3 domains with 3 PCEs

1.2. Terminology

ABR: Area Border Routers. Routers used to connect two IGP areas (areas in OSPF or levels in IS-IS).

AS: Autonomous System

ASBR: Autonomous System Border Router. Router used to connect together ASes of the same or different service providers via one or more inter-AS links.

Border Node (BN): a boundary node is either an ABR in the context of inter-area Traffic Engineering or an ASBR in the context of inter-AS Traffic Engineering.

BN-en(i): Entry BN of domain(i) connecting domain(i-1) to domain(i) along a determined sequence of domains. Multiple entry BN-en(i) could be used to connect domain(i-1) to domain(i).

BN-ex(i): Exit BN of domain(i) connecting domain(i) to domain(i+1) along a determined sequence of domains. Multiple exit BN-ex(i) could be used to connect domain(i) to domain(i+1).

Domains: Autonomous System (AS) or IGP Area. An Autonomous System is composed by one or more IGP area.

ERO(i): The Explicit Route Object scoped to domain(i)

IGP-TE: Interior Gateway Protocol with Traffic Engineering support. Both OSPF-TE and IS-IS-TE are identified in this category.

Inter-domain path: A path that crosses two or more domains through a pair of Border Node (BN-ex, BN-en).

LK(i): A Link that connect BN-ex(i-1) to BN-en(i). Note that BN-ex(i-1) could be connected to BN-en(i) by more than one link. LK(i) serves to identify which of the multiple links will be used for the inter-domain LSP setup.

Local LSP tunnel: A LSP tunnel that do not cross a domain. It is setup between entry BN-en to output BN-ex, any source to output BN-ex or entry BN-en to any destination of the same domain. This LSP could be enforce by means of RSVP-TE signaling or Segment Routing labels stack.

Local LSP tunnel(i): A local LSP tunnel of domain(i)

PLSP-ID(i): A PLSP-ID that identify the local tunnel part of an inter-domain tunnel in the domain(i).

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.

PCE(i) is a PCE with the scope of domain(i).

PST: Path Setup Type

R(i,j): The router j of domain i

Stitching Label (SL): A dedicated label that is used to stitch two RSVP-TE tunnels or two Segment Routing paths.

SL(i): A Stitching Label that link domain(i-1) to domain(i).

2. Stitching Label

This section introduce the concept of Stitching Label that allows stitching and nesting of local LSP tunnels in order to form inter-domain path that cross several different domains.

2.1. Definition

The operation of stitch or nest a local LSP tunnel(i) to a local LSP tunnel(i+1) in order to form an inter-domain path simply consist in defining the label that the output BN-ex(i) will use to send its traffic to the entry BN-en(i+1). Indeed, the entry BN-en(i+1) needs to identify the incoming traffic i.e. IP packets, in order to know if this traffic must follow the local LSP tunnel(i+1) or not. Forwarding Equivalent Class (FEC) could be used for that purpose. But, when stitching or nesting tunnels, the FEC is reduced to the incoming label that the entry BN-en(i+1) has chosen for the local LSP tunnel(i+1).

In this memo, we introduce the named of 'Stitching Label (SL)' to designate this label. Such label is usually exchanged between output BN-ex(i) and entry BN-en(i+1) with the RSVP-TE signaling. But, as we want to avoid to use RSVP-TE signaling due to operational constraints, this Stitching Label will be conveyed by PCEP protocol. In fact, the Explicit Route Object (ERO) and the Record Route Object (RRO) are defined in order to transport this Stitching Label in the RSVP-TE signaling. As PCEP protocol used RSVP-TE Objects, and in particular the ERO and RRO, it is able to convey the Stitching Label without any modification of the PCEP protocol nor the PCE or RSVP-TE Objects.

As per RFC4003 [RFC4003], the Stitching Label will be conveyed as a companion of an IP address. In our case, this is one of the IP address of the link LK(i) which connects BN-ex(i) to BN-en(i+1) and carries the traffic from the domain(i) to domain(i+1). It is left to implementation to select which of the two IP addresses of the link LK(i) is used.

2.2. Inter-domain LSP-TYPE

However, even if PCEP could convey the Stitching Label, a PCC is not aware that a PCE requests or provides such label. For that purpose, this memo proposes to use the PST as defined in [RFC8408] with new values (See IANA section of this memo) defined as follows:

- o TBD1: Inter-Domain Traffic engineering end-to-end path is setup using Backward Recursive or Hierarchical method. This new PST value MUST be set in a PCInitiate message sent by a PCE(i) to its neighbor PCE(i+1) in the Backward Recursive method or by the Parent PCE to the Child PCE(i) to initiate a new inter-domain path. In turn, neighbor PCE(i+1) or Child PCE(i) MUST return a Stitching Label SL with the IP address of the associated link in the RRO of the PCRep message to PCE(i) or Parent PCE.

- o TBD2: Inter-Domain Traffic engineering local path is setup using RSVP-TE. This new PST value MUST be set in the PCInitiate message sends by a PCE(i) requesting to a PCC of domain(i) to initiate a new local LSP tunnel(i) which is part of an inter-domain path. This PST value MUST be used by the PCE(i) only after receiving a PCInitiate message with an PST equal to TBD1 from a neighbor PCE(i+1) in the Backward Recursive method or Parent PCE in the Hierarchical method. In turn, the PCC of domain(i) MUST return a Stitching Label SL with the IP address of associated link in the RRO of the PCRpt message.
- o TBD3: Inter-Domain Traffic engineering local path is setup using Segment Routing. This new PST value MUST be set in the PCInitiate message sends by a PCE(i) requesting to a PCC of domain(i) to initiate a new Segment Routing path which is part of an inter-domain Segment Routing path. This PST value MUST be used by the PCE(i) only after receiving a PCInitiate message with an PST equal to TBD1 from a neighbor PCE(i+1). In turn, the PCC MUST return a Stitching Label SL with the IP address of the associated link in the RRO of the PCRpt message.

3. Backward Recursive PCInitiate procedure

This section describes how to setup inter-domain paths than cross several different domains by using a Backward Recursive method which is compatible to inter-domain path computation by means of the BRPC procedure as describe in RFC5441 [RFC5441].

3.1. Mode of operation

This section describes how PCInitiate and PCRpt messages are combined between PCE in order to setup inter-domain paths between a source domain(1) to a destination domain(n). S and D are respectively the source and destination of the inter-domain path. Domain(1) and domain(n) are different and connected through 0 or more intermediate domains denoted domain(i) with $i = (2, n-1)$. Domains are directly connected when $n = 2$.

First, the PCE(1) runs standard BRPC algorithm as per RFC5441 [RFC5441] with its neighbor PCEs in order to compute the inter-domain path from S to D, where S and D are respectively a node in the domain(1) and domain(n). Path Key confidentiality as per RFC5520 [RFC5520] SHOULD be used to obfuscate the detailed ERO(i) of the different domains(i). The resulting ERO is of the form {S, PKS(1), BN-ex(1), ..., BN-en(i), PKS(i), BN-ex(i), ..., BN-en(n), PKS(n), D} when Path Key is used and of the form {S, R(1,1), ..., R(1,k), BN-ex(1), ..., BN-en(i), R(i,1), ..., R(i,l), BN-ex(i), ..., BN-en(n), R(n,1), ..., R(n,m), D} otherwise. As subsequent domains are not

aware about the final computed ERO in case of multiple VSPTs, the final ERO selected by the PCE(1) MUST be sent in the PCInitiate message to indicate to the subsequent PCEs which solution has been finally chosen. PCE(1) MUST ensure that this ERO is self comprehensive by subsequent PCEs. Indeed, when a PCE(i) receives the ERO, it MUST be able to verify that it is in the scope of this ERO and to determine the PCE(i+1). When Path Key is used, PCEs MUST encode the Path Key with a reachable IP address in order for previous PCEs in the AS chain to join them. When Path Key is not used, the PCEs MUST be able to retrieve IP address of the next PCE from the ERO.

The complete procedure with Path Key follow the different steps described below:

Steps 1: Initialization

Once ERO(S, D) is computed, PCE(1) sends a PCInitiate message to PCE(2) containing an ERO equal to {S, PKS(2), ..., PKS(i), ..., PKS(n), D}, PST = TBD1 and End-Points Object = (S, D). The ERO corresponds to the one PCE(1) has received from PCE(2) during the BRPC process in which only Path Key are kept. In case of multiple EROs, i.e. VSPT, PCE(1) has chosen one of them and used the selected one for the PCInitiate message. PKS(i) could be replaced by the full ERO description if Path Key is not used by PCE(i).

When PCE(i) receives a PCInitiate message from domain(i-1) with PST = TBD1 and ERO = {PKS(i), PKS(i+1), ..., PKS(n), D}, it sends a PCInitiate message to PCE(i+1) with a popped ERO and records its received PKS(i) part. All PCE(i)s generate the appropriate PCInitiate message to PCE(i+1) up to PCE(n), i.e. to the destination domain(n).

Steps 2: Actions taken at the destination domain(n) by PCE(n)

When PCInitiate message propagation reach the destination domain(n), PCE(n) retrieves the ERO from the PKS(n) if necessary and sends to BN-en(n) a PCInitiate message with the ERO(n) = {BN-en(n), ..., D}, PST = TBD2 and End-Points Object = {BN(n), D} in order to inform the PCC BN-en(n) that this local LSP tunnel(n) is part of an inter-domain path. When the PCC BN-en(n) received the PCInitiate message from its PCE(n), it setup the local LSP tunnels from entry BN-en(n) to D by means of RSVP-TE signaling with the given ERO(n). Once the tunnel setup, it chooses a free label for the Stitching Label SL(n) and add a new entry in its MPLS L(F)IB with this SL(n) label. Then, it sends a PCRpt message to its PCE(n) with an RRO equal to {[LK(n), SL(n)], RRO(n)} and PLSP-ID(n). Once PCE(n) receives the PCRpt from the PCC BN-en(n) with the RRO, PLSP-ID and PST = TBD2, it sends to the

PCE(n-1) a PCRpt containing the RRO equal to {[LK(n), SL(n)]} and PLSP-ID(n). PCE(n) MAY add {PKS(n), D} in the RRO.

Steps i: Actions performed by all intermediate domains(i), for i = 2 to n-1

1. When the PCE(i) receives a PCRpt message from domain(i+1) with PST = TBD1, RRO = {[LK(i+1), SL(i+1)]} and PLSP-ID(i+1), it retrieves the ERO(i) from the PKS(i), recorded in step 1, and sends to the PCC BN-en(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, PST = TBD2 and End-Points Object = {BN-en(i), BN-ex(i)} in order to inform the PCC BN-en(i) that this local LSP tunnel(i) is part of an inter-domain path.
2. When the PCC BN-en(i) received the PCInitiate message from its PCE(i), it setup the local LSP tunnels from BN-en(i) to BN-ex(i) by means of RSVP-TE signaling with the given ERO(i).
3. Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is used to instruct the egress node of domain(i), i.e. BN-ex(i), to forward packets belonging to this tunnel with the Stitching Label. Both Stitching Label and IP address of outgoing interface are carried in the ERO = {..., [LK(i+1), SL(i+1)]} as the last SubObject in conformance to [RFC4003]. So that, BN-ex(i) installs in its MPLS L(F)IB the SWAP instruction to label SL(i+1) with forward to LK(i+1).
4. Once the tunnel setup, PCC BN-en(i) chooses a free label for the Stitching Label SL(i) and adds a new entry in its MPLS L(F)IB with this SL(i) label. Then, it sends a PCRpt message to its PCE(i) with an RRO equal to {[LK(i), SL(i)], RRO(i)} and PLSP-ID(i).
5. Once PCE(i) receives the PCRpt from the PCC BN-en(i) with the RRO and PST = TBD2, it sends to the PCE(i-1) a PCRpt message containing the RRO equal to {[LK(i), SL(i)]} and the PLSP-ID(i). PCE(i) MAY add {PKS(i), ..., PKS(n)} in the RRO.

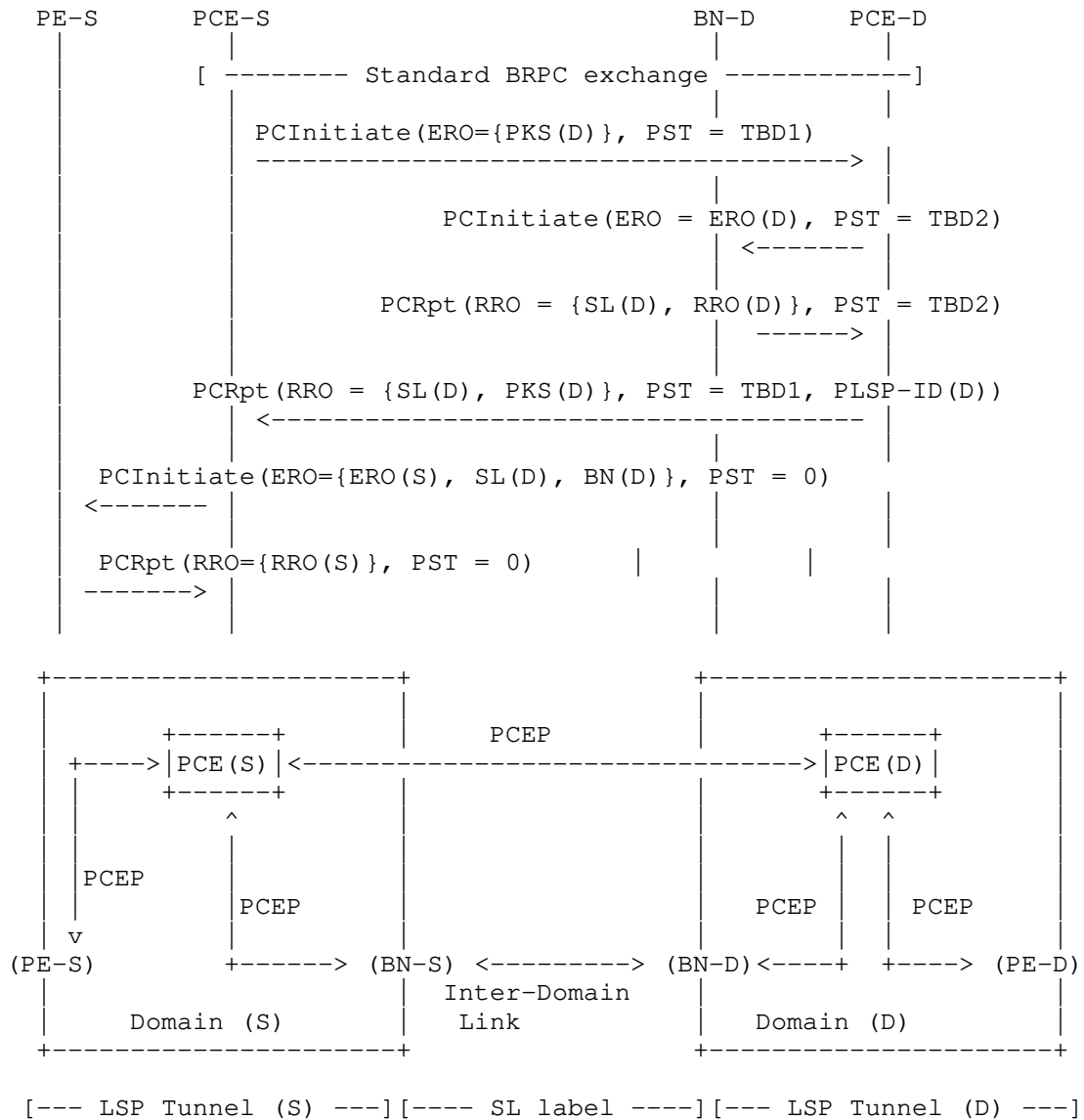
Steps n: Actions performed at the source domain(1) by PCE(1)

Once PCE(1) received the PCRpt message from PCE(2) with the RRO containing the label SL(2), it sends a PCInitiate message to PCC node S with ERO equal to {ERO(1), [LK(2), SL(2)]}, PST = 0 and End-Points Object = {S, BN-ex(1)}. This time, the PST is equal to 0 as the PCC S does not need to return a Stitching Label SL, i.e. it is the head-end of the inter-domain path. Standard PCRpt message is sent back to PCE(1) by the PCC node S.

3.2. Example

In the figure below, two different domains S and D are interconnected through BN respectively BN-S and BN-D. PE-S and PE-D are edge routers. All routers in the figure are connected to their respective PCE through PCEP protocol. In this example, PCE(S) would setup an inter-domain path between PE-S and PE-D acting as source and destination of the tunnel. Intermediate routers between (PE-S, BN-S), (BN-D and PE-D) as well as RSVP-TE messages are not represented to simplify the figure. But they are all presents. The following notation is used in the figure (note that, in this example, we use the PKS for the sake of simplicity):

- o PKS(D) = Path Key corresponding to the path from BN(D) to PE-D
- o ERO(D) = Explicit Route Object corresponding to the path from BN(D) to PE-D retrieves from PKS(D)
- o RRO(D) = Record Route Object of local LSP tunnel(D) from BN(D) to PE-D
- o SL(D) = Stitching Label for local LSP tunnel from BN(D) to PE-D
- o ERO(S) = Explicit Route Object corresponding to the path from PE-S to BN(S)
- o RRO(S) = Record Route Object of local LSP tunnel(S) from PE-S to BN(S)



Example of inter-domain path setup between two domains

3.3. Inter-domain LSP setup procedure completion failure

In case of error during LSP setup, PCRpt and or PCErr messages MUST be used to signal the problem to the neighbor PCE domain backward. In particular, if new PST values defined in this memo are not

supported by the neighbor PCE or the PCC, the PCE, receptively the PCC, MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its neighbor PCE. If a PCE(i) receives a PCInitiate message from its peer PCE(i-1) without PST set to TBD1 or PST set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its peer PCE(i-1).

If a PCC or a PCE don't return an RRO or an RRO without the Stitching Label SL with the IP address of the associated link following a PCInitiate message with PST set to TBD1, the PCE MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = TBD5 (No Mandatory Stitching Label is present in the RRO).

In case of completion failure, the PCE(i) MUST propagate the PCErr message up to the PCE(1). In turn, PCE(1) MUST send a PCInitiate message (R flag set in the SRP Object as per draft pce initiated lsp [RFC8281]) to delete this inter-domain path to its neighbor PCEs. PCE(i) MUST propagate the PCInitiate message and remove their local LSP tunnel by means of PCInitiate message to their PCC BN-en(i) and send back PCRpt message to PCE(i-1).

In case of error in domain(i+1), PCE(i) MAY add the AS number of domain(i+1) in the RRO to identify the faulty domain.

4. Hierarchical PCInitiate procedure

This section describes how to setup inter-domain paths than cross several different domains by using a Hierarchical method which is compatible to inter-domain path computation as describe in [RFC6805].

4.1. Mode of operation

This section describes how PCInitiate and PCRpt messages are combined between PCE in order to setup inter-domain paths between a source domain(1) to a destination domain(n). S and D are respectively the source and destination of the inter-domain path. Domain(1) and domain(n) are different and connected through 0 or more intermediate domains denoted domain(i) with i = (2, n-1). Domains are directly connected when n = 2.

First, the Parent PCE contacts its Child PCE as per [RFC6805] in order to compute the inter-domain path from S to D, where S and D are respectively a node in the domain(1) and domain(n). Path Key confidentiality as per RFC5520 [RFC5520] SHOULD be used to obfuscate the detailed ERO(i) of the different domains(i). The resulting ERO

is of the form (S, PKS(1), BN-ex(1), ..., BN-en(i), PKS(i), BN-ex(i), ..., BN-en(n), PKS(n), D) when Path Key is used and of the form {S, R(1,1), ..., R(1,k), BN-ex(1), ..., BN-en(i), R(i,1), ..., R(i,l), BN-ex(i), ..., BN-en(n), R(n,1), ..., R(n,m), D} otherwise.

The complete procedure with Path Key follow the different steps described below:

Step 1: Initialization

Parent PCE sends a PCInitiate message to child PCE(n) with an ERO = {PKS(n)} and End-Points = {BN-en(n), D}. Then, PCE(n) retrieves the ERO from the PKS(n) if necessary and sends to BN-en(n) a PCInitiate message with the ERO(n) = {BN-en(n), ..., D}, PST = TBD2 and End-Points Object = {BN-en(n), D} in order to inform the PCC BN-en(n) that this local LSP tunnel(n) is part of an inter-domain path. When the PCC BN-en(n) received the PCInitiate message from its PCE(n), it setup the local LSP tunnel from entry BN-en(n) to D by means of RSVP-TE signaling with the given ERO(n). Once the tunnel setup, it chooses a free label for the Stitching Label SL(n) and add a new entry in its MPLS L(F)IB with this SL(n) label. Then, it sends a PCRpt message to its PCE(n) with an RRO equal to {[LK(n), SL(n)], RRO(n)} and PLSP-ID(n). Once PCE(n) receives the PCRpt from the PCC BN-en(n) with the RRO, PLSP-ID and PST = TBD2, it sends to its Parent PCE a PCRpt containing the RRO equal to {[LK(n), SL(n)]} and PLSP-ID(n). PCE(n) MAY add PKS(n) in the RRO.

Steps i: Actions performed for all intermediate domains(i), for i = n-1 to 2

1. Parent PCE sends a PCInitiate message to Child PCE(i) with PST = TBD1, ERO = {PKS(i), [LK(i+1), SL(i+1)]} and End-Points = {BN-en(i), BN-ex(i)}
2. Then, PCE(i) retrieves the ERO from the PKS(i) if necessary and sends to the PCC BN-en(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, PST = TBD2 and End-Points Object = {BN-en(i), BN-ex(i)} in order to inform the PCC BN-en(i) that this local LSP tunnel(i) is part of an inter-domain path.
3. When the PCC BN-en(i) received the PCInitiate message from its PCE(i), it setup the local LSP tunnel from BN-en(i) to BN-ex(i) by means of RSVP-TE signaling with the given ERO(i).
4. Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is used to instruct the egress node of domain(i), i.e. BN-ex(i) to forward packets belonging to this tunnel with the Stitching Label. Both Label Stitching and IP address of outgoing interface

are carried in the ERO = {..., [LK(i+1), SL(i+1)]} as the last SubObject in conformance to [RFC4003]. So that, BN-ex(i) installs in its MPLS L(F)IB the SWAP instruction to label SL(i+1) with forward to LK(i+1) instead of the usual POP instruction.

5. Once the tunnel setup, PCC BN-en(i) chooses a free label for the Stitching Label SL(i) and add a new entry in its MPLS L(F)IB with this SL(i) label. Then, it sends a PCRpt message to its PCE(i) with an RRO equal to {[LK(i), SL(i)], RRO(i)} and PLSP-ID(i).
6. Once PCE(i) receives the PCRpt from the PCC BN-en(i) with the RRO and PST = TBD2, it sends to its Parent PCE a PCRpt message containing the RRO equal to {[LK(i), SL(i)]} and the PLSP-ID(i). PCE(i) MAY add PKS(i) in the RRO.
7. Once Parent PCE receives the PCRpt from the Child PCE(i), it stores the corresponding PLSP-ID for this inter-domain tunnel part

Steps n: Actions performed to the source domain(1)

Finally, Parent PCE sends a last PCInitiate message to Child PCE(1) with PST = TBD1, ERO = {PKS(1), [LK(2), SL(2)]} and End-Points = {S, BN-ex(1)}. In turn, Child PCE(1) sends a PCInitiate message to PCC node S with ERO equal to {ERO(1), [LK(2), SL(2)]}, PST = 0 and End-Points Object = {S, BN-ex(1)}. This time, the PST is equal to 0 as the PCC S does not need to return a Stitching Label SL, i.e. it is the head-end of the inter-domain path. Standard PCRpt message is sent back to PCE(1) by the PCC node S. In turn, Child PCE(1) send a final PCRpt message to the Parent PCE with the PSLP-ID(1). PCE(1) MAY adds {S, BN-ex(1)} in the RRO as loose path.

4.2. Inter-domain LSP setup procedure completion failure

In case of error during LSP setup, PCRpt and or PCErr messages MUST be used to signal the problem to the Parent PCE. In particular, if new PST values defined in this memo are not supported by the Child PCE or the PCC, the Child PCE, receptively the PCC, MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its Parent PCE. If Child PCE(i) receives a PCInitiate message from its Parent PCE without PST set to TBD1 or PST set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = 1 (Unsupported path setup type) to its Parent PCE.

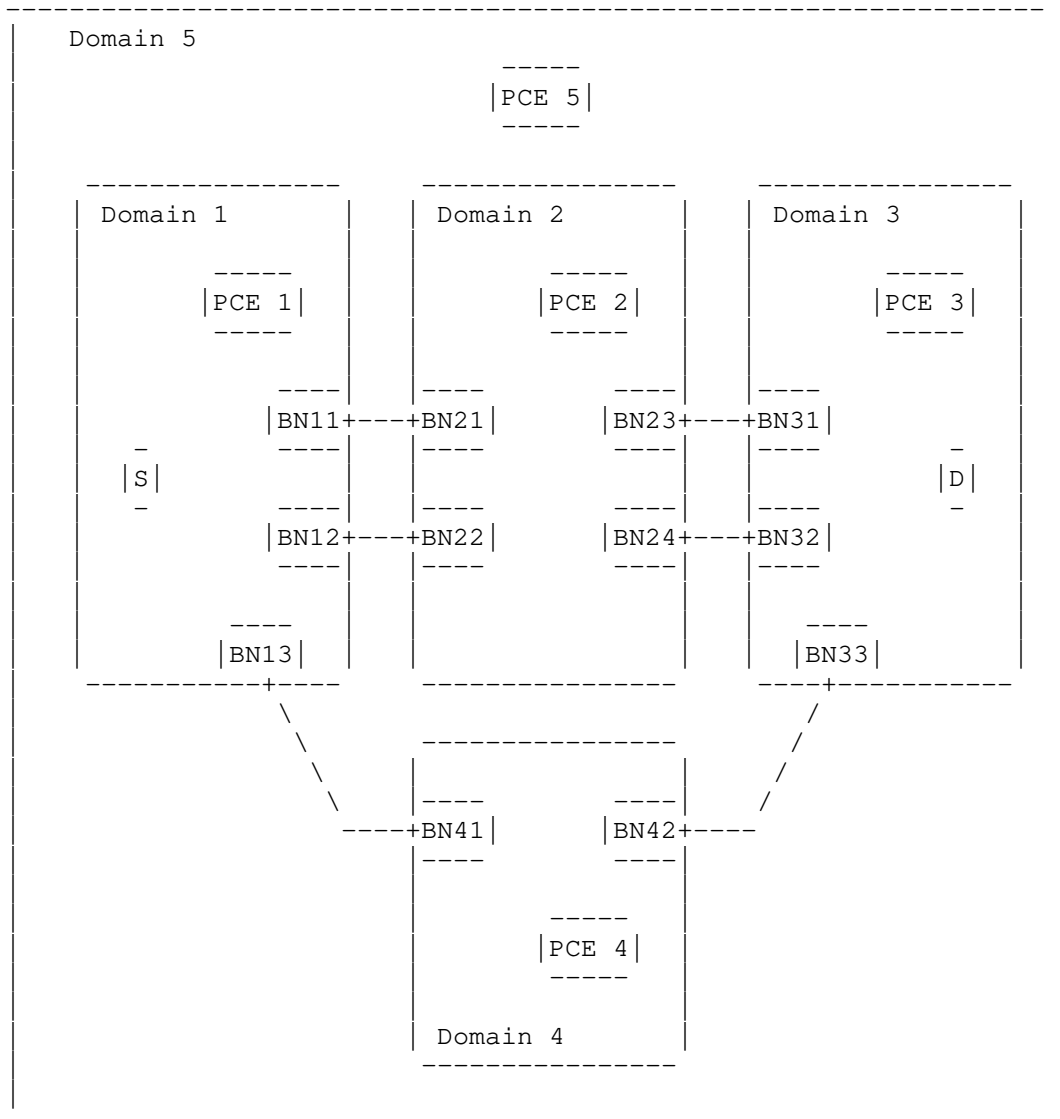
If a Child PCE or a PCC don't return an RRO or an RRO without the Stitching Label SL with the IP address of the associated link

following a PCInitiate message with PST set to TBD1, the Parent PCE, respectively the Child PCE, MUST return a PCErr message with Error-Type = 21 (Traffic engineering path setup error) and Error-Value = TBD5 (No Mandatory Stitching Label is present in the RRO).

In case of completion failure, the Parent PCE MUST send a PCInitiate message (R flag set in the SRP Object as per draft pce initiated lsp [RFC8281]) to delete this inter-domain path to the Child PCEs that already setup their respective part of the inter-domain tunnel. Child PCE(i) MUST remove their local LSP tunnel by means of PCInitiate message with R flag set to 1 to their PCC BN-en(i) and send back PCRpt message to the Parent PCE.

4.3. Example for Stateful H-PCE Sticking procedure

Taking the sample hierarchical domain topology example from [RFC6805] as the reference topology for the entirety of this section.



Hierarchical domain topology from RFC6805

Section 3.3.1 of [I-D.ietf-pce-stateful-hpce] describes the per-domain stitched LSP mode and list all the steps needed. To support SL based stitching, using the reference architecture described in Figure above, the steps are modified as follows (note that we do not use PKS in this example for simplicity):

Step 1: initialization

The P-PCE (PCE5) is requested to initiate a LSP. Steps 4 to 10 of section 4.6.2 of [RFC6805] are executed to determine the end to end path, which are broken into per-domain LSPs e.g. {S-BN41, BN41-BN33, BN33-D}

Step 2: LSP (BN33-D) at PCE3:

1. The P-PCE (PCE5) sends the initiate request to the child PCE (PCE3) via PCInitiate message for LSP (BN33-D) with ERO=(BN33..D) and PST = TBD1
2. The PCE3 further propagates the initiate message to BN33 with the ERO and PST = TBD2/TBD3 based on setup type
3. BN33 initiates the setup of the LSP as per the path and reports to the PCE3 the LSP status ("GOING-UP")
4. The PCE3 further reports the status of the LSP to the P-PCE (PCE5)
5. The node BN33 notifies the LSP state to PCE3 when the state is "UP" it also carry the stitching label (SL33) in RRO as (SL33,BN33..D)
6. The PCE3 further reports the status of the LSP to the P-PCE (PCE5) as well as carry the stitching label (SL33) in RRO as (LK33,SL33,BN33..D)

Step 3: LSP (BN41-BN33) at PCE4

1. The P-PCE (PCE5) sends the initiate request to the child PCE (PCE4) via PCInitiate message for LSP (BN41-BN33) with ERO=(BN41..BN42,LK33,SL33,BN33) and PST = TBD1
2. The PCE4 further propagates the initiate message to BN41 with the ERO and PST = TBD2/TBD3 based on setup type. In case of RSVP_TE, the node BN41 encode the stitching label SL33 as part of the ERO to make sure the node BN42 uses the label SL33 towards node BN33. In case of SR, the label SL33 is part of the label stack pushed at node BN41
3. BN41 initiates the setup of the LSP as per the path and reports to the PCE4 the LSP status ("GOING-UP")
4. The PCE4 further reports the status of the LSP to the P-PCE (PCE5)

5. The node BN41 notifies the LSP state to PCE4 when the state is "UP" it also carry the stitching label (SL41) in RRO as (LK41,SL41,BN41..BN33)
6. The PCE4 further reports the status of the LSP to the P-PCE (PCE5) as well as carry the stitching label (SL41) in RRO as (LK41,SL41,BN41..BN33)

Step 3: LSP (S-BN41) for PCE1

1. The P-PCE (PCE5) sends the initiate request to the child PCE (PCE1) via PCInitiate message for LSP (S-BN41) with ERO=(S..BN13,LK41,SL41,BN41)
2. The PCE1 further propagates the initiate message to node S with the ERO. In case of RSVP_TE, the node S encode the stitching label SL41 as part of the ERO to make sure the node BN13 uses the label SL41 towards node BN41. In case of SR, the label SL41 is part of the label stack pushed at node S
3. S initiates the setup of the LSP as per the path and reports to the PCE1 the LSP status ("GOING-UP")
4. The PCE1 further reports the status of the LSP to the P-PCE (PCE5)
5. The node S notifies the LSP state to PCE1 when the state is "UP"
6. The PCE1 further reports the status of the LSP to the P-PCE (PCE5)

In this way, per-domain LSP are stitched together using the stitching label (SL). The per-domain LSP MUST be setup from the destination domain towards the source domain one after the other.

Once the per-domain LSP is setup, the entry BN chooses a free label for the Stitching Label SL and add a new entry in its MPLS L(F)IB with this SL label. The SL from the destination domain is propagated to adjacent transit domain, towards the source domain at each step. This happens through the entry BN to C-PCE to the P-PCE and vice-versa. In case of RSVP-TE, the entry BN further propagates the SL label to the exit BN via RSVP-TE. In case of SR, the SL label is pushed as part of the SR label stack.

5. Inter-domain LSP Management

This section describe how inter-domain LSPs could be manage.

5.1. Identification of inter-domain tunnels

First, in order to manage inter-domain tunnels composed by the stitching or nesting of local tunnels, it is important to identify them. For this purpose, PLSP-ID managed by PCEs are combined to one provided by PCCs to form global identifier as follow:

- o PCE(i) in the Backward Recursive method or the Child PCE in Hierarchical method MUST create a new unique PLSP-ID for this inter-domain LSP part and MUST send it in the PCRpt message, to the PCE(i-1), respectively the Parent PCE. In addition this new PLSP-ID MUST be associated to the one received from the PCC that instantiate the local tunnel part for further reference.
- o In Hierarchical mode, Parent PCE MUST store and associate the different PLSP-ID(i)s received from the different Child PCE(i)s in order to identify the different part of the inter-domain paths.
- o In Backward Recursive method, PCE(i) MUST store and associate its PLSP-ID(i) and the PLSP-ID(i+1) it received from the PCE(i+1). PCE(n) i.e. the last one in the chain, don't need to perform such association.

Further reference to the inter-domain tunnel will use this PLSP-ID(i). In Backward Recursive method, PCE(i) MUST replace the PLSP-ID(i) by PLSP-ID(i+1) in the PCUpd, PCRpt or PCinitiate message before propagating it to PCE(i+1) and PCE(i) MUST replace the PLSP-ID(i+1) by PLSP-ID(i) in the PCRpt message before propagating it to the PCE(i-1). In Hierarchical method, Parent PCE MUST use the corresponding PLSP-ID(i) of the Child PCE(i).

5.2. Inter-domain association group

In case of failure, a PCE(i) will received PCRpt messages from its PCCs and neighbors PCE(i+1) to synchronize the Inter-domain LSPs. In addition, it may received PCInitiate messages from its previous neighbors PCE(i-1) to re-initiate inter-domain tunnel part. As the PCE(i) may loose the PLSP-ID association, a new association group (within Association Object) is used to ease the association of the different part of the inter-domain tunnel: the local parts and the PCE to PCE parts. The use of the Association Object is MANDATORY in the Backward Recursive method and OPTIONAL in the Hierarchical method.

For that purpose, a new Inter-Domain Association Type with value TBD4 is defined. The first PCE in the Backward Recursive chain (the one which received the initial request) MUST send the PCInitiate message with an Association Object as follows:

- o Association Type field MUST be set to new value TBD4
- o Association ID MUST be set to a unique value. In case of Association ID field is too short or wraps, the first PCE MAY use the Extended Association ID to increase the number of association groups. The Association ID is managed locally by the PCE and does not need to be coordinated with neighbor or remote PCEs.
- o IPv4 or IPv6 association source MUST be set to the IP address which identifies PCE(1) in domain(1).
- o The Global Association Source TLV MUST be present and set with the ASN number of domain(1). It allows to create a globally unique association scope without putting constraint on operator's IP association source. Thus the IP Association Source is associated with the Global Association source to form a unique identifier.
- o Extended Association ID MAY be present and MANDATORY if association ID is too short or wraps.

Subsequent PCE(i) for $i = 2$ to n , MUST send this Association Object as is to the local PCC and the neighbor PCE(i+1).

In case of error with the association group, PCErr message MUST be raised with Error = 26 (Association Error) and Error value set accordingly. A new Error value TBD6 is defined to identify association of inter-domain LSPs.

In Hierarchical method, parent PCE MAY act as initiator of the Association and send to the Child PCEs an Association Object that follows the same rules as for the Backward Recursive method. In turn, Child PCEs MUST propagate the Association Object to the local PCCs as is.

5.3. Inter-domain LSP management

For the Backward Recursive method, each domain manages their respective local LSP tunnel part of an inter-domain path independently of each other. In particular, Stitching Label(i) is managed by domain(i) and is of interest of domain(i-1) only. Thus, Stitching Label SL(i) is not supposed to be propagated to other domains. The same behavior apply to PLSP-ID(i). In Hierarchical method, the Parent PCE MUST ensure the correct distribution of

Stitching Label SL(i) to Child PCE(i-1). The PLSP-ID(i) is kept for the usage of the Parent PCE and thus is not propagated. Only the Association Object defined in section 5.2 is propagated if it is present.

If a PCE(i) needs to modify its local LSP tunnel(i) with a PCUpd message to the PCC BN-en(i), once PCRpt message received by the PCC BN-en(i), it MUST send a new PCRpt message to its neighbor PCE(i-1) in Backward Recursive method, respectively to Parent PCE in Hierarchical method, to advertise PCE(i-1) of the modification. In this case PLSP-ID(i) is used to identify the inter-domain tunnel. PCE(i-1), respectively the Parent PCE, MUST propagate the PCRpt message if the modification implies the previous domain e.g. if the PCRpt indicates that the Stitching Label SL(i) has changed.

PCE(1), respectively Parent PCE, could modify the inter-domain path. For that purpose, it MUST send a PCUpd message to its neighbor PCEs, respectively Child PCE, using the PLSP-ID it received. Each PCE(i) MUST process PCUpd message the same way they process PCInitiate message as defined in section 3.1 for Backward Recursive method and in section 4.1 for Hierarchical method.

In case a failure appears in domain(i), e.g. tunnel becoming down, PCE(i) MUST send a PCRpt message to its neighbor PCE(i-1), respectively its Parent PCE to advertise it of the problem in its local part of the inter-domain path. Once PCE(1), respectively Parent PCE, receives this PCRpt message indicating that the tunnel is down, it is up to the PCE(1), respectively Parent PCE to take appropriate correction e.g. start a new path computation to update the ERO.

5.4. Modification of inter-domain LSP

Modification of local LSP tunnel, BN-en(i) and BN-ex(i) is left for further study.

5.5. Removal of inter-domain LSP

Deletion of inter-domain LSP is only possible by the inter-domain tunnel initiator i.e. PCE(1). For Backward Recursive method, PCInitiate message with R flag set to 1, PLSP-ID set accordingly to section 5.1 and the Association Object with R flag set to 1, is sent by PCE(1) to PCE(n) through PCE(i) and processed the same way as described in section 3.1. For Hierarchical method, PCInitiate message with R flag set to 1 is sent by the Parent PCE to each Child PCE(i) with corresponding PLSP-ID(i) and processed accordingly to section 4.1. Each domain PCE(i) is responsible to delete its part of the tunnel and PCC MUST remove the Stitching label SL in its L(F)IB in addition

to the tunnel when it receives the PCInitiate message with the R flag set to 1 and corresponding PLSP-ID. The Association Group MUST also be removed by the PCC and PCE(i).

6. Applicability

The newly introduce Stitching Label SL serves to stitch or nest part of local LSP tunnels to form an inter-domain path. Each domain is free to decide if the tunnel is stitched or nested and how the tunnel is enforced e.g. tough RSVP-TE or Segment Routing. However, the Stitching Label principle is only compatible with MPLS data plane. At the peering point, the Border Node BN-ex(i) MUST encapsulated the packet with the Stitching Label i.e. the MPLS label prior to send them to the next Border Node BN-en(i+1). Thus, only RSVP-TE and Segment Routing over MPLS technology are detailed in the following sections.

6.1. RSVP-TE

In case of RSVP-TE, the Border Node BN-ex(i) needs to received the Stitching Label through the RSVP-TE message and install in its L(F)IB a SWAP instruction to the Stitching Label and forward it to the next Border Node BN-en(i+1). For that purpose, the Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is RECOMMENDED to instruct the Border Node BN-ex(i) of this action. Other mechanisms to program the L(F)IB could be used e.g. NetConf.

As the Stitching Label could serves to stitch or nest tunnels, a domain(i) may decided to nest the incoming local LSP tunnel into a higher hierarchy of tunnel for Traffic Engineering purpose. A PCE(i) may also decided to group local LSP tunnels part of inter-domain paths into a higher hierarchical tunnel to carry all these local LSP tunnels from one BN-en(i) to one BN-ex(i).

6.2. Segment Routing

To use Segment Routing instead of RSVP-TE to setup the local LSP tunnels as defined in draft pce segment routing [I-D.ietf-pce-segment-routing], PCE(i) MUST send PCInitiate message with PST = TBD3 instead of TBD2 to advertise their respective PCC that the local LSP tunnels is enforce by means of Segment Routing.

Stitching Label SL(i) will be inserted in the label stack in order to become the top label in the stack when the packet reach BN-en(i+1): Thus, the Stitching Label SL(i) serves as entry FEC for BN-en(i+1) to identify the packets that follow the next Segment Path. For that purpose, BN-en(i+1) MUST install in its MPLS L(F)IB an instruction to replace the incoming Stitching Label SL(i) by the label stack given

by the ERO(i+1) plus the Stitching Label SL(i+1). When a packet reaches BN-ex(i), the last label in the stack before the label SL(i+1) corresponds to a SID that allows to reach BN-en(i+1).

However, BN-ex(i) needs to know how to send the packets to BN-en(i+1), in particular when there are multiple interfaces between Border Nodes. Similar to the Egress Control mechanism used with RSVP-TE, it is RECOMMENDED to use the inter-domain SID defined as per draft Egress Peer Engineering [I-D.ietf-idr-bgpls-segment-routing-epe] for that purpose. The inter-domain SID is announced by BN-ex(i) to PCE(i) through BGP-LS for each interface that connects BN-ex(i) to neighbors BN-en(i+1). Thus, the label stack will end with {BN-ex(i) SID, Inter-Domain SID, SL(i+1)} and processes as follows:

- o Penultimate router pops its node SID, and sends the packet to the next node designated by the top label in the label stack i.e. the node SID of BN-ex(i)
- o BN-ex(i) pops its node SID and looks up the next label in the stack, i.e. the inter-domain SID which corresponds to the interface to BN-en(i+1). BN-ex(i) pops again this inter-domain SID and send the packet to BN-en(i) through the interface that corresponds to the inter-domain SID.
- o BN-en(i+1) pops the Stitching Label SL(i+1) and replaces it by the sub-sequent label stack.

Other mechanisms, e.g. NetConf, could be used to configure the inter-domain SID on exit Border Nodes.

6.3. Mixing technology

During the instantiation procedure, if PCE(i) decides to reuse a local tunnel which is not yet part of an inter-domain tunnel, it SHOULD send a PCUpd message with PST = TBD2 to the PCC BN-en(i) in order to request a Stitching Label SL(i) and new ERO(i) to include the Stitching Label SL(i+1) and the associated link to the previous ERO.

[RFC8453] describes framework for Abstraction and Control of TE Networks (ACTN), where each Physical Network Controller (PNC) is equivalent to C-PCE and P-PCE is the Multi-Domain Service Coordinator (MDSC). The Per domain stitched LSP as per the Hierarchical PCE architecture described in Section 3.3.1 and Section 4.1 of [I-D.ietf-pce-stateful-hpce] is well suited for ACTN. The stitching label (SL) mechanism as described in this document is well suited for ACTN when per domain LSP needs to be stitched to form an E2E tunnel

or a VN Member. It is to be noted that certain VNs require isolation from other clients. The stitching label mechanism described in this document can be applicable to the VN isolation use-case by uniquely identifying the concatenated stitching labels across multi-domain only to a certain VN member or an E2E tunnel.

As each operator is free to enforce the tunnel with its technology choice, it is a local policy decision for PCE(i) to instantiate the local part of the end to end tunnel by either RSVP-TE or Segment Routing. Thus, the PST value (i.e. TBD2 or TBD3) used in the PCInitiate message sends by the PCE(i) to the local PCC is determined by the local policy. How the local policy decision is set in PCE is out of scope of this memo. This flexibility is allowed because the stitching label principle allows to mix (data plane) technologies between domains. For example, a domain(i) could used RSVP-TE while domain(i+1) used Segment Routing, reciprocally. The Stitching Label SL could serves to stitch indifferently Segment Path and RSVP-TE tunnel. Indeed, Stitching Label SL will be part of the label stack in order to become the top label in the stack when reaching the BN-en(i+1). This Stitching Label could be swap as usual if the next domain uses RSVP-TE tunnel. When the previous domain uses a RSVP-TE tunnel, the Stitching Label will serve as key for the BN-en(i+1) to determine which label stack it must use on top of the packet for a Segment Routing path.

7. IANA Considerations

7.1. Path Setup Type values

[RFC8408] defines the PATH-SETUP-TYPE TLV and requests that IANA creates a registry to manage the value of the PATH_SETUP_TYPE TLV's PST field. IANA is requested to allocate a new code point in the PCEP PATH_SETUP_TYPE TLV PST field registry, as follows:

Value	Description	Reference
TBD1	Inter-Domain Traffic engineering end-to-end path is setup using Backward Recursive method	This Document
TBD2	Inter-Domain Traffic engineering local path is setup using RSVP-TE	This Document
TBD3	Inter-Domain Traffic engineering local path is setup using Segment Routing	This Document

7.2. Association Type value

Draft pce association group [I-D.ietf-pce-association-group] defines the ASSOCIATION Object and requests that IANA creates a registry to manage the value of the Association Type value. IANA is requested to allocate a new code point in the PCEP ASSOCIATION GROUP TLV Association Type field registry, as follows:

Association Type	Description
TBD4	Inter-domain Association Group

7.3. PCEP Error values

IANA is requested to allocate code-points in the PCEP-ERROR Object Error Values registry for a new error-value of Error-Type 21 Invalid traffic engineering path setup and new error-value of Error-Type 26 Association Error:

Error-Type	Error-Value	Description
21	TBD5	Missing Mandatory Stitching Label in RRO
26	TBD6	Error in association of Inter-domain LSPs

8. Security Considerations

No modification of PCE protocol (PCEP) has been requested by this draft which not introduce any issue regarding security. Concerning the PCEP session between PCEs, authors recommend to use the secure version of PCEP as defined in PCEPS [RFC8253] or use any other secure tunnel mechanism e.g. IPsec tunnel to transport PCEP session between PCE.

9. Acknowledgements

The authors want to thanks PCE's WG members, and in particular Dhruv Dhody who greatly contributed to the Hierarchical section of this document.

10. Disclaimer

This work has been performed in the framework of the H2020-ICT-2014 project 5GEx (Grant Agreement no. 671636), which is partially funded by the European Commission. This information reflects the consortium's view, but neither the consortium nor the European Commission are liable for any use that may be done of the information contained therein.

11. References

11.1. Normative References

- [I-D.ietf-pce-association-group]
Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "PCEP Extensions for Establishing Relationships Between Sets of LSPs", draft-ietf-pce-association-group-07 (work in progress), December 2018.
- [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>>.
- [RFC5441] Vasseur, JP., Ed., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, DOI 10.17487/RFC5441, April 2009, <<https://www.rfc-editor.org/info/rfc5441>>.
- [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>>.

- [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>>.

11.2. Informative References

- [I-D.dong-pce-discovery-proto-bgp]
Dong, J., Chen, M., Dhody, D., Tantsura, J., Kumaki, K., and T. Murai, "BGP Extensions for Path Computation Element (PCE) Discovery", draft-dong-pce-discovery-proto-bgp-07 (work in progress), July 2017.
- [I-D.ietf-idr-bgpls-segment-routing-epe]
Previdi, S., Talaulikar, K., Filsfils, C., Patel, K., Ray, S., and J. Dong, "BGP-LS extensions for Segment Routing BGP Egress Peer Engineering", draft-ietf-idr-bgpls-segment-routing-epe-17 (work in progress), October 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-pce-stateful-hpce]
Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., King, D., and O. Dios, "Hierarchical Stateful Path Computation Element (PCE).", draft-ietf-pce-stateful-hpce-06 (work in progress), October 2018.
- [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>>.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <<https://www.rfc-editor.org/info/rfc3473>>.
- [RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", RFC 4003, DOI 10.17487/RFC4003, February 2005, <<https://www.rfc-editor.org/info/rfc4003>>.

- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, DOI 10.17487/RFC4206, October 2005, <<https://www.rfc-editor.org/info/rfc4206>>.
- [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>>.
- [RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, DOI 10.17487/RFC5150, February 2008, <<https://www.rfc-editor.org/info/rfc5150>>.
- [RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, DOI 10.17487/RFC5520, April 2009, <<https://www.rfc-editor.org/info/rfc5520>>.
- [RFC6805] King, D., Ed. and A. Farrel, Ed., "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, DOI 10.17487/RFC6805, November 2012, <<https://www.rfc-editor.org/info/rfc6805>>.
- [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>>.
- [RFC8453] Ceccarelli, D., Ed. and Y. Lee, Ed., "Framework for Abstraction and Control of TE Networks (ACTN)", RFC 8453, DOI 10.17487/RFC8453, August 2018, <<https://www.rfc-editor.org/info/rfc8453>>.

Authors' Addresses

Olivier Dugeon
Orange Labs
2, Avenue Pierre Marzin
Lannion 22307
France

Email: olivier.dugeon@orange.com

Julien Meuric
Orange Labs
2, Avenue Pierre Marzin
Lannion 22307
France

Email: julien.meuric@orange.com

Young Lee
Huawei Technologies
5340 Legacy Drive, Building 3
Plano TX 75023
USA

Email: leeyoung@huwaei.com

Daniele Ceccarelli
Ericsson
Torshamnsgatan, 48
Stockholm
Sweden

Email: daniele.ceccarelli@ericsson.com

Path Computation Element Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 11, 2021

O. Dugeon
J. Meuric
Orange Labs
Y. Lee
Huawei Technologies
D. Ceccarelli
Ericsson
July 10, 2020

PCEP Extension for Stateful Inter-Domain Tunnels
draft-dugeon-pce-stateful-interdomain-04

Abstract

This document specifies how to combine a Backward Recursive or Hierarchical method with inter-domain paths in the context of stateful Path Computation Element (PCE). It relies on the PCInitiate message to set up independent paths per domain. Combining these different paths together enables to operate them as end-to-end inter-domain paths without the need for a signaling session between inter-domain border routers. A new Stitching Label is defined, new Path Setup Types, a new Association Type and a new PCEP communication Protocol (PCEP) Capability are considered for that purpose.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 11, 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. General Assumptions	5
1.2. Terminology	6
2. Stitching Label	8
2.1. Definition	8
2.2. Inter-domain LSP-TYPE	9
3. Backward Recursive PCInitiate Procedure	9
3.1. Mode of Operation	10
3.2. Example	12
3.3. Completion Failure of Inter-domain Path Setup Procedure	14
4. Hierarchical PCInitiate Procedure	14
4.1. Mode of Operation	14
4.2. Completion Failure of Inter-domain Path Setup Procedure	17
4.3. Example for Stateful H-PCE Sticking Procedure	17
5. Inter-domain Path Management	21
5.1. Stitching Label PCE Capabilities	21
5.2. Identification of Inter-domain Paths	22
5.3. Inter-domain Association Group	23
5.4. Modification of Inter-domain Paths	24
5.5. Modification of Inter-domain Paths	25
5.6. Tear-Down of Inter-domain Paths	25
6. Applicability	25
6.1. RSVP-TE	25
6.2. Segment Routing	26
6.3. Mixing Technologies	27
6.4. Inter-Area	27
7. IANA Considerations	28
7.1. Path Setup Type Values	28
7.2. Association Type Value	28
7.3. PCEP Error Values	29
7.4. PCEP TLV Type Indicators	29

7.5. Stitching Label PCE Capability	29
8. Security Considerations	30
9. Acknowledgements	30
10. Disclaimer	30
11. References	30
11.1. Normative References	30
11.2. Informative References	31
Authors' Addresses	33

1. Introduction

The PCE working group has produced a set of RFCs to standardize the behavior of the Path Computation Element as a tool to help MultiProtocol Label Switching - Traffic Engineering (MPLS-TE)/Generalized MPLS (GMPLS) Label Switched Paths (LSPs) and Segment Routing paths placement. This also includes the ability to compute inter-domain LSPs or Segment Routing paths following a distributed or hierarchical approach. To complement the original stateless mode, a stateful mode has been added and supports both passive and active control models. In particular, the new PCInitiate message allows a PCE to directly ask a PCC to set up an MPLS-TE/GMPLS LSP or a Segment Routing path. However, once computed, the inter-domain LSPs or Segment Routing paths are hard to set up in the underlying network. Especially, in operational networks, RSVP-TE signaling is usually not enabled between AS border routers. But, such RSVP-TE signaling is mandatory to set up contiguous LSP tunnels or to stitch or nest independent LSP tunnels to form the end-to-end inter-domain paths.

Looking at the different RFCs that describe the PCE architecture and in particular the PCE-based architecture [RFC4655], the PCE communication Protocol [RFC5440], BRPC [RFC5441] and H-PCE [RFC6805], the PCE is able to compute inter-domain paths, thus complementing the intra-domain computation. Such inter-domain paths could then serve as an Explicit Route Object (ERO) input for the RSVP-TE signaling to set up the tunnels within the underlying network. Three kinds of inter-domain paths could be established:

- o Contiguous tunnel ([RFC3209] and [RFC3473]): The RSVP-TE signaling crosses the boundary between two domains, e.g. between two AS Border Routers (ASBRs) as if they were two routers of the same domain. This kind of tunnel is not recommended mostly for security and scalability purpose. In addition, the initiating domain imposes huge constraints on subsequent domains, because they undergo the tunnel request without being able to control it.
- o Stitching tunnel ([RFC5150]): Each domain establishes in its own network the corresponding part of the inter-domain path independently. Then, a second end-to-end RSVP-TE Path message is

sent by the initiating domain to stitch the different tunnel parts to form the inter-domain path. In fact, this second RSVP-TE Path message is used by border nodes to request the label that must be used by the previous domain to send the traffic in order that the MPLS packets follow the next LSP in the downstream domain. These labels are conveyed in the RSVP-TE Resv message.

- o Nesting tunnel ([RFC4206]): This is similar to the stitching mode but, this time, with the possibility to set up tunnel hierarchy. For example, an LSP between two edge domains crossing a transit domain could be carried over a tunnel of a higher level in the transit domain. Again, a second end-to-end RSVP-TE Path message is sent from the source to the destination. Labels that must be used by the ASBRs of transit domains to identify flows to be nested are carried by the RSVP-TE Resv message.

In all cases, RSVP-TE signaling must be exchanged between the different domains. However, from an operational point of view, looking to different networks under the responsibility of different administrative entities, typically only BGP sessions are set up and configured between ASBRs. Technologically speaking, this is possible and many RFCs describe how to use RSVP-TE for inter-domain. But, due to security, scalability, management and contract constraints, RSVP-TE is not exposed at the network boundary. To address some of the security concerns, RSVP-TE can be carried inside an IPsec tunnel between ASBRs, but, this does not eliminate the scalability aspect nor the constraints imposed by setting up inter-domain paths.

For Segment Routing, issues are different as there is no signaling between routers. Here, the main problem comes from label stacking. The first issue concerns the size of the labels stack which is limited due to hardware constraint. The PCEP Extensions for Segment Routing [RFC8664] takes into account this limitation within the PCEP Capability when the PCEP session is established. Thus, taking into account Maximum Stack Depth (MSD), a PCE may be unable to find a solution when it computes an end-to-end inter-domain path. The second issue is related to the path confidentiality. With SR-TE, to express an explicit path, all Node-SID must be stacked by the head end router while some of the Node-SIDs are associated to routers of the next domains. It is clear that operators would not disclose details of their network, which includes Node-SIDs. Thus, it is not possible to stack remote labels for an end-to-end inter-domain path even if MSD constraint is respected.

The purpose of this memo is to take the benefit of active stateful PCE [RFC8231] and PCE-Initiated [RFC8281] modes to stitch or nest inter-domain paths directly using PCEP between domains' PCEs. This avoids using another signaling (e.g. RSVP-TE) at the inter-domain

border nodes, while keeping each operator free to independently set up their respective part of the inter-domain paths. The PCInitiate message is used in a Backward Recursive way like the PCReq message in BRPC [RFC5441], to recursively set up the end-to-end tunnel. PCRep message is used to automatically stitch or nest the different local LSPs. And, PCRep in conjunction with PCUpd messages are used to report, maintain, modify and remove inter-domain paths. This method is also applicable to Segment Routing to build inter-domain segment paths.

H-PCE [RFC6805] describes a Hierarchical PCE architecture which can be used for computing end-to-end paths for inter-domain MPLS-TE and GMPLS LSPs. Within this architecture, the Parent PCE (P-PCE) is used to compute a multi-domain path based on the domain connectivity information. A Child PCE (C-PCE) may be responsible for a single domain or multiple domains, it is used to compute the intra-domain path based on its domain topology information.

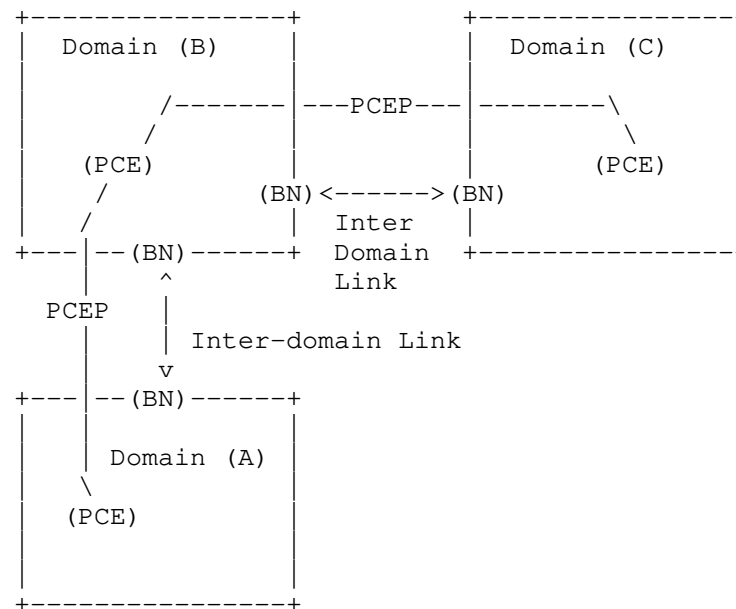
Stateful H-PCE [RFC8751] presents general considerations for stateful PCE(s) in the hierarchical PCE architecture. In particular, the behavior extends the existing stateful PCE mechanisms (including PCE-initiated LSP setup and active PCE usage) in the context of networks using the H-PCE architecture. Section 3.3.1 [RFC8751] describes the per-domain stitched LSP mode, where the individual per-domain LSPs are stitched together. PCInitiate message is also used to stitch the end-to-end tunnel. See section 4 for details.

1.1. General Assumptions

In the remainder of this document, the same references as per BRPC [RFC5441] are used and the following set of assumptions are made (see figure below):

- o Domain refers to administrative partitions, i.e. an IGP area or an Autonomous System (AS).
- o Inter-domain path is used to refer to a path that crosses two or more different domains as defined previously,
- o At least one PCE is deployed in each domain. These PCEs are all active stateful-capable and can request to enforce LSPs in their respective domain by means of PCInitiate messages.
- o LSRs, including border nodes, are PCC-enabled and support active stateful mode. PCEP sessions are established between these routers and their domains' PCE.

- o Each PCE establishes a PCEP session with its respective neighbor domains' PCEs. The way a PCE discovers its neighboring PCEs is out of the scope of this document. This information could be administratively configured or automatically discovered through, for example, [I-D.dong-pce-discovery-proto-bgp].
- o PCEs are able to compute an end-to-end path as per BRPC procedure [RFC5441] or as per H-PCE procedure (stateless [RFC6805] or stateful [RFC8751]).
- o "Path" is a generic term to refer to both LSP setup by mean of RSVP-TE or Segment Path in a Segment Routing network.



Example of the representation of 3 domains with 3 PCEs

1.2. Terminology

ABR: Area Border Routers. Routers used to connect two IGP areas (areas in OSPF or levels in IS-IS).

AS: Autonomous System

ASBR: Autonomous System Border Router. Router used to connect together ASes (of the same or different service providers) via one or more inter-AS links.

Border Node (BN): a boundary node is either an ABR in the context of inter-area TE or an ASBR in the context of inter-AS TE.

BN-en(i): Entry BN of domain(i) connecting domain(i-1) to domain(i) along a determined sequence of domains. Multiple entry BN-en(i) could be used to connect domain(i-1) to domain(i).

BN-ex(i): Exit BN of domain(i) connecting domain(i) to domain(i+1) along a determined sequence of domains. Multiple exit BN-ex(i) could be used to connect domain(i) to domain(i+1).

Domains: Autonomous System (AS) or IGP Area. An Autonomous System is composed by one or more IGP area.

ERO(i): The Explicit Route Object scoped to domain(i)

IGP-TE: Interior Gateway Protocol with Traffic Engineering support. Both OSPF-TE and IS-IS-TE are identified in this category.

Inter-domain path: A path that crosses two or more domains through a pair of Border Node (BN-ex, BN-en).

LK(i): A Link that connect BN-ex(i-1) to BN-en(i). Note that BN-ex(i-1) could be connected to BN-en(i) by more than one link. LK(i) identifies which of the multiple links will be used for the inter-domain path setup. For inter-AS scenario, LK(i) represents the link between ASBR of domain i to the ASBR of domain i-1. For inter-area scenario, LK(i) is present only in IS-IS networks and represents the link between ABR of region L1, reciprocally L2, to the ABR of region L2, reciprocally L1.

Local path: A path that does not cross a domain border. It is set up either from entry BN-en, to output BN-ex or between both. This path could be enforce by means of RSVP-TE signaling or Segment Routing labels stack.

Local path(i): A Local path of domain(i)

PLSP-ID(i): A PLSP-ID that identifies, in the domain(i), the local part of an inter-domain path.

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.

PCE(i) is a PCE within the scope of domain(i).

PST: Path Setup Type

$R(i,j)$: The router j of domain i

Stitching Label (SL): A dedicated label that is used to stitch two RSVP-TE LSPs or two Segment Routing paths.

$SL(i)$: A Stitching Label that links domain($i-1$) to domain(i).

2. Stitching Label

This section introduces the concept of Stitching Label that allows stitching and nesting of local paths in order to form an inter-domain path that cross several different domains.

2.1. Definition

The operation of stitch or nest a local path(i) to a local path($i+1$) in order to form an inter-domain path mainly consists in defining the label that the output BN-ex(i) will use to send its traffic to the entry BN-en($i+1$). Indeed, the entry BN-en($i+1$) needs to identify the incoming traffic (e.g. IP packets), in order to know if this traffic must follow the local path($i+1$) or not. Forwarding Equivalent Class (FEC) could be used for that purpose. But, when stitching or nesting tunnels, the FEC is reduced to the incoming label that the entry BN-en($i+1$) has chosen for the local path($i+1$).

In this memo, we introduce the term of "Stitching Label (SL)" to refer to this label. Such label is usually exchanged between output BN-ex(i) and entry BN-en($i+1$) with the RSVP-TE signaling. But, as we want to avoid to use RSVP-TE signaling due to operational constraints, and allow compatibility support for Segment Routing, this Stitching Label is here conveyed by PCEP. In fact, the Explicit Route Object (ERO) and the Record Route Object (RRO) are already defined in order to transport (G)MPLS labels (for RSVP-TE or Segment Routing) in the PCEP signaling. Thus, the Stitching Label could be conveyed in the ERO and RRO without any modification of PCEP nor PCEP Objects.

As per RFC4003 [RFC4003], the Stitching Label will be conveyed as a companion of a link identifier (e.g. an IP address for numbered links). In our case, this is one of the endpoint IDs of the link $LK(i)$ which connects BN-ex(i) to BN-en($i+1$) and carries the traffic from the domain(i) to domain($i+1$). It is left to implementation to select which of the two endpoint IDs of the link $LK(i)$ is used.

2.2. Inter-domain LSP-TYPE

Even if PCEP could convey the Stitching Label, a PCC is not aware that a PCE requests or provides such a label. For that purpose, this specification relies on the use of the PST as defined in [RFC8408] with new values (See IANA section of this memo) defined as follow:

- o TBD1: Inter-Domain TE end-to-end path is set up using Backward Recursive or Hierarchical method. This new PST value MUST be set in a PCInitiate messages sends by a PCE(i-1) to its neighbor PCE(i) in the Backward Recursive method or by the Parent PCE to the Child PCE(i) to initiate a new inter-domain path. In its response, the neighbor PCE(1) or Child PCE(i) MUST return a Stitching Label SL with an identifier of the associated link in the RRO of the PCRpt message to PCE(i-1) or Parent PCE.
- o TBD2: Inter-Domain TE local path is set up using RSVP-TE. This new PST value MUST be set in the PCInitiate message sends by a PCE(i) requesting to a PCC of domain(i) to initiate a new local path(i) which is part of an inter-domain path. This PST value MUST be used by the PCE(i) only after receiving a PCInitiate message with an PST equal to TBD1 from a neighbor PCE(i-1) in the Backward Recursive method or Parent PCE in the Hierarchical method. In its response, the PCC of domain(i) MUST return a Stitching Label SL with the an identifier of associated link in the RRO of the PCRpt message.
- o TBD3: Inter-Domain TE local path is set up using Segment Routing. This new PST value MUST be set in the PCInitiate message sends by a PCE(i) requesting to a PCC of domain(i) to initiate a new Segment Routing path which is part of and inter-domain Segment Routing path. This PST value MUST be used by the PCE(i) only after receiving a PCInitiate message with an PST equal to TBD1 from a neighbor PCE(i-1). In its response, the PCC MUST return a Stitching Label SL with an identifier of the associated link in the RRO of the PCRpt message.

3. Backward Recursive PCInitiate Procedure

This section describes how to set up inter-domain paths that cross different domains by using a Backward Recursive method. It is compatible with the inter-domain path computation by means of the BRPC procedure as describe in RFC5441 [RFC5441].

3.1. Mode of Operation

This section describes how PCInitiate and PCRpt messages are combined between PCE in order to set up inter-domain paths between a source domain(1) to a destination domain(n). S and D are respectively the source and destination of the inter-domain path. Domain(1) and domain(n) are different and connected through 0 (i.e. direct connection when n = 2) or more intermediate domains denoted domain(i) with i = [2, n-1].

First, the PCE(1) runs standard BRPC algorithm as per RFC5441 [RFC5441] with its neighbor PCEs in order to compute the inter-domain path from S to D, where S and D are respectively a node in the domain(1) and domain(n). Path Key confidentiality as per RFC5520 [RFC5520] SHOULD be used to obfuscate the detailed ERO(i) of the different domains(i). The resulting ERO is in the form {S, PKS(1), BN-ex(1), ..., BN-en(i), PKS(i), BN-ex(i), ..., BN-en(n), PKS(n), D} when Path Key is used and of the form {S, R(1,1), ..., R(1,k), BN-ex(1), ..., BN-en(i), R(i,1), ..., R(i,l), BN-ex(i), ..., BN-en(n), R(n,1), ..., R(n,m), D} otherwise. As subsequent domains are not aware about the computed end-to-end ERO in case of Virtual Source Path trees (VSPTs), the final ERO selected by the PCE(1) MUST be sent in the PCInitiate message to indicate to the subsequent PCEs which path has been finally chosen. PCE(1) MUST ensure that this ERO is self comprehensive by subsequent PCEs. Indeed, when a PCE(i) receives the ERO, it MUST be able to verify that this ERO matches its own scope and to determine the PCE(i+1). When Path Key is used, PCEs MUST encode the Path Key with a reachable IP address so that previous PCEs in the AS chain are able to join them. When Path Key is not used, the PCEs MUST be able to retrieve an IP address of the next PCE corresponding to the ERO (e.g., relying on a per prefix table).

The complete procedure with Path Key follows the different steps described below:

Steps 1: Initialization

Once ERO(S, D) is computed, PCE(1) sends a PCInitiate message to PCE(2) containing an ERO equal to {S, PKS(2), ..., PKS(i), ..., PKS(n), D}, PST = TBD1 and End-Points Object = (S, D). The ERO corresponds to the one PCE(1) has received from PCE(2) during the BRPC process in which only Path Key are kept. In case of multiple EROs, i.e. VSPT, PCE(1) has chosen one of them and used the selected one for the PCInitiate message. PKS(i) could be replaced by the full ERO description if Path Key is not used by PCE(i).

When PCE(i) receives a PCInitiate message from domain(i-1) with PST = TBD1 and ERO = {PKS(i), PKS(i+1), ..., PKS(n), D}, it sends a

PCInitiate message to PCE(i+1) with a popped ERO and records its received PKS(i) part. All PCE(i)s generate the appropriate PCInitiate message to PCE(i+1) up to PCE(n), i.e. to the destination domain(n).

Steps 2: Actions taken at the destination domain(n) by PCE(n)

1. When a PCInitiate message reaches the destination domain(n), PCE(n) retrieves the ERO from the PKS(n) if necessary and sends to BN-en(n) a PCInitiate message with the ERO(n) = {BN-en(n), ..., D}, PST = TBD2 and End-Points Object = {BN(n), D} in order to inform the PCC BN-en(n) that this local path(n) is part of an inter-domain path.
2. When the PCC BN-en(n) receives the PCInitiate message from its PCE(n), it sets up the local path from entry BN-en(n) to D by means of RSVP-TE signaling with the given ERO(n).
3. Once the tunnel is set up, BN-en(n) chooses a free label for the Stitching Label SL(n) and adds a new entry in its MPLS L(F)IB with this SL(n) label. Then, it sends a PCRpt message to its PCE(n) with an RRO equal to {[LK(n), SL(n)], RRO(n)} and PLSP-ID(n).
4. Once PCE(n) receives the PCRpt from the PCC BN-en(n) with the RRO, PLSP-ID and PST = TBD2, it sends to the PCE(n-1) a PCRpt containing the RRO equal to {[LK(n), SL(n)]} and PLSP-ID(n). PCE(n) MAY add {PKS(n), D} in the RRO.

Steps i: Actions performed by all intermediate domains(i), for i = 2 to n-1

1. When the PCE(i) receives a PCRpt message from domain(i+1) with PST = TBD1, RRO = {[LK(i+1), SL(i+1)]} and PLSP-ID(i+1), it retrieves the ERO(i) from the PKS(i), recorded in step 1, and sends to the PCC BN-en(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, PST = TBD2 and End-Points Object = {BN-en(i), BN-ex(i)} in order to inform the PCC BN-en(i) that this local path(i) is part of an inter-domain path.
2. When the PCC BN-en(i) receives the PCInitiate message from its PCE(i), it sets up the local path from BN-en(i) to BN-ex(i) by means of RSVP-TE signaling with the given ERO(i).
3. Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is used to instruct the egress node of domain(i), i.e. BN-ex(i), to forward packets belonging to this tunnel with the Stitching Label. Both the Stitching Label and the identifier of the

interface are carried in the ERO = {..., [LK(i+1), SL(i+1)]} as the last SubObject in conformance to [RFC4003]. As a result, BN-ex(i) installs in its MPLS L(F)IB the SWAP instruction to label SL(i+1) with forward to LK(i+1).

4. Once the tunnel is set up, PCC BN-en(i) chooses a free label for the Stitching Label SL(i) and adds a new entry in its MPLS L(F)IB with this SL(i) label. Then, it sends a PCRpt message to its PCE(i) with an RRO equal to {[LK(i), SL(i)], RRO(i)} and PLSP-ID(i).
5. Once PCE(i) receives the PCRpt from the PCC BN-en(i) with the RRO and PST = TBD2, it sends to PCE(i-1) a PCRpt message containing the RRO equal to {[LK(i), SL(i)]} and the PLSP-ID(i). PCE(i) MAY add {PKS(i), ..., PKS(n)} in the RRO.

Steps n: Actions performed at the source domain(1) by PCE(1)

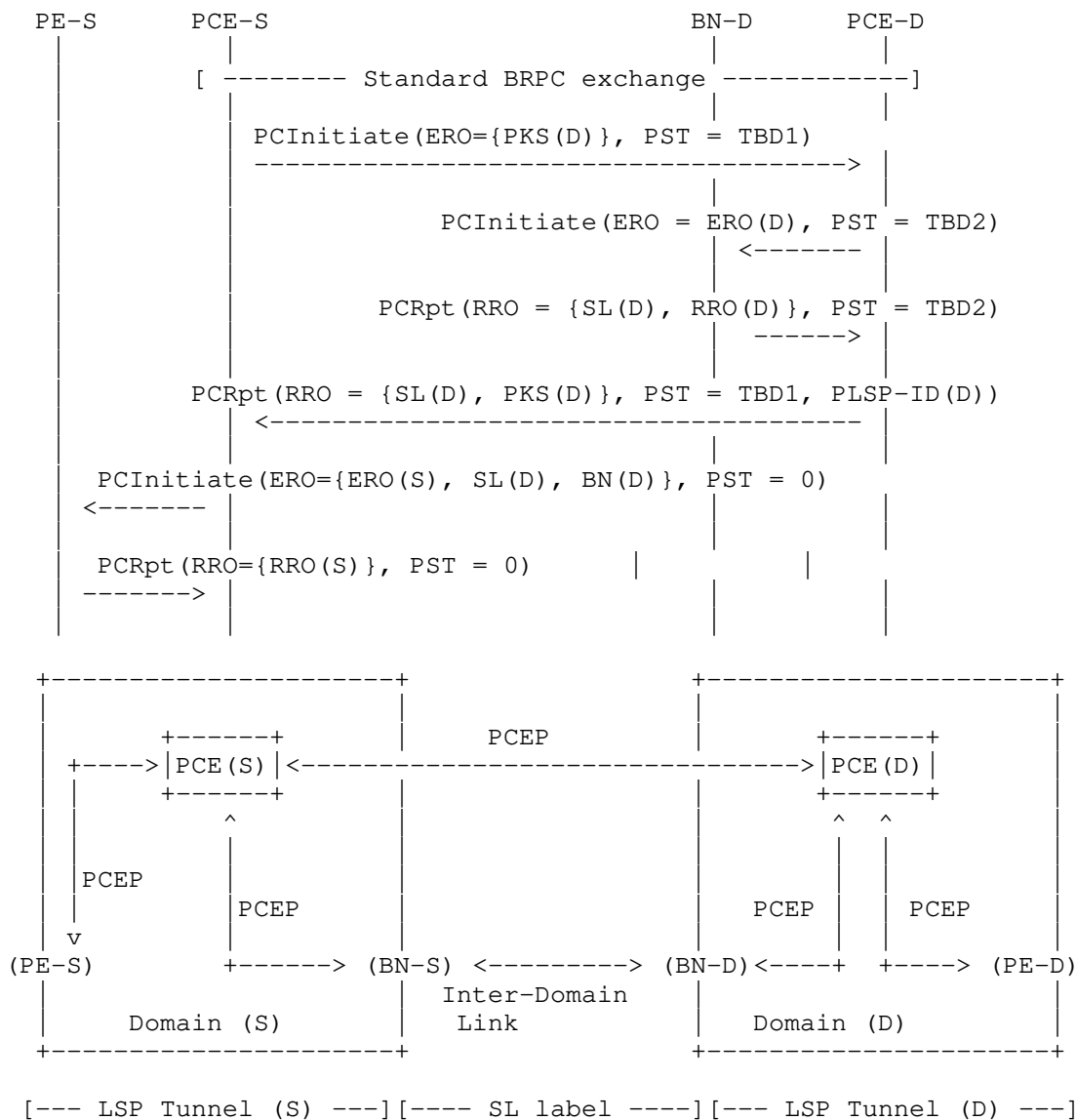
Once PCE(1) receives the PCRpt message from PCE(2) with the RRO containing the label SL(2), it sends a PCInitiate message to PCC node S with ERO equal to {ERO(1), [LK(2), SL(2)]}, PST = 0 and End-Points Object = {S, BN-ex(1)}. This time, the PST is equal to 0 as the PCC S does not need to return a Stitching Label SL, because it is the head-end of the inter-domain path. A usual PCRpt message is sent back to PCE(1) by the PCC node S.

3.2. Example

In the figure below, two different domains S and D are interconnected through BN respectively BN-S and BN-D. PE-S and PE-D are edge routers. All routers in the figure are connected to their respective PCE through PCEP. In this example, we consider that PCE(S) needs to set up an inter-domain path between PE-S and PE-D acting as source and destination of the path. To simplify the figure, neither intermediate routers between (PE-S, BN-S), (BN-D and PE-D), nor RSVP-TE messages are represented, but they are all presents. The following notation is used (in this example, we use the PKS for the sake of simplicity):

- o PKS(D) = Path Key corresponding to the path from BN(D) to PE-D
- o ERO(D) = Explicit Route Object corresponding to the path from BN(D) to PE-D, retrieved from PKS(D)
- o RRO(D) = Record Route Object of the local path(D) from BN(D) to PE-D
- o SL(D) = Stitching Label for the local path from BN(D) to PE-D

- o ERO(S) = Explicit Route Object corresponding to the path from PE-S to BN(S)
- o RRO(S) = Record Route Object of local path(S) from PE-S to BN(S)



Example of inter-domain path setup between two domains

3.3. Completion Failure of Inter-domain Path Setup Procedure

In case of error during path setup, PCRpt and or PCErr messages MUST be used to signal the problem to the neighbor PCE domain backward. In particular, if the new PST values defined in this memo are not supported by the neighbor PCE or the PCC, the PCE, respectively the PCC, MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = 1 (Unsupported path setup type) to its neighbor PCE. If a PCE(i) receives a PCInitiate message from its peer PCE(i-1) without PST set to TBD1 or PST set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = 1 (Unsupported path setup type) to its peer PCE(i-1).

Following a PCInitiate message with PST set to TBD1, if a PCC or a PCE returns no RRO, or an RRO without the Stitching Label SL and an identifier of the associated link, the PCE MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = TBD5 (Mandatory Stitching Label missing in the RRO).

In case of completion failure, the PCE(i) MUST propagate the PCErr message up to the PCE(1). In turn, PCE(1) MUST send a PCInitiate message (R flag set in the SRP Object as per [RFC8281]) to tear down this inter-domain path from its neighbor PCEs. PCE(i) MUST propagate the PCInitiate message and remove its local path by means of PCInitiate message to its PCC BN-en(i) and send back PCRpt message to PCE(i-1).

In case of error in domain(i+1), PCE(i) MAY add the AS number of domain(i+1) in the RRO to identify the faulty domain.

4. Hierarchical PCInitiate Procedure

This section describes how to set up inter-domain paths that cross different domains by using a hierarchical method. It is compatible with inter-domain path computation as described in [RFC6805].

4.1. Mode of Operation

This section describes how PCInitiate and PCRpt messages are combined between PCEs in order to set up inter-domain paths between a source domain(1) to a destination domain(n). S and D are respectively the source and destination of the inter-domain path. Domain(1) and domain(n) are different and connected through 0 or more intermediate domains denoted domain(i) with $i = (2, n-1)$. Domains are directly connected when $n = 2$.

First, the Parent PCE contacts its Child PCE as per [RFC6805] in order to compute the inter-domain path from S to D, where S and D are respectively a node in the domain(1) and domain(n). Path Key confidentiality as per RFC5520 [RFC5520] SHOULD be used to obfuscate the detailed ERO(i) of the different domains(i). The resulting ERO is of the form (S, PKS(1), BN-ex(1), ..., BN-en(i), PKS(i), BN-ex(i), ..., BN-en(n), PKS(n), D) when Path Key is used and of the form {S, R(1,1), ..., R(1,k), BN-ex(1), ..., BN-en(i), R(i,1), ..., R(i,l), BN-ex(i), ..., BN-en(n), R(n,1), ..., R(n,m), D} otherwise.

The complete procedure with Path Key follow the different steps described below:

Step 1: Initialization

1. The Parent PCE sends a PCInitiate message to Child PCE(n) with an ERO = {PKS(n)} and End-Points = {BN-en(n), D}. Then, PCE(n) retrieves the ERO from the PKS(n) (if necessary) and sends to BN-en(n) a PCInitiate message with the ERO(n) = {BN-en(n), ..., D}, PST = TBD2 and End-Points Object = {BN-en(n), D} in order to inform the PCC BN-en(n) that this local path(n) is part of an inter-domain path.
2. When the PCC BN-en(n) receives the PCInitiate message from its PCE(n), it sets up the local path from the entry BN-en(n) to D by means of RSVP-TE signaling with the given ERO(n).
3. Once the path is set up, it chooses a free label for the Stitching Label SL(n) and adds a new entry in its MPLS L(F)IB with this SL(n) label. Then, it sends a PCRpt message to its PCE(n) with an RRO equal to {[LK(n), SL(n)], RRO(n)} and PLSP-ID(n).
4. Once PCE(n) receives the PCRpt from the PCC BN-en(n) with the RRO, PLSP-ID and PST = TBD2, it sends to its Parent PCE a PCRpt containing the RRO equal to {[LK(n), SL(n)]} and PLSP-ID(n). PCE(n) MAY add PKS(n) in the RRO.

Steps i: Actions performed for all intermediate domains(i), for i = n-1 to 2

1. The Parent PCE sends a PCInitiate message to Child PCE(i) with PST = TBD1, ERO = {PKS(i), [LK(i+1), SL(i+1)]} and End-Points = {BN-en(i), BN-ex(i)}
2. Then, PCE(i) retrieves the ERO from the PKS(i) if necessary and sends to the PCC BN-en(i) a PCInitiate message with ERO = {ERO(i), [LK(i+1), SL(i+1)]}, PST = TBD2 and End-Points Object =

{BN-en(i), BN-ex(i)} in order to inform the PCC BN-en(i) that this local path(i) is part of an inter-domain path.

3. When the PCC BN-en(i) receives the PCInitiate message from its PCE(i), it sets up the local path from BN-en(i) to BN-ex(i) by means of RSVP-TE signaling with the given ERO(i).
4. Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is used to instruct the egress node of domain(i), i.e. BN-ex(i) to forward packets belonging to this tunnel with the Stitching Label. Both the Label Stitching and an identifier of the outgoing interface are carried in the ERO = {..., [LK(i+1), SL(i+1)]} as the last SubObject in conformance to [RFC4003]. So that, BN-ex(i) installs in its MPLS L(F)IB the SWAP instruction to label SL(i+1) with forward to LK(i+1) instead of the usual POP instruction.
5. Once the tunnel is set up, PCC BN-en(i) chooses a free label for the Stitching Label SL(i) and adds a new entry in its MPLS L(F)IB with this SL(i) label. Then, it sends a PCRpt message to its PCE(i) with an RRO equal to {[LK(i), SL(i)], RRO(i)} and PLSP-ID(i).
6. Once PCE(i) receives the PCRpt from the PCC BN-en(i) with the RRO and PST = TBD2, it sends to its Parent PCE a PCRpt message containing the RRO equal to {[LK(i), SL(i)]} and the PLSP-ID(i). PCE(i) MAY add PKS(i) in the RRO.
7. Once the Parent PCE receives the PCRpt from the Child PCE(i), it stores the corresponding PLSP-ID for this inter-domain path part.

Steps n: Actions performed to the source domain(1)

Finally, the Parent PCE sends a last PCInitiate message to its Child PCE(1) with PST = TBD1, ERO = {PKS(1), [LK(2), SL(2)]} and End-Points = {S, BN-ex(1)}. In turn, Child PCE(1) sends a PCInitiate message to PCC node S with ERO equal to {ERO(1), [LK(2), SL(2)]}, PST = 0 and End-Points Object = {S, BN-ex(1)}. This time, the PST is equal to 0 as the PCC S does not need to return a Stitching Label SL, because it is the head-end of the inter-domain path. A usual PCRpt message is sent back to PCE(1) by the PCC node S. In turn, Child PCE(1) sends a final PCRpt message to the Parent PCE with the PSLP-ID(1). PCE(1) MAY add {S, BN-ex(1)} in the RRO as a loose path.

4.2. Completion Failure of Inter-domain Path Setup Procedure

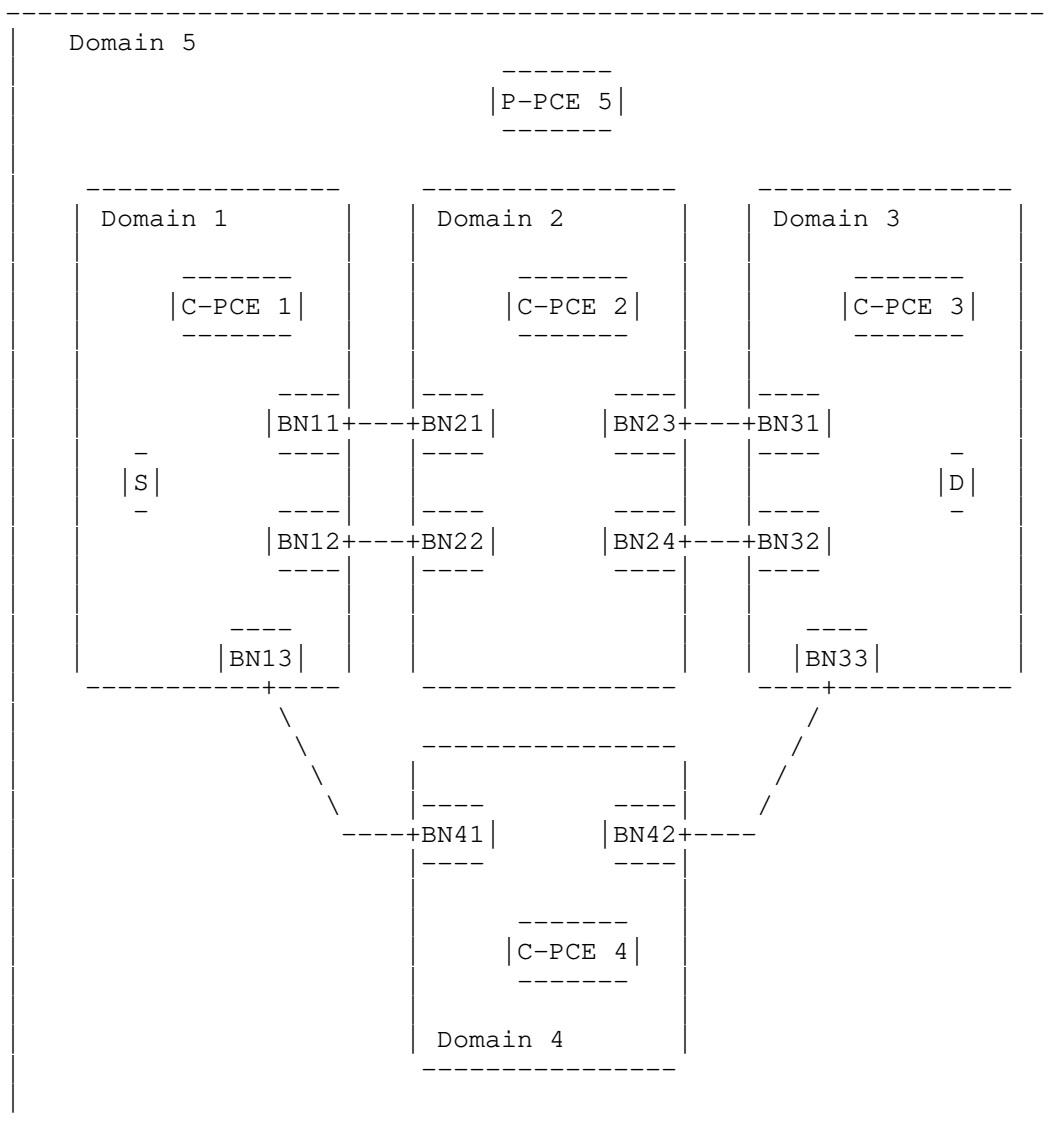
In case of error during path set up, PCRpt and or PCErr messages MUST be used to signal the problem to the Parent PCE. In particular, if the new PST values defined in this memo are not supported by the Child PCE or the PCC, the Child PCE, respectively the PCC, MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = 1 (Unsupported path setup type) to its Parent PCE. If Child PCE(i) receives a PCInitiate message from its Parent PCE without PST set to TBD1 or PST set to a value different from TBD1, it MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = 1 (Unsupported path setup type) to its Parent PCE.

Following a PCInitiate message with PST set to TBD1, if a Child PCE or a PCC returns no RRO, or an RRO without the Stitching Label SL and an identifier of the associated link, the Parent PCE, respectively the Child PCE, MUST return a PCErr message with Error-Type = 21 (TE path setup error) and Error-Value = TBD5 (Mandatory Stitching Label missing in the RRO).

In case of completion failure, the Parent PCE MUST send a PCInitiate message (R flag set in the SRP Object as per [RFC8281]) to tear down this inter-domain path from the Child PCEs that already set up their respective part of the inter-domain path. Child PCE(i) MUST remove its local path by means of PCInitiate message with R flag set to 1 to its PCC BN-en(i) and send back a PCRpt message to the Parent PCE.

4.3. Example for Stateful H-PCE Sticking Procedure

Taking the sample hierarchical domain topology example from [RFC6805] as the reference topology for the entirety of this section.



Hierarchical domain topology from RFC6805

Section 3.3.1 of [RFC8751] describes the per-domain stitched LSP mode and list all the steps needed. To support SL-based stitching, using the reference architecture described in the figure above, the steps are modified as follows (note that we do not use PKS in this example for simplicity):

Step 1: initialization

The P-PCE (PCE5) is requested to initiate a path. Steps 4 to 10 of section 4.6.2 of [RFC6805] are executed to determine the end-to-end path, which are split into per-domain paths, e.g. {S-BN41, BN41-BN33, BN33-D}.

Step 2: Path (BN33-D) at C-PCE3:

1. The P-PCE (P-PCE5) sends the initiate request to the C-PCE (C-PCE3) via PCInitiate message for path (BN33-D) with ERO={BN33..D} and PST = TBD1.
2. C-PCE3 further propagates the initiate message to BN33 with the ERO and PST = TBD2/TBD3 based on the setup type.
3. BN33 initiates the setup of the path and reports to the status ("GOING-UP") to C-PCE3.
4. C-PCE3 further reports the status of the path to the P-PCE (P-PCE5)
5. The node BN33 notifies the path state to C-PCE3 when the state is "UP"; it also sends the Stitching Label (SL33) in the RRO as {SL33,BN33..D}.
6. C-PCE3 further reports the status of the path to the P-PCE (P-PCE5) as well as sends the Stitching Label (SL33) in the RRO as {LK33,SL33,BN33..D}.

Step 3: Path (BN41-BN33) at C-PCE4

1. The P-PCE (P-PCE5) sends the initiate request to the C-PCE (C-PCE4) via PCInitiate message for path (BN41-BN33) with ERO={BN41..BN42,LK33,SL33,BN33} and PST = TBD1.
2. C-PCE4 further propagates the initiate message to BN41 with the ERO and PST = TBD2/TBD3 based on the setup type. In case of RSVP_TE, the node BN41 encode the Stitching Label SL33 as part of the ERO to make sure the node BN42 uses the label SL33 towards node BN33. In case of SR, the label SL33 is part of the label stack pushed at node BN41.
3. BN41 initiates the setup of the path and reports the path status ("GOING-UP") to C-PCE4.
4. C-PCE4 further reports the status of the path to the P-PCE (P-PCE5).

5. The node BN41 notifies the path state to C-PCE4 when the state is "UP"; it also sends the Stitching Label (SL41) in RRO as {LK41,SL41,BN41..BN33}.
6. C-PCE4 further reports the status of the to the P-PCE (P-PCE5) as well as sends the Stitching Label (SL41) in the RRO as {LK41,SL41,BN41..BN33}.

Step 3: Path (S-BN41) at C-PCE1

1. The P-PCE (P-PCE5) sends the initiate request to the C-PCE (C-PCE1) via PCInitiate message for path (S-BN41) with ERO={S..BN13,LK41,SL41,BN41}.
2. C-PCE1 further propagates the initiate message to node S with the ERO. In case of RSVP-TE, node S encodes the Stitching Label SL41 as part of the ERO to make sure the node BN13 uses the label SL41 towards node BN41. In case of SR, the label SL41 is part of the label stack pushed at node S.
3. S initiates the setup of the path and reports the path status ("GOING-UP") to C-PCE1.
4. C-PCE1 further reports the status of the path to the P-PCE (P-PCE5)
5. The node S notifies the path state to C-PCE1 when the state is "UP".
6. C-PCE1 further reports the status of the path to the P-PCE (P-PCE5).

In this way, per-domain paths are stitched together using the Stitching Label (SL). The per-domain paths MUST be set up from the destination domain towards the source domain one after the other.

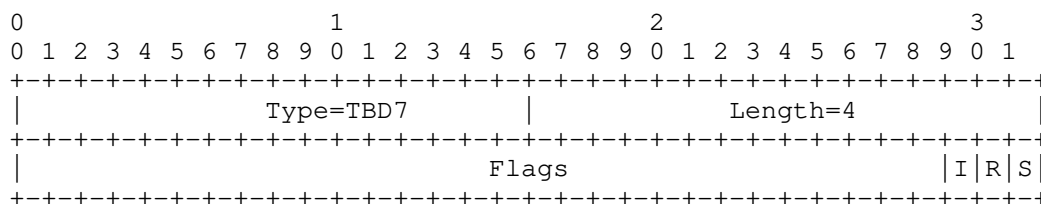
Once the per-domain path is set up, the entry BN chooses a free label for the Stitching Label SL and adds a new entry in its MPLS L(F)IB with this SL label. The SL from the destination domain is propagated to adjacent transit domain, towards the source domain at each step. This happens from the entry BN to C-PCE then to the P-PCE, and vice-versa. In case of RSVP-TE, the entry BN further propagates the SL label to the exit BN via RSVP-TE. In case of SR, the SL label is pushed as part of the SR label stack.

5. Inter-domain Path Management

This section describes how inter-domain paths could be managed.

5.1. Stitching Label PCE Capabilities

A PCE needs to know if its neighbor PCEs as well as PCCs are able to configure and provide a Stitching Label. The STITCHING-LABEL-PCE-CAPABILITY TLV is an optional TLV for use in the OPEN object for Stitching Label PCE capability advertisement. Its format is shown in the following figure:



STITCHING-LABEL-PCE-CAPABILITY TLV Format

The Type (16 bits) of the TLV is TBD7. The Length field is 16 bits long and has a fixed value of 4.

The value comprises a single 32 bits "Flags" field:

R (RSVP-TE-STITCHING-LABEL-CAPABILITY - 1 bit): if set to 1 by a PCC, the R flag indicates that the PCC is able to provide Stitching Labels, for RSVP-TE inter-domain paths, when requested by a PCE. If set to 1 by a PCE, the R flag indicates that the domain controlled by this PCE is able to set up inter-domain paths by means of RSVP-TE signaling.

S (SEGMENT-ROUTING-STITCHING-LABEL-CAPABILITY - 1 bit): if set to 1 by a PCC, the S flag indicates that the PCC is able to provide Stitching Labels, for Segment-Routing inter-domain paths, when requested by a PCE. If set to 1 by a PCE, the R flag indicates that the domain controlled by this PCE is able to set up inter-domain paths by means of Segment Routing.

I (INTER-DOMAIN-STITCHING-LABEL-CAPABILITY - 1 bit): if set to 1 by a PCE, the I flag indicates that the domain is supporting Stitching Label to set up inter-domain paths. This flag is reserved for PCEP session established between PCEs and MUST be kept unset by a PCC.

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

PCCs MUST set the R and/or S flags and MUST NOT set the I flag when adding the Stitching Label Capability to the PCEP Open Message. The RSVP-TE-STITCHING-LABEL-CAPABILITY, respectively SEGMENT-ROUTING-STITCHING-LABEL-CAPABILITY, flag must be set by both the PCC and PCE in order to enable the configuration of Stitching Labels with RSVP-TE, respectively with Segment-Routing.

A PCE MUST set the I flag when establishing a PCEP session with a neighbor PCE when adding Stitching Label Capability to the PCEP Open Message. It MAY set R and/or S flags depending if the operator would like to keep confidential the technology used to set up inter-domain paths or not. The INTER-DOMAIN-STITCHING-LABEL-CAPABILITY flag must be set by both PCEs in order to enable inter-domain paths instantiation by means of Stitching Label.

5.2. Identification of Inter-domain Paths

First, in order to manage inter-domain paths composed by the stitching or nesting of local paths, it is important to identify them. For this purpose, the PLSP-ID managed by the PCEs are combined to one provided by PCCs to form a global identifier as follow:

- o PCE(i) in the Backward Recursive method or the Child PCE in Hierarchical method MUST create a new unique PLSP-ID for this inter-domain path part and MUST send it in the PCRpt message, to the PCE(i-1), respectively the Parent PCE. In addition this new PLSP-ID MUST be associated to the one received from the PCC that instantiates the local path part for further reference.
- o In the Hierarchical mode, the Parent PCE MUST store and associate the different PLSP-ID(i)s received from the different Child PCE(i)s in order to identify the different part of the inter-domain paths.
- o In the Backward Recursive method, PCE(i) MUST store and associate its PLSP-ID(i) and the PLSP-ID(i+1) it received from the PCE(i+1). PCE(n), i.e. the last one in the chain, does not need to perform such association.

Further reference to the inter-domain path will use this PLSP-ID(i). In the Backward Recursive method, PCE(i) MUST replace the PLSP-ID(i) by PLSP-ID(i+1) in the PCUpd, PCRpt or PCinitiate message before propagating it to PCE(i+1); and PCE(i) MUST replace the PLSP-ID(i+1) by PLSP-ID(i) in the PCRpt message before propagating it to the

PCE(i-1). In the Hierarchical method, the Parent PCE MUST use the corresponding PLSP-ID(i) of the Child PCE(i).

5.3. Inter-domain Association Group

In case of failure, a PCE(i) will received PCRpt messages from its PCCs and neighbors PCE(i+1) to synchronize the Inter-domain paths. In addition, it may received PCInitiate messages from its previous neighbors PCE(i-1) to re-initiate its inter-domain path part. As the PCE(i) may loose the PLSP-ID association, a new association group (within Association Object) is used to ease the association of the different parts of the inter-domain path: the local part and the PCE-to-PCE part. The use of the Association Object is MANDATORY in the Backward Recursive method and OPTIONAL in the Hierarchical method.

For that purpose, a new Inter-Domain Association Type with value TBD4 is defined. The first PCE in the Backward Recursive chain (the one which received the initial request) MUST send the PCInitiate message with an Association Object as follows:

- o Association Type field MUST be set to new value TBD4
- o Association ID MUST be set to a unique value. In case the Association ID field is too short or wraps, the first PCE MAY use the Extended Association ID to increase the number of association groups. The Association ID is managed locally by the PCE and does not need to be coordinated with neighbor or remote PCEs.
- o IPV4 or IPV6 association source MUST be set to the IP address which identifies PCE(1) in domain(1).
- o The Global Association Source TLV MUST be present and set with the ASN number of domain(1). It allows to create a globally unique association scope without putting constraint on operator's IP association source. Thus the IP Association Source is associated with the Global Association source to form a unique identifier.
- o Extended Association ID MAY be present and MANDATORY if association ID is too short or wraps.

Subsequent PCE(i), for i = 2 to n, MUST send this Association Object as is to the local PCC and the neighbor PCE(i+1).

In case of error with the association group, a PCErr message MUST be raised with Error = 26 (Association Error) and Error value set accordingly. A new Error value TBD6 is defined to identify association of inter-domain paths.

In the Hierarchical method, the Parent PCE MAY act as the initiator of the Association and send to the Child PCEs an Association Object that follows the same rules as for the Backward Recursive method. In turn, Child PCEs MUST propagate the Association Object to the local PCCs as is.

5.4. Modification of Inter-domain Paths

For the Backward Recursive method, each domain manages their respective local path part of an inter-domain path independently of each other. In particular, Stitching Label(i) is managed by domain(i) and is of interest of domain(i-1) only. Thus, Stitching Label SL(i) is not supposed to be propagated to other domains. The same behavior apply to PLSP-ID(i). In the Hierarchical method, the Parent PCE MUST ensure the correct distribution of Stitching Label SL(i) to Child PCE(i-1). The PLSP-ID(i) is kept for the usage of the Parent PCE and thus is not propagated. Only the Association Object defined in section 5.2 is propagated if it is present.

If PCE(i) needs to modify its local path(i) with a PCUpd message to the PCC BN-en(i), once the PCRpt message received from the PCC BN-en(i), it MUST send a new PCRpt message to advertise the modification. This message is targeted to its neighbor PCE(i-1) in the Backward Recursive method, respectively to the Parent PCE in the Hierarchical method. In this case PLSP-ID(i) is used to identify the inter-domain path. PCE(i-1), respectively the Parent PCE, MUST propagate the PCRpt message if the modification implies the upstream domain, e.g. if the PCRpt indicates that the Stitching Label SL(i) has changed.

PCE(1), respectively the Parent PCE, could modify the inter-domain path. For that purpose, it MUST send a PCUpd message to its neighbor PCEs, respectively Child PCE, using the PLSP-ID it received. Each PCE(i) MUST process the PCUpd message the same way they process the PCInitiate message as define in section 3.1 for the Backward Recursive method and in section 4.1 for the Hierarchical method.

In case a failure appear in domain(i), e.g. path becoming down, PCE(i) MUST send a PCRpt message to its neighbor PCE(i-1), respectively its Parent PCE to advertise the problem in its local part of the inter-domain path. Once PCE(1), respectively the Parent PCE, receives this PCRpt message indicating that the path is down, it is up to the PCE(1), respectively the Parent PCE to take appropriate correction e.g. start a new path computation to update the ERO.

5.5. Modification of Inter-domain Paths

Modification of local path, BN-en(i) and BN-ex(i) is left for further study.

5.6. Tear-Down of Inter-domain Paths

The tear-down of an inter-domain path is only possible by the inter-domain path initiator i.e. PCE(1). For the Backward Recursive method, a PCInitiate message with R flag set to 1, PLSP-ID set accordingly to section 5.1 and the Association Object with R flag set to 1, is sent by PCE(1) to PCE(n) through PCE(i), and processed the same way as described in section 3.1. For the Hierarchical method, a PCInitiate message with R flag set to 1 is sent by the Parent PCE to each Child PCE(i) with corresponding PLSP-ID(i), and processed according to section 4.1. Each domain PCE(i) is responsible to tear down its part of the path and the PCC MUST release both the Stitching label SL in its L(F)IB and the path when it receives the PCInitiate message with the R flag set to 1 and the corresponding PLSP-ID. The Association Group MUST also be removed by the PCC and PCE(i).

6. Applicability

The newly introduce Stitching Label SL serves to stitch or nest part of local paths to form an inter-domain path. Each domain is free to decide if the incoming path is stitched or nested and how the path is enforced, e.g. through RSVP-TE or Segment Routing. At the peering point, the Border Node BN-ex(i) MUST encapsulate the packet with the Stitching Label, i.e. the MPLS label prior to send them to the next Border Node BN-en(i+1). Thus, only RSVP-TE and Segment Routing over MPLS technology are detailed in the following sections.

6.1. RSVP-TE

In case of RSVP-TE, the Border Node BN-ex(i) needs to received the Stitching Label from BN-en(i) through the RSVP-TE message and install in its L(F)IB a SWAP instruction to the Stitching Label and forward it to the next Border Node BN-en(i+1). For that purpose, the Egress Control mechanism, as per RFC4003 section 2.1 [RFC4003], is RECOMMENDED to instruct the Border Node BN-ex(i) of this action. Other mechanisms to program the L(F)IB could be used, e.g. NETCONF.

As the Stitching Label could serves to stitch or nest tunnels, a domain(i) may decide to nest the incoming LSPs into a higher hierarchy of LSPs for a Traffic Engineering purpose. A PCE(i) may also decide to group local LSPs part of inter-domain paths into a higher hierarchical LSP to carry all these local paths from a BN-en(i) to a BN-ex(i).

6.2. Segment Routing

To use Segment Routing instead of RSVP-TE to set up the local LSP tunnels as defined in [RFC8664], PCE(i) MUST send a PCInitiate message with PST = TBD3 instead of TBD2 to advertise its respective PCC that the local path is enforced by means of Segment Routing.

The Stitching Label SL(i+1) will be inserted into the label stack in order to become the top label in the stack when the packet reaches BN-en(i+1). Thus, the Stitching Label SL(i+1) serves as a FEC entry for BN-en(i+1) to identify the packets that follow the next Segment Path. For that purpose, BN-en(i+1) MUST install in its MPLS L(F)IB an instruction to replace the incoming Stitching Label SL(i+1) by the label stack given by the ERO(i+1) plus the Stitching Label SL(i+2), if any.

When a packet reaches BN-ex(i), the last label in the stack before the label SL(i+1) corresponds to a SID that allows to reach BN-en(i+1). When there are multiple interfaces between Border Nodes, BN-ex(i) needs to know how to send the packets to BN-en(i+1). Similarly to the Egress Control mechanism used with RSVP-TE, it is RECOMMENDED to use the inter-domain SID defined as per draft Egress Peer Engineering [I-D.ietf-idr-bgpls-segment-routing-epe] for that purpose. The inter-domain SID is announced by BN-ex(i) to PCE(i) through BGP-LS for each interface that connects BN-ex(i) to neighbors BN-en(i+1). Thus, the label stack will end with {BN-ex(i) SID, Inter-Domain SID, SL(i+1)} and should be processed as follows:

- o The penultimate router of domain(i) pops its node SID, and sends the packet to the next node designated by the top label in the label stack, i.e. the node SID of BN-ex(i) or the adjacency SID of the link between the router and BN-ex(i).
- o BN-ex(i) pops its node SID or its adjacency SID and looks up the next label in the stack, i.e. the inter-domain SID which corresponds to the interface to BN-en(i+1). BN-ex(i) pops this inter-domain SID as well and sends the packet to BN-en(i) through the corresponding interface.
- o BN-en(i+1) looks up the top label which is the Stitching Label SL(i+1), pops it and replaces it by the sub-sequent label stack.

Other mechanisms, e.g. NETCONF, could be used to configure the inter-domain SID on exit Border Nodes.

6.3. Mixing Technologies

During the instantiation procedure, if PCE(i) decides to reuse a local tunnel which is not yet part of an inter-domain tunnel, it SHOULD send a PCUpd message with PST = TBD2 to the PCC BN-en(i), in order to request a Stitching Label SL(i), and new ERO(i) to add the Stitching Label SL(i+1) and the associated link to the previous ERO.

[RFC8453] describes framework for Abstraction and Control of TE Networks (ACTN), where each Physical Network Controller (PNC) is equivalent to C-PCE and the Multi-Domain Service Coordinator (MDSC) to the P-PCE. The per-domain stitched LSP as per the Hierarchical PCE architecture described in Section 3.3.1 and Section 4.1 of [RFC8751] is well suited for ACTN. The Stitching Label mechanism as described in this document is well suited for ACTN when per-domain LSPs need to be stitched to form an E2E tunnel or a VN Member. It is to be noted that certain VNs require isolation from other clients. The SL mechanism described in this document can be applicable to the VN isolation use-case by uniquely identifying the concatenated stitching labels across multi-domain only to a certain VN member or an E2E tunnel.

As each operator is free to enforce the tunnel with its technology choice, it is a local policy decision for PCE(i) to instantiate the local part of the end-to-end tunnel by either RSVP-TE or Segment Routing. Thus, the PST value (i.e. TBD2 or TBD3) used in the PCInitiate message sent by the PCE(i) to the local PCC is determined by the local policy. How the local policy decision is set in the PCE is out of the scope of this memo. This flexibility is allowed because the SL principle allows to mix (data plane) technologies between domains. For example, a domain(i) could use RSVP-TE while domain(i+1) uses SR. The SL could serve to stitch indifferently Segment Paths and RSVP-TE tunnels. Indeed, the SL will be part of the label stack in order to become the top label in the stack when reaching the BN-en(i+1). This SL could be swapped as usual if the next domain uses RSVP-TE tunnels. When the upstream domain uses an RSVP-TE tunnel, the SL will serve as a key for the BN-en(i+1) to determine which label stack it must use on top of the packet for a Segment Routing path.

6.4. Inter-Area

If use cases for inter-AS are easily identifiable, this is less evident for inter-area. However, two scenarios have been identified:

- o Paths between levels for IS-IS networks.
- o Reduction of labels stack depth for Segment Routing.

Thus, the SL could be used to stitch or nest independent tunnels deployed through different IS-IS levels, even if there are controlled by the same PCE. IS-IS levels are considered as domains but under the control of the same PCE. In this scenario, there is no exchange between PCEs (it remains internal and implementation matter) and new TLVs are only applicable between the PCE and PCCs. The PCE requests to the different PCCs it identifies (i.e. BNs of the different IS-IS levels) to set up SLs and propagated them.

In large-scale networks, MSD could constraints the path computation in the possibility of path selection i.e. explicit expression of a path could exceeded the MSD. The SL could be used to split a too long explicit path regarding the MSD constraints. In this scenario, there is also no communications between PCEs and new TLVs are only used between PCE and PCCs.

7. IANA Considerations

7.1. Path Setup Type Values

[RFC8408] defines the PATH-SETUP-TYPE TLV. IANA is requested to allocate new code points in the PCEP PATH-SETUP-TYPE TLV PST field registry, as follows:

Value	Description	Reference
TBD1	Inter-domain TE end-to-end path is set up using the Backward Recursive method	This Document
TBD2	Inter-domain TE local path is set up using RSVP-TE signaling	This Document
TBD3	Inter-domain TE local path is set up using Segment Routing	This Document

7.2. Association Type Value

PCE Association Group [RFC8697] defines the ASSOCIATION Object and requests that IANA creates a registry to manage the value of the Association Type value. IANA is requested to allocate a new code point in the PCEP ASSOCIATION GROUP TLV Association Type field registry, as follows:

Association Type	Description
TBD4	Inter-domain Association Group

7.3. PCEP Error Values

IANA is requested to allocate code-points in the PCEP-ERROR Object Error Values registry for a new error-value of Error-Type 21 Invalid TE path setup and new error-value of Error-Type 26 Association Error:

Error-Type	Error-Value	Description
21	TBD5	Mandatory Stitching Label missing in the RRO
26	TBD6	Error in association of Inter-domain LSPs

7.4. PCEP TLV Type Indicators

IANA is requested to allocate a new TLV Type Indicator for the "Stitching Label PCE Capability" within the "PCEP TLV Type Indicators" subregistry of the "Path Computation Element Protocol (PCEP) Numbers" registry:

Value	Description	Reference
TBD7	STITCHING-LABEL-PCE-CAPABILITY	This Document

7.5. Stitching Label PCE Capability

IANA is requested to allocate a new subregistry, named "STITCHING-LABEL-PCE-CAPABILITY TLV Flag Field", within the "Path Computation Element Protocol (PCEP) Numbers" registry, to manage the Flag field in the STITCHING-LABEL-PCE-CAPABILITY TLV of the PCEP OPEN object (class = 1). New values are 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

Value	Description	Reference
31	RSVP-TE-STITCHING-CAPABILITY	This Document
30	SEGMENT-ROUTING-STITCHING-CAPABILITY	This Document
29	INTER-DOMAIN-STITCHING-CAPABILITY	This Document

8. Security Considerations

No modification of PCE protocol (PCEP) has been requested by this draft which does not introduce any issue regarding security. Concerning the PCEP session between PCEs, authors recommend to use the secured version of PCEP as defined in PCEPS [RFC8253] or use any other secured tunnel mechanism, e.g. IPsec tunnel to transport PCEP session between PCEs.

9. Acknowledgements

The authors want to thanks PCE's WG members, and in particular Dhruv Dhody who greatly contributed to the Hierarchical section of this document and Quan Xiong for his advice.

10. Disclaimer

This work has been performed in the framework of the H2020-ICT-2014 project 5GEx (Grant Agreement no. 671636), which is partially funded by the European Commission. This information reflects the consortium's view, but neither the consortium nor the European Commission are liable for any use that may be done of the information contained therein.

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>>.

- [RFC5441] Vasseur, JP., Ed., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, DOI 10.17487/RFC5441, April 2009, <<https://www.rfc-editor.org/info/rfc5441>>.
- [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>>.
- [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>>.
- [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>>.

11.2. Informative References

- [I-D.dong-pce-discovery-proto-bgp]
Dong, J., Chen, M., Dhody, D., Tantsura, J., Kumaki, K., and T. Murai, "BGP Extensions for Path Computation Element (PCE) Discovery", draft-dong-pce-discovery-proto-bgp-07 (work in progress), July 2017.
- [I-D.ietf-idr-bgppls-segment-routing-epe]
Previdi, S., Talaulikar, K., Filsfils, C., Patel, K., Ray, S., and J. Dong, "BGP-LS extensions for Segment Routing BGP Egress Peer Engineering", draft-ietf-idr-bgppls-segment-routing-epe-19 (work in progress), May 2019.

- [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>>.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <<https://www.rfc-editor.org/info/rfc3473>>.
- [RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", RFC 4003, DOI 10.17487/RFC4003, February 2005, <<https://www.rfc-editor.org/info/rfc4003>>.
- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", RFC 4206, DOI 10.17487/RFC4206, October 2005, <<https://www.rfc-editor.org/info/rfc4206>>.
- [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>>.
- [RFC5150] Ayyangar, A., Kompella, K., Vasseur, JP., and A. Farrel, "Label Switched Path Stitching with Generalized Multiprotocol Label Switching Traffic Engineering (GMPLS TE)", RFC 5150, DOI 10.17487/RFC5150, February 2008, <<https://www.rfc-editor.org/info/rfc5150>>.
- [RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, DOI 10.17487/RFC5520, April 2009, <<https://www.rfc-editor.org/info/rfc5520>>.
- [RFC6805] King, D., Ed. and A. Farrel, Ed., "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, DOI 10.17487/RFC6805, November 2012, <<https://www.rfc-editor.org/info/rfc6805>>.

- [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>>.
- [RFC8453] Ceccarelli, D., Ed. and Y. Lee, Ed., "Framework for Abstraction and Control of TE Networks (ACTN)", RFC 8453, DOI 10.17487/RFC8453, August 2018, <<https://www.rfc-editor.org/info/rfc8453>>.
- [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>>.
- [RFC8751] Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., and D. King, "Hierarchical Stateful Path Computation Element (PCE)", RFC 8751, DOI 10.17487/RFC8751, March 2020, <<https://www.rfc-editor.org/info/rfc8751>>.

Authors' Addresses

Olivier Dugeon
Orange Labs
2, Avenue Pierre Marzin
Lannion 22307
France

Email: olivier.dugeon@orange.com

Julien Meuric
Orange Labs
2, Avenue Pierre Marzin
Lannion 22307
France

Email: julien.meuric@orange.com

Young Lee
Huawei Technologies
5340 Legacy Drive, Building 3
Plano TX 75023
USA

Email: leeyoung@huawei.com

Daniele Ceccarelli
Ericsson
Torshamnsgatan, 48
Stockholm
Sweden

Email: daniele.ceccarelli@ericsson.com

PCE Working Group
Internet-Draft
Updates: 5440 (if approved)
Intended status: Standards Track
Expires: August 19, 2019

H. Pouyllau
Alcatel-Lucent
R. Theillaud
Marben Products
J. Meuric
France Telecom Orange
H. Zheng (Editor)
X. Zhang
Huawei Technologies
February 15, 2019

Extensions to the Path Computation Element Communication Protocol for
Enhanced Errors and Notifications
draft-ietf-pce-enhanced-errors-05.txt

Abstract

This document defines new error and notification TLVs for the PCE Communication Protocol (PCEP) [RFC5440]. It identifies the possible PCEP behaviors in case of error or notification. Thus, this draft defines types of errors and how they are disclosed to other PCEs in order to support predefined PCEP behaviors.

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 19, 2019.

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. Terminology	3
2. Conventions used in this document	3
3. Introduction	3
3.1. Examples	4
3.1.1. Error use-case	4
3.1.2. Notification use-case	4
4. PCEP Behaviors	4
4.1. PCEP Behaviors in Case of Error	5
4.2. PCEP Behaviors in Case of Notification	6
4.3. PCE Peer Identification	6
5. PCEP Extensions for Error and Notification Handling	7
5.1. Propagation TLV	7
5.2. Error-criticality TLV	7
5.3. Behaviors and TLV combinations	8
5.4. Propagation Restrictions TLVs	9
5.4.1. Time-To-Live (TTL) TLV	9
5.4.2. DIFFUSION-LIST TLV	9
5.4.3. Rules Applied to Existing Errors and Notifications	11
6. Future Extension Consideration	15
7. Backward Compatibility Consideration	15
8. Error and Notification Scenarios	15
8.1. Error Behavior Type 1	15
8.2. Error Behavior Type 2	16
8.3. Error Behavior Type 4	17
8.4. Error Behavior Type 5	17
9. Security Considerations	18
10. IANA Considerations	18
10.1. PCEP TLV Type Indicators	18
10.2. New DIFFUSION-LIST TLV	19
11. References	19
11.1. Normative References	19
11.2. Informational References	21
Authors' Addresses	22

1. Terminology

PCE terminology is defined in [RFC4655].

PCEP Peer: An element involved in a PCEP session (i.e. a PCC or a PCE).

Source PCC: the PCC, for a given path computation query, initiating the first PCEP request, which may then trigger a chain of successive requests.

Target PCE: the PCE that can compute a path to the destination without having to query any other PCE.

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. Introduction

The PCE Communication Protocol [RFC5440] is designed to be flexible and extensible in order to allow future evolutions or specific constraint support such as proposed in [RFC7470]. Crossing different PCE implementations (e.g. from different providers or due to different releases), a PCEP request may encounter unknown errors or notification messages. In such a case, the PCEP RFC [RFC5440] specifies to send a specific error code to the PCEP peer. This document updates [RFC5440] by introducing mechanism to propagate the error message, with specifying error and notification TLVs.

In the context of path computation crossing different routing domains or autonomous systems, the number of different PCE system specificities is potentially high, thus possibly leading to divergent and unstable situations. Such phenomenon can also occur in homogeneous cases since PCE systems have their own policies that can introduce differences in requests treatment even for requests having the same destination. In order to generalize PCEP behaviors in the case of heterogeneous PCE systems, new objects have to be defined. Dealing with heterogeneity is a major challenge considering PCE applicability, particularly in multi-layer, multi-domain and H-PCE contexts [I-D.ietf-pce-stateful-hpce]. Thus, extending such error codes and PCEP behaviors accordingly would improve interoperability among different PCEP implementations and would solve some of these issues. However, some of them would still remain (e.g. the divergences in request treatment introduced by different policies).

The purpose of this draft is to identify and specify new optional TLVs and objects in order to generalize PCEP behaviors.

3.1. Examples

The two following scenarios underline the need for a normalization of the PCEP behaviors according to existing error or notification types.

3.1.1. Error use-case

PCE(i-1) has sent a request to PCE(i) which has also sent a request to PCE(i+1). PCE(i-1) and PCE(i+1) have the same error semantic but not PCE(i). If PCE(i+1) throws an error type and value unknown by PCE(i). PCE(i) could then adopt any other behaviors and sends back to PCE(i-1) an error of type 2 (Capability not supported), 3 (Unknown Object) or 4 (Not supported Object) for instance. As a consequence, the path request would be cancelled but the error has no meaning for PCE(i-1) whereas if PCE(i) had simply forwarded the error sent by PCE(i+1), it would have been understood by PCE(i-1).

3.1.2. Notification use-case

PCE(i-1) has sent a request to PCE(i) which has also sent a request to PCE(i+1) but PCE(i+1) is overloaded. Without extensions, PCE(i+1) should send a notification of type 2 and a value flag giving its estimated congestion duration. PCE(i) can choose to stop the path computation and send a NO_PATH reply to PCE(i-1). Hence, PCE(i-1) ignores the congestion duration on PCE(i+1) and could seek it for further requests.

4. PCEP Behaviors

One of the purposes of the PCE architecture is to compute paths across networks, but an added value is to compute such paths in inter-area/layer/domain environments. The PCE Communication Protocol [RFC5440] is based on the Transport Communication Protocol (TCP). Thus, to compute a path within the PCE architecture, several TCP/PCEP sessions have to be set up, in a peer-to-peer manner, along a set of identified PCEs.

When the PCEP session is up for two PCEP peers, the PCC of the first PCE System (the source PCC) sends a PCReq message. If the PCC does not receive any reply before the dead timer is out, then it goes back to the idle state. A PCC can expect two kinds of replies: a PCRep message containing one or more valid paths (EROs) or a negative PCRep message containing a NO-PATH object.

Beside PCReq and PCRep messages, notification and error messages, named respectively PCNtf and PCErr, can be sent. There are two types of notification messages: type 1 is for cancelling pending requests and type 2 for signaling a congestion of the PCE. Several error values are described in [RFC5440]. The error types concerning the session phase begin at 2, error type 1 values are dedicated to the initialization phase.

As the PCE Communication Protocol is built to work in a peer-to-peer manner (i.e. supported by a TCP Connection), it supposes that the "deadtimer" of the source PCC is long enough to support the end-to-end distributed path computation process.

The exchange of messages in the PCE Communication Protocol is described in details when PCEP is in states OpenWait and KeepWait in [RFC5440]. When the session is up, message exchange is defined in [RFC5440]. [RFC5441] describes the Backward Recursive Path Computation (BRPC) procedure, and, because it considers an inter-domain path computation, gives a bigger picture of the possible behaviors when the session is up. Detailed behavior is mostly left free to any specific implementation. The following sections identify the PCEP behaviors in case of error or notification and also introduce the requirement of PCEP peer identification in both cases.

4.1. PCEP Behaviors in Case of Error

[RFC5440] specifies that "a PCEP Error message is sent in several situations: when a protocol error condition is met or the request is not compliant with the PCEP specification". On this basis, and according to the other RFCs, the identified PCEP behaviors are the followings:

- o "Propagation": the received message requires to be propagated forwardly or backwardly (depending on which PCEP peer has sent the message) to a set of PCEP peers;
- o "Criticality level": in different RFCs, error-types affects the state of the PCEP request or session in different manners; hence, different level of criticality can be observed:
- o
 - * Low-level of criticality: the received message does not affect the PCEP connection and further answer can still be expected;

- * Medium-level of criticality: the received message does not affect the PCEP connection but the request(s) is(are) cancelled;
- * High-level of criticality: the received message indicates that the PCEP peer will close the session with its peer (and so pending requests associated by the error, if any, are cancelled.)

The high-level of criticality has been extracted from [RFC5440] which associates such a behavior to error-type of 1 (errors raised during the PCEP session establishment). Hence, such errors are quite specific. For the sake of completeness, they have been included in this document.

4.2. PCEP Behaviors in Case of Notification

Notification messages can be employed in two different manners: during the treatment of a PCEP request, or independently from it to advertise information (in [RFC5440], the request ID list within a PCNtf message is optional). Hence, three different types of behaviors can be identified:

- o "Local": the notification does not imply any forward or backward propagation of the message;
- o "Request-specific propagation": the received message requires to be propagated forwardly or backwardly (depending on which peer has sent the message) to the PCEP peers;
- o "Non request-specific propagation": the received message must be propagated to any known peers (e.g. if PCE discovery is activated) or to a list of identified peers.

4.3. PCE Peer Identification

The propagation of errors and notifications affects the state of the PCEP peers along the chain. In some cases, for instance a notification that a PCE is overloaded, the identification of the PCEP peer - or that the sender PCE is not the direct neighbor - might be an important information for the PCEP peers receiving the message. The ID of sender PCE is not carried in the error TLVs, but can be achieved via the speaker entity ID TLV during state synchronization. An example can be found in [RFC8232].

5. PCEP Extensions for Error and Notification Handling

This section describes extensions to support error and notification with respect to the PCEP behavior description defined in Section 4. This document does not intend to modify errors and notification types previously defined in existing documents (e.g. [RFC5440], [RFC5441], etc.). Error related TLVs have been specified in this section, while the notification functionality can be achieved via using PCNtf message with RP object with no need to extend further notification type.

5.1. Propagation TLV

To support the propagation behavior mentioned in Section 4.1 and Section 4.2, a new optional TLV is defined, which can be carried in PCEP-ERROR and NOTIFICATION objects, to indicate whether a message has to be propagated or not. The allocation from the "PCEP TLV Type Indicators" sub-registry will be assigned by IANA and the request is documented in Section 10.

The description is "Propagation", the length value is 2 bytes and the value field is 1 byte. The value field is set to 0 meaning that the message MUST NOT be propagated. If the value field is set to 1, the message MUST be propagated. Section 5.4 specifies the destination and to limit the number of messages.

5.2. Error-criticality TLV

To support the shutdown behavior mentioned in Section 4.1, we extend the PCEP-ERROR object by creating a new optional TLV to indicate whether an error is recoverable or not. The allocation from the "PCEP TLV Type Indicators" sub-registry will be assigned by IANA and the request is documented in Section 10.

The description is "Error-criticality", the length value is 2 bytes and the value field is 1 byte. The value field is set to 0 meaning that the error has a low-level of criticality (so further messages can be expected for this request). If the value field is set to 1, the error has a medium-level of criticality and requests whose identifiers appear in the same message MUST be cancelled (so no further messages can be expected for these requests). If the value field is set to 2, the error has a high-level of criticality, the connection for this PCEP session is closed by the sender PCE peer.

5.3. Behaviors and TLV combinations

The propagation behavior MAY be combined with all criticality levels, thus leading to 6 different behaviors. In the case of a criticality level of 2, the session is closed by the PCE peer which sends the message. Hence, the criticality level is purely informative for the PCE peer which receives the message. If it is combined with a propagation behavior, then the PCE propagating the message MUST indicate the same level of criticality if it closes the session. Otherwise, it MUST use a criticality level of 1 if it does not close the session.

For a PCErr message, all the possible behaviors described in Section 4.1 can be covered with TLVs included in a PCEP-ERROR object. The following table captures all combinations of error behaviors:

Error criticality\ Value	0 (No Propagation)	1 (Propagation Required)
0 (low)	Type 1	Type 4
1 (medium)	Type 2	Type 5
2 (high)	Type 3	Type 6

- o "Error Behavior Type 1" : Local Error with a low level of criticality;
- o "Error Behavior Type 2": Local Error with a medium level of criticality;
- o "Error Behavior Type 3": Local Error with a high level of criticality;
- o "Error Behavior Type 4": Propagated Error with a low level of criticality;
- o "Error Behavior Type 5": Propagated Error with a medium level of criticality;
- o "Error Behavior Type 6": Propagated Error with a high level of criticality;

5.4. Propagation Restrictions TLVs

In order to limit the propagation of errors and notifications, the following mechanisms SHOULD be used:

A Time-To-Live(TTL) TLV: to limit the number of PCEP peers that will recursively receive the message;

A DIFFUSION-LIST TLV: to specify the PCEP peer addresses or domains of PCEP peers the message must be propagate to;

History mechanism: if a PCEP peer keeps track of the messages it has relayed, it could avoid propagating an error or notification it has already received.

Such mechanisms SHOULD be used jointly or independently depending the error or notification behaviors they are associated to. The conditions of use for the TTL and DIFFUSION-LIST TLVs are described in sections below.

5.4.1. Time-To-Live (TTL) TLV

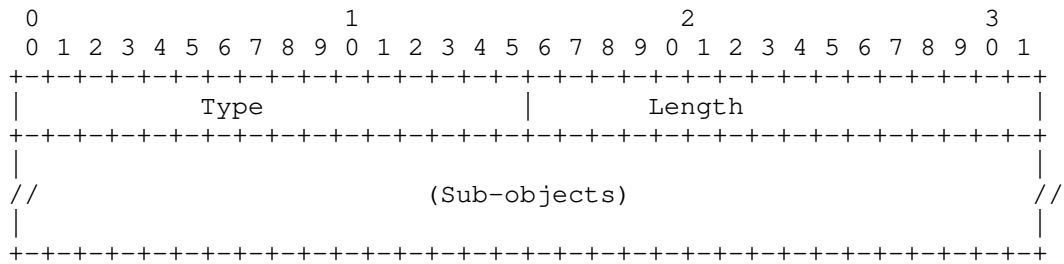
The TTL value is set to any integer value to indicate the number of PCEP peers that will recursively receive the message. The TTL TLV SHOULD be used with propagated errors or notifications ("Propagation" TLV with value 1 in PCEP-ERROR or NOTIFICATION objects). Each PCEP peer MUST decrement the TTL value before propagating the message. When the TTL value becomes 0, the message is no more propagated.

If the message to be propagated is request-specific and there is no TTL or DIFFUSION-LIST TLVs included, the message MUST reach the source PCC (or alternatively the target PCE).

5.4.2. DIFFUSION-LIST TLV

The DIFFUSION-LIST TLV can be carried within either the error object of a PCErr message, or the notification object of a PCNtf message. It can either be used in a message sent by a PCC to a PCE or vice versa. The DIFFUSION-LIST MAY be used with propagated errors (TLV "Propagation" at value 1 in PCEP-ERROR object).

The format of the DIFFUSION-LIST object body is as follows:



Type (16 bits): restricts the diffusion to certain peers. The following values are currently defined:

- 0: Any PCEP peer indicated in the list must be reached.
- 1: Only PCEs must be reached (and not PCC).
- 2: All PCEP peers with which a session is still opened must be reached.

The value of DIFFUSION-LIST is made of sub-objects similar to the IRO defined in [RFC5440]. The following sub-object types are supported.

Type Sub-object

- 1 IPv4 address
- 2 IPv6 address
- 4 Unnumbered Interface ID
- 5 4-byte AS number
- 6 OSPF area ID
- 7 IS-IS Area ID
- 32 Autonomous System number
- 33 Explicit eXclusion Route Sub-object (EXRS)

If the error or notification codes target specific PCEP peers, a DIFFUSION-LIST TLV avoids partially flooding all PCEP peers. Any PCEP peer receiving a PCErr or PCNTf message containing a PCEP-ERROR or a NOTIFICATION object with a TLV "Propagation" at value 1 and where a DIFFUSION-LIST appears, MUST remove the addresses of the PCEP peers from the DIFFUSION-LIST, before sending the message to any other PCEP peers. This is performed by adding the PCEP peer addresses to the Explicit eXclusion Route Sub-object of the DIFFUSION-LIST. If a DIFFUSION-LIST value is empty, the PCEP peer MUST NOT propagate the message to any peer.

Note that, a Diffusion-List could contain strict or loose addresses to refer to a network domain (e.g. an Autonomous System number, an OSPF area, an IP address). Hence, the PCEP peers targeted by the message would be the PCEP peers covering the corresponding domain. If an address is loose, each time a PCEP peer forwards a message to another PCEP peer of this address, it MUST add its own address to the Explicit eXclusion Route Sub-object (EXRS) of the Diffusion-List for any forwarded messages. Hence, a PCE SHOULD avoid forwarding the same message repeated to the same set of peers. Finally, when an address is loose, the forwarding SHOULD be restrained indicating what type of PCEP peers are targeted (i.e. PCE and/or PCC).

5.4.3. Rules Applied to Existing Errors and Notifications

Many existing normative references states on error definitions (see for instance [RFC5440], [RFC5441], [RFC5455], [RFC5521], [RFC5557], [RFC5886], [RFC6006], [RFC8231], [RFC8232], [RFC8253], [RFC8281], [RFC8306], [RFC8408], [I-D.ietf-pce-association-group]). This section provides processing rules for existing error types handling, as a recommendation. According to the definitions provided in this document, the following rules are applicable:

Error-type 1, described in [RFC5440], relates to PCEP session establishment failures. All errors of this type are local and not propagated. Hence, if a "Propagation" TLV is added to the error message it is recommended to be set to value 0. Error-values 1,2,6,7 have a high level of criticality. Hence, if the "Error-criticality" TLV is included within a PCERR message of type 1 and value 1,2,6 or 7, it is recommended to have a value of 2.

Error-type 2,3,4, "Capability not supported", "Unknown object" and "Not supported object" respectively, described in [RFC5440]: errors of this type MAY be propagated using the TLV "Propagation". Their level of criticality is defined as leading to cancel the path computation request [RFC5440]. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The error-value 4 of error-type 4 ("Unsupported parameter") associated to the BRPC procedure [RFC5441] is suggested to contain the "Propagation" TLV with a DIFFUSION-LIST requesting a propagation to the PCC at the origin of the request.

Error-type 5 refers to "Policy violation", error values for this type have been defined in [RFC5440], [RFC5541], [RFC5557], [RFC5886] and [RFC6006]. In [RFC5440], it is specified that the path computation request MUST be cancelled when an error of type 5 occurs. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. As such errors might be conveyed to several PCEs, the "Propagation" TLV MAY be used.

Error-type 6 described as "Mandatory object missing" in [RFC5440], leads to the cancellation of the path computation request. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors. The error-value of 4 for Monitoring object missing defined in [RFC5886] is no exception to the rule.

Error-type 7 is described as "synchronized path computation request missing". In [RFC5440], it is specified that the reffered synchronized path computation request MUST be cancelled when an error of type 5 occurs. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors.

Error-type 8 is raised when a PCE receives a PCRep with an unknown request reference. If the "Propagation" TLV is used with error-type 8, it is recommended to be set at a value of 0. The "Error-criticality" TLV is not particularly relevant for error-type 8. Hence, it usually have the value of 0 if used.

Error-type 9 is raised when a PCE attempts to establish a second PCEP session. The existing session must be preserved. Hence, if the "Error-criticality" TLV is included, it usually have a value of 0. By definition, such an error message SHOULD NOT be propagated. Thus, if the "Propagation" TLV is used with error-type 9, it is usually set to a value of 0.

Error-type 10 which refers to the reception of an invalid object as described in [RFC5440] no indication is provided on the cancellation of the path computation request. Hence, if the "Error-criticality" TLV is included, it usually have a value of 0. The "Propagation" TLV MAY be used with such errors with any value depending on the expected behavior.

Error-type 11 relates to "Unrecognized EXRS subobject" and is described in [RFC5521]. No path computation request cancellation is required by [RFC5521]. Hence, if the "Error-criticality" TLV is included, it usually have a value of 0. The "Propagation" TLV MAY be used with such errors with any value depending on the expected behavior.

Error-type 12 refers to "Diffserv-aware TE error" and is described in [RFC5455]. Such errors are raised when the CLASSTYPE object of a PCReq is recognized but not supported by a PCE. [RFC5455] does not state about the path computation request when such errors are met. Hence, both "Propagation" and "Error-criticality" TLVs COULD be used within such error-types' messages and set to any specified values.

Error-type 13 on "BRPC procedure completion failure" is described in [RFC5441]. [RFC5441] states that in such cases, the PCErr message MUST be relayed to the PCC. Hence, such messages SHOULD contain a "Propagation" TLV and a DIFFUSION-LIST with a Target-Type of 0 and corresponding addresses or with a Target-Type of 2. It is not specified in [RFC5441] whether the path computation request should be canceled or not. If the procedure is not supported, it does not necessarily imply to cancel the path computation request if another procedure is able to read and write VSPT objects. Thus, the "Error-criticality" TLV MAY be used with any value depending on the expected behavior.

Error-type 15 refers to "Global Concurrent Optimization Error" defined in [RFC5557]. [RFC5557] states that the corresponding global concurrent path optimization MUST be cancelled at the PCC. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors.

Error-type 16 relates to "P2MP Capability Error" defined in [RFC6006]. Such errors lead to the cancellation of the path computation request. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors.

Error-type 17, titled "P2MP END-POINTS Error" is defined [RFC6006]. Such errors are thrown when a PCE tries to add or prune nodes to or from a P2MP Tree. [RFC6006] does not specify if such errors lead to cancel the path computation request. Hence, the "Error-criticality" and "Propagation" TLVs MAY be used with this type of error with any value depending on the expected behavior.

Error-type 18 of "P2MP Fragmentation Error" is described [RFC6006] which does not specify whether the path computation request should be cancelled. But, as messages are fragmented, it is natural to think that the PCE should wait at least a bit for further messages. The "Error-criticality" TLV MAY be included in such error messages and is particularly adapted to differ the semantic of the same error-type message: if it is included with a value of 0 then the PCE will still wait for further fragmented messages, when this waiting time ends, the TLV can be included with a value of 1 in order to finally cancel the request. The "Propagation" TLV MAY also be used with such errors.

Error-type 19 of "Invalid Operation" is described in [RFC8231] and [RFC8281], which implies a wrong capability description for PCEP session. In this case, the PCErr message MUST be returned to PCC, and this message usually contain a "Propagation" TLV and a

DIFFUSION-LIST with a Target-Type of 0 or 2. The "Error-criticality" TLV is recommended be set to 2 in order to guarantee the termination of PCEP session.

Error-type 20 of "LSP State Synchronization Error" is described in [RFC8231] and [RFC8232], which cannot successfully sync up the LSP states. In this case, the "Error-criticality" TLV should be set to 2 in order to guarantee the termination of PCEP session. The "Propagation" TLV MAY also be used with such errors.

Error-type 21 of "Invalid traffic engineering path setup type" is described in [RFC8408] . Such errors failed to find a matched path setup type and the PCEP sessions MUST be closed. In this case, the "Error-criticality" TLV is usually set to 2 in order to guarantee the termination of PCEP session. The "Propagation" TLV MAY also be used with such errors.

Error-type 23 of "Bad parameter value" is described in [RFC8281] . Such errors occur when there is a conflict in path name of C flag not set for PCE initiation. In this case, the "Error-criticality" TLV may be set to either 0 or 1 to indicate whether the request is still valid, with the PCEP session open. The "Propagation" TLV MAY also be used with such errors.

Error-type 24 of "LSP instantiation error" is described in [RFC8281] . Such errors occur when PCC detects problems when establishing the path, the message MUST relay to the PCE, therefore the "Propagation" TLV is usually contained. The "Error-criticality" TLV may be set to either 0 or 1 to indicate whether the request is still valid, with the PCEP session open.

Error-type 25 of "PCEP StartTLS failure" is described in [RFC8253]. Such errors indicate the security issue in transport layer. In this case, the "Error-criticality" TLV is usually set to 2 in order to close the PCEP session. The "Propagation" TLV MAY also be used with such errors, depending on the detailed security conditions.

Error-type 26 of "Association Error " is described in [I-D.ietf-pce-association-group] . Such errors occur when there is problem for LSP association. In this case, the "Error-criticality" TLV should be set to either 0 or 1 to indicate whether the request is still valid, with the PCEP session open. The "Propagation" TLV MAY also be used with such errors.

6. Future Extension Consideration

The procedures specified in this work applies to all the existing error types. For future PCE protocol extension who gives new error types, it is requested to provide description on the applicability of "Propagation" TLV and "Error-criticality" TLV.

7. Backward Compatibility Consideration

There would be backward compatibility issue if there are multiple PCEs with different level understanding of error message. In a scenario that PCE(i) propagate the error message to PCE (i+1), it is possible that PCE (i+1) is not capable to extract the message correctly, then such error message would be ignored and not be further propagated.

There can be potential approach to avoid these problem, such as recognizing the incapable PCE and avoiding propagation. However, these approach is not in the scope of this document.

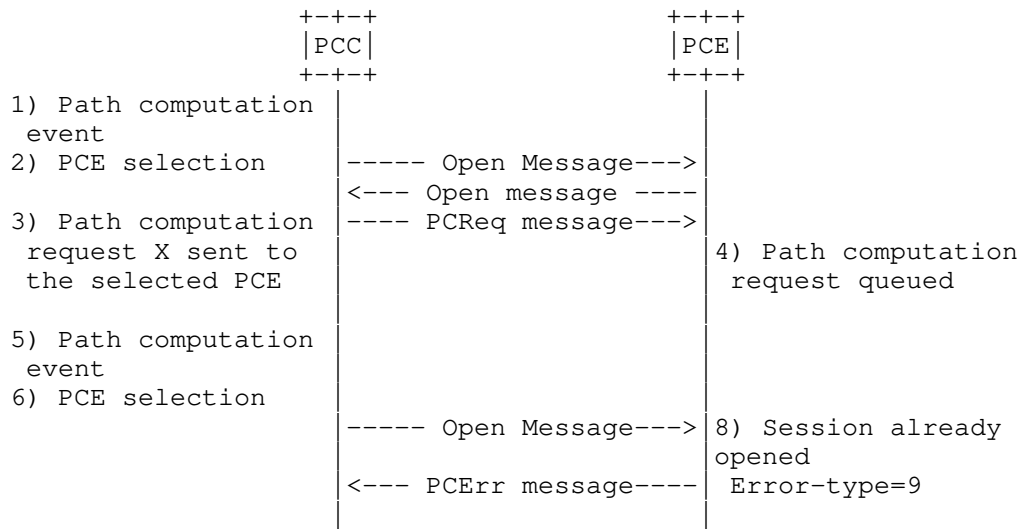
8. Error and Notification Scenarios

[Editor Note] This section will be moved to appendix for publication.

This section provides some examples depicting how the error described above can be used in a PCEP session. The origin of the errors or notifications is only illustrative and has no normative purpose. Sometimes the PCE features behind may be implementation-specific (e.g. detection of flooding). This section does not provide scenarios for errors with a high-level of critcity (i.e., Error behaviors 3 and 6) since such errors are very specific and until now have been normalized only during the session establishment (error-type of 1).

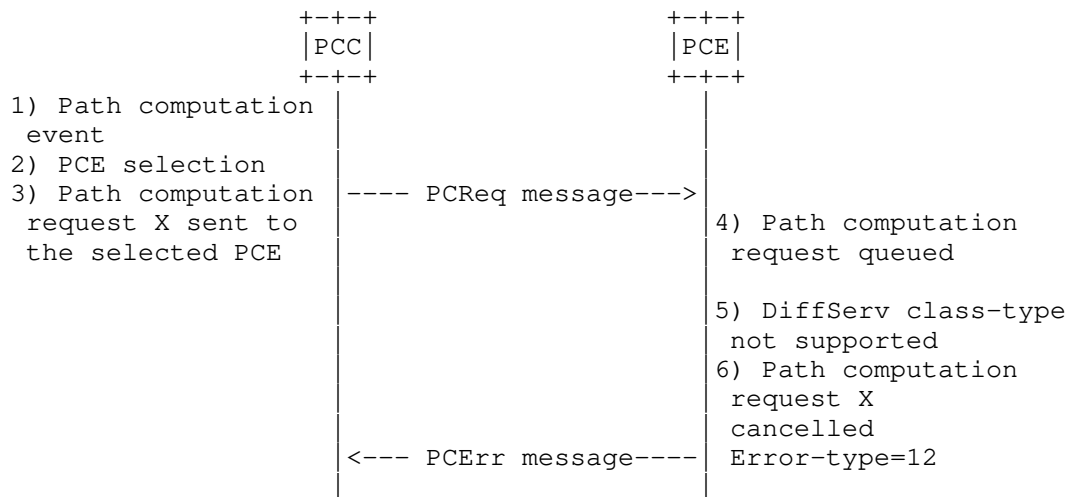
8.1. Error Behavior Type 1

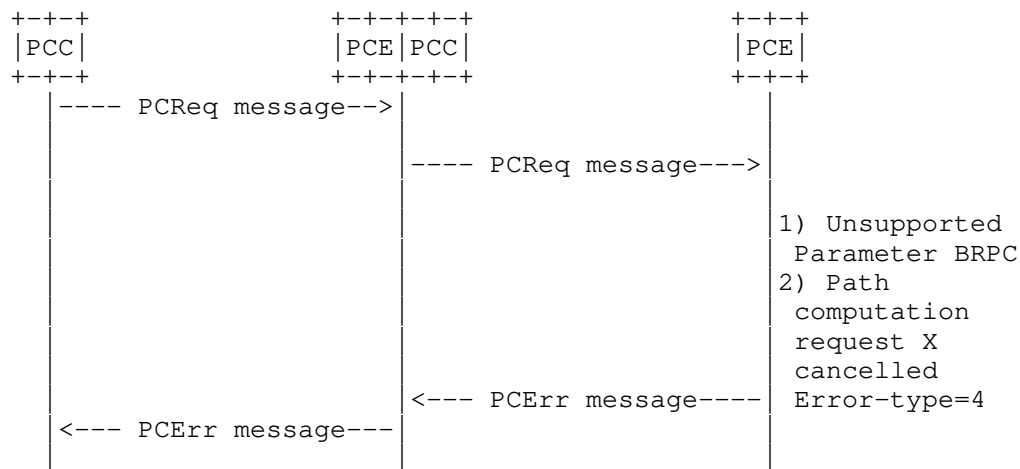
In this example, a PCC attempts to establish a second PCEP session with the same PCE for another request. Consequently the PCE sends back an error message with error-type 9. This error stays local and does not affect the former session. The second session is ignored. If the "Propagation" TLV and "Error-criticality" TLV are used, they should be both set to value 0.



8.2. Error Behavior Type 2

In this example, the PCC sends a DiffServ-aware path computation request. If the PCE receiving the request does not support the indicated class-type, it thus sends back a PCErr message with error-type=12 and error-value=1. If the "Propagation" TLV and "Error-criticality" TLV are present, they should carry value 0 and value 1 respectively. Consequently, the request is cancelled.





9. Security Considerations

Within the introduced set of TLVs, the "Propagation" TLV affects PCEP security considerations since it forces propagation behaviors. Thus, a PCEP implementation SHOULD activate stateful mechanism when receiving PCEP-ERROR or NOTIFICATION object including this TLV in order to avoid DoS attacks.

10. IANA Considerations

IANA maintains a registry of PCEP parameters. This includes a sub-registry for PCEP Objects.

IANA is requested to make an allocation from the sub-registry as follows. The values here are suggested for use by IANA.

10.1. PCEP TLV Type Indicators

As described in Section 5.4 the newly defined TLVs allows a PCE to enforce specific error and notification behaviors within PCEP-ERROR and NOTIFICATION objects. IANA is requested to make the following allocations from the "PCEP TLV Type Indicators" sub-registry.

Value	Description	Reference
TBD	Propagation	this document
TBD	Error-criticality	this document

10.2. New DIFFUSION-LIST TLV

Type	Value	Meaning	Reference
	0	Any PCEP peers	this document
	1	PCEs but excludes PCC-only peers	this document
	2	PCEs and PCCs with which a session is still opened	this document
Subobjects			Reference
	1: IPv4 prefix		this document
	2: IPv6 prefix		this document
	4: Unnumbered Interface ID		this document
	5: OSPF Area ID		this document
	6 OSPF area ID		this document
	7 IS-IS Area ID		this document
	32: Autonomous system number		this document
	33: Explicit Exclusion Route subobject (EXRS)		this document

11. References

11.1. Normative References

- [I-D.ietf-pce-association-group]
Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "PCEP Extensions for Establishing Relationships Between Sets of LSPs", draft-ietf-pce-association-group-07 (work in progress), December 2018.
- [I-D.ietf-pce-stateful-hpce]
Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., King, D., and O. Dios, "Hierarchical Stateful Path Computation Element (PCE).", draft-ietf-pce-stateful-hpce-06 (work in progress), October 2018.
- [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>>.
- [RFC5441] Vasseur, JP., Ed., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, DOI 10.17487/RFC5441, April 2009, <<https://www.rfc-editor.org/info/rfc5441>>.
- [RFC5455] Sivabalan, S., Ed., Parker, J., Boutros, S., and K. Kumaki, "Diffserv-Aware Class-Type Object for the Path Computation Element Communication Protocol", RFC 5455, DOI 10.17487/RFC5455, March 2009, <<https://www.rfc-editor.org/info/rfc5455>>.
- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, DOI 10.17487/RFC5521, April 2009, <<https://www.rfc-editor.org/info/rfc5521>>.
- [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>>.
- [RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, DOI 10.17487/RFC5557, July 2009, <<https://www.rfc-editor.org/info/rfc5557>>.
- [RFC5886] Vasseur, JP., Ed., Le Roux, JL., and Y. Ikejiri, "A Set of Monitoring Tools for Path Computation Element (PCE)-Based Architecture", RFC 5886, DOI 10.17487/RFC5886, June 2010, <<https://www.rfc-editor.org/info/rfc5886>>.
- [RFC6006] Zhao, Q., Ed., King, D., Ed., Verhaeghe, F., Takeda, T., Ali, Z., and J. Meuric, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 6006, DOI 10.17487/RFC6006, September 2010, <<https://www.rfc-editor.org/info/rfc6006>>.

- [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>>.
- [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>>.
- [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>>.
- [RFC8306] Zhao, Q., Dhody, D., Ed., Palletti, 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>>.

11.2. Informational 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>>.
- [RFC7470] Zhang, F. and A. Farrel, "Conveying Vendor-Specific Constraints in the Path Computation Element Communication Protocol", RFC 7470, DOI 10.17487/RFC7470, March 2015, <<https://www.rfc-editor.org/info/rfc7470>>.

Authors' Addresses

Helia Pouyllau
Alcatel-Lucent
Route de Villejust
NOZAY 91620
FRANCE

Phone: + 33 (0)1 30 77 63 11
Email: helia.pouyllau@alcatel-lucent.com

Remi Theillaud
Marben Products
176 rue Jean Jaures
Puteaux 92800
FRANCE

Phone: + 33 (0)1 79 62 10 22
Email: remi.theillaud@marben-products.com

Julien Meuric
France Telecom Orange
2, avenue Pierre Marzin
Lannion 22307
FRANCE

Email: julien.meuric@orange-ftgroup.com

Haomian Zheng (Editor)
Huawei Technologies
H1-1-A043S Huawei Industrial Base, Songshanhu
Dongguan, Guangdong 523808
P.R.China

Email: zhenghaomian@huawei.com

Xian Zhang
Huawei Technologies
G1-2, Huawei Industrial Base, Bantian, Longgang District
Shenzhen, Guangdong 518129
P.R.China

Email: zhang.xian@huawei.com

PCE Working Group
Internet-Draft
Updates: 5440 (if approved)
Intended status: Standards Track
Expires: 8 September 2022

H.P. Pouyllau
Alcatel-Lucent
R.T. Theillaud
Marben Products
J.M. Meuric
Orange
H. Zheng (Editor)
X. Zhang
Huawei Technologies
7 March 2022

Extensions to the Path Computation Element Communication Protocol for
Enhanced Errors and Notifications
draft-ietf-pce-enhanced-errors-11

Abstract

This document defines new error and notification TLVs for the PCE Communication Protocol (PCEP) specified in RFC5440, and will update it. It identifies the possible PCEP behaviors in case of error or notification. Thus, this draft defines types of errors and how they are disclosed to other PCEs in order to support predefined PCEP behaviors.

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 8 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. Terminology	3
2. Conventions used in this document	3
3. Introduction	3
3.1. Examples	4
3.1.1. Error use-case	4
3.1.2. Notification use-case	4
4. PCEP Behaviors	4
4.1. PCEP Behaviors in Case of Error	5
4.2. PCEP Behaviors in Case of Notification	6
4.3. PCE Peer Identification	6
5. PCEP Extensions for Error and Notification Handling	7
5.1. Propagation TLV	7
5.2. Error-criticality TLV	7
5.3. Behaviors and TLV combinations	8
5.4. Propagation Restrictions TLVs	9
5.4.1. Time-To-Live (TTL) TLV	9
5.4.2. DIFFUSION-LIST TLV	9
5.4.3. Rules Applied to Existing Errors and Notifications	11
6. Error Handling Guidelines for Future PCEP Extension	15
7. Backward Compatibility Consideration	15
8. Implementation Status	15
9. Security Considerations	16
10. IANA Considerations	16
10.1. PCEP TLV Type Indicators	16
10.2. New DIFFUSION-LIST TLV	17
11. References	17
11.1. Normative References	17
11.2. Informational References	19
Appendix A. Error and Notification Scenarios	20
A.1. Error Behavior Type 1	20
A.2. Error Behavior Type 2	21
A.3. Error Behavior Type 4	21
A.4. Error Behavior Type 5	22
Authors' Addresses	22

1. Terminology

PCE terminology is defined in [RFC4655].

PCEP Peer: An element involved in a PCEP session (i.e. a PCC or a PCE).

Source PCC: the PCC, for a given path computation query, initiating the first PCEP request, which may then trigger a chain of successive requests.

Target PCE: the PCE that can compute a path to the destination without having to query any other PCE.

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. Introduction

The PCE Communication Protocol [RFC5440] is designed to be flexible and extensible in order to allow future evolutions or specific constraint support such as proposed in [RFC7470]. Crossing different PCE implementations (e.g. from different providers or due to different releases), a PCEP request may encounter unknown errors or notification messages. In such a case, the PCEP RFC [RFC5440] specifies to send a specific error code to the PCEP peer. This document updates [RFC5440] by introducing mechanism to propagate the error message, with specifying error and notification TLVs.

In the context of path computation crossing different routing domains or autonomous systems, the number of different PCE system specificities is potentially high, thus possibly leading to divergent and unstable situations. Such phenomenon can also occur in homogeneous cases since PCE systems have their own policies that can introduce differences in requests treatment even for requests having the same destination. In order to generalize PCEP behaviors in the case of heterogeneous PCE systems, new objects have to be defined. Dealing with heterogeneity is a major challenge considering PCE applicability, particularly in multi-layer, multi-domain and H-PCE contexts [RFC8751]. Thus, extending such error codes and PCEP behaviors accordingly would improve interoperability among different PCEP implementations and would solve some of these issues. However, some of them would still remain (e.g. the divergences in request treatment introduced by different policies).

The purpose of this draft is to identify and specify new optional TLVs and objects in order to generalize PCEP behaviors.

3.1. Examples

The two following scenarios underline the need for a normalization of the PCEP behaviors according to existing error or notification types.

3.1.1. Error use-case

PCE(i-1) has sent a request to PCE(i) which has also sent a request to PCE(i+1). PCE(i-1) and PCE(i+1) have the same error semantic but not PCE(i). If PCE(i+1) throws an error type and value unknown by PCE(i). PCE(i) could then adopt any other behaviors and sends back to PCE(i-1) an error of type 2 (Capability not supported), 3 (Unknown Object) or 4 (Not supported Object) for instance. As a consequence, the path request would be cancelled but the error has no meaning for PCE(i-1) whereas if PCE(i) had simply forwarded the error sent by PCE(i+1), it would have been understood by PCE(i-1).

3.1.2. Notification use-case

PCE(i-1) has sent a request to PCE(i) which has also sent a request to PCE(i+1) but PCE(i+1) is overloaded. Without extensions, PCE(i+1) should send a notification of type 2 and a value flag giving its estimated congestion duration. PCE(i) can choose to stop the path computation and send a NO_PATH reply to PCE(i-1). Hence, PCE(i-1) ignores the congestion duration on PCE(i+1) and could seek it for further requests.

4. PCEP Behaviors

One of the purposes of the PCE architecture is to compute paths across networks, but an added value is to compute such paths in inter-area/layer/domain environments. The PCE Communication Protocol [RFC5440] is based on the Transport Communication Protocol (TCP). Thus, to compute a path within the PCE architecture, several TCP/PCEP sessions have to be set up, in a peer-to-peer manner, along a set of identified PCEs.

When the PCEP session is up for two PCEP peers, the PCC of the first PCE System (the source PCC) sends a PCReq message. If the PCC does not receive any reply before the dead timer is out, then it goes back to the idle state. A PCC can expect two kinds of replies: a PCRep message containing one or more valid paths (EROs) or a negative PCRep message containing a NO-PATH object.

Beside PCReq and PCRep messages, notification and error messages, named respectively PCNtf and PCErr, can be sent. There are two types of notification messages: type 1 is for cancelling pending requests and type 2 for signaling a congestion of the PCE. Several error values are described in [RFC5440]. The error types concerning the session phase begin at 2, error type 1 values are dedicated to the initialization phase.

As the PCE Communication Protocol is built to work in a peer-to-peer manner (i.e. supported by a TCP Connection), it supposes that the "deadtimer" of the source PCC is long enough to support the end-to-end distributed path computation process.

The exchange of messages in the PCE Communication Protocol is described in details when PCEP is in states OpenWait and KeepWait in [RFC5440]. When the session is up, message exchange is defined in [RFC5440]. [RFC5441] describes the Backward Recursive Path Computation (BRPC) procedure, and, because it considers an inter-domain path computation, gives a bigger picture of the possible behaviors when the session is up. Detailed behavior is mostly left free to any specific implementation. The following sections identify the PCEP behaviors in case of error or notification and also introduce the requirement of PCEP peer identification in both cases.

4.1. PCEP Behaviors in Case of Error

[RFC5440] specifies that "a PCEP Error message is sent in several situations: when a protocol error condition is met or the request is not compliant with the PCEP specification". On this basis, and according to the other RFCs, the identified PCEP behaviors are the followings:

- * "Propagation": the received message requires to be propagated forwardly or backwardly (depending on which PCEP peer has sent the message) to a set of PCEP peers;
- * "Criticality level": in different RFCs, error-types affects the state of the PCEP request or session in different manners; hence, different level of criticality can be observed:
 - Low-level of criticality: the received message does not affect the PCEP connection and further answer can still be expected;
 - Medium-level of criticality: the received message does not affect the PCEP connection but the request(s) is(are) cancelled;

- High-level of criticality: the received message indicates that the PCEP peer will close the session with its peer (and so pending requests associated by the error, if any, are cancelled.)

The high-level of criticality has been extracted from [RFC5440] which associates such a behavior to error-type of 1 (errors raised during the PCEP session establishment). Hence, such errors are quite specific. For the sake of completeness, they have been included in this document.

4.2. PCEP Behaviors in Case of Notification

Notification messages can be employed in two different manners: during the treatment of a PCEP request, or independently from it to advertise information (in [RFC5440], the request ID list within a PCNtf message is optional). Hence, three different types of behaviors can be identified:

- * "Local": the notification does not imply any forward or backward propagation of the message;
- * "Request-specific propagation": the received message requires to be propagated forwardly or backwardly (depending on which peer has sent the message) to the PCEP peers;
- * "Non request-specific propagation": the received message must be propagated to any known peers (e.g. if PCE discovery is activated) or to a list of identified peers.

4.3. PCE Peer Identification

The propagation of errors and notifications affects the state of the PCEP peers along the chain. In some cases, for instance a notification that a PCE is overloaded, the identification of the PCEP peer - or that the sender PCE is not the direct neighbor - might be an important information for the PCEP peers receiving the message. The ID of sender PCE is not carried in the error TLVs, but can be achieved via the speaker entity ID TLV during state synchronization. An example can be found in [RFC8232].

5. PCEP Extensions for Error and Notification Handling

This section describes extensions to support error and notification with respect to the PCEP behavior description defined in Section 4. This document does not intend to modify errors and notification types previously defined in existing documents (e.g. [RFC5440], [RFC5441], etc.). Error related TLVs have been specified in this section, while the notification functionality can be achieved via using PCNtf message with RP object with no need to extend further notification type.

5.1. Propagation TLV

To support the propagation behavior mentioned in Section 4.1 and Section 4.2, a new optional TLV is defined, which can be carried in PCEP-ERROR and NOTIFICATION objects, to indicate whether a message has to be propagated or not. The allocation from the "PCEP TLV Type Indicators" sub-registry will be assigned by IANA and the request is documented in Section 10.

The description is "Propagation", the length value is 2 bytes and the value field is 1 byte. The value field is set to 0 meaning that the message MUST NOT be propagated. If the value field is set to 1, the message MUST be propagated. Section 5.4 specifies the destination and to limit the number of messages.

5.2. Error-criticality TLV

To support the shutdown behavior mentioned in Section 4.1, we extend the PCEP-ERROR object by creating a new optional TLV to indicate whether an error is recoverable or not. The allocation from the "PCEP TLV Type Indicators" sub-registry will be assigned by IANA and the request is documented in Section 10.

The description is "Error-criticality", the length value is 2 bytes and the value field is 1 byte. The value field is set to 0 meaning that the error has a low-level of criticality (so further messages can be expected for this request). If the value field is set to 1, the error has a medium-level of criticality and requests whose identifiers appear in the same message MUST be cancelled (so no further messages can be expected for these requests). If the value field is set to 2, the error has a high-level of criticality, the connection for this PCEP session is closed by the sender PCE peer.

5.3. Behaviors and TLV combinations

The propagation behavior MAY be combined with all criticality levels, thus leading to 6 different behaviors. In the case of a criticality level of 2, the session is closed by the PCE peer which sends the message. Hence, the criticality level is purely informative for the PCE peer which receives the message. If it is combined with a propagation behavior, then the PCE propagating the message MUST indicate the same level of criticality if it closes the session. Otherwise, it MUST use a criticality level of 1 if it does not close the session.

For a PCErr message, all the possible behaviors described in Section 4.1 can be covered with TLVs included in a PCEP-ERROR object. The following table captures all combinations of error behaviors:

Error criticality\ Value value \	0 (No Propagation)	1 (Propagation Required)
0 (low)	Type 1	Type 4
1 (medium)	Type 2	Type 5
2 (high)	Type 3	Type 6

- * "Error Behavior Type 1" : Local Error with a low level of criticality;
- * "Error Behavior Type 2": Local Error with a medium level of criticality;
- * "Error Behavior Type 3": Local Error with a high level of criticality;
- * "Error Behavior Type 4": Propagated Error with a low level of criticality;
- * "Error Behavior Type 5": Propagated Error with a medium level of criticality;
- * "Error Behavior Type 6": Propagated Error with a high level of criticality;

5.4. Propagation Restrictions TLVs

In order to limit the propagation of errors and notifications, the following mechanisms SHOULD be used:

A Time-To-Live(TTL) RLV: to limit the number of PCEP peers that will recursively receive the message;

A DIFFUSION-LIST TLV: to specify the PCEP peer addresses or domains of PCEP peers the message must be propagate to;

History mechanism: if a PCEP peer keeps track of the messages it has relayed, it could avoid propagating an error or notification it has already received.

Such mechanisms SHOULD be used jointly or independently depending the error or notification behaviors they are associated to. The conditions of use for the TTL and DIFFUSION-LIST TLVs are described in sections below.

5.4.1. Time-To-Live (TTL) TLV

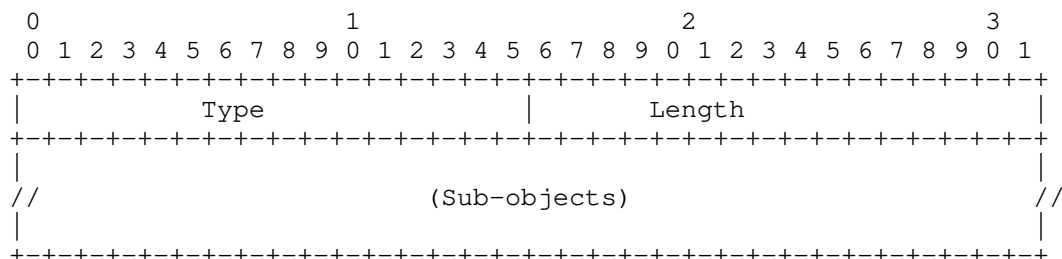
The TTL value is set to any integer value to indicate the number of PCEP peers that will recursively receive the message. The TTL TLV SHOULD be used with propagated errors or notifications ("Propagation" TLV with value 1 in PCEP-ERROR or NOTIFICATION objects). Each PCEP peer MUST decrement the TTL value before propagating the message. When the TTL value becomes 0, the message is no more propagated.

If the message to be propagated is request-specific and there is no TTL or DIFFUSION-LIST TLVs included, the message MUST reach the source PCC (or alternatively the target PCE).

5.4.2. DIFFUSION-LIST TLV

The DIFFUSION-LIST TLV can be carried within either the error object of a PCErr message, or the notification object of a PCNtf message. It can either be used in a message sent by a PCC to a PCE or vice versa. The DIFFUSION-LIST MAY be used with propagated errors (TLV "Propagation" at value 1 in PCEP-ERROR object).

The format of the DIFFUSION-LIST object body is as follows:



Type (16 bits): restricts the diffusion to certain peers. The following values are currently defined:

0: Any PCEP peer indicated in the list must be reached.

1: Only PCEs must be reached (and not PCC).

2: All PCEP peers with which a session is still opened must be reached.

The value of DIFFUSION-LIST is made of sub-objects similar to the IRO defined in [RFC5440]. The following sub-object types are supported.

Type Sub-object

- 1 IPv4 address
- 2 IPv6 address
- 4 Unnumbered Interface ID
- 5 4-byte AS number
- 6 OSPF area ID
- 7 IS-IS Area ID
- 32 Autonomous System number
- 33 Explicit eXclusion Route Sub-object (EXRS)

If the error or notification codes target specific PCEP peers, a DIFFUSION-LIST TLV avoids partially flooding all PCEP peers. Any PCEP peer receiving a PCErr or PCNTf message containing a PCEP-ERROR or a NOTIFICATION object with a TLV "Propagation" at value 1 and where a DIFFUSION-LIST appears, MUST remove the addresses of the PCEP peers from the DIFFUSION-LIST, before sending the message to any other PCEP peers. This is performed by adding the PCEP peer addresses to the Explicit eXclusion Route Sub-object of the DIFFUSION-LIST. If a DIFFUSION-LIST value is empty, the PCEP peer MUST NOT propagate the message to any peer.

Note that, a Diffusion-List could contain strict or loose addresses to refer to a network domain (e.g. an Autonomous System number, an OSPF area, an IP address). Hence, the PCEP peers targeted by the message would be the PCEP peers covering the corresponding domain. If an address is loose, each time a PCEP peer forwards a message to another PCEP peer of this address, it MUST add its own address to the Explicit eXclusion Route Sub-object (EXRS) of the Diffusion-List for any forwarded messages. Hence, a PCE SHOULD avoid forwarding the same message repeated to the same set of peers. Finally, when an address is loose, the forwarding SHOULD be restrained indicating what type of PCEP peers are targeted (i.e. PCE and/or PCC).

5.4.3. Rules Applied to Existing Errors and Notifications

Many existing normative references states on error definitions (see for instance [RFC5440], [RFC5441], [RFC5455], [RFC5521], [RFC5557], [RFC5886], [RFC8231], [RFC8232], [RFC8253], [RFC8281], [RFC8306], [RFC8408], [RFC8697]). This section provides processing rules for existing error types handling, as a recommendation. According to the definitions provided in this document, the following rules are applicable:

Error-type 1, described in [RFC5440], relates to PCEP session establishment failures. All errors of this type are local and not propagated. Hence, if a "Propagation" TLV is added to the error message it is recommended to be set to value 0. Error-values 1,2,6,7 have a high level of criticality. Hence, if the "Error-criticality" TLV is included within a PCErr message of type 1 and value 1,2,6 or 7, it is recommended to have a value of 2.

Error-type 2,3,4, "Capability not supported", "Unknown object" and "Not supported object" respectively, described in [RFC5440]: errors of this type MAY be propagated using the TLV "Propagation". Their level of criticality is defined as leading to cancel the path computation request [RFC5440]. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The error-value 4 of error-type 4 ("Unsupported parameter") associated to the BRPC procedure [RFC5441] is suggested to contain the "Propagation" TLV with a DIFFUSION-LIST requesting a propagation to the PCC at the origin of the request.

Error-type 5 refers to "Policy violation", error values for this type have been defined in [RFC5440], [RFC5541], [RFC5557], [RFC5886] and [RFC8306]. In [RFC5440], it is specified that the path computation request MUST be cancelled when an error of type 5 occurs. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. As such errors might be conveyed to several PCEs, the "Propagation" TLV MAY be used.

Error-type 6 described as "Mandatory object missing" in [RFC5440], leads to the cancellation of the path computation request. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors. The error-value of 4 for Monitoring object missing defined in [RFC5886] is no exception to the rule.

Error-type 7 is described as "synchronized path computation request missing". In [RFC5440], it is specified that the reffered synchronized path computation request MUST be cancelled when an error of type 5 occurs. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors.

Error-type 8 is raised when a PCE receives a PCRep with an unknown request reference. If the "Propagation" TLV is used with error-type 8, it is recommended to be set at a value of 0. The "Error-criticality" TLV is not particularly relevant for error-type 8. Hence, it usually have the value of 0 if used.

Error-type 9 is raised when a PCE attempts to establish a second PCEP session. The existing session must be preserved. Hence, if the "Error-criticality" TLV is included, it usually have a value of 0. By definition, such an error message SHOULD NOT be propagated. Thus, if the "Propagation" TLV is used with error-type 9, it is usually set to a value of 0.

Error-type 10 which refers to the reception of an invalid object as described in [RFC5440] no indication is provided on the cancellation of the path computation request. Hence, if the "Error-criticality" TLV is included, it usually have a value of 0. The "Propagation" TLV MAY be used with such errors with any value depending on the expected behavior.

Error-type 11 relates to "Unrecognized EXRS subobject" and is described in [RFC5521]. No path computation request cancellation is required by [RFC5521]. Hence, if the "Error-criticality" TLV is included, it usually have a value of 0. The "Propagation" TLV MAY be used with such errors with any value depending on the expected behavior.

Error-type 12 refers to "Diffserv-aware TE error" and is described in [RFC5455]. Such errors are raised when the CLASSTYPE object of a PCReq is recognized but not supported by a PCE. [RFC5455] does not state about the path computation request when such errors are met. Hence, both "Propagation" and "Error-criticality" TLVs COULD be used within such error-types' messages and set to any specified values.

Error-type 13 on "BRPC procedure completion failure" is described in [RFC5441]. [RFC5441] states that in such cases, the PCErr message MUST be relayed to the PCC. Hence, such messages SHOULD contain a "Propagation" TLV and a DIFFUSION-LIST with a Target-Type of 0 and corresponding addresses or with a Target-Type of 2. It is not specified in [RFC5441] whether the path computation request should be canceled or not. If the procedure is not supported, it does not necessarily imply to cancel the path computation request if another procedure is able to read and write VSPT objects. Thus, the "Error-criticality" TLV MAY be used with any value depending on the expected behavior.

Error-type 15 refers to "Global Concurrent Optimization Error" defined in [RFC5557]. [RFC5557] states that the corresponding global concurrent path optimization MUST be cancelled at the PCC. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors.

Error-type 16 relates to "P2MP Capability Error" defined in [RFC8306]. Such errors lead to the cancellation of the path computation request. Hence, if the "Error-criticality" TLV is included, it usually have a value of 1. The "Propagation" TLV MAY be used with such errors.

Error-type 17, titled "P2MP END-POINTS Error" is defined [RFC8306]. Such errors are thrown when a PCE tries to add or prune nodes to or from a P2MP Tree. [RFC8306] does not specify if such errors lead to cancel the path computation request. Hence, the "Error-criticality" and "Propagation" TLVs MAY be used with this type of error with any value depending on the expected behavior.

Error-type 18 of "P2MP Fragmentation Error" is described [RFC8306] which does not specify whether the path computation request should be cancelled. But, as messages are fragmented, it is natural to think that the PCE should wait at least a bit for further messages. The "Error-criticality" TLV MAY be included in such error messages and is particularly adapted to differ the semantic of the same error-type message: if it is included with a value of 0 then the PCE will still wait for further fragmented messages, when this waiting time ends, the TLV can be included with a value of 1 in order to finally cancel the request. The "Propagation" TLV MAY also be used with such errors.

Error-type 19 of "Invalid Operation" is described in [RFC8231] and [RFC8281], which implies a wrong capability description for PCEP session. In this case, the PCErr message MUST be returned to PCC, and this message usually contain a "Propagation" TLV and a

DIFFUSION-LIST with a Target-Type of 0 or 2. The "Error-criticality" TLV is recommended be set to 2 in order to guarantee the termination of PCEP session.

Error-type 20 of "LSP State Synchronization Error" is described in [RFC8231] and [RFC8232], which cannot successfully sync up the LSP states. In this case, the "Error-criticality" TLV should be set to 2 in order to guarantee the termination of PCEP session. The "Propagation" TLV MAY also be used with such errors.

Error-type 21 of "Invalid traffic engineering path setup type" is described in [RFC8408]. Such errors failed to find a matched path setup type and the PCEP sessions MUST be closed. In this case, the "Error-criticality" TLV is usually set to 2 in order to guarantee the termination of PCEP session. The "Propagation" TLV MAY also be used with such errors.

Error-type 23 of "Bad parameter value" is described in [RFC8281]. Such errors occur when there is a conflict in path name of C flag not set for PCE initiation. In this case, the "Error-criticality" TLV may be set to either 0 or 1 to indicate whether the request is still valid, with the PCEP session open. The "Propagation" TLV MAY also be used with such errors.

Error-type 24 of "LSP instantiation error" is described in [RFC8281]. Such errors occur when PCC detects problems when establishing the path, the message MUST relay to the PCE, therefore the "Propagation" TLV is usually contained. The "Error-criticality" TLV may be set to either 0 or 1 to indicate whether the request is still valid, with the PCEP session open.

Error-type 25 of "PCEP StartTLS failure" is described in [RFC8253]. Such errors indicate the security issue in transport layer. In this case, the "Error-criticality" TLV is usually set to 2 in order to close the PCEP session. The "Propagation" TLV MAY also be used with such errors, depending on the detailed security conditions.

Error-type 26 of "Association Error" is described in [RFC8697]. Such errors occur when there is problem for LSP association. In this case, the "Error-criticality" TLV should be set to either 0 or 1 to indicate whether the request is still valid, with the PCEP session open. The "Propagation" TLV MAY also be used with such errors.

6. Error Handling Guidelines for Future PCEP Extension

Error and Notification handling in this document should be considered in PCE documents that include new errors and notifications. A requirement for the authors of these drafts is to evaluate the applicability of the procedure in this document and provide details about the "Error-criticality" TLV and "Propagation" TLV for errors and notifications defined in the draft. Example text is provided as follow.

Error-type XX (fill in value of the Error-type) of " XXXX " (fill in name of the Error-type) is described in [RFCYYYY] (fill in the document reference of the Error-type). Such errors occur when ZZZZ (fill in typical scenario). In this case, the "Error-criticality" TLV should be set to X (fill in the recommended value) to indicate whether the request is still valid, with the PCEP session open. The error messages SHOULD/MAY (select the mandatory level) contain a "Propagation" TLV and a DIFFUSION-LIST with a Target-Type of A(fill in the recommended value).

7. Backward Compatibility Consideration

There would be backward compatibility issue if there are multiple PCEs with different level understanding of error message. In a scenario that PCE(i) propagate the error message to PCE (i+1), it is possible that PCE (i+1) is not capable to extract the message correctly, then such error message would be ignored and not be further propagated.

There can be potential approach to avoid these problem, such as recognizing the incapable PCE and avoiding propagation. However, these approach is not in the scope of this document.

8. Implementation Status

[NOTE TO RFC EDITOR : This whole section and the reference to [RFC7942] is to be removed before publication as an RFC]

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".

At the time of posting the -08 version of this document, there are no known implementations of this mechanism. It is believed that two vendors are considering prototype implementations, but these plans are too vague to make any further assertions.

9. Security Considerations

Within the introduced set of TLVs, the "Propagation" TLV affects PCEP security considerations since it forces propagation behaviors. Thus, a PCEP implementation SHOULD activate stateful mechanism when receiving PCEP-ERROR or NOTIFICATION object including this TLV in order to avoid DoS attacks.

10. IANA Considerations

IANA maintains a registry of PCEP parameters. This includes a sub-registry for PCEP Objects.

IANA is requested to make an allocation from the sub-registry as follows. The values here are suggested for use by IANA.

10.1. PCEP TLV Type Indicators

As described in Section 5.4 the newly defined TLVs allows a PCE to enforce specific error and notification behaviors within PCEP-ERROR and NOTIFICATION objects. IANA is requested to make the following allocations from the "PCEP TLV Type Indicators" sub-registry.

Value	Description	Reference
TBD	Propagation	this document
TBD	Error-criticality	this document

10.2. New DIFFUSION-LIST TLV

Type	Value	Meaning	Reference
	0	Any PCEP peers	this document
	1	PCEs but excludes PCC-only peers	this document
	2	PCEs and PCCs with which a session is still opened	this document

Subobjects	Reference
1: IPv4 prefix	this document
2: IPv6 prefix	this document
4: Unnumbered Interface ID	this document
5: 4-byte AS number	this document
6 OSPF area ID	this document
7 IS-IS Area ID	this document
32: Autonomous system number	this document
33: Explicit Exclusion Route subobject (EXRS)	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>>.

- [RFC5441] Vasseur, JP., Ed., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, DOI 10.17487/RFC5441, April 2009, <<https://www.rfc-editor.org/info/rfc5441>>.
- [RFC5455] Sivabalan, S., Ed., Parker, J., Boutros, S., and K. Kumaki, "Diffserv-Aware Class-Type Object for the Path Computation Element Communication Protocol", RFC 5455, DOI 10.17487/RFC5455, March 2009, <<https://www.rfc-editor.org/info/rfc5455>>.
- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, DOI 10.17487/RFC5521, April 2009, <<https://www.rfc-editor.org/info/rfc5521>>.
- [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>>.
- [RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, DOI 10.17487/RFC5557, July 2009, <<https://www.rfc-editor.org/info/rfc5557>>.
- [RFC5886] Vasseur, JP., Ed., Le Roux, JL., and Y. Ikejiri, "A Set of Monitoring Tools for Path Computation Element (PCE)-Based Architecture", RFC 5886, DOI 10.17487/RFC5886, June 2010, <<https://www.rfc-editor.org/info/rfc5886>>.
- [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>>.
- [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>>.
- [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>>.
- [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>>.
- [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>>.

11.2. Informational 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>>.
- [RFC7470] Zhang, F. and A. Farrel, "Conveying Vendor-Specific Constraints in the Path Computation Element Communication Protocol", RFC 7470, DOI 10.17487/RFC7470, March 2015, <<https://www.rfc-editor.org/info/rfc7470>>.
- [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>>.

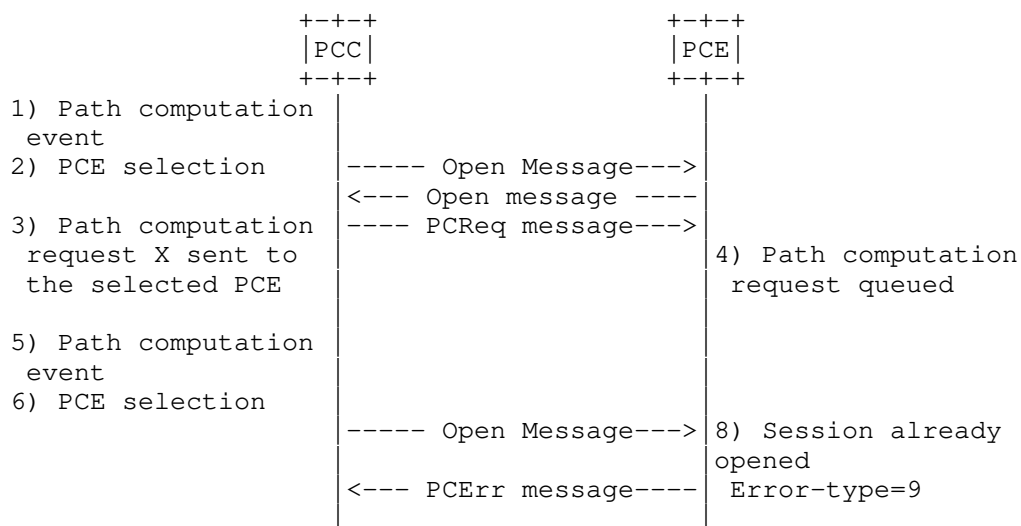
[RFC8751] Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., and D. King,
 "Hierarchical Stateful Path Computation Element (PCE)",
 RFC 8751, DOI 10.17487/RFC8751, March 2020,
 <<https://www.rfc-editor.org/info/rfc8751>>.

Appendix A. Error and Notification Scenarios

This section provides some examples depicting how the error described above can be used in a PCEP session. The origin of the errors or notifications is only illustrative and has no normative purpose. Sometimes the PCE features behind may be implementation-specific (e.g. detection of flooding). This section does not provide scenarios for errors with a high-level of criticality (i.e., Error behaviors 3 and 6) since such errors are very specific and until now have been normalized only during the session establishment (error-type of 1).

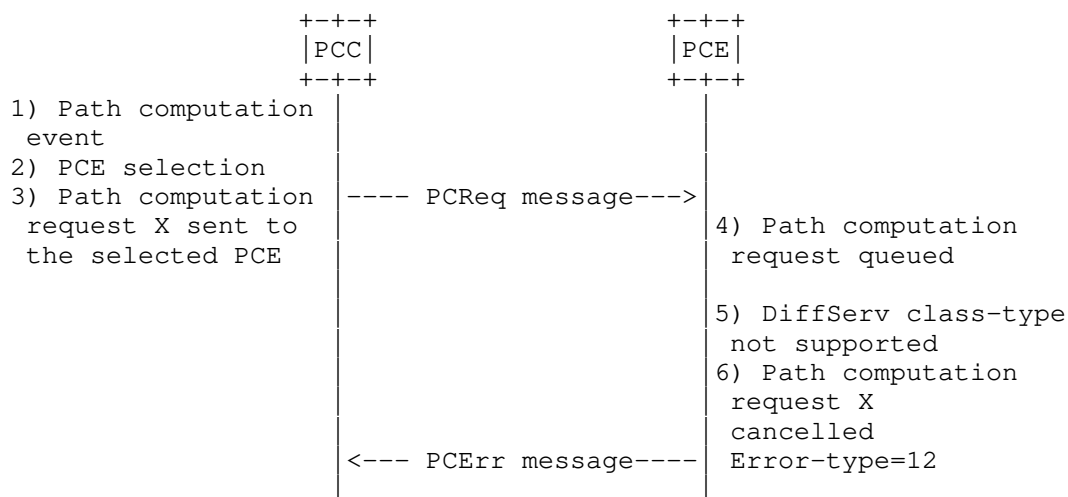
A.1. Error Behavior Type 1

In this example, a PCC attempts to establish a second PCEP session with the same PCE for another request. Consequently the PCE sends back an error message with error-type 9. This error stays local and does not affect the former session. The second session is ignored. If the "Propagation" TLV and "Error-criticality" TLV are used, they should be both set to value 0.



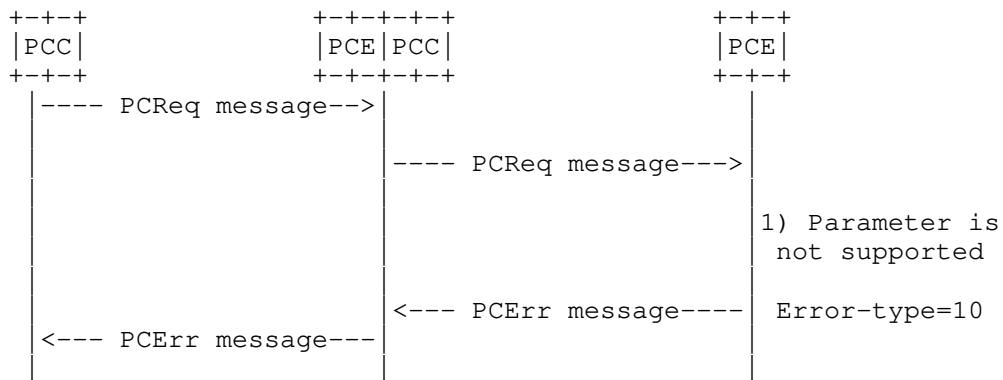
A.2. Error Behavior Type 2

In this example, the PCC sends a DiffServ-aware path computation request. If the PCE receiving the request does not support the indicated class-type, it thus sends back a PCErr message with error-type=12 and error-value=1. If the "Propagation" TLV and "Error-criticality" TLV are present, they should carry value 0 and value 1 respectively. Consequently, the request is cancelled.



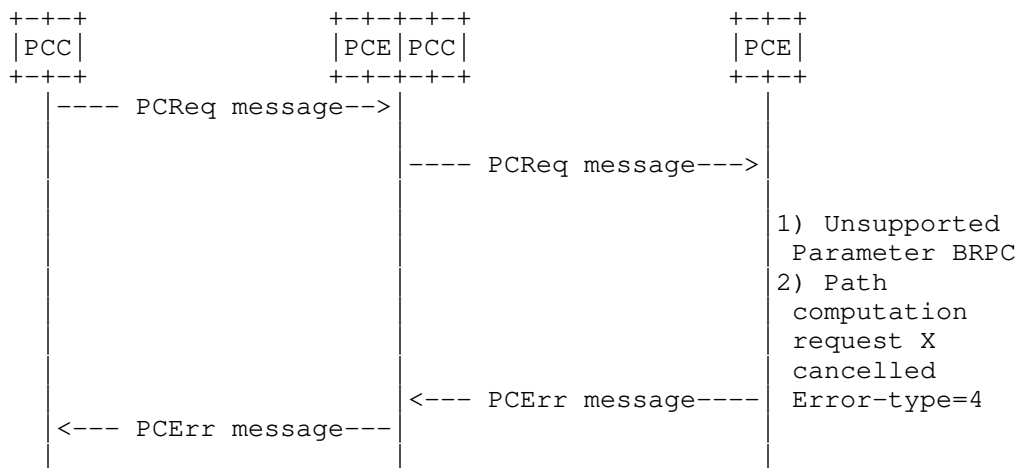
A.3. Error Behavior Type 4

In this example, a PCC sends a path computation requests with no P flag set (e.g. END-POINT object with P-flag cleared). This is detected by another PCE in the sequence. The path computation request can thus be treated but the P-Flag will be ignored. Hence, this error is not critical but the source PCC should be informed of this fact. So, a PCErr message with error-type 10 ("Reception of an invalid object"). The PCEP-ERROR object of the message contains a "Propagation" TLV at value 1 and a "Error-criticality" TLV at value 0. It is hence propagated backwardly to the source PCC.



A.4. Error Behavior Type 5

In this example, PCEs are using the BRPC procedure to treat a path computation request [RFC5441]. However, one of the PCEs does not support a parameter of the request. Hence, a PCErr message with error-type 4 and error-value 4 is sent by this PCE and has to be forwarded to the source PCC. The PCEP-ERROR object includes a "Propagation" TLV at value 1 and "Error-criticality" TLV at value 1 and the message is propagated backwardly to the source PCC. Consequently, the request is cancelled.



Authors' Addresses

Helia Pouyllau
Alcatel-Lucent
Route de Villejust
91620 NOZAY
France
Phone: + 33 (0)1 30 77 63 11
Email: helia.pouyllau@alcatel-lucent.com

Remi Theillaud
Marben Products
176 rue Jean Jaures
92800 Puteaux
France
Phone: + 33 (0)1 79 62 10 22
Email: remi.theillaud@marben-products.com

Julien Meuric
Orange
2, avenue Pierre Marzin
22307 Lannion
France
Email: julien.meuric@orange.com

Haomian Zheng (Editor)
Huawei Technologies
H1, Xiliu Beipo Village, Songshan Lake,
Dongguan
Guangdong, 523808
China
Email: zhenghaomian@huawei.com

Xian Zhang
Huawei Technologies
A10, Huawei Industrial Base, Bantian, Longgang District
Shenzhen
Guangdong, 518129
China
Email: zhang.xian@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 9, 2019

Q. Zhao
Z. Li
M. Negi
Huawei Technologies
C. Zhou
Cisco Systems
February 5, 2019

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) of LSPs
draft-ietf-pce-pcep-extension-for-pce-controller-01

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 August 9, 2019.

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	12
5.4.2.2. Label Cleanup	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. Security Considerations	26
8.1. Malicious PCE	26
9. Manageability Considerations	26
9.1. Control of Function and Policy	26
9.2. Information and Data Models	26
9.3. Liveness Detection and Monitoring	26
9.4. Verify Correct Operations	26
9.5. Requirements On Other Protocols	26
9.6. Impact On Network Operations	27
10. IANA Considerations	27
10.1. PCEP TLV Type Indicators	27
10.2. New Path Setup Type Registry	27
10.3. PCEP Object	27
10.4. CCI Object Flag Field	27
10.5. PCEP-Error Object	28
11. Acknowledgments	28
12. References	28
12.1. Normative References	28
12.2. Informative References	30
Appendix A. Contributor Addresses	32
Authors' Addresses	33

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] and [I-D.ietf-teas-pcecc-use-cases].

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 MUST have the capability to advertise its PCECC capability to its peers.

2. PCEP speaker not supporting this draft MUST be able to reject PCECC related extensions with a error reason code that indicates that this feature is not supported.
 3. PCEP speaker MUST provide a means to identify PCECC based LSP in the PCEP messages.
 4. PCEP procedures SHOULD provide a means to update (or cleanup) the label- download entry to the PCC.
 5. PCEP procedures SHOULD 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:

(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.

(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.

(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 object 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. 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 = TBD: 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=TBD 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.

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=TBD (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 SHOULD send a PCErr message with Error-Type=19 (Invalid Operation) and Error-value=TBD (stateful PCE capability was not advertised) and terminate the session.

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. 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 PCEP. 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 SHOULD send the PCUpd message to the Ingress PCC. The 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.

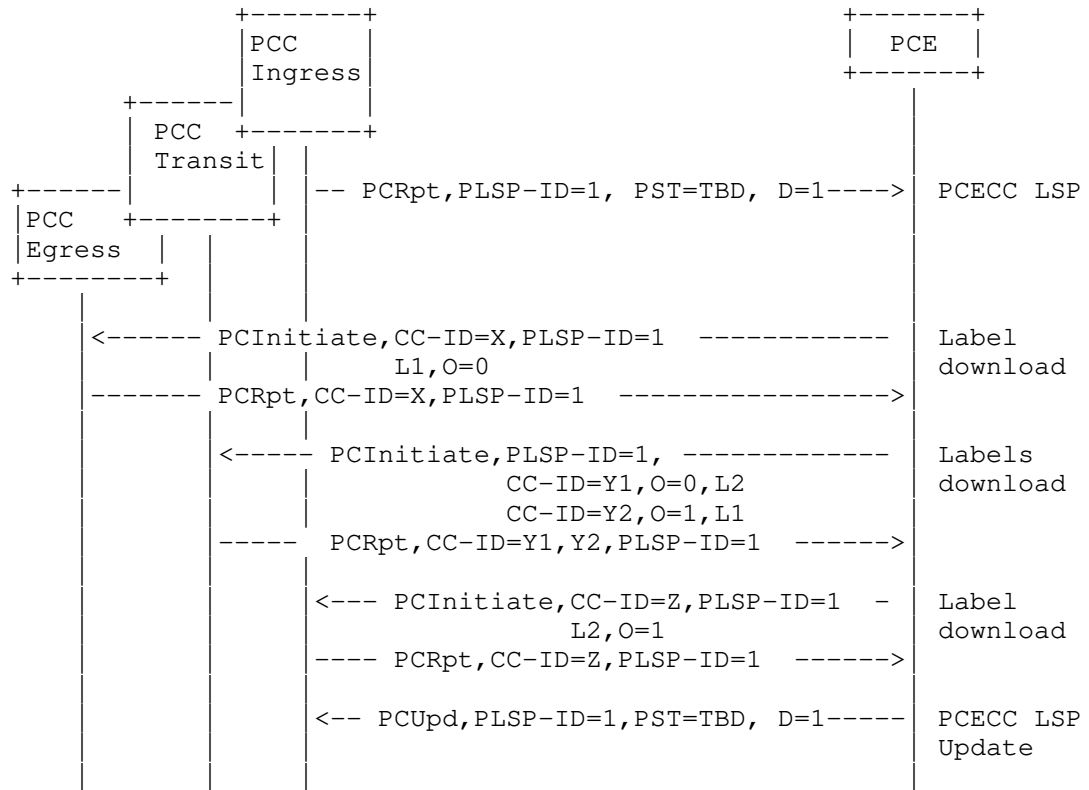
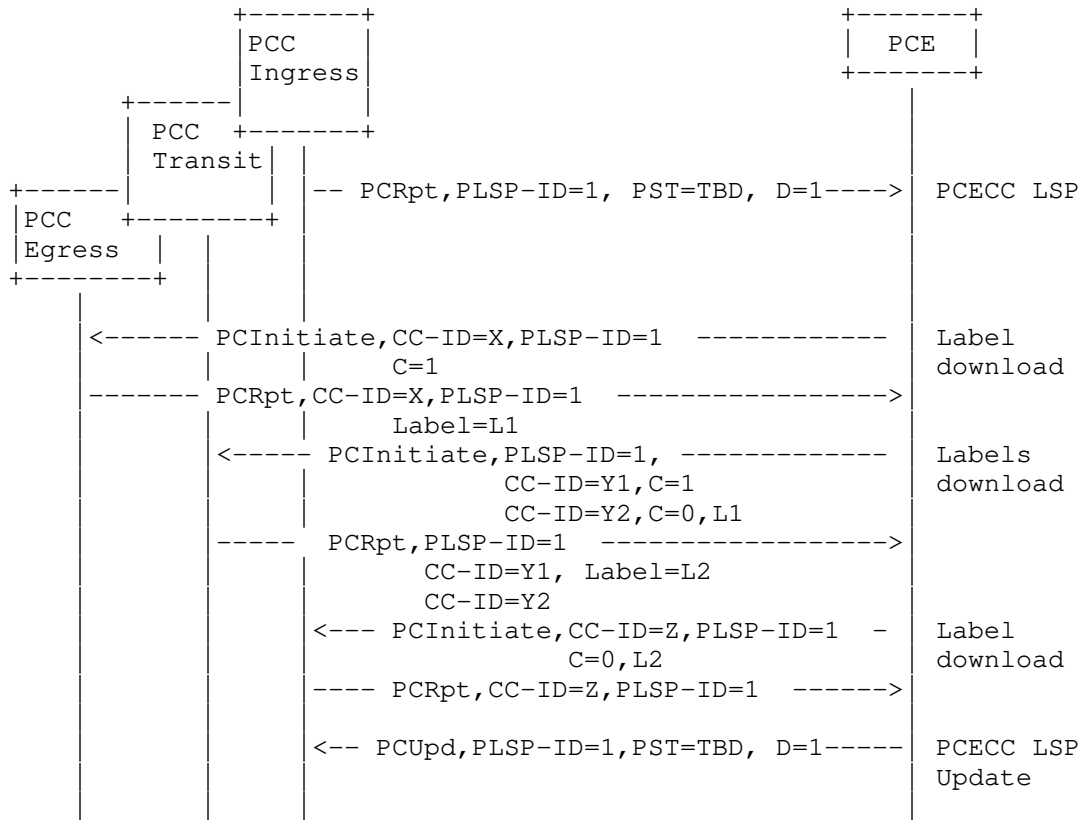


Figure 2: Basic PCECC LSP setup

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 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 -



- 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

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=TBD (PCECC failure) and Error-value=TBD (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=TBD (PCECC failure) and Error-value=TBD (instruction failed) and MUST include the SRP object to specify the error is for the corresponding label update via PCInitiate message.

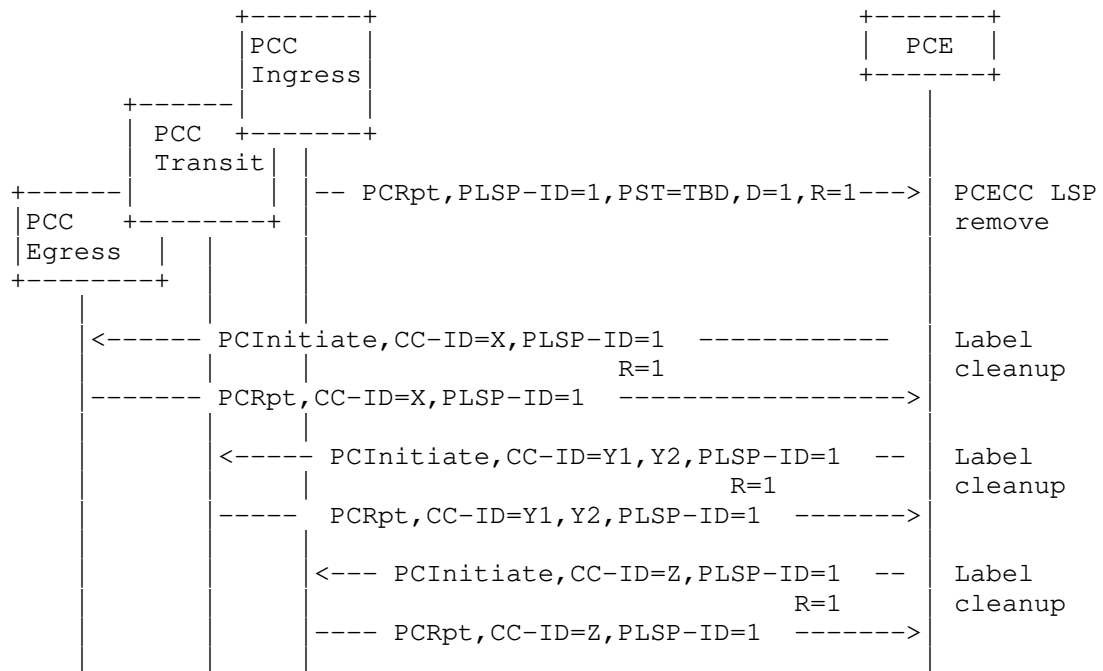
New PCEP object for central control instructions (CCI) is defined in Section 7.3.

5.4.2.2. Label Cleanup

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=TBD, "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].

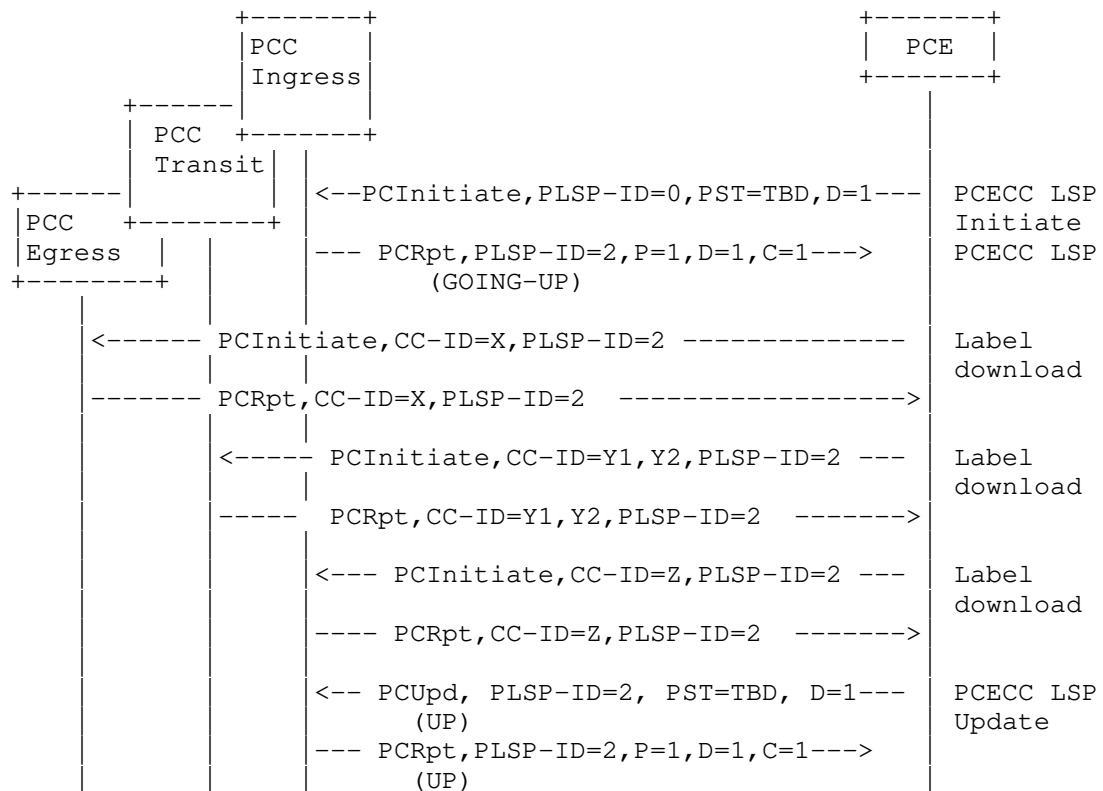
In order to setup a PCE Initiated LSP based on the PCECC mechanism, a PCE sends PCInitiate message with Path Setup Type set for PCECC (see Section 7.2) to the Ingress PCC.

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 -



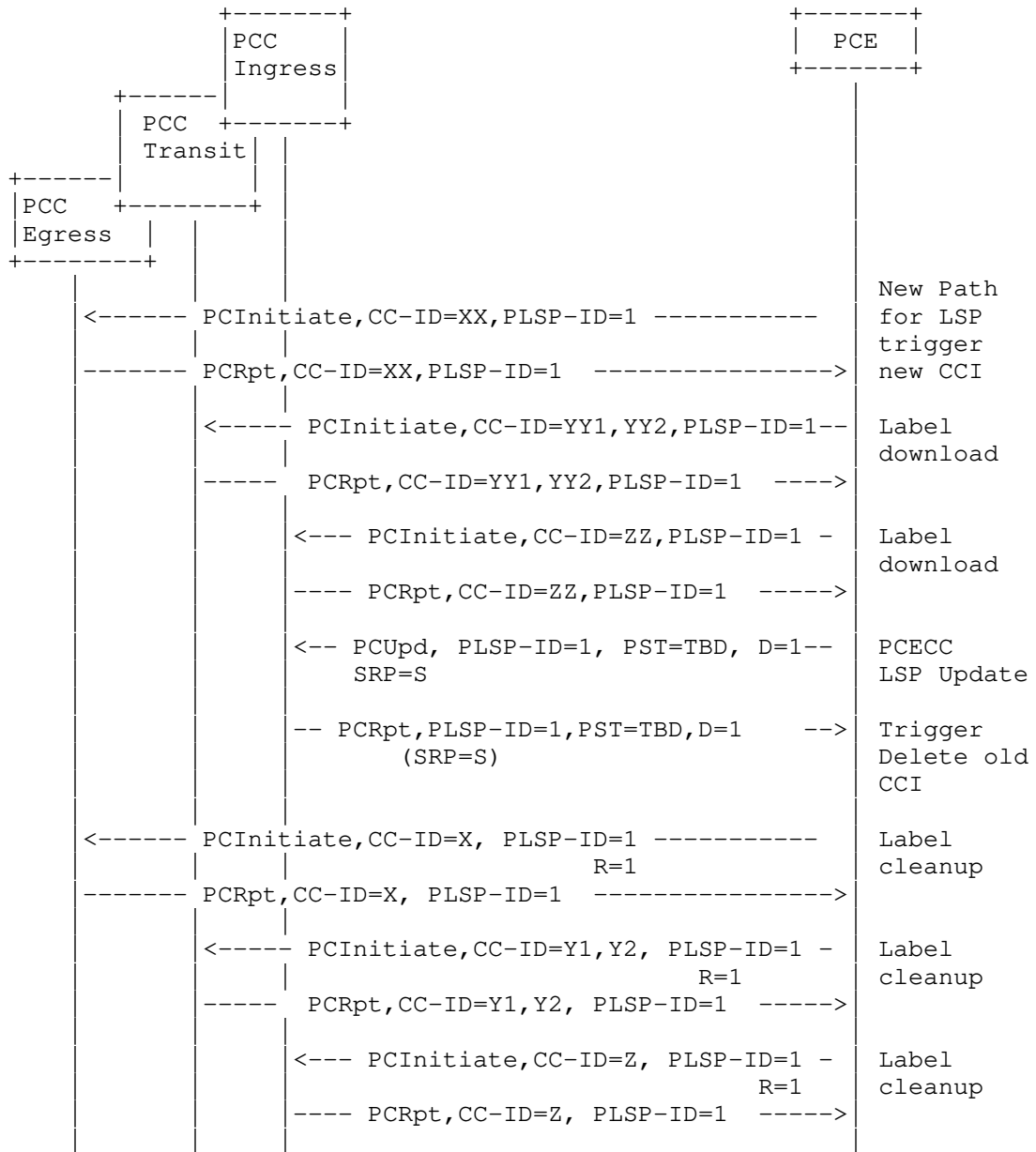
Once the label operations are completed, the PCE SHOULD send the PCUpd message to the Ingress PCC. The PCUpd message is as per [RFC8231].

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 = TBD ("PCECC failure") and Error Value = TBD ("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 = 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, 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. The same flag can also be 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=TBD (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=TBD (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].

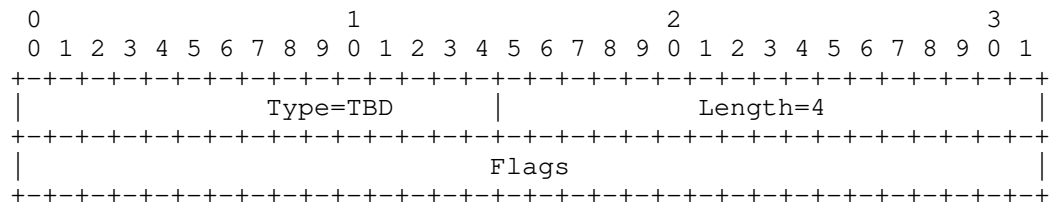
7.1. OPEN Object

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 TBD 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 = TBD: Path is setup via PCECC mode.

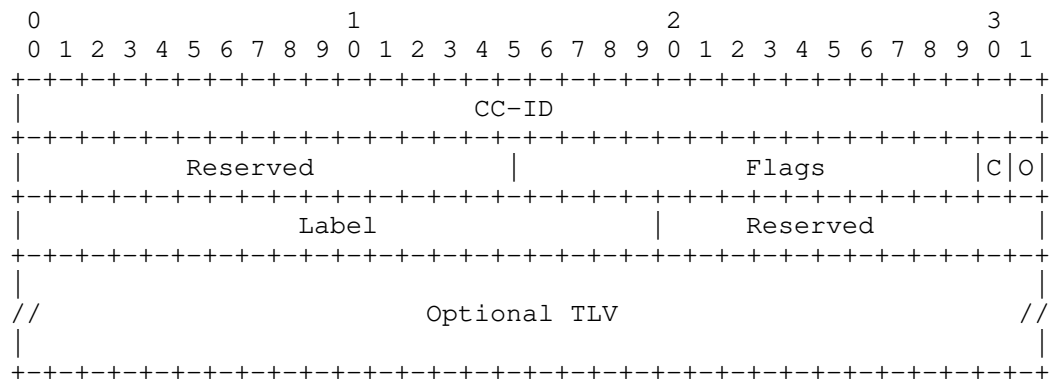
On a PCRpt/PCUpd/PCInitiate message, the PST=TBD in PATH-SETUP-TYPE TLV in SRP object indicates that this LSP was setup via a PCECC based mechanism.

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 TBD.

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=TBD                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     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=TBD                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     //                                     //
|                                     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=TBD                               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Node-ID                                |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     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 pair of Node ID / Interface ID tuples.

8. 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.

8.1. 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].

9. Manageability Considerations

9.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC 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 capability status.

The PCEP YANG module [I-D.ietf-pce-pcep-yang] could be extended to enable/disable PCECC 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 extensions defined in this document do not put new requirements on network operations.

10. IANA Considerations

10.1. PCEP TLV Type Indicators

IANA is requested to confirm the early allocation of the following TLV Type Indicator values within the "PCEP TLV Type Indicators" sub-registry of the PCEP Numbers registry, and to update the reference in the registry to point to this document, when it is an RFC:

Value	Meaning	Reference
TBD	PCECC-CAPABILITY	This document
TBD	IPV4-ADDRESS TLV	This document
TBD	IPV6-ADDRESS TLV	This document
TBD	UNNUMBERED-IPV4-ID-ADDRESS TLV	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	Traffic engineering path is setup using PCECC mode	This document

10.3. PCEP Object

IANA is requested to allocate new registry for CCI PCEP object.

Object-Class Value	Name	Reference
TBD	CCI Object-Type	This document
	1	MPLS Label

10.4. CCI Object Flag Field

IANA is requested to create a registry to manage the Flag field of the CCI object.

One bit to be defined for the CCI Object flag field in this document:

Codespace of the Flag field (CCI Object)

Bit	Description	Reference
7	Specifies label is out label	This document

10.5. 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 :	Attempted PCECC operations when PCECC capability was not advertised
	Error-value = TBD :	Stateful PCE capability was not advertised
	Error-value = TBD :	Unknown Label
6	Mandatory Object missing.	
	Error-value = TBD :	CCI object missing
TBD	PCECC failure.	
	Error-value = TBD :	Label out of range.
	Error-value = TBD :	Instruction failed.
	Error-value = TBD :	Invalid CCI.
	Error-value = TBD :	Unable to allocate the specified CCI.

11. Acknowledgments

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

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

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>>.
- [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-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-02 (work in progress), October 2018.
- [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-09 (work in progress), October 2018.

- [I-D.zhao-pce-pcep-extension-pce-controller-sr]
Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A.,
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-03 (work
in progress), June 2018.
- [I-D.li-pce-controlled-id-space]
Li, C., Chen, M., Dong, J., Li, Z., and A. Wang, "PCE
Controlled ID Space", draft-li-pce-controlled-id-space-01
(work in progress), December 2018.
- [I-D.sivabalan-pce-binding-label-sid]
Sivabalan, S., Filsfils, C., Tantsura, J., Hardwick, J.,
Previdi, S., and D. Dhody, "Carrying Binding Label/
Segment-ID in PCE-based Networks.", draft-sivabalan-pce-
binding-label-sid-05 (work in progress), October 2018.
- [I-D.li-pce-sr-path-segment]
Li, C., Chen, M., Dhody, D., Cheng, W., Dong, J., Li, Z.,
and R. Gandhi, "Path Computation Element Communication
Protocol (PCEP) Extension for Path Segment in Segment
Routing (SR)", draft-li-pce-sr-path-segment-03 (work in
progress), October 2018.

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

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: 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: mahendrasingh@huawei.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.

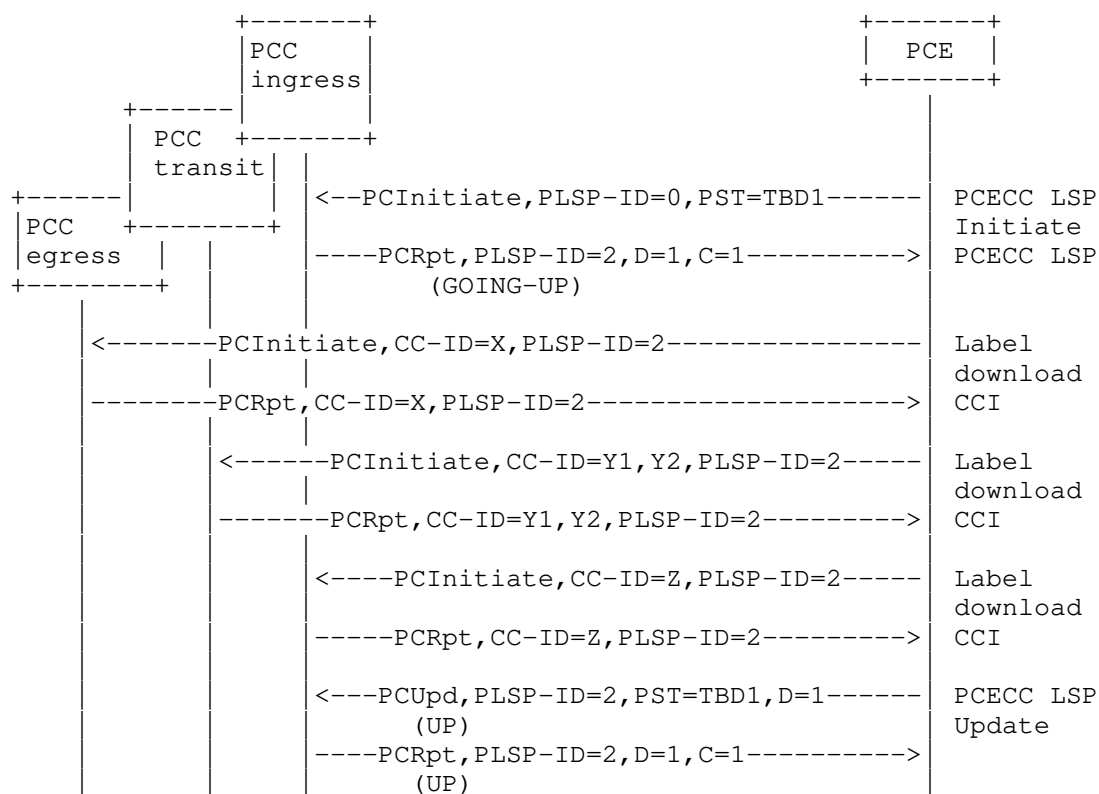


Figure 1: PCE-Initiated PCECC LSP

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.

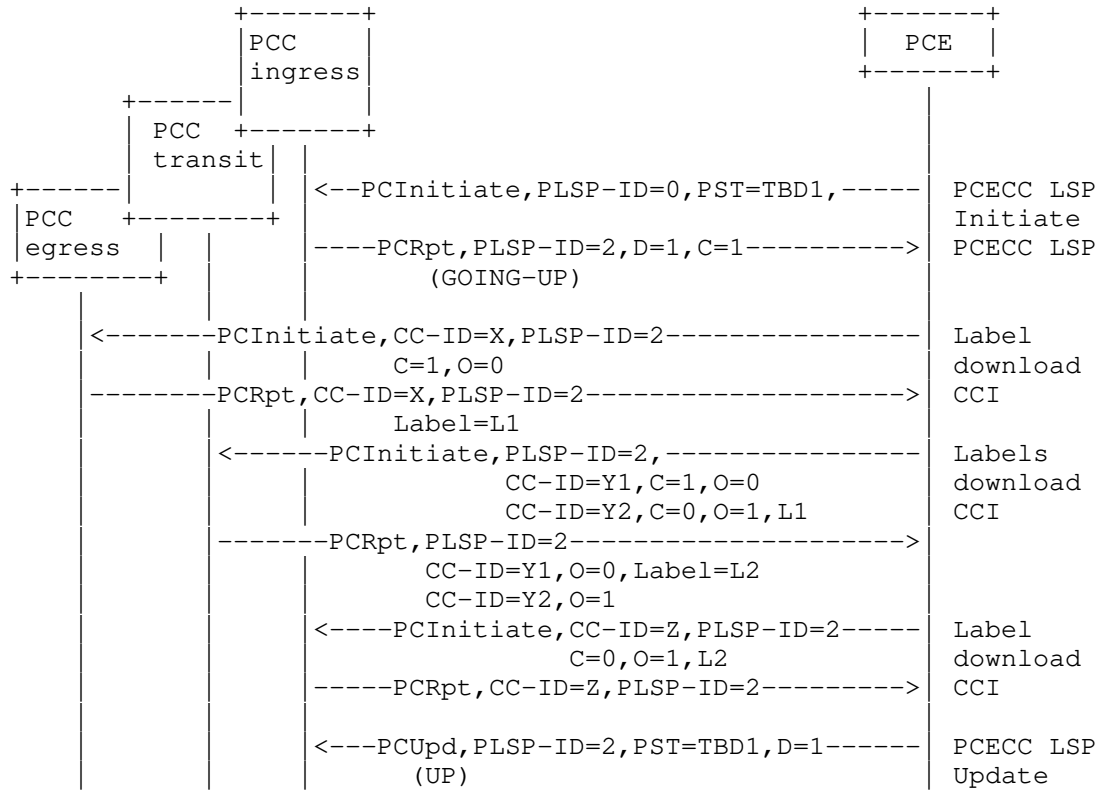


Figure 2: PCE-Initiated PCECC LSP (PCC allocation)

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.

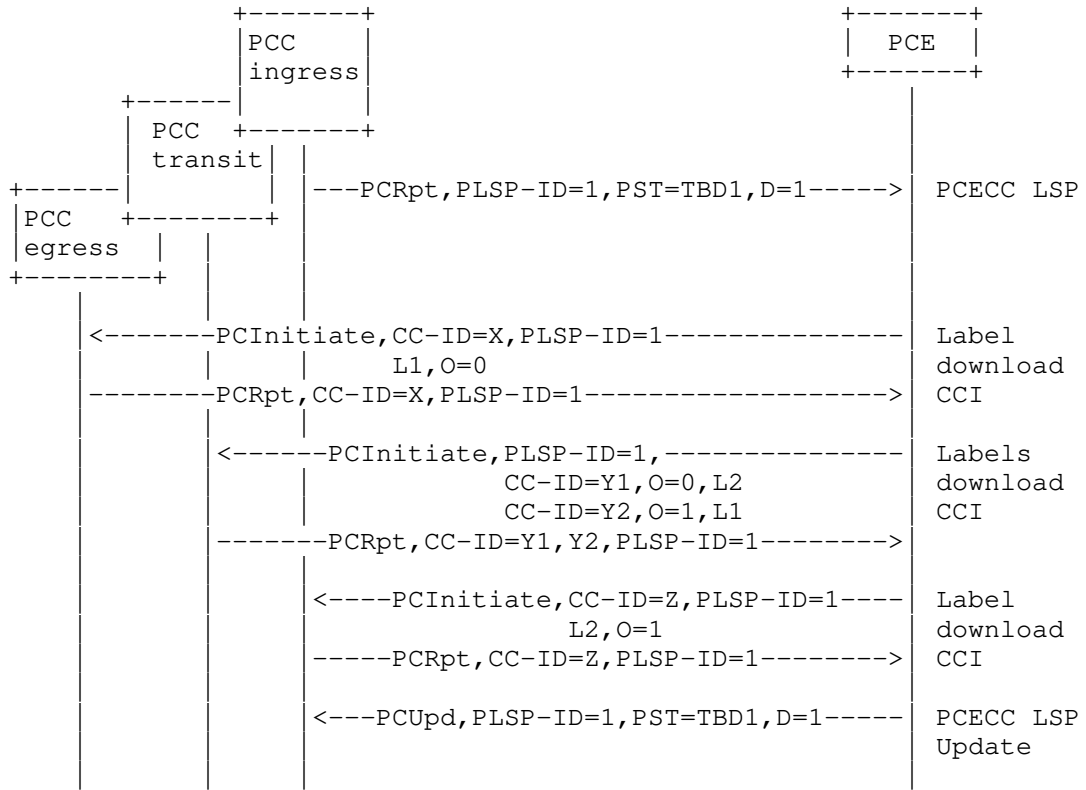
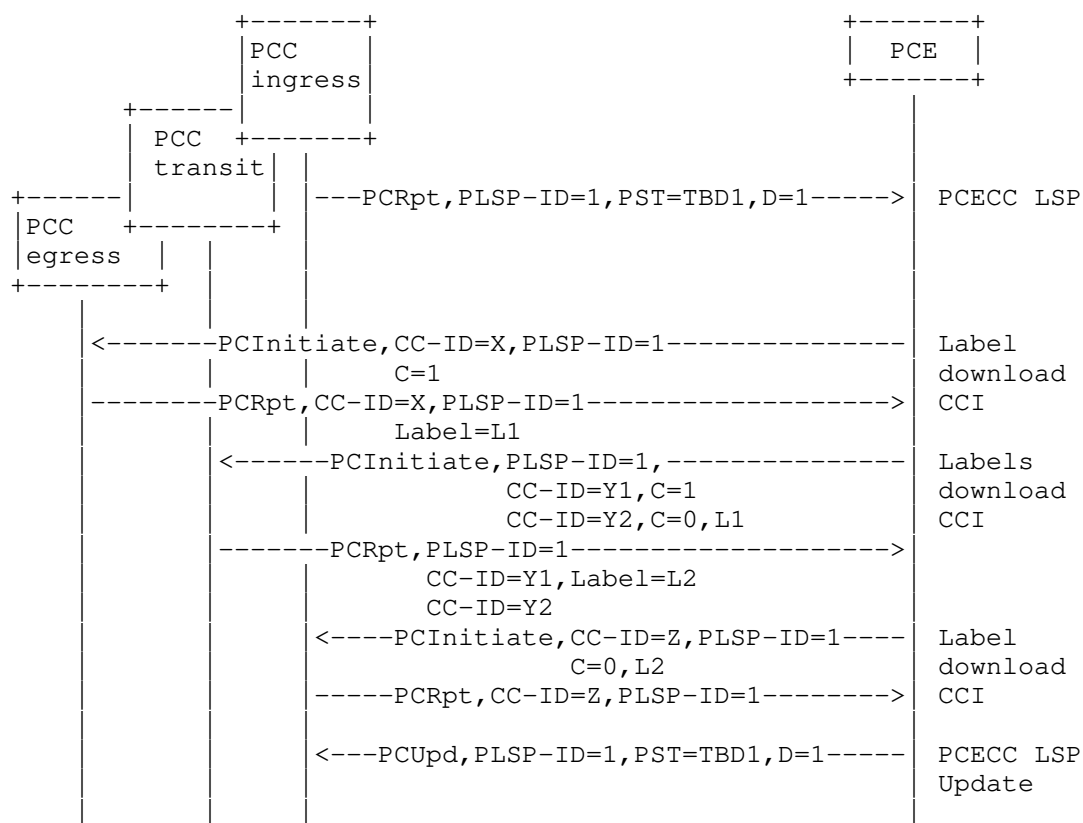


Figure 3: PCC-Initiated PCECC LSP

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.

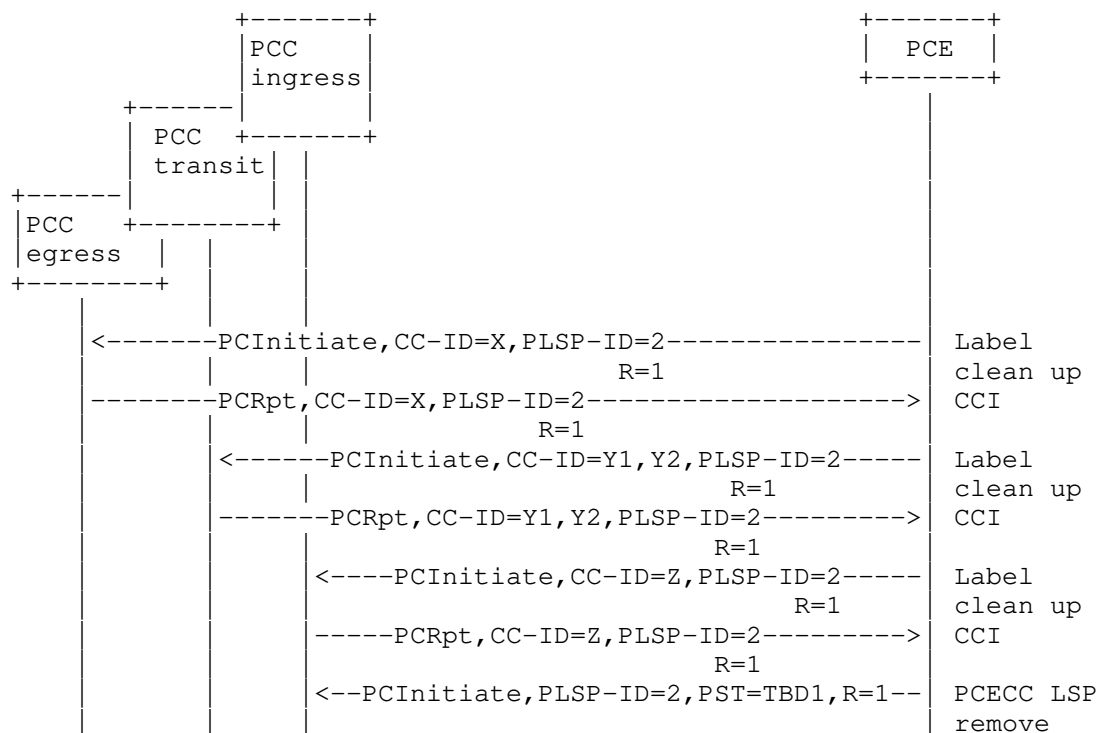


Figure 5: Label Cleanup

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.

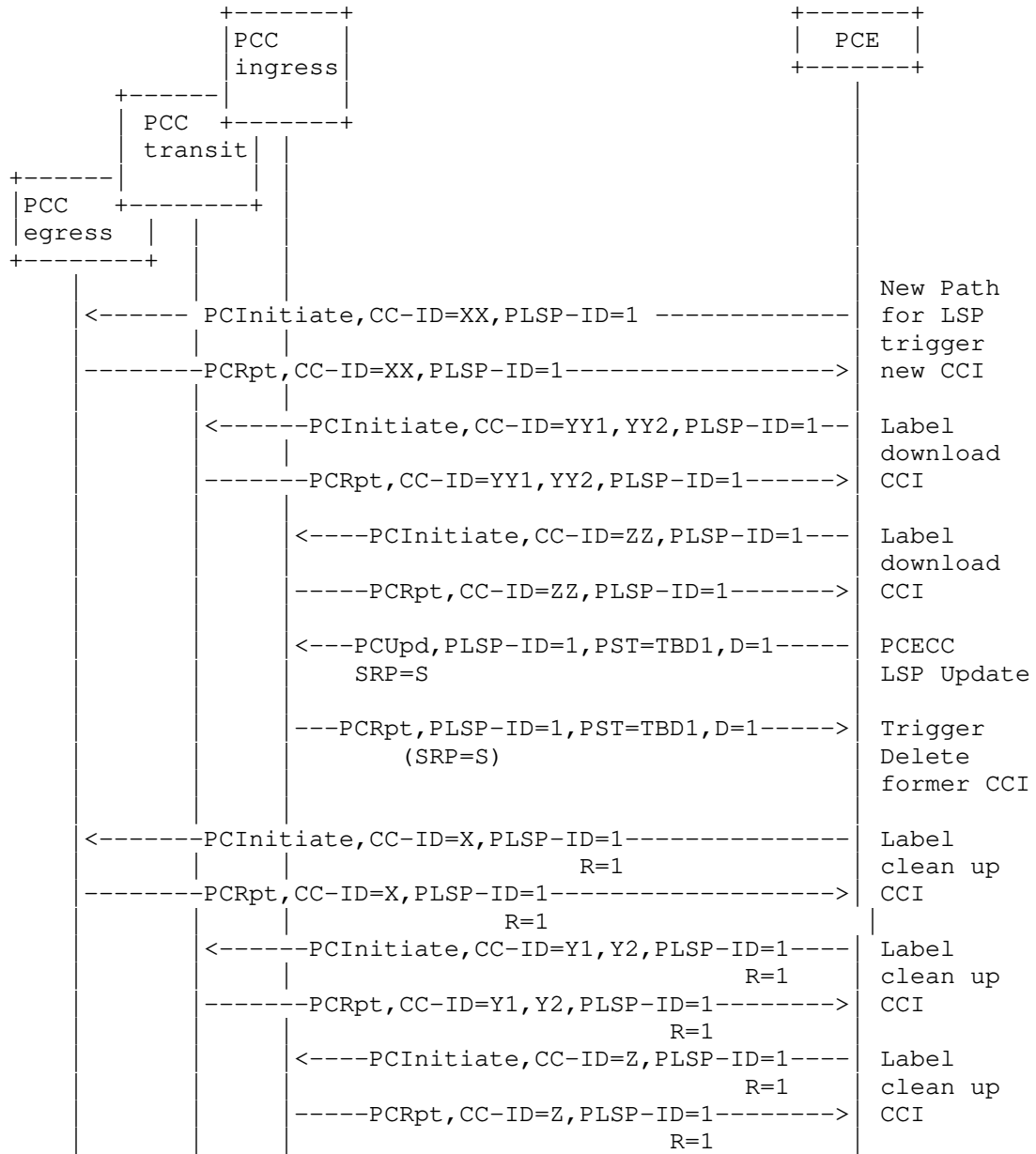


Figure 6: PCECC LSP Update

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.

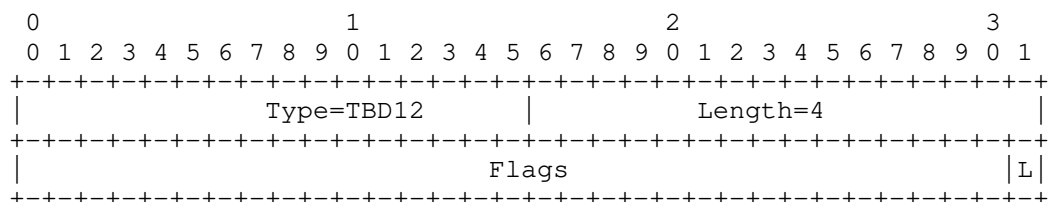


Figure 7: PCECC Capability sub-TLV

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.

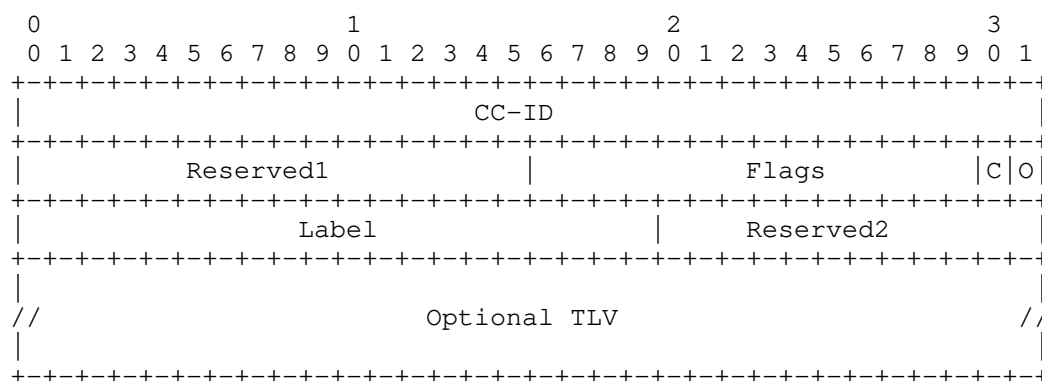


Figure 8: CCI Object

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

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 9, 2019

A. Wang
China Telecom
B. Khasanov
Huawei
S. Cheruathur
Juniper Networks
C. Zhu
ZTE Corporation
S. Fang
Huawei
March 8, 2019

PCEP Extension for Native IP Network
draft-ietf-pce-pcep-extension-native-ip-03

Abstract

This document defines the PCEP extension for CCDR application in Native IP network. The scenario and architecture of CCDR in native IP is described in [I-D.ietf-teas-native-ip-scenarios] and [I-D.ietf-teas-pce-native-ip]. This draft describes the key information that is transferred between PCE and PCC to accomplish the end2end traffic assurance in Native IP network under central control mode.

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 9, 2019.

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. CCI Objects	3
4. CCI Object associated TLV	4
4.1. Peer Address List TLV	4
4.2. Peer Prefix Association TLV	5
4.2.1. Prefix sub TLV	6
4.3. Explicit Peer Route TLV	7
5. Management Consideration	7
6. Security Considerations	7
7. IANA Considerations	8
7.1. CCI Object Type	8
7.2. CCI Object Associated TLV	8
8. Acknowledgement	8
9. Normative References	8
Authors' Addresses	9

1. Introduction

Traditionally, MPLS-TE traffic assurance requires the corresponding network devices support MPLS or the complex RSVP/LDP/Segment Routing etc. technologies to assure the end-to-end traffic performance. But in native IP network, there will be no such signaling protocol to synchronize the action among different network devices. It is necessary to use the central control mode that described in [RFC8283] to correlate the forwarding behavior among different network devices. Draft [I-D.ietf-teas-pce-native-ip] describes the architecture and solution philosophy for the end2end traffic assurance in Native IP network via Dual/Multi BGP solution. This draft describes the corresponding PCEP extensions to transfer the key information about peer address list, peer prefix association and the explicit peer route on on-path router.

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 RFC 2119 [RFC2119].

3. CCI Objects

Draft [I-D.ietf-pce-pcep-extension-for-pce-controller] introduces the CCI object which is included in the PCInitiate and PCRpt message to transfer the centrally control instruction and status between PCE and PCC. This object is extended to include the construction for native IP solution. Additional TLVs are defined and included in this extended CCI object.

CCI Object-Class is TBD, should be same as that defined in draft [I-D.ietf-pce-pcep-extension-for-pce-controller]

CCI Object-Type is TBD for Native IP network

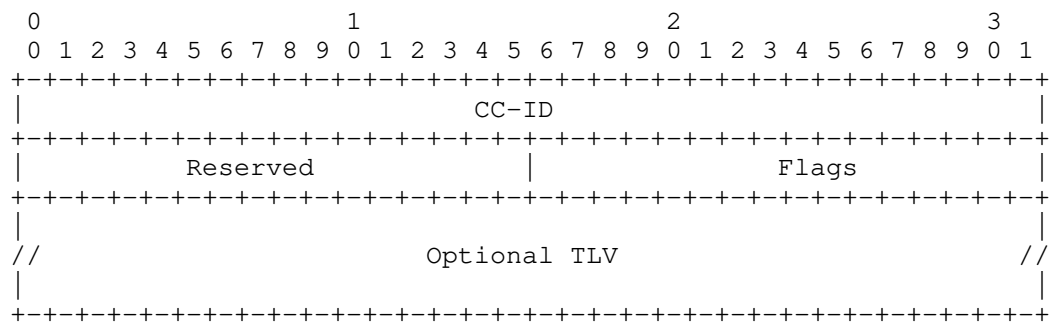


Figure 1: CCI Object Format

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.

Optional TLV: Additional TLVs that are associated with the Native IP construction.

4. CCI Object associated TLV

Three new TLVs are defined in this draft:

- o PAL TLV: Peer Address List TLV, used to tell the network device which peer it should be peered with dynamically
- o PPA TLV: Peer Prefix Association TLV, used to tell which prefixes should be advertised via the corresponding peer
- o EPR TLV: Explicit Peer Route TLV, used to point out which route should be taken to arrive to the peer.

4.1. Peer Address List TLV

The Peer Address List TLV is defined to specify the IP address of peer that the received network device should establish the BGP relationship with. This TLV should only be included and sent to the head and end router of the end2end path in case there is no RR involved. If the RR is used between the head and end routers, then such information should be sent to head router, RR and end router respectively.

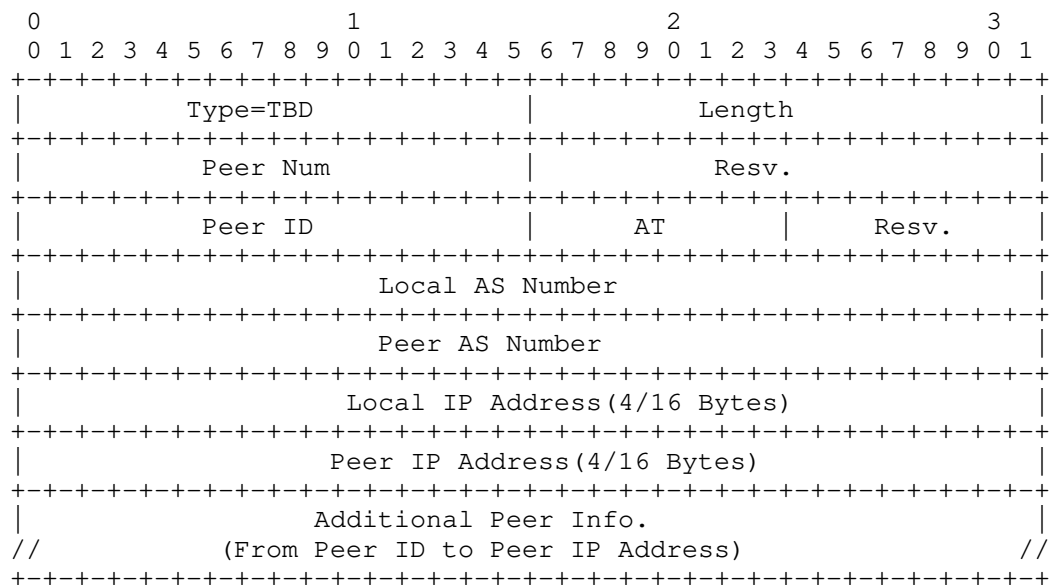


Figure 2: Peer Address List TLV Format

Type: 2 Bytes, value is TBD.

Length: 2 Bytes, the length of the following fields.

Peer Num : 2 Bytes, Peer Address Number on the advertised router.

Peer-ID: 2 Bytes, to distinguish the different peer pair, will be referenced in Peer Prefix Association, if the PCE use multi-BGP solution for different QoS assurance requirement.

AT: 1 Bytes, Address Type. To indicate the address type of Peer. Equal to 4, if the following IP address of peer is belong to IPv4; Equal to 6 if the following IP address of peer is belong to IPv6.

Resv: 1 Bytes, Reserved for future use.

Local AS Number: 4 Bytes, to indicate the AS number of the Local Peer.

Peer AS Number: 4 Bytes, to indicate the AS number of Remote Peer.

Local IP Address(4/16 Bytes): IPv4 address of the local router, used to peer with other end router. When AT equal to 4, length is 32bit; when AT equal to 16, length is 128bit.

Peer IP Address(4/16 Bytes): IPv4 address of the peer router, used to peer with the local router. When AT equal to 4, length is 32bit; IPv6 address of the peer when AT equal to 16, length is 128bit;

4.2. Peer Prefix Association TLV

The Peer Prefix Association TLV is defined to specify the IP prefixes that should be advertised by the corresponding Peer. This TLV should only be included and sent to the head/end router of the end2end path in case there is no RR involved. If the RR is used between the head and end routers, then such information should be sent to head router,RR and end router respectively.

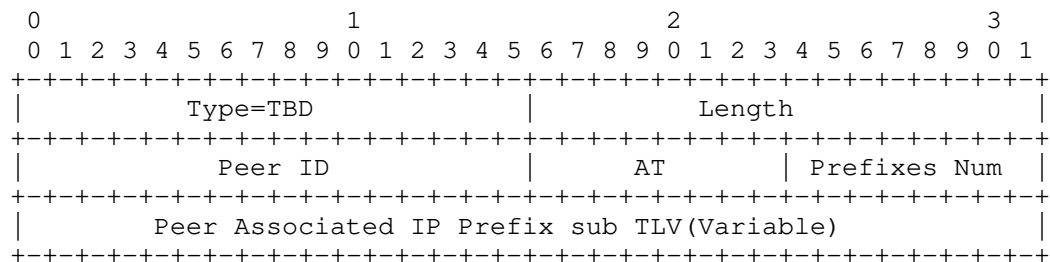


Figure 3: Peer Prefix Association TLV Format

Type: 2 Bytes, value is TBD

Length: 2 Bytes, the length of the following fields.

Peer-ID: 2 Bytes, to indicate which peer should be used to advertise the following IP Prefix TLV. This value is assigned in the Peer Address List object and is referred in this object.

AT: 2 Bytes, Address Type. To indicate the address type of Peer. Equal to 4, if the following IP address of peer is belong to IPv4; Equal to 6 if the following IP address of peer is belong to IPv6.

Prefixes Num: 2 Bytes, number of prefixes that advertised by the corresponding Peer. It should be equal to number of the following IP prefix sub TLV.

Peer Associated IP Prefix sub TLV: Variable Length, indicate the advertised IP Prefix.

4.2.1. Prefix sub TLV

Prefix sub TLV is used to carry the prefix information, which has the following format:

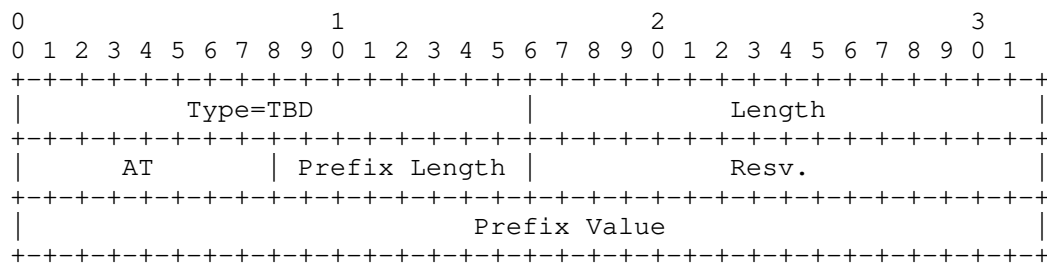


Figure 4: Prefix sub TLV Format

Type: 2 Bytes, value is TBD

Length: 2 Bytes, the length of the following fields.

AT: 1 Byte, Address Type. To indicate the address type of Peer. Equal to 4, if the following "Prefix address" belong to IPv4; Equal to 6 if the following "Prefix address" belong to IPv6.

Prefix Length: 1 Byte, the length of the following prefix. For example, for 10.0.0.0/8, this field will be equal to 8.

Prefix Value: Variable length, the value of the prefix. For example, for 10.0.0./8, this field will be 10.0.0.0

4.3. Explicit Peer Route TLV

The Explicit Peer Route TLV is defined to specify the explicit peer route to the corresponding peer address on each device that is on the end2end assurance path. This TLV should be sent to all the devices that locates on the end2end assurance path that calculated by PCE.

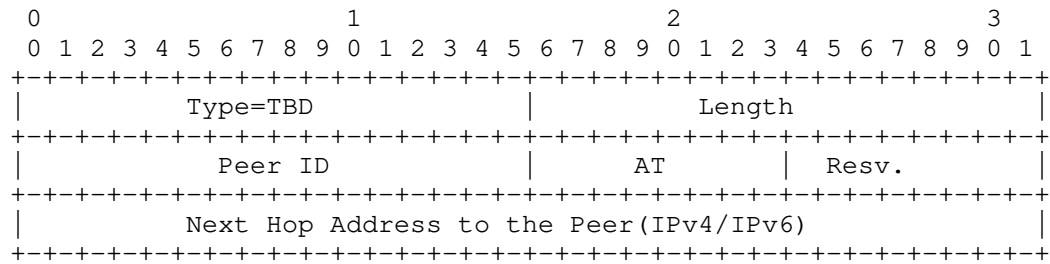


Figure 5: Explicit Peer Route TLV

Type: 2 Bytes, value is TBD

Length: 2 Bytes, the length of following fields.

Peer-ID: 2 Bytes, to indicate the peer that the following next hop address point to. This value is assigned in the Peer Address List object and is referred in this object.

AT: 1 Byte, Address Type. To indicate the address type of explicit peer route. Equal to 4, if the following next hop address to the peer belongs to IPv4; Equal to 6 if the following next hop address to the peer belongs to IPv6.

Resv.: 1 Byte, reservation for future use.

Next Hop Address to the Peer: Variable Length, to indicate the next hop address to the corresponding peer that indicated by the Peer-ID. If AT=4, the length will be 4 bytes, if AT=6, the length will be 16 bytes.

5. Management Consideration

TBD

6. Security Considerations

TBD

7. IANA Considerations

7.1. CCI Object Type

IANA is requested to allocate new registry for the CCI Object Type:

Object-Type Value	CCI Object Name	Reference
3	Native IP	This document

7.2. CCI Object Associated TLV

IANA is requested to confirm the early allocation of the following TLV Type Indicator values within the "PCEP TLV Type Indicator" sub-registry of the PCEP Numbers registry, and to update the reference in the registry to point to this document, when it is an RFC:

Value	Meaning	Reference
TBD	Peer Address List TLV	This document
TBD	Peer Prefix Association TLV	This document
TBD	Explicit Peer Route TLV	This document
TBD	Prefix sub TLV	This document

8. Acknowledgement

Thanks Dhruv Dhody for his valuable suggestions and comments.

9. Normative References

[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-01 (work in progress), February 2019.

[I-D.ietf-teas-native-ip-scenarios]

Wang, A., Huang, X., Qou, C., Li, Z., and P. Mi, "Scenario, Simulation and Suggestion of PCE in Native IP Network", draft-ietf-teas-native-ip-scenarios-02 (work in progress), October 2018.

[I-D.ietf-teas-pce-native-ip]

Wang, A., Zhao, Q., Khasanov, B., Chen, H., and R. Mallia, "PCE in Native IP Network", draft-ietf-teas-pce-native-ip-02 (work in progress), October 2018.

- [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>>.
- [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>>.

Authors' Addresses

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

Email: wangaj.bri@chinatelecom.cn

Boris Khasanov
Huawei Technologies, Co., Ltd
Moskovskiy Prospekt 97A
St. Petersburg 196084
Russia

Email: khasanov.boris@huawei.com

Sudhir Cheruathur
Juniper Networks
1133 Innovation Way
Sunnyvale, California 94089
USA

Email: scheruathur@juniper.net

Chun Zhu
ZTE Corporation
50 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: zhu.chun1@zte.com.cn

Sheng Fang
Huawei Technologies, Co., Ltd
Huawei Bld., No.156 Beiqing Rd.
Beijing
China

Email: fsheng@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: 22 September 2022

A. Wang
China Telecom
B. Khasanov
Yandex LLC
S. Fang
R. Tan
Huawei Technologies, Co., Ltd
C. Zhu
ZTE Corporation
21 March 2022

PCEP Extension for Native IP Network
draft-ietf-pce-pcep-extension-native-ip-18

Abstract

This document defines the Path Computation Element Communication Protocol (PCEP) extension for Central Control Dynamic Routing (CCDR) based application in Native IP network. The scenario and framework of CCDR in native IP is described in [RFC8735] and [RFC8821]. This draft describes the key information that is transferred between Path Computation Element (PCE) and Path Computation Clients (PCC) to accomplish the End to End (E2E) traffic assurance in Native IP network under central control mode.

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 22 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. Conventions used in this document	3
3. Terminology	3
4. Capability Advertisemnt	4
4.1. Open message	4
5. PCEP messages	4
5.1. The PCInitiate message	5
5.2. The PCRpt message	6
6. PCECC Native IP TE Procedures	7
6.1. BGP Session Establishment Procedures	7
6.2. Explicit Route Establish Procedures	9
6.3. BGP Prefix Advertisement Procedures	12
7. New PCEP Objects	14
7.1. CCI Object	14
7.2. BGP Peer Info Object	15
7.3. Explicit Peer Route Object	17
7.4. Peer Prefix Advertisement Object	20
8. End to End Path Protection	21
9. Re-Delegation and Clean up	21
10. BGP Considerations	22
11. New Error-Types and Error-Values Defined	22
12. Deployment Considerations	23
13. Implementation Status	24
13.1. Proof of Concept based on ODL	24
14. Security Considerations	25
15. IANA Considerations	25
15.1. Path Setup Type Registry	25
15.2. PCECC-CAPABILITY sub-TLV's Flag field	25
15.3. PCEP Object Types	25
15.4. PCEP-Error Object	26
16. Contributor	27
17. Acknowledgement	27
18. Normative References	27
Authors' Addresses	29

1. Introduction

Generally, Multiprotocol Label Switching Traffic Engineering (MPLS-TE) requires the corresponding network devices support Multiprotocol Label Switching (MPLS) or Resource ReSerVation Protocol (RSVP)/Label Distribution Protocol (LDP) technologies to assure the End-to-End (E2E) traffic performance. In Segment Routing either IGP extensions or BGP are used to steer a packet through an SR Policy instantiated as an ordered list of instructions called "segments". But in native IP network, there will be no such signaling protocol to synchronize the action among different network devices. It is necessary to use the central control mode that described in [RFC8283] to correlate the forwarding behavior among different network devices. [RFC8821] describes the architecture and solution philosophy for the E2E traffic assurance in Native IP network via Multi Border Gateway Protocol (BGP) solution. This draft describes the corresponding Path Computation Element Communication Protocol (PCEP) extensions to transfer the key information about BGP peer info, peer prefix advertisement and the explicit peer route on on-path routers.

2. Conventions used in this document

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. Terminology

This document uses the following terms defined in [RFC5440]: PCE, PCEP

The following terms are defined in this document:

- * CCDR: Central Control Dynamic Routing
- * E2E: End to End
- * BPI: BGP Peer Info
- * EPR: Explicit Peer Route
- * PPA: Peer Prefix Advertisement
- * QoS: Quality of Service

4. Capability Advertisement

4.1. Open message

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

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

- * PST = TBD1: Path is a Native IP path as per [RFC8821].

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.

[RFC9050] defined the PCECC-CAPABILITY sub-TLV to exchange information about their PCECC capability. A new flag is defined in PCECC-CAPABILITY sub-TLV for Native IP:

N (NATIVE-IP-TE-CAPABILITY - 1 bit - TBD2): If set to 1 by a PCEP speaker, it indicates that the PCEP speaker is capable for TE in Native IP network as specified in this document. The flag MUST be set by both the PCC and PCE in order to support this extension.

If a PCEP speaker receives the PATH-SETUP-TYPE-CAPABILITY TLV with the newly defined path setup type, but without the N bit set in PCECC-CAPABILITY sub-TLV, it MUST:

- * Send a PCErr message with Error-Type=10(Reception of an invalid object) and Error-Value TBD3(PCECC NATIVE-IP-TE-CAPABILITY bit is not set).
- * Terminate the PCEP session

5. PCEP messages

PCECC Native IP TE solution utilizing the existing PCE LSP Initiate Request message(PCInitiate) [RFC8281], and PCE Report message(PCRppt) [RFC8281] to accomplish the multi BGP sessions establishment, E2E TE path deployment, and route prefixes advertisement among different BGP sessions. A new PST for Native-IP is used to indicate the path setup based on TE in Native IP networks.

The extended PCInitiate message described in [RFC9050] is used to download or cleanup central controller's instructions (CCIs). [RFC9050] specifies an object called CCI for the encoding of central controller's instructions. This document specifies a new CCI object-

type for Native IP. The PCEP messages are extended in this document to handle the PCECC operations for Native IP. Three new PCEP Objects (BGP Peer Info (BPI) Object, Explicit Peer Route (EPR) Object and Peer Prefix Advertisement (PPA) Object) are defined in this document. Refer to Section 7 for detail object definitions.

5.1. The PCInitiate message

The PCInitiate Message defined in [RFC8281] and extended in [RFC9050] is further extended to support Native-IP CCI.

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>|
                                           ((<BPI>|<EPR>|<PPA>)
                                            <CCI>))

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

```

Where:

<cci-list> is as per
 [I-D.ietf-pce-pcep-extension-for-pce-controller].
 <PCE-initiated-lsp-instantiation> and
 <PCE-initiated-lsp-deletion> are as per
 [RFC8281].

The LSP and SRP objects are defined in [RFC8231].

When PCInitiate message is used create Native IP instructions, the SRP, LSP and CCI objects MUST be present. The error handling for missing SRP, LSP or CCI object is as per [RFC9050]. Further only one of BPI, EPR, or PPA object MUST be present. The PLSP-ID within the

LSP object should be set by PCC uniquely according to the Symbolic Path Name TLV that included in the CCI object. The Symbolic Path Name is used by the PCE/PCC to identify uniquely the E2E native IP TE path.

If none of them are present, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD4 (Native IP object missing). If there are more than one of BPI, EPR or PPA object are presented, the receiving PCC MUST send a PCErr message with Error-type=19 (Invalid Operation) and Error-value=TBD5 (Only one of the BPI, EPR or PPA object can be included in this message).

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

5.2. The PCRpt message

The PCRpt message is used to acknowledge the Native-IP instructions received from the central controller (PCE).

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>|
                              ((<BPI>|<EPR>|<PPA>)
                               <CCI>))
```

Where:

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

The error handling for missing CCI object is as per [RFC9050]. Further only one of BPI, EPR, or PPA object MUST be present.

If none of them are present, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=TBD4 (Native IP object missing). If there are more than one of BPI, EPR or PPA object are presented, the receiving PCE MUST send a PCErr message with Error-type=19(Invalid Operation) and Error-value=TBD5(Only one of the BPI, EPR or PPA object can be included in this message).

6. PCECC Native IP TE Procedures

The detail procedures for the TE in native IP environment are described in the following sections.

6.1. BGP Session Establishment Procedures

The PCInitiate message can be used to configure the parameters for a BGP peer session using the PCInitiate and PCRpt message pair. This pair of PCE messages is exchanged with a PCE function attached to each BGP peer which needs to be configured. After the BGP peer session has been configured via this pair of PCE messages the BGP session establishment process operates in a normal fashion. All BGP peers are configured for peer to peer communication whether the peers are E-BGP peers or I-BGP peers. One of the IBGP topologies requires that multiple I-BGPs peers operate in a route-reflector I-BGP peer topology. The example below shows two I-BGP route reflector clients interacting with one Route Reflector (RR), but Route Reflector topologies may have up to 100s of clients. Centralized configuration via PCE provides mechanisms to scale auto-configuration of small and large topologies.

The PCInitiate message should be sent to PCC which acts as BGP router and/or route reflector(RR).

The route reflector topology for a single AS is shown in Figure 1. The BGP routers R1, R3, and R7 are within a single AS. R1 and R7 are BGP router-reflector clients, and R3 is a Route Reflector. The PCInitiate message should be sent all of the BGP routers that need to be configured R1 (M3), R3 (M2 & M3), and R7 (M4).

PCInitiate message creates an auto-configuration function for these BGP peers providing the indicated Peer AS and the Local/Peer IP Address.

When PCC receives the BPI and CCI object (with the R bit set to 0 in SRP object) in PCInitiate message, the PCC should try to establish the BGP session with the indicated Peer AS and Local/Peer IP address.

When PCC creates successfully the BGP session that is indicated by the associated information, it should report the result via the PCRpt messages, with BPI object and the corresponding SRP and CCI object included.

When PCC receives this message with the R bit set to 1 in SRP object in PCInitiate message, the PCC should clear the BGP session that indicated by the BPI object.

When PCC clears successfully the specified BGP session, it should report the result via the PCRpt message, with the BPI object included, and the corresponding SRP and CCI object.

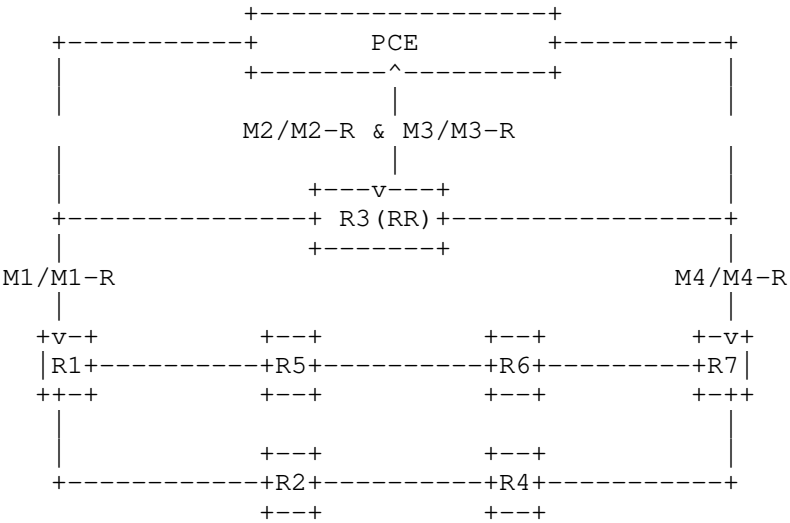


Figure 1: BGP Session Establishment Procedures(R3 act as RR)

The message number, message peers, message type and message key parameters in the above figures are shown in below table:

Table 1: Message Information

No.	Peers	Type	Message Key Parameters
M1 M1-R	PCE/R1	PCInitiate PCRpt	CC-ID=X1 (Symbolic Path Name=Class A) BPI Object (Local_IP=R1_A, Peer_IP=R3_A)
M2 M2-R	PCE/R3	PCInitiate PCRpt	CC-ID=X2 (Symbolic Path Name=Class A) BPI Object (Local_IP=R3_A, Peer_IP=R1_A)
M3 M3-R	PCE/R3	PCInitiate PCRpt	CC-ID=X3 (Symbolic Path Name=Class A) BPI Object (Local_IP=R3_A, Peer_IP=R7_A)
M4 M4-R	PCE/R7	PCInitiate PCRpt	CC-ID=X4 (Symbolic Path Name=Class A) BPI Object (Local_IP=R7_A, Peer_IP=R3_A)

If the PCC cannot establish the BGP session that required by this object, it should report the error values via PCErr message with the newly defined error type (Error-type=TBD6) and error value (Error-value=TBD7, Peer AS not match; or Error-Value=TBD8, Peer IP can't be reached), which is indicated in Section 11

If the Local IP Address or Peer IP Address within BPI object is used in other existing BGP sessions, the PCC should report such error situation via PCErr message with Err-type=TBD6 and error value (Error-value=TBD9, Local IP is in use; Error-value=TBD10, Remote IP is in use).

6.2. Explicit Route Establish Procedures

The explicit route establishment procedures can be used to install a route via PCE in the PCC/BGP Peer, using PCInitiate and PCRpt message pair. Although the BGP policy might redistribute the routes installed by explicit route, the PCE-BGP implementation needs to prohibit the redistribution of the explicit route. PCE explicit routes operate similar to static routes installed by network management protocols (netconf/restconf) but the routes are associated with the PCE routing module. Explicit route installations (like NM static routes) must carefully install and uninstall static routes in an specific order so that the pathways are established without loops.

The PCInitiate message should be sent to the on-path routers respectively. In the example, for explicit route from R1 to R7, the PCInitiate message should be sent to R1 (M1), R2 (M2) and R4 (M3), as shown in Figure 2. For explicit route from R7 to R1, the PCInitiate message should be sent to R7 (M1), R4 (M2) and R2 (M3), as shown in Figure 3.

When PCC receives the EPR and the CCI object (with the R bit set to 0 in SRP object) in PCInitiate message, the PCC should install the explicit route to the the peer.

When PCC install successfully the explicit route to the peer, it should report the result via the PCRpt messages, with EPR object and the corresponding SRP and CCI object included.

When PCC receives the EPR and the CCI object with the R bit set to 1 in SRP object in PCInitiate message, the PCC should clear the explicit route to the peer that indicated by the EPR object.

When PCC clear successfully the explicit route that indicated by this object, it should report the result via the PCRpt message, with the EPR object included, and the corresponding SRP and CCI object.

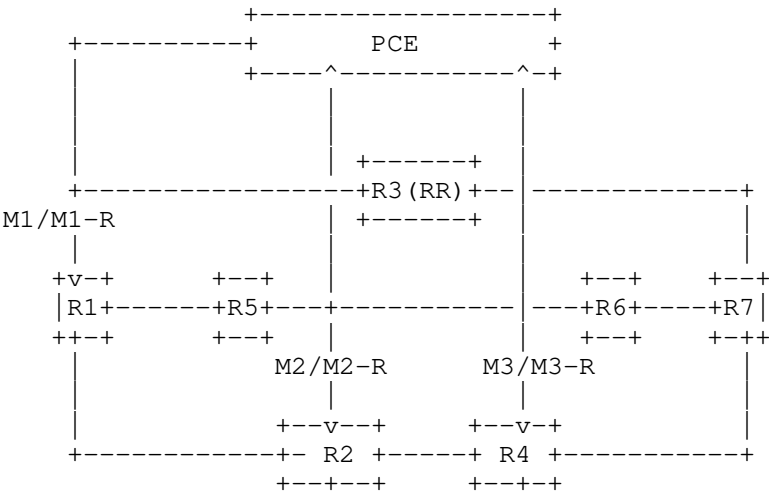


Figure 2: Explicit Route Establish Procedures (From R1 to R7)

The message number, message peers, message type and message key parameters in the above figures are shown in below table:

Table 2: Message Information

No.	Peers	Type	Message Key Parameters
M1 M1-R	PCE/R1	PCInitiate PCRpt	CC-ID=X1 (Symbolic Path Name=Class A) EPR Object (Peer Address=R7_A, Next Hop=R2_A)
M2 M2-R	PCE/R2	PCInitiate PCRpt	CC-ID=X2 (Symbolic Path Name=Class A) EPR Object (Peer Address=R7_A, Next Hop=R4_A)
M3 M3-R	PCE/R4	PCInitiate PCRpt	CC-ID=X3 (Symbolic Path Name=Class A) EPR Object (Peer Address=R7_A, Next Hop=R7_A)

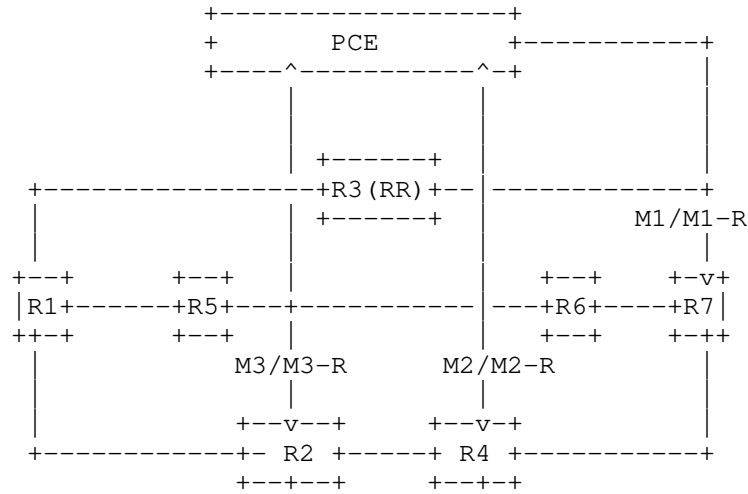


Figure 3: Explicit Route Establish Procedures (From R7 to R1)

The message number, message peers, message type and message key parameters in the above figures are shown in below table:

Table 3: Message Information

No.	Peers	Type	Message Key Parameters
M1 M1-R	PCE/R7	PCInitiate PCRpt	CC-ID=X1 (Symbolic Path Name=Class A) EPR Object (Peer Address=R1_A, Next Hop=R4_A)
M2 M2-R	PCE/R4	PCInitiate PCRpt	CC-ID=X2 (Symbolic Path Name=Class A) EPR Object (Peer Address=R1_A, Next Hop=R2_A)
M3 M3-R	PCE/R2	PCInitiate PCRpt	CC-ID=X3 (Symbolic Path Name=Class A) EPR Object (Peer Address=R1_A, Next Hop=R1_A)

In order to avoid the transient loop during the deploy of explicit peer route, the EPR object should be sent to the PCCs in the reverse order of the E2E path. To remove the explicit peer route, the EPR object should be sent to the PCCs in the same order of E2E path.

To accomplish ECMP effects, the PCE can send multiple EPR objects to the same node, with the same route priority and peer address value but different next hop addresses.

The PCC should verify that the next hop address is reachable. Upon the error occurs, the PCC SHOULD send the corresponding error via PCErr message, with an error information (Error-type=TBD6, Error-value=TBD12, Explicit Peer Route Error) that defined in Section 11.

When the peer info is not the same as the peer info that indicated in BPI object in PCC for the same path that is identified by Symbolic Path Name TLV, an error (Error-type=TBD6, Error-value=17, EPR/BPI Peer Info mismatch) should be reported via the PCErr message.

6.3. BGP Prefix Advertisement Procedures

The detail procedures for BGP prefix advertisement are shown below, using PCInitiate and PCRpt message pair.

The PCInitiate message should be sent to PCC that acts as BGP peer router only. In the example, it should be sent to R1(M1) or R7(M2) respectively.

When PCC receives the PPA and the CCI object (with the R bit set to 0 in SRP object) in PCInitiate message, the PCC should send the prefixes indicated in this object to the appointed BGP peer.

When PCC sends successfully the prefixes to the appointed BGP peer, it should report the result via the PCRpt messages, with PPA object and the corresponding SRP and CCI object included.

When PCC receives the PPA and the CCI object with the R bit set to 1 in SRP object in PCInitiate message, the PCC should withdraw the prefixes advertisement to the peer that indicated by this object.

When PCC withdraws successfully the prefixes that indicated by this object, it should report the result via the PCRpt message, with the PPA object included, and the corresponding SRP and CCI object.

The allowed AFI/SAFI for the IPv4 BGP session should be 1/1(IPv4 prefix) and the allowed AFI/SAFI for the IPv6 BGP session should be 2/1(IPv6 prefix). If mismatch occur, an error(Error-type=TBD6, Error-value=TBD18, BPI/PPR address family mismatch) should be reported via PCErr message.

When the peer info is not the same as the peer info that indicated in BPI object in PCC for the same path that is identified by Symbolic Path Name TLV, an error (Error-type=TBD6, Error-value=TBD19, PPA/BPI peer info mismatch) should be reported via the PCErr message.

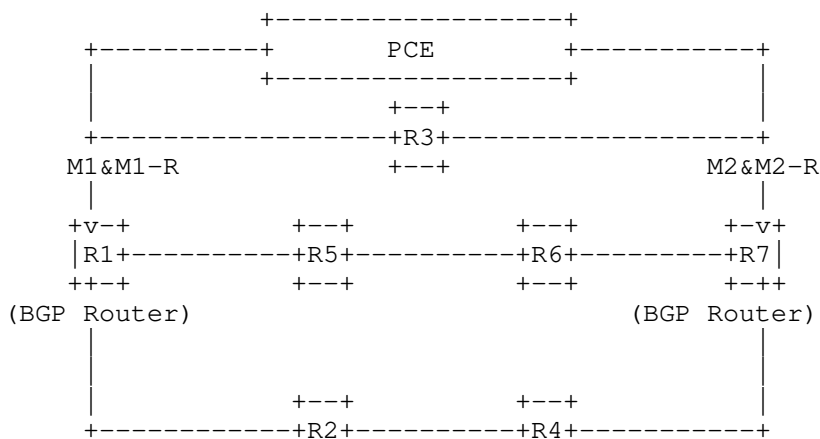


Figure 4: BGP Prefix Advertisement Procedures

Table 4: Message Information

No.	Peers	Type	Message Key Parameters
M1 M1-R	PCE/R1	PCInitiate PCRpt	CC-ID=X1 (Symbolic Path Name=Class A) PPA Object (Peer IP=R7_A, Prefix=1_A)
M2 M2-R	PCE/R7	PCInitiate PCRpt	CC-ID=X2 (Symbolic Path Name=Class A) PPA Object (Peer IP=R1_A, Prefix=7_A)

7. New PCEP Objects

One new CCI Object and three new PCEP objects are defined in this draft. All new PCEP objects are as per [RFC5440]

7.1. CCI Object

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

CCI Object-Type is TBD13 for Native-IP as below

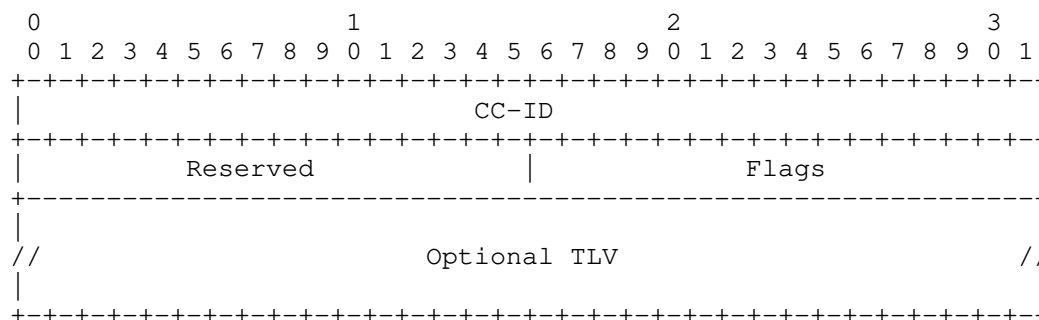


Figure 5: CCI Object for Native IP

Figure 1

The field CC-ID is as described in [RFC9050]. Following fields are defined for CCI Object-Type TBD13

Reserved: is set to zero while sending, ignored on receipt.

Flags: is used to carry any additional information pertaining to the CCI. Currently no flag bits are defined.

The Symbolic Path Name TLV [RFC8231] MUST be included in the CCI Object-Type TBD13 to identify the E2E TE path in Native IP environment and MUST be unique.

7.2. BGP Peer Info Object

The BGP Peer Info object is used to specify the information about the peer that the PCC should establish the BGP relationship with. This object should only be included and sent to the head and end router of the E2E path in case there is no Route Reflection (RR) involved. If the RR is used between the head and end routers, then such information should be sent to head router, RR and end router respectively.

By default, there MUST be no prefix be distributed via such BGP session that established by this object.

By default, the Local/Peer IP address SHOULD be dedicated to the usage of native IP TE solution, and SHOULD NOT be used by other BGP sessions that established by manual or non PCE initiated configuration.

BGP Peer Info Object-Class is TBD14

BGP Peer Info Object-Type is 1 for IPv4 and 2 for IPv6

The format of the BGP Peer Info object body for IPv4 (Object-Type=1) is as follows:

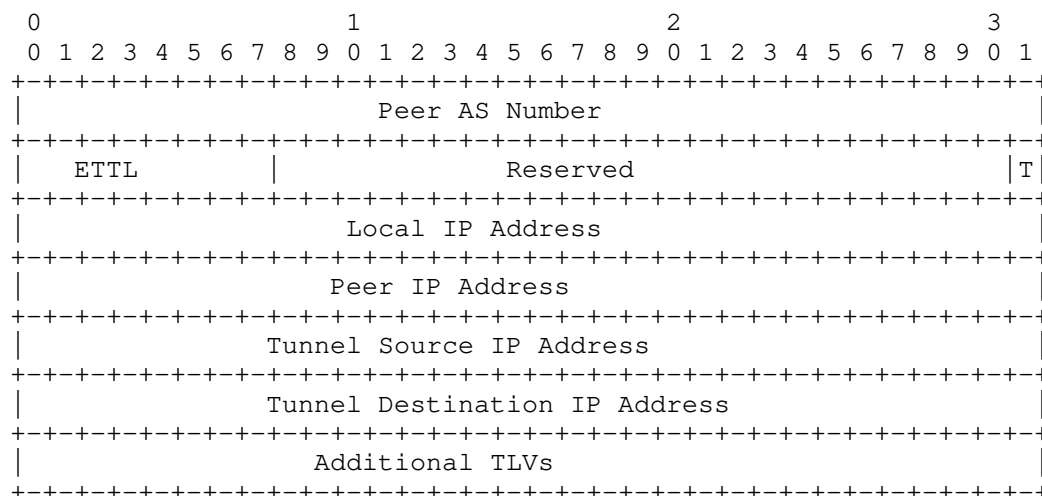


Figure 6: BGP Peer Info Object Body Format for IPv4

The format of the BGP Peer Info object body for IPv6(Object-Type=2) is as follows:

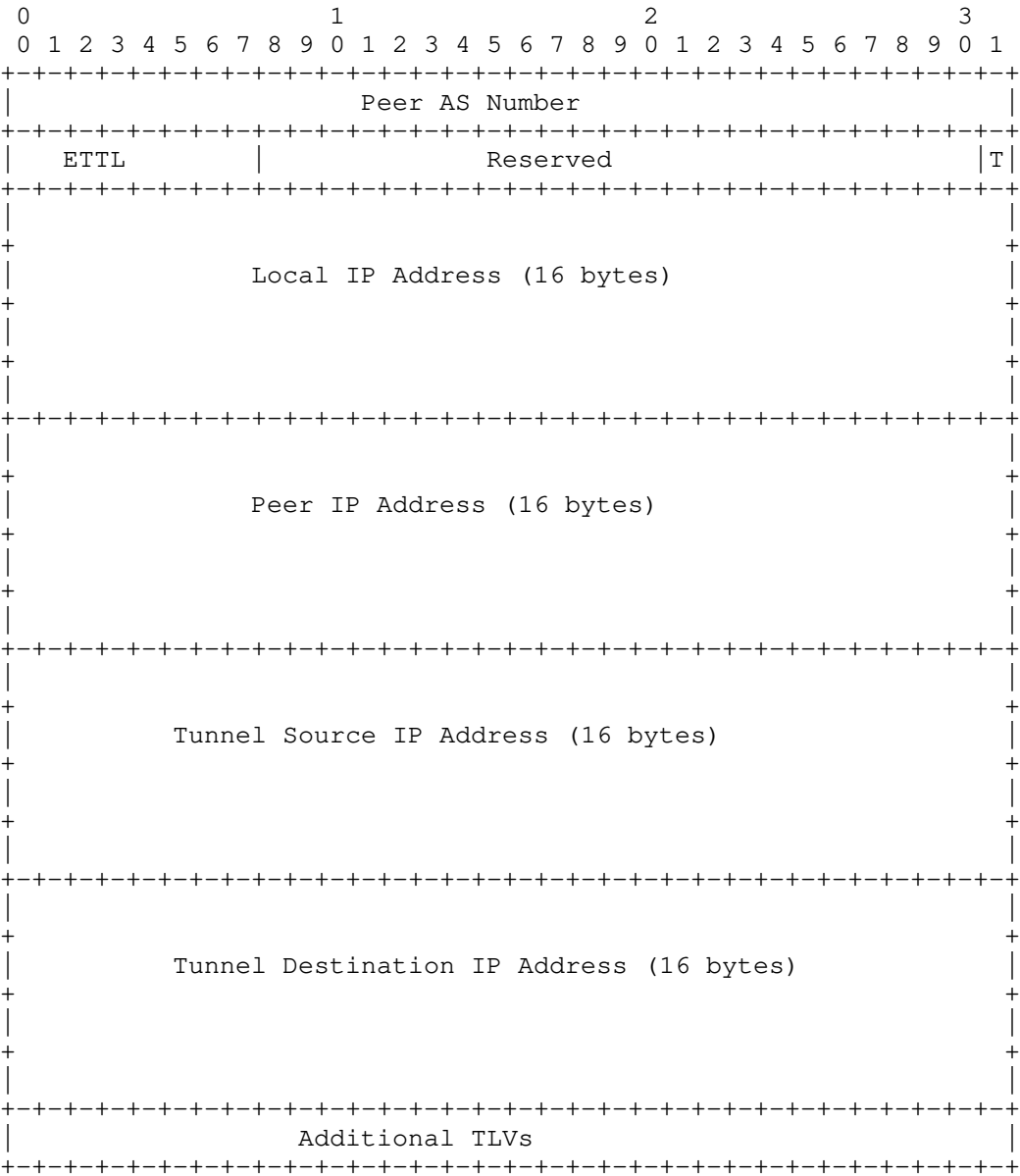


Figure 7: BGP Peer Info Object Body Format for IPv6

Peer AS Number: 4 Bytes, to indicate the AS number of Remote Peer.

ETTL: 1 Byte, to indicate the multihop count for EBGp session. It should be 0 and ignored when Local AS and Peer AS is same.

Reserved: is set to zero while sending, ignored on receipt.

T bit: Indicates whether the traffic that associated with the prefixes advertised via this BGP session is transported via IPinIP tunnel (when T bit is set) or not (when T bit is clear).

Local IP Address(4/16 Bytes): IP address of the local router, used to peer with other end router. When Object-Type is 1, length is 4 bytes; when Object-Type is 2, length is 16 bytes.

Peer IP Address(4/16 Bytes): IP address of the peer router, used to peer with the local router. When Object-Type is 1, length is 4 bytes; when Object-Type is 2, length is 16 bytes;

Tunnel Source IP Address(4/16 Bytes): IP address of the tunnel source, should be owned by the local router. When Object-Type is 1, length is 4 bytes; when Object-Type is 2, length is 16 bytes.

Tunnel Destination IP Address(4/16 Bytes): IP address of the tunnel destination, should be owned by the peer router. When Object-Type is 1, length is 4 bytes; when Object-Type is 2, length is 16 bytes. Should be different from the Peer IP Address.

Additional TLVs: TLVs that associated with this object, can be used to convey other necessary information for dynamic BGP session establishment. Their definition are out of the current document.

When PCC receives BPI object, with Object-Type=1, it should try to establish BGP session with the peer in AFI/SAFI=1/1; when PCC receives BPI object with Object-Type=2, it should try to establish the BGP session with the peer in AFI/SAFI=2/1. Other BGP capabilities, for example, Graceful Restart (GR) that enhance the BGP performance should also be negotiated and used by default.

7.3. Explicit Peer Route Object

The Explicit Peer Route object is defined to specify the explicit peer route to the corresponding peer address on each device that is on the E2E assurance path. This Object should be sent to all the devices that locates on the E2E assurance path that calculated by PCE.

The path established by this object should have higher priority than other path calculated by dynamic IGP protocol, but should be lower priority than the static route configured by manual or NETCONF or by other means.

Explicit Peer Route Object-Class is TBD15.

Explicit Peer Route Object-Type is 1 for IPv4 and 2 for IPv6

The format of Explicit Peer Route object body for IPv4 (Object-Type=1) is as follows:

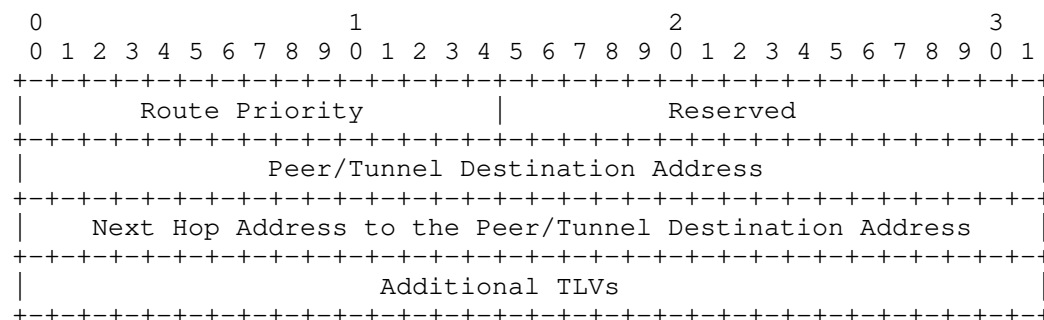


Figure 8: Explicit Peer Route Object Body Format for IPv4

The format of Explicit Peer Route object body for IPv6 (Object-Type=2) is as follows:

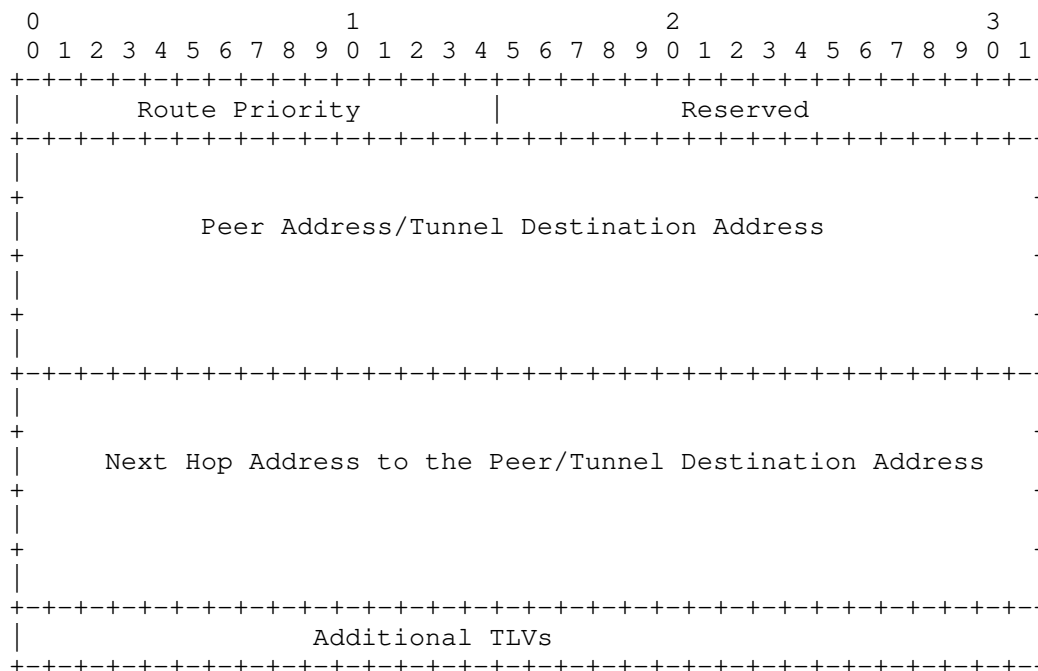


Figure 9: Explicit Peer Route Object Body Format for IPv6

Route Priority: 2 Bytes, The priority of this explicit route. The higher priority should be preferred by the device. This field is used to indicate the backup path at each hop.

Reserved: is set to zero while sending, ignored on receipt.

Peer/Tunnel Destination Address: To indicate the peer address(4/16 Bytes). When T bit is set in the associated BPI object, use the tunnel destination address in BPI object; when T bit is clear, use the peer address in BPI object.

Next Hop Address to the Peer/Tunnel Destination Address: To indicate the next hop address(4/16 Bytes) to the corresponding peer/tunnel destination address.

Additional TLVs: TLVs that associated with this object, can be used to convey other necessary information for explicit peer path establishment. Their definitions are out of the current document.

7.4. Peer Prefix Advertisement Object

The Peer Prefix Advertisement object is defined to specify the IP prefixes that should be advertised to the corresponding peer. This object should only be included and sent to the head/end router of the end2end path.

The prefixes information included in this object MUST only be advertised to the indicated peer, MUST NOT be advertised to other BGP peers.

Peer Prefix Advertisement Object-Class is TBD16

Peer Prefix Advertisement Object-Type is 1 for IPv4 and 2 for IPv6

The format of the Peer Prefix Advertisement object body is as follows:

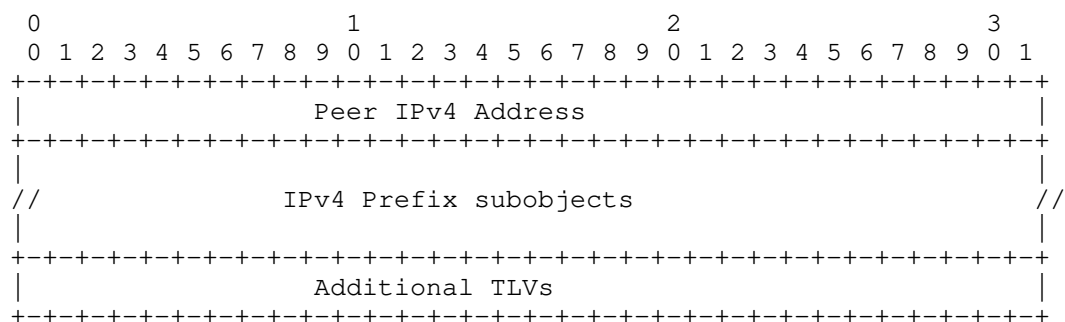


Figure 10: Peer Prefix Advertisement Object Body Format for IPv4

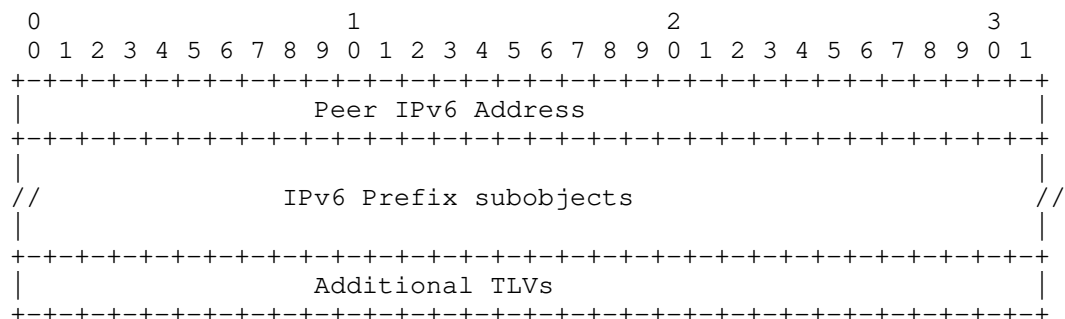


Figure 11: Peer Prefix Advertisement Object Body Format for IPv6

Peer IPv4 Address: 4 Bytes. Identifies the peer IPv4 address that the associated prefixes will be sent to.

IPv4 Prefix subobjects: List of IPv4 Prefix subobjects that defined in [RFC3209], identify the prefixes that will be sent to the peer that identified by Peer IPv4 Address List.

Peer IPv6 Address: 16 Bytes. Identifies the peer IPv6 address that the associated prefixes will be sent to.

IPv6 Prefix subobjects: List of IPv6 Prefix subobjects that defined in [RFC3209], identify the prefixes that will be sent to the peer that identified by Peer IPv6 Address List.

Additional TLVs: TLVs that associated with this object, can be used to convey other necessary information for prefixes advertisement. Their definitions are out of the current document.

8. End to End Path Protection

[RFC8697] defines the path associations procedures between sets of Label Switched Path (LSP). Such procedures can also be used for the E2E path protection. To accomplish this, the PCE should attach the ASSOCIATION object with the EPR object in the PCInitiate message, with the association type set to 1 (Path Protection Association). The Extended Association ID that included within the Extended Association ID TLV, which is included in the ASSOCIATION object, should be set to the Symbolic Path Name of different E2E path. This PCInitiate should be sent to the head-end of the E2E path.

The head-end of the path can use the existing path detection mechanism(for example, Bidirectional Forwarding Detection [RFC5880]), to monitor the status of the active path. Once it detects the failure, it can switch the backup protection path immediately.

9. Re-Delegation and Clean up

In case of a PCE failure, a new PCE can gain control over the central controller instructions. As per the PCEP procedures in [RFC8281], the State Timeout Interval timer is used to ensure that a PCE failure does not result in automatic and immediate disruption for the services. Similarly, as per [RFC9050], the central controller instructions are not removed immediately upon PCE failure. Instead, they could be re-delegated to the new PCE before the expiration of this timer, or be cleaned up on the expiration of this timer. The 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.

10. BGP Considerations

This draft defines the procedures and objects to create the BGP sessions and advertises the associated prefixes dynamically. Only the key information, for example peer IP addresses, peer AS number are exchanged via the PCEP protocol. Other parameters that are needed for the BGP session setup should be derived from their default values, as described in Section 7.2. Upon receives such key information, the BGP module on the PCC should try to accomplish the task that appointed by the PCEP protocol and report the status to the PCEP modules.

There is no influence to current implementation of BGP Finite State Machine(FSM). The PCEP cares only the success and failure status of BGP session, and act upon such information accordingly.

The error handling procedures related to incorrect BGP parameters are specified in Section 6.1, Section 6.2, and Section 6.3. The handling of the dynamic BGP sessions and associated prefixes on PCE failure is described in Section 9.

11. New Error-Types and Error-Values Defined

A PCEP-ERROR object is used to report a PCEP error and is characterized by an Error-Type that specifies that type of error and an Error-value that provides additional information about the error. An additional Error-Type and several Error-values are defined to represent some the errors related to the newly defined objects, which are related to Native IP TE procedures.

Error-Type	Meaning	Error-value
TBD6	Native IP TE failure	
		0: Unassigned
		TBD7: Peer AS not match
		TBD8:Peer IP can't be reached
		TBD9:Local IP is in use
		TBD10:Remote IP is in use
		TBD11:Exist BGP session broken
		TBD12:Explicit Peer Route Error
		TBD17:EPR/BPI Peer Info mismatch
		TBD18:BPI/PPA Address Family mismatch
		TBD19:PPA/BPI Peer Info mismatch

Figure 12: Newly defined Error-Type and Error-Value

12. Deployment Considerations

The information transferred in this draft is mainly used for the light weight BGP session setup, explicit route deployment and the prefix distribution. The planning, allocation and distribution of the peer addresses within IGP should be accomplished in advanced and they are out of the scope of this draft.

[RFC8232] describes the state synchronization procedure between stateful PCE and PCC. The communication of PCE and PCC described in this draft should also follow this procedures, treat the three newly defined objects that associated with the same symbolic path name as the attribute of the same path in the LSP-DB.

When PCE detects one or some of the PCCs are out of control, it should recompute and redeploy the traffic engineering path for native IP on the active PCCs. When PCC detects that it is out of control of the PCE, it should clear the information that initiated by the PCE. The PCE should assure the avoidance of possible transient loop in such node failure when it deploy the explicit peer route on the PCCs.

If the established BGP session is broken after some time, the PCC should also report such error via PCErr message with Err-type=TBD6 and error value(Error-value=TBD11, Existing BGP session is broken). Upon receiving such PCErr message, the PCE should clear the prefixes advertisement on the previous BGP session, clear the explicit peer route to the previous peer address; select other Local_IP/Peer_IP pair to establish the new BGP session, deploy the explicit peer route to the new peer address, and advertises the prefixes on the new BGP session.

13. 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".

13.1. Proof of Concept based on ODL

.At the time of posting the -18 version of this document, there are no known implementations of this mechanism. A proof of concept for the overall design has been verified using another SBI protocol on the Open DayLight (ODL) controller.

14. Security Considerations

The setup of BGP sessions, prefix advertisement, and explicit peer route establishment are all controlled by the PCE. See [RFC4271] and [RFC4272] for BGP security considerations. Security consideration part in [RFC5440] and [RFC8231] should be considered. To prevent a bogus PCE sending harmful messages to the network nodes, the network devices should authenticate the validity of the PCE and ensure a secure communication channel between them. Mechanisms described in [RFC8253] should be used.

15. IANA Considerations

15.1. 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	Native IP TE Path	This document

15.2. PCECC-CAPABILITY sub-TLV's Flag field

[RFC9050] created a sub-registry within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the value of the PCECC-CAPABILITY sub-TLV's 32-bits Flag field. IANA is requested to allocate a new bit position within this registry, as follows:

Value	Description	Reference
TBD2(N)	NATIVE-IP-TE-CAPABILITY	This document

15.3. PCEP Object Types

IANA is requested to allocate new registry for the PCEP Object Type:

Object-Class Value	Name	Reference
44	CCI Object Object-Type TBD13: Native IP	This document
TBD14	BGP Peer Info Object-Type 1: IPv4 address 2: IPv6 address	This document
TBD15	Explicit Peer Route Object-Type 1: IPv4 address 2: IPv6 address	This document
TBD16	Peer Prefix Advertisement Object-Type 1: IPv4 address 2: IPv6 address	This document

15.4. 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	Error-value
		Reference
6	Mandatory Object missing	TBD4:Native IP object missing This document
10	Reception of an invalid object	TBD3:PCECC NATIVE-IP-TE-CAPABILITY bit is not set This document
19	Invalid Operation	TBD5:Only one of the BPI,EPR or PPA object can be included in this message This document
TBD6	Native IP TE failure	This document TBD7:Peer AS not match TBD8:Peer IP can't be reached TBD9:Local IP is in use TBD10:Remote IP is in use TBD11:Exist BGP session broken TBD12:Explicit Peer Route Error TBD17:EPR/BPI Peer Info mismatch TBD18:BPI/PPA Address Family mismatch TBD19:PPA/BPI Peer Info mismatch

16. Contributor

Dhruv Dhody has contributed the contents of this draft.

17. Acknowledgement

Thanks Mike Koldychev, Susan Hares, Siva Sivabalan, Adam Simpson for his valuable suggestions and comments.

18. 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>>.
- [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>>.
- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.
- [RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, DOI 10.17487/RFC4272, January 2006, <<https://www.rfc-editor.org/info/rfc4272>>.
- [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>>.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.
- [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>>.

- [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>>.
- [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>>.
- [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>>.
- [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>>.
- [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>>.
- [RFC8735] Wang, A., Huang, X., Kou, C., Li, Z., and P. Mi, "Scenarios and Simulation Results of PCE in a Native IP Network", RFC 8735, DOI 10.17487/RFC8735, February 2020, <<https://www.rfc-editor.org/info/rfc8735>>.
- [RFC8821] Wang, A., Khasanov, B., Zhao, Q., and H. Chen, "PCE-Based Traffic Engineering (TE) in Native IP Networks", RFC 8821, DOI 10.17487/RFC8821, April 2021, <<https://www.rfc-editor.org/info/rfc8821>>.

[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>>.

Authors' Addresses

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing
Beijing, 102209
China
Email: wangaj3@chinatelecom.cn

Boris Khasanov
Yandex LLC
Ulitsa Lva Tolstogo 16
Moscow
Email: bhassanov@yahoo.com

Sheng Fang
Huawei Technologies, Co., Ltd
Huawei Bld., No.156 Beiqing Rd.
Beijing
China
Email: fsheng@huawei.com

Ren Tan
Huawei Technologies, Co., Ltd
Huawei Bld., No.156 Beiqing Rd.
Beijing
China
Email: tanren@huawei.com

Chun Zhu
ZTE Corporation
50 Software Avenue, Yuhua District
Nanjing
Jiangsu, 210012
China
Email: zhu.chun1@zte.com.cn

PCE Working Group
Internet-Draft
Intended status: Standards Track

Young Lee (Editor)
Fatai Zhang
Huawei
Ramon Casellas
CTTC
Oscar Gonzalez de Dios
Telefonica I+D
Zafar Ali
Cisco Systems

March 7, 2019

Path Computation Element (PCE) Protocol Extensions for Stateful PCE
Usage in GMPLS-controlled Networks

draft-ietf-pce-pcep-stateful-pce-gmpls-10

Abstract

The Path Computation Element (PCE) facilitates Traffic Engineering (TE) based path calculation in large, multi-domain, multi-region, or multi-layer networks. The PCE communication Protocol (PCEP) has been extended to support stateful PCE functions where the PCE retains information about the paths already present in the network, but those extensions are technology-agnostic. This memo provides extensions required for PCEP so as to enable the usage of a stateful PCE capability in GMPLS-controlled networks.

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 September 7, 2019.

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

Table of Contents.....	2
1. Introduction.....	3
2. Conventions used in this document.....	3
3. Context of Stateful PCE and PCEP for GMPLS.....	4
4. Main Requirements.....	4
5. PCEP Extensions.....	5
5.1. LSP Update in GMPLS-controlled Networks.....	5
5.2. LSP Synchronization in GMPLS-controlled Networks.....	5
5.3. Modification of Existing PCEP Messages and Procedures.....	7
5.3.1. Modification for LSP Re-optimization.....	7
5.3.2. Modification for Route Exclusion.....	8
5.3.3. Modification for SRP Object to indicate Bi-directional LSP.....	9
5.4. Object Encoding.....	9
6. IANA Considerations.....	9
6.1. New PCEP Error Codes.....	9
6.2. New Subobject for the Exclude Route Object.....	10
6.3. New "B" Flag in the SRP Object.....	10
7. Manageability Considerations.....	10
7.1. Requirements on Other Protocols and Functional Components	10

8. Security Considerations.....	11
9. Acknowledgement.....	11
10. References.....	11
10.1. Normative References.....	11
10.2. Informative References.....	12
11. Contributors' Address.....	12
Authors' Addresses.....	13

1. Introduction

[RFC4655] presents the architecture of a Path Computation Element (PCE)-based model for computing Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs). To perform such a constrained computation, a PCE stores the network topology (i.e., TE links and nodes) and resource information (i.e., TE attributes) in its TE Database (TED). Such a PCE is usually referred as a stateless PCE. To request path computation services to a PCE, [RFC5440] defines the PCE communication Protocol (PCEP) for interaction between a Path Computation Client (PCC) and a PCE, or between two PCEs. PCEP as specified in [RFC 5440] mainly focuses on MPLS networks and the PCEP extensions needed for GMPLS-controlled networks are provided in [PCEP-GMPLS].

Stateful PCEs are shown to be helpful in many application scenarios, in both MPLS and GMPLS networks, as illustrated in [RFC8051]. Further discussion of concept of a stateful PCE can be found in [RFC7399]. In order for these applications to able to exploit the capability of stateful PCEs, extensions to PCEP are required.

[RFC8051] describes how a stateful PCE can be applicable to solve various problems for MPLS-TE and GMPLS networks and the benefits it brings to such deployments.

[RFC8231] provides the fundamental extensions needed for stateful PCE to support general functionality, but leaves out the specification for technology-specific objects/TLVs. This document focuses on the extensions that are necessary in order for the deployment of stateful PCEs in GMPLS-controlled networks.

2. Conventions used in this document

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. Context of Stateful PCE and PCEP for GMPLS

This document is built on the basis of Stateful PCE [RFC8231] and PCEP for GMPLS [PCEP-GMPLS].

There are two types of LSP operation for Stateful PCE.

For Active Stateful PCE, PCUpd message is sent from PCE to PCC to update the LSP state for the LSP delegated to PCE. Any changes to the delegated LSPs generate a PCRpt message by the PCC to PCE to convey the changes of the LSP. Any modifications to the Objects/TLVs that are identified in this document to support GMPLS technology-specific attributes will be carried in the PCRpt and PCUpd messages.

For Passive Stateful PCEs, PCReq/PCRep messages are used to convey path computation instructions. GMPLS-technology specific Objects and TLVs are defined in [PCEP-GMPLS], so this document just points at that work and only adds the stateful PCE aspects where applicable. Passive Stateful PCE makes use of PCRpt messages when reporting LSP State changes sent by PCC to PCEs. Any modifications to the Objects/TLVs that are identified in this document to support GMPLS technology-specific attributes will be carried in the PCRpt message.

[PCEP-GMPLS] defines GMPLS-technology specific Objects/TLVs and this document makes use of these Objects/TLVs without modifications where applicable. Some of these Objects/TLVs may require modifications to incorporate stateful PCE element where applicable.

4. Main Requirements

This section notes the main functional requirements for PCEP extensions to support stateful PCE for use in GMPLS-controlled networks, based on the description in [RFC8051]. Many requirements are common across a variety of network types (e.g., MPLS-TE networks and GMPLS networks) and the protocol extensions to meet the requirements are already described in [RFC8231]. This document does not repeat the description of those protocol extensions. This document presents protocol extensions for a set of requirements which are specific to the use of a stateful PCE in a GMPLS-controlled network.

The basic requirements are as follows:

- o Advertisement of the stateful PCE capability. This generic requirement is covered in Section 5.4. of [RFC8231]. This

document assumes that STATEFUL-PCE-CAPABILITY TLV can be used for GMPLS Stateful PCE capability and therefore does not provide any further extensions.

- o LSP delegation is already covered in Section 5.7. of [RFC8231]. Section 2.2. of this document does not provide any further extensions.
- o Active LSP update is covered in Section 6.2 of [RFC8231]. Section 4.1. of this document provides extension for its application in GMPLS-controlled networks.
- o LSP state synchronization and LSP state report. This is a generic requirement already covered in Section 5.6. of [RFC8231]. However, there are further extensions required specifically for GMPLS-controlled networks and discussed in Section 4.2.

5. PCEP Extensions

5.1. LSP Update in GMPLS-controlled Networks

[RFC8231] defines the Path Computation LSP Update Request (PCUpd) message to enable to update the attributes of an LSP. However, that document does not define technology-specific parameters.

A key element of the PCUpd message is the attribute-list construct defined in [RFC5440] and extended by many other PCEP specifications.

For GMPLS purposes we note that the BANDWIDTH object used in the attribute-list is defined in [PCEP-GMPLS]. Furthermore, additional TLVs are defined for the LSPA object in [PCEP-GMPLS] and MAY be included to indicate technology-specific attributes. There are other technology-specific attributes that need to be conveyed in the <intended-attribute-list> of the <path> construct in the PCUpd message. Note that these path details in the PCUpd message are the same as the <attribute-list> of the PCRep message. See Section 4.2 for the details.

5.2. LSP Synchronization in GMPLS-controlled Networks

PCCs need to report the attributes of LSPs to the PCE to enable stateful operation of a GMPLS network. This process is known as LSP state synchronization. The LSP attributes include bandwidth, associated route, and protection information etc., are stored by the PCE in the LSP database (LSP-DB). Note that, as described in [RFC8231], the LSP state synchronization covers both the bulk reporting of LSPs at initialization as well the reporting of new or modified LSP during normal operation. Incremental LSP-DB

synchronization may be desired in a GMPLS-controlled network and it is specified in [RFC8232].

[RFC8231] describes mechanisms for LSP synchronization using the Path Computation State Report (PCRpt) message, but does not cover reporting of technology-specific attributes. As stated in [RFC8231], the <path> construct is further composed of a compulsory Explicit Route Object (ERO) and a compulsory attribute-list and an optional Record Route Object (RRO). In order to report LSP states in GMPLS networks, this specification allows the use within a PCRpt message both of technology- and GMPLS-specific attribute objects and TLVs defined in [PCEP-GMPLS] as follows:

- o Include Route Object (IRO)/ Exclude Route Object (XRO) Extensions to support the inclusion/exclusion of labels and label sub-objects for GMPLS. (See Section 2.6 and 2.7 in [PCEP-GMPLS])
- o END-POINTS (Generalized END-POINTS Object Type. See Section 2.5 in [PCEP-GMPLS])
- o BANDWIDTH (Generalized BANDWIDTH Object Type. See Section 2.3 in [PCEP-GMPLS])
- o LSPA (PROTECTION ATTRIBUTE TLV, See Section 2.8 in [PCEP-GMPLS]).

The END-POINTS object SHOULD be carried within the attribute-list to specify the endpoints pertaining to the reported LSP. The XRO object MAY be carried to specify the network resources that the reported LSP avoids and a PCE SHOULD consider avoid these network resources during the process of re-optimizing after this LSP is delegated to the PCE. To be more specific, the <attribute-list> is updated as follows using the notations of [RFC5511]:

```
<attribute-list> ::= [<END-POINTS>]
                        [<LSPA>]
                        [<BANDWIDTH>]
                        [<metric-list>]
                        [<IRO>]
                        [<XRO>]

<metric-list> ::= <METRIC> [<metric-list>]
```

If the LSP being reported protects another LSP, the PROTECTION-ATTRIBUTE TLV [PCEP-GMPLS] MUST be included in the LSPA object to

describe its attributes and restrictions. Moreover, if the status of the protecting LSP changes from non-operational to operational, the PCC SHOULD synchronize the state change of the LSPs to the stateful PCE using a PCRpt message. This use case arises, for example, when the protecting LSP becomes operational due to the failure of the primary LSP.

5.3. Modification of Existing PCEP Messages and Procedures

One of the advantages mentioned in [RFC8051] is that the stateful nature of a PCE simplifies the information conveyed in PCEP messages, notably between PCC and PCE, since it is possible to refer to PCE managed state for active LSPs. To be more specific, with a stateful PCE, it is possible to refer to an LSP with a unique identifier in the scope of the PCC-PCE session and thus use such identifier to refer to that LSP. Note this is also applicable to packet networks.

5.3.1. Modification for LSP Re-optimization

The Request Parameters (RP) object on a Path Computation Request (PCReq) message carries the R bit. When set, this indicates that the PCC is requesting re-optimization of an existing LSP. Upon receiving such a PCReq, a stateful PCE SHOULD perform the re-optimization in the following cases:

- o The existing bandwidth and route information of the LSP to be re-optimized is provided in the PCReq message using the BANDWIDTH object and the ERO.
- o The existing bandwidth and route information is not supplied in the PCReq message, but can be found in the PCE's LSP-DB. In this case, the LSP MUST be identified using an LSP identifier carried in the PCReq message, and that fact requires that the LSP identifier was previously supplied either by the PCC in a PCRpt message or by the PCE in a PCRep message. [RFC8231] defines how this is achieved using a combination of the per-node LSP identifier (PLSP-ID) and the PCC's address.

If no LSP state information is available to carry out re-optimization, the stateful PCE should report the error "LSP state information unavailable for the LSP re-optimization" (Error Type = TBD1, Error value= TBD2).

5.3.2. Modification for Route Exclusion

[RFC5521] defines a mechanism for a PCC to request or demand that specific nodes, links, or other network resources are excluded from paths computed by a PCE. A PCC may wish to request the computation of a path that avoids all link and nodes traversed by some other LSP.

To this end this document defines a new sub-object for use with route exclusion defined in [RFC5521]. The LSP exclusion sub-object is as follows:

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																												
X Type (TBD3)										Length										Attributes										Flag																													
//										Symbolic Path Name																																								//									

X bit and Attribute fields are defined in [RFC5521].

Type: Subobject Type for an LSP exclusion sub-object. Value of TBD3. To be assigned by IANA.

Length: The Length contains the total length of the subobject in bytes, including the Type and Length fields.

Flags: This field may be used to further specify the exclusion constraint with regard to the LSP. Currently, no values are defined.

Symbolic Path Name: This is the identifier given to an LSP and is unique in the context of the PCC address as defined in [RFC8231].

Reserved: MUST be transmitted as zero and SHOULD be ignored on receipt.

This sub-object is OPTIONAL in the exclude route object (XRO) and can be present multiple times. When a stateful PCE receives a PCReq message carrying this sub-object, it SHOULD search for the identified LSP in its LSP-DB and then exclude from the new path computation all resources used by the identified LSP. If the stateful PCE cannot recognize one or more of the received LSP identifiers, it should send an error message PCErr reporting "The LSP state information for route exclusion purpose cannot be found"

(Error-type = TBD1, Error-value = TBD4). Optionally, it may provide with the unrecognized identifier information to the requesting PCC using the error reporting techniques described in [RFC5440].

5.3.3. Modification for SRP Object to indicate Bi-directional LSP

The format of the SRP object is defined in [RFC8231]. The object is used in PCUpd and PCInit messages for GMPLS.

This document defines a new flag to be carried in the Flags field of the SRP object. This flag indicates a bidirectional co-routed LSP setup operation initiated by the PCE as follows:

- o B (Bidirectional LSP -- 1 bit): If set to 0, it indicates a request to create a uni-directional LSP. If set to 1, it indicates a request to create a bidirectional co-routed LSP.

The bit position is TBD5 as assigned by IANA (see Section 5.3)

5.4. Object Encoding

Note that, as is stated in Section 7 of [RFC8231], the P flag and the I flag of the PCEP objects used on PCUpd and PCRpt messages SHOULD be set to 0 on transmission and SHOULD be ignored on receipt since these flags are exclusively related to path computation requests.

6. IANA Considerations

6.1. New PCEP Error Codes

IANA is requested to make the following allocation in the "PCEP-ERROR Object Error Types and Values" registry.

Error Type	Meaning	Reference
TBD1	LSP state information missing	[This.I-D]
Error-value TBD2:	LSP state information unavailable for the LSP re-optimization	[This.I-D]
Error-value TBD4:	LSP state information for route exclusion purpose cannot be found	[This.I-D]

6.2. New Subobject for the Exclude Route Object

IANA maintains the "PCEP Parameters" registry containing a subregistry called "PCEP Objects". This registry has a subregistry for the XRO (Exclude Route Object) listing the sub-objects that can be carried in the XRO. IANA is requested to assign a further sub-object that can be carried in the XRO as follows:

Value	Description	Reference
-----+-----+-----		
TBD3	LSP identifier sub-object	[This.I-D]

6.3. New "B" Flag in the SRP Object

IANA maintains a subregistry, named the "SRP Object Flag Field", within the "Path Computation Element Protocol (PCEP) Numbers" registry, to manage the Flag field of the SRP object.

IANA is requested to make an assignment from this registry as follows:

Bit	Description	Reference
---	-----	-----
TDB5	Bi-directional co-routed LSP	[This.I-D]

7. Manageability Considerations

The description and functionality specifications presented related to stateful PCEs should also comply with the manageability specifications covered in Section 8 of [RFC4655]. Furthermore, a further list of manageability issues presented in [RFC8231] should also be considered.

Additional considerations are presented in the next section.

7.1. Requirements on Other Protocols and Functional Components

When the detailed route information is included for LSP state synchronization (either at the initial stage or during LSP state report process), this requires the ingress node of an LSP carry the RRO object in order to enable the collection of such information.

8. Security Considerations

This draft provides additional extensions to PCEP so as to facilitate stateful PCE usage in GMPLS-controlled networks, on top of [RFC8231]. The PCEP extensions to support GMPLS-controlled networks should be considered under the same security as for MPLS networks, as noted in [RFC7025]. Therefore, the security considerations elaborated in [RFC5440] still apply to this draft. Furthermore, [RFC8231] provides a detailed analysis of the additional security issues incurred due to the new extensions and possible solutions needed to support for the new stateful PCE capabilities and they apply to this document as well.

9. Acknowledgement

We would like to thank Adrian Farrel and Cyril Margaria for the useful comments and discussions.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.
- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5440] Vasseur, J.-P., and Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC8174] B. Leiba, "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", RFC 8174, May 2017.
- [RFC8231] Crabbe, E., Medved, J., Varga, R., Minei, I., "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, September 2017.
- [PCEP-GMPLS] Margaria, C., Gonzalez de Dios, O., Zhang, F., "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions, work in progress.

10.2. Informative References

- [RFC5511] A. Farrel, "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.
- [RFC8051] Zhang, X., Minei, I., et al, "Applicability of Stateful Path Computation Element (PCE) ", RFC 8051, January 2017.
- [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, September 2017.

11. Contributors' Address

Xian Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972645
Email: zhang.xian@huawei.com

Dhruv Dhody
Huawei Technology
India

Email: dhruv.ietf@gmail.com

Yi Lin
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972914
Email: yi.lin@huawei.com

Authors' Addresses

Young Lee (Editor)
Huawei
5340 Legacy Drive, Suite 170
Plano, TX 75023
US

Phone: +1 469 278 5838
EMail: leeyoung@huawei.com

Fatai Zhang
Huawei
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
P.R. China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Phone:
Email: ramon.casellas@cttc.es

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Zafar Ali
Cisco Systems
Email: zali@cisco.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 10, 2022

Y. Lee
Samsung
H. Zheng
Huawei Technologies
O. G. de Dios
Telefonica
Victor Lopez
Nokia
Z. Ali
Cisco Systems
February 10, 2022

Path Computation Element (PCE) Protocol Extensions for Stateful PCE
Usage in GMPLS-controlled Networks

draft-ietf-pce-pcep-stateful-pce-gmpls-17

Abstract

The Path Computation Element (PCE) facilitates Traffic Engineering (TE) based path calculation in large, multi-domain, multi-region, or multi-layer networks. The PCE communication Protocol (PCEP) has been extended to support stateful PCE functions where the PCE retains information about the paths already present in the network, but those extensions are technology-agnostic. This memo provides extensions required for PCEP so as to enable the usage of a stateful PCE capability in GMPLS-controlled networks.

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. 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."

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 August 10, 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 (<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

Table of Contents	2
1. Introduction	3
2. Conventions used in this document	4
3. General Context of Stateful PCE and PCEP for GMPLS	4
4. Main Requirements	5
5. Overview of Stateful PCEP Extensions for GMPLS Networks	6
5.1. Capability Advertisement for Stateful PCEP in GMPLS	6
5.2. LSP Synchronization	6
5.3. LSP Delegation and Cleanup	7
5.4. LSP Operations	7
6. Extension of Existing PCEP Messages	7
6.1. The PCRpt Message	7
6.2. The PCUpd Message	9
6.3. The PCInitiate Message	9
7. PCEP Object Extensions	11
7.1. Existing Extensions used for Stateful GMPLS	11
7.2. New Extensions	11
7.2.1. OPEN Object Extension GMPLS-CAPABILITY TLV	11
7.2.2. New LSP Exclusion Sub-object in the XRO	12
7.2.3. SRP Extension	13

8. Update to Error Handling	13
8.1. Error Handling in LSP Re-optimization	13
8.2. Error Handling in Route Exclusion	13
8.3. Error Handling for generalized END-POINTS	14
9. Implementation	14
9.1. Huawei Technologies	14
10. IANA Considerations.....	15
10.1. New GMPLS-CAPABILITY	15
10.2. New Sub-object for the Exclude Route Object	15
10.3. Flag Field for new XRO Sub-object	15
10.4. New "B" Flag in the SRP Object	16
10.5. New PCEP Error Codes	16
11. Manageability Considerations	16
11.1. Requirements on Other Protocols	17
12. Security Considerations	17
13. Acknowledgement	17
14. References	17
14.1. Normative References	17
14.2. Informative References	18
15. Contributors' Address	19
Authors' Addresses	21

1. Introduction

[RFC4655] presents the architecture of a Path Computation Element (PCE)-based model for computing Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs). To perform such a constrained computation, a PCE stores the network topology (i.e., TE links and nodes) and resource information (i.e., TE attributes) in its TE Database (TED). Such a PCE is usually referred as a stateless PCE. To request path computation services to a PCE, [RFC5440] defines the PCE communication Protocol (PCEP) for interaction between a Path Computation Client (PCC) and a PCE, or between two PCEs. PCEP as specified in [RFC5440] mainly focuses on MPLS networks and the PCEP extensions needed for GMPLS-controlled networks are provided in [RFC8779].

Stateful PCEs are shown to be helpful in many application scenarios, in both MPLS and GMPLS networks, as illustrated in [RFC8051]. Further discussion of concept of a stateful PCE can be found in [RFC7399]. In order for these applications to able to exploit the capability of stateful PCEs, extensions to PCEP are required.

[RFC8051] describes how a stateful PCE can be applicable to solve various problems for MPLS-TE and GMPLS networks and the benefits it brings to such deployments.

[RFC8231] provides the fundamental extensions needed for stateful PCE to support general functionality. Furthermore, [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. However, both the documents left out the specification for technology-specific objects/TLVs, and do not cover the GMPLS networks (e.g., WSON, OTN, SONET/ SDH, etc. technologies).

This document focuses on the extensions that are necessary in order for the deployment of stateful PCEs and the requirements for remote-initiated LSPs in GMPLS-controlled networks. Section 3 provides General context of Stateful PCE and PCEP for GMPLS are provided in Section 3, and PCE initiation requirement for GMPLS is provided in section 4. Protocol extensions are included in section 5, as a solution to address such requirements.

2. Conventions used in this document

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. General Context of Stateful PCE and PCEP for GMPLS

This section is built on the basis of Stateful PCE in [RFC8231] and PCEP for GMPLS in [RFC8779].

The operation for Stateful PCE on LSPs can be divided into two types, active stateful PCE and passive stateful PCE.

For active stateful PCE, a PCUpd message is sent from PCE to PCC to update the LSP state for the LSP delegated to the PCE. Any changes to the delegated LSPs generate a PCRpt message from the PCC to PCE to convey the changes of the LSP. Any modifications to the Objects/TLVs that are identified in this document to support GMPLS technology-specific attributes will be carried in the PCRpt and PCUpd messages.

For passive stateful PCEs, PCReq/PCRep messages are used to convey path computation instructions. GMPLS-technology specific Objects and TLVs are defined in [RFC8779], so this document just points at that work and only adds the stateful PCE aspects where applicable. Passive Stateful PCE makes use of PCRpt messages when reporting LSP State changes sent by PCC to PCEs. Any modifications to the

Objects/TLVs that are identified in this document to support GMPLS technology-specific attributes will be carried in the PCRpt message.

Furthermore, the LSP Initiation function of PCEP is defined in [RFC8281] to allow the PCE to initiate LSP establishment after the path is computed. PCInitiate messages are used to trigger the end node to set up the LSP. Any modifications to the Objects/TLVs that are identified in this document to support GMPLS technology-specific attributes will be carried in the PCInitiate messages.

[RFC8779] defines GMPLS-technology specific Objects/TLVs in stateless PCEP, and this document makes use of these Objects/TLVs without modifications where applicable. Where these Objects/TLVs require modifications to incorporate stateful PCE, they are described in this document. The remote-initiated LSP would follow the principle specified in [RFC8281], and GMPLS-specific extensions are also included in this document.

4. Main Requirements

This section notes the main functional requirements for PCEP extensions to support stateful PCE for use in GMPLS-controlled networks, based on the description in [RFC8051]. Many requirements are common across a variety of network types (e.g., MPLS-TE networks and GMPLS networks) and the protocol extensions to meet the requirements are already described in [RFC8231]. This document does not repeat the description of those protocol extensions. This document presents protocol extensions for a set of requirements which are specific to the use of a stateful PCE in a GMPLS-controlled network.

The requirements for GMPLS-specific stateful PCE are as follows:

- o Advertisement of the stateful PCE capability. This generic requirement is covered in Section 5.4 of [RFC8231]. The GMPLS CAPABILITY TLV in section 2.1 of [RFC8779] and its extension in this document MUST be advertised as well.
- o LSP operations, including LSP update, delegation and state synchronization/report are covered in [RFC8231]. This document provides extensions for its application in GMPLS-controlled networks.
- o All the PCEP messages need to be capable of indicating GMPLS-specific switching capabilities a per TE link basis. GMPLS LSP creation/modification/deletion requires knowledge of LSP switching capability (e.g., TDM, L2SC, OTN-TDM, LSC, etc.) and the generalized payload (G-PID) to be used according to

[RFC3471], [RFC3473]. It also requires the specification of data flow specific traffic parameters (also known as TSpec), which are technology specific. Such information would need to be included in various PCEP messages.

- o In some technologies, path calculation is tightly coupled with label selection along the route. For example, path calculation in a WDM network may include lambda continuity and/or lambda feasibility constraints and hence a path computed by the PCE is associated with a specific lambda (label). Hence, in such networks, the label information needs to be provided to a PCC in order for a PCE to initiate GMPLS LSPs under the active stateful PCE model, i.e., explicit label control may be required.
- o Stateful PCEP messages also need to indicate the protection context information for the LSP specified by GMPLS, as defined in [RFC4872], [RFC4873].

5. Overview of Stateful PCEP Extensions for GMPLS Networks

5.1. Capability Advertisement for Stateful PCEP in GMPLS

Capability Advertisement has been specified in [RFC8231], and can be achieved by using the "STATEFUL-PCE-CAPABILITY" in the PCEP TLV Type Indicators. Another GMPLS-CAPABILITY TLV in the PCEP TLV Type Indicators has been defined in [RFC8779]. According to [RFC8779], IANA created a registry to manage the value of the GMPLS-CAPABILITY TLV's Flag field. New bits, LSP-UPDATE-CAPABILITY (TBD1) and LSP-INSTITUTION-CAPABILITY (TBD2), are introduced as flags to indicate the capability for LSP update and remote LSP initiation in GMPLS networks.

5.2. LSP Synchronization

PCCs need to report the attributes of LSPs to the PCE to enable stateful operation of a GMPLS network. This process is known as LSP state synchronization. The LSP attributes including bandwidth, associated route, and protection information etc., are stored by the PCE in the LSP database (LSP-DB). Note that, as described in [RFC8231], the LSP state synchronization covers both the bulk reporting of LSPs at initialization as well the reporting of new or modified LSPs during normal operation. Incremental LSP-DB synchronization may be desired in a GMPLS-controlled network and it is specified in [RFC8232].

The END-POINTS object is extended for GMPLS in [RFC8779]. The END-POINTS object is carried in the PCRpt message as specified in

[RFC8623]. The END-POINTS object type for GMPLS is included in the PCRpt message as per the same.

The BANDWIDTH, LSPA, IRO and XRO objects are extended for GMPLS in [RFC8779]. These objects are carried in the PCRpt message as specified in [RFC8231] (as the attribute-list defined in Section 6.5 of [RFC5440] and extended by many other documents that define PCEP extensions for specific scenarios).

The SWITCH-LAYER object is defined in [RFC8282]. This object is carried in PCRpt message as specified in section 3.2 of [RFC8282].

5.3. LSP Delegation and Cleanup

LSP delegation and cleanup procedure specified in [RFC8231] are equally applicable to GMPLS LSPs and this document does not modify the associated usage.

5.4. LSP Operations

Both passive and active stateful PCE mechanisms in [RFC8231] are applicable in GMPLS-controlled networks. Remote LSP Initiation in [RFC8281] is also applicable in GMPLS-controlled networks.

6. Extension of Existing PCEP Messages

This section describes how the PCEP messages are extended by using Routing Backus-Naur Form (RBNF) [RFC5511] formats. Contents in this section are for informative purpose.

6.1. The PCRpt Message

According to [RFC8231], the PCRpt Message is used to report the current state of an LSP. This document extends the message in reporting the status of LSPs with GMPLS characteristics.

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> ::= [<SRP>]
```

<LSP>

<path>

Where:

<path>::= <intended-path>

[<actual-attribute-list><actual-path>]

<intended-attribute-list>

<actual-attribute-list>::=[<BANDWIDTH>]

[<metric-list>]

Where:

<intended-path> is represented by the ERO object defined in Section 7.9 of [RFC5440], augmented in [RFC8779] with explicit label control (ELC) and Path Keys.

<actual-attribute-list> consists of the actual computed and signaled values of the <BANDWIDTH> and <metric-lists> objects defined in [RFC5440]. GENERALIZED-BANDWIDTH object has been defined in [RFC8779] to address the limitation of the BANDWIDTH object, with supporting the following:

- o Asymmetric bandwidth (different bandwidth in forward and reverse direction), as described in [RFC6387].
- o Technology specific GMPLS parameters (e.g., TSpec for SDH/SONET, G.709, ATM, MEF, etc.).

<actual-path> is represented by the RRO object defined in Section 7.10 of [RFC5440].

<intended-attribute-list> is the attribute-list defined in Section 6.5 of [RFC5440] and extended by many other documents that define PCEP extensions for specific scenarios.

The SRP object is OPTIONAL, and the usage is extended in the section 7.2.3 of this document.

6.2. The PCUpd Message

The format of a PCUpd message is as follows:

```
<PCUpd Message> ::= <Common Header>
                        <update-request-list>
```

Where:

```
<update-request-list> ::= <update-request> [<update-request-
list>]
```

```
<update-request> ::= <SRP>
                        <LSP>
                        <path>
```

Where:

```
<path> ::= <intended-path> <intended-attribute-list>
```

Where:

<intended-path> is represented by the ERO object defined in Section 7.9 of [RFC5440], augmented in [RFC8779] with explicit label control (ELC) and Path Keys.

<intended-attribute-list> is the attribute-list defined in [RFC5440] and extended by many other documents that define PCEP extensions for specific scenarios.

The SRP object is OPTIONAL, and the usage is extended in the section 7.2.3 of this document.

6.3. The PCInitiate Message

According to [RFC8281], the PCInitiate Message is used allow remote LSP Initiation. This document extends the message in initiating LSPs with GMPLS characteristics. The format of a 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-instantiation> ::= <SRP>
                                       <LSP>
                                       [<END-POINTS>]
                                       <ERO>
                                       [<attribute-list>]

<PCE-initiated-lsp-deletion> ::= <SRP>
                                   <LSP>
```

The format of the PCInitiate message is unchanged from Section 5.1 of [RFC8281]. However, note the following:

- o The END-POINTS object was been extended by [RFC8779] to include a new object type called "Generalized Endpoint". A PCInitiate message used to trigger a GMPLS LSP instantiation MUST use that extension.
- o A PCInitiate message sent by a PCE to a PCC to trigger a GMPLS LSP instantiation MUST include the END-POINTS with Generalized Endpoint object type (even though it is marked as optional in the message definition).
- o The END-POINTS object MUST contain a "label request" TLV per [RFC8779]. The label request TLV is used to specify the switching type, encoding type and G-PID of the LSP being instantiated by the PCE.
- o If unnumbered endpoint addresses are used for the LSP being instantiated by the PCE, the unnumbered endpoint TLV [RFC8779] MUST be use to specify the unnumbered endpoint addresses.
- o The END-POINTS MAY contain other TLVs defined in [RFC8779].

7. PCEP Object Extensions

7.1. Existing Extensions used for Stateful GMPLS

Existing extensions defined in [RFC8779] can be used in the Stateful PCEP with no changes or slightly changes for GMPLS network control, including the following:

- o END-POINTS: Generalized END-POINTS was specified in [RFC8779] to include GMPLS capabilities. Stateful PCEP messages MUST include the END-POINTS with Generalized Endpoint object type, containing the "label request" TLV.
- o BANDWIDTH: Generalized BANDWIDTH was specified in [RFC8779] to represent GMPLS features, including asymmetric bandwidth and G-PID information.
- o LSPA: LSPA Extensions in Section 2.8 of [RFC8779] is applicable in Stateful PCEP for GMPLS networks.
- o IRO: IRO Extensions in Section 2.6 of [RFC8779] is applicable in Stateful PCEP for GMPLS networks.
- o XRO: XRO Extensions in Section 2.7 of [RFC8779] is applicable in Stateful PCEP for GMPLS networks. A new flag is defined in section 7.2.2 of this document.
- o ERO: The ERO was not extended in [RFC8779], and not in this document as well.
- o SWITCH-LAYER: SWITCHING-LAYER definition in Section 3.2 of [RFC8282] is applicable in Stateful PCEP messages for GMPLS networks.

7.2. New Extensions

7.2.1. OPEN Object Extension GMPLS-CAPABILITY TLV

In [RFC8779], IANA has allocated value 45 (GMPLS-CAPABILITY) from the "PCEP TLV Type Indicators" sub-registry. The TLV is extended with two flags to indicate the Stateful and remote initiate capability.

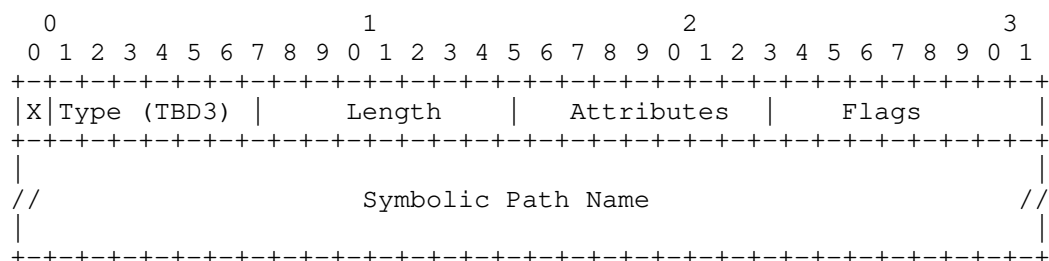
S (LSP-UPDATE-CAPABILITY(TBD1) -- 1 bit): if set to 1 by a PCC, the S flag indicates that the PCC allows modification of LSP parameters; if set to 1 by a PCE, the S flag indicates that the PCE is capable of updating LSP parameters. The LSP-UPDATE-CAPABILITY flag must be advertised by both a PCC and a PCE for PCUpd messages to be allowed on a PCEP session.

I (LSP-INSTANTIATION-CAPABILITY(TBD2) -- 1 bit): If set to 1 by a PCC, the I flag indicates that the PCC allows instantiation of an LSP by a PCE. If set to 1 by a PCE, the I flag indicates that the PCE supports instantiating LSPs. The LSP-INSTANTIATION-CAPABILITY flag must be set by both the PCC and PCE in order to enable PCE-initiated LSP instantiation.

7.2.2. New LSP Exclusion Sub-object in the XRO

[RFC5521] defines a mechanism for a PCC to request or demand that specific nodes, links, or other network resources are excluded from paths computed by a PCE. A PCC may wish to request the computation of a path that avoids all link and nodes traversed by some other LSP.

To this end this document defines a new sub-object for use with route exclusion defined in [RFC5521]. The LSP exclusion sub-object is as follows:



X bit and Attribute fields are defined in [RFC5521].

Type: Sub-object Type for an LSP exclusion sub-object. Value of TBD3. To be assigned by IANA.

Length: The Length contains the total length of the sub-object in bytes, including the Type and Length fields.

Flags: This field may be used to further specify the exclusion constraint with regard to the LSP. Currently, no values are defined.

Symbolic Path Name: This is the identifier given to an LSP and is unique in the context of the PCC address as defined in [RFC8231].

This sub-object is OPTIONAL in the exclude route object (XRO) and can be present multiple times. When a stateful PCE receives a PCReq message carrying this sub-object, it MUST search for the identified

LSP in its LSP-DB and then exclude from the new path computation all resources used by the identified LSP.

7.2.3. SRP Extension

The format of the SRP object is defined in [RFC8231]. The object is used in PCUpd and PCInitiate messages for GMPLS.

This document defines a new flag to be carried in the Flags field of the SRP object. This flag indicates a bidirectional co-routed LSP setup operation initiated by the PCE as follows:

- o B (Bidirectional LSP -- 1 bit): If set to 0, it indicates a request to create a uni-directional LSP. If set to 1, it indicates a request to create a bidirectional co-routed LSP.

The bit position is TBD4 as assigned by IANA.

8. Update to Error Handling

A PCEP-ERROR object is used to report a PCEP error and is characterized by an Error-Type that specifies the type of error and an Error-value that provides additional information about the error. In this document the following Error-Type and Error-Value are introduced.

8.1. Error Handling in LSP Re-optimization

A stateful PCE performs the re-optimization when the R bit is set in RP object. If no LSP state information is available to carry out re-optimization, the stateful PCE SHOULD report the error "LSP state information unavailable for the LSP re-optimization" (Error Type = TBD5, Error value= TBD6). The PCE MAY suppress this error message on a configurable threshold.

8.2. Error Handling in Route Exclusion

This sub-object in XRO defined in section 7.2.2 of this document is OPTIONAL and can be present multiple times. When a stateful PCE receives a PCReq message carrying this sub-object, it searches for the identified LSP in its LSP-DB and then excludes from the new path computation all resources used by the identified LSP. If the stateful PCE cannot recognize one or more of the received LSP identifiers, it SHOULD send an error message PCErr reporting "The LSP state information for route exclusion purpose cannot be found"

(Error-type = TBD5, Error-value = TBD7). Optionally, it may also provide with the unrecognized identifier information to the requesting PCC using the error reporting techniques described in [RFC5440]. However, the PCE MAY suppress this error message on a configurable threshold.

8.3. Error Handling for generalized END-POINTS

If the END-POINTS Object of type Generalized Endpoint is missing the label request TLV, the PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value= TBD8 (label request TLV missing).

9. Implementation

[NOTE TO RFC EDITOR : This whole section and the reference to RFC 7942 is to be removed before publication as an RFC]

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".

9.1. Huawei Technologies

- o Organization: Huawei Technologies, Co. LTD
- o Implementation: Huawei NCE-T
- o Description: PCRpt, PCUpd and PCInitiate messages for GMPLS Network

- o Maturity Level: Production
- o Coverage: Full
- o Contact: zhenghaomian@huawei.com

10. IANA Considerations

10.1. New GMPLS-CAPABILITY

[RFC8231] defines the STATEFUL-PCE-CAPABILITY TLV; per that RFC, IANA created a registry to manage the value of the STATEFUL-PCE-CAPABILITY TLV's Flag field. IANA has allocated a new bit in the STATEFUL-PCE-CAPABILITY TLV Flag Field registry, as follows:

Bit	Description	Reference
---	-----	-----
TBD1	LSP-UPDATE-CAPABILITY (S)	[This.I-D]
TBD2	LSP-INstantiation-CAPABILITY (I)	[This.I-D]

10.2. New Sub-object for the Exclude Route Object

IANA maintains the "PCEP Parameters" registry containing a subregistry called "PCEP Objects". This registry has a subregistry for the XRO (Exclude Route Object) listing the sub-objects that can be carried in the XRO. IANA is requested to assign a further sub-object that can be carried in the XRO as follows:

Value	Description	Reference
-----+-----+-----		
TBD3	LSP Exclusion sub-object	[This.I-D]

10.3. Flag Field for new XRO Sub-object

IANA has created a registry to manage the Flag field of the LSP Exclusion sub-object in XRO object. No Flag is currently defined for this flag field in this document.

Codespace of the Flag field (LSP Exclusion sub-object)

Bit	Description	Reference
0-7	Unassigned	[This.I-D]

10.4. New "B" Flag in the SRP Object

IANA maintains a subregistry, named the "SRP Object Flag Field", within the "Path Computation Element Protocol (PCEP) Numbers" registry, to manage the Flag field of the SRP object.

IANA is requested to make an assignment from this registry as follows:

Bit ---	Description -----	Reference -----
TBD4	Bi-directional co-routed LSP	[This.I-D]

10.5. New PCEP Error Codes

IANA is requested to make the following allocation in the "PCEP-ERROR Object Error Types and Values" registry.

Error Type	Meaning	Reference
TBD5	LSP state information missing	[This.I-D]
Error-value TBD6:	LSP state information unavailable for the LSP re-optimization	[This.I-D]
Error-value TBD7:	LSP state information for route exclusion purpose cannot be found	[This.I-D]

This document defines the following new Error-Value:

Error-Type	Error-Value	Reference
6	Error-value TBD8: Label Request TLV missing	[This.I-D]

11. Manageability Considerations

The description and functionality specifications presented related to stateful PCEs should also comply with the manageability specifications covered in Section 8 of [RFC4655]. Furthermore, a further list of manageability issues presented in [RFC8231] should also be considered.

11.1. Requirements on Other Protocols

When the detailed route information is included for LSP state synchronization (either at the initial stage or during LSP state report process), this requires the ingress node of an LSP carry the RRO object in order to enable the collection of such information.

12. Security Considerations

This draft provides additional extensions to PCEP so as to facilitate stateful PCE usage in GMPLS-controlled networks, on top of [RFC8231]. The PCEP extensions to support GMPLS-controlled networks should be considered under the same security as for MPLS networks, as noted in [RFC7025]. Therefore, the security considerations elaborated in [RFC5440] still apply to this draft. Furthermore, [RFC8231] provides a detailed analysis of the additional security issues incurred due to the new extensions and possible solutions needed to support for the new stateful PCE capabilities and they apply to this document as well.

13. Acknowledgement

We would like to thank Adrian Farrel, Cyril Margaria, George Swallow and Jan Medved for the useful comments and discussions.

14. References

14.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.
- [RFC5440] Vasseur, J.-P., and Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, April 2009.
- [RFC8174] B. Leiba, "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", RFC 8174, May 2017.
- [RFC8231] Crabbe, E., Medved, J., Varga, R., Minei, I., "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, September 2017.

- [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, December 2017.
- [RFC8779] Margaria, C., Gonzalez de Dios, O., Zhang, F., "Path Computation Element Communication Protocol (PCEP) extensions for GMPLS", RFC 8779, July 2020.

14.2. Informative References

- [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>>.
- [RFC8051] Zhang, X., Minei, I., et al, "Applicability of Stateful Path Computation Element (PCE) ", RFC 8051, January 2017.
- [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, September 2017.
- [RFC8282] Oki, E., Takeda, T., Farrel, A., and F. Zhang, "Extensions to the Path Computation Element communication Protocol (PCEP) for Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 8282, December 2017.
- [RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4872] Lang, J., Ed., Rekhter, Y., Ed., and D. Papadimitriou, Ed., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.

- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, May 2007.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC5511, April 2005.
- [RFC6387] Takacs, A., Berger, L., Caviglia, D., Fedyk, D., and J. Meuric, "GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths (LSPs)", RFC 6387, September 2011.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, September 2013,
- [RFC7399] Farrel, A., King, D., "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, October 2014.
- [RFC8623] Palle, U., Dhody, D., Tanaka, Y., Beeram, V., "Stateful Path Computation Element (PCE) Protocol Extensions for Usage with Point-to-Multipoint TE Label Switched Paths (LSPs)" June 2019.

15. Contributors' Address

Xian Zhang
Huawei Technologies
Email: zhang.xian@huawei.com

Dhruv Dhody
Huawei Technology
India
Email: dhruv.ietf@gmail.com

Yi Lin
Huawei Technologies
Email: yi.lin@huawei.com

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain
Email: ramon.casellas@cttc.es

Siva Sivabalan
Cisco Systems
Email: msiva@cisco.com

Clarence Filsfils
Cisco Systems
Email: cfilsfil@cisco.com

Robert Varga
Pantheon Technologies
Email: nite@hq.sk

Authors' Addresses

Young Lee
Samsung
Email: younglee.tx@gmail.com

Haomian Zheng
Huawei Technologies
H1, Huawei Xiliu Beipo Village, Songshan Lake
Dongguan, Guangdong 523808
China
Email: zhenghaomian@huawei.com

Oscar Gonzalez de Dios
Telefonica
Phone: +34 913374013
Email: oscar.gonzalezdedios@telefonica.com

Victor Lopez
Nokia
Email: victor.lopez@nokia.com

Zafar Ali
Cisco Systems
Email: zali@cisco.com

Network Working Group
Internet-Draft
Intended status: Experimental
Expires: September 9, 2019

C. Li
M. Chen
J. Dong
Z. Li
Huawei Technologies
A. Wang
China Telecom
C. Zhou
Cisco System
March 8, 2019

PCE Controlled ID Space
draft-li-pce-controlled-id-space-02

Abstract

The Path Computation Element Communication Protocol (PCEP) provides mechanisms for Path Computation Elements (PCEs) to perform path computations in response to Path Computation Clients (PCCs) requests. The Stateful PCE extensions allow stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Label Switched Paths (LSPs) using PCEP. Furthermore, PCEP can be used for computing paths in SR networks.

Stateful PCE provide active control of MPLS-TE LSPs via PCEP, for a model where the PCC delegates control over one or more locally configured LSPs to the PCE. Further, stateful PCE could also create and delete PCE-initiated LSPs itself. A PCE-based central controller (PCECC) simplify the processing of a distributed control plane by integrating with elements of Software-Defined Networking (SDN).

In some use cases, such as PCECC or Binding Segment Identifier (SID) for Segment Routing (SR), there are requirements for a stateful PCE to make allocation of labels, SIDs, etc. These use cases require PCE aware of various identifier spaces where to make allocations on behalf of PCC. This document describes a mechanism for PCC to inform the PCE of the identifier space under its control via PCEP. The identifier could be MPLS label, SID or any other to-be-defined identifier to be allocated by a PCE.

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 9, 2019.

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
2. Terminology	4
2.1. Requirements Language	4
3. Use cases	4
3.1. PCE-based Central Control	4
3.1.1. PCECC for MPLS/SR-MPLS	4
3.1.2. PCECC for SRv6	5
3.2. Binding SID Allocation	5
4. Overview	5
5. Objects	7
5.1. Open Object	7
5.1.1. LABEL-CONTROL-SPACE TLV	7
5.1.2. FUNCT-ID-CONTROL-SPACE TLV	8
6. Other Considerations	10
7. IANA Considerations	11
8. Security Considerations	11
9. Acknowledgements	11
10. References	11
10.1. Normative References	11
10.2. Informative References	12
Appendix A. Contributors	14

Authors' Addresses	14
------------------------------	----

1. Introduction

[RFC5440] defines the stateless Path Computation Element communication Protocol (PCEP) for Path Computation Elements (PCEs) to perform path computations in response to Path Computation Clients (PCCs) requests. For supporting stateful operations, [RFC8231] specifies a set of extensions to PCEP to enable stateful control of LSPs within and across PCEP sessions in compliance with [RFC4657]. Furthermore, [RFC8281] describes the setup, maintenance, and teardown of PCE-initiated LSPs under the stateful PCE model, without the need for local configuration on the PCC, thus allowing for a dynamic network that is centrally controlled and deployed.

[RFC8283] introduces the architecture for PCE as a central controller, it examines the motivations and applicability for PCEP as a control protocol in this environment, and introduces the implications for the protocol. Also, [I-D.ietf-pce-pcep-extension-for-pce-controller] specifies the procedures and PCEP protocol extensions for using the PCE as the central controller, where LSPs are calculated/setup/initiated and label forwarding entries are downloaded through extending PCEP. However, the document assumes that label range to be used by a PCE is known and set on both PCEP peers. This extension adds the capability to advertise the range via a PCEP extension.

Similarly, [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. However, the document assumes that label range to be used by a PCE is known and set on both PCEP peers. This extension adds the capability to advertise the range (from SRGB or SRLB of the node) via a PCEP extension.

In addition, [I-D.dhody-pce-pcep-extension-pce-controller-srv6] specifies the procedures and PCEP protocol extensions of PCECC for SRv6. An SRv6 SID is represented as LOC:FUNCT where LOC is the L most significant bits and FUNCT is the 128-L least significant bits. The FUNCT part of the SID is an opaque identification of a local function bound to the SID. This extension adds the capability to advertise the range of Function ID (FUNCT part) via a PCEP extension.

Once the PCC/node has given control over an ID space (for example labels), the PCC/node MUST NOT allocate the ID from this ID space.

For example, a PCC/node MUST NOT use this labels from the PCE controlled label space to make allocation for VPN Prefix distributed via BGP or labels used for LDP/RSVP-TE signalling. This is done to make sure that the PCE control over ID space does not conflict with the existing node allocation.

The use case are described in Section 3. The ID space range information can be advertised via the TLVs in the Open message. The detailed procedures are described in Section 4, and the objects' format is specified in Section 5.

2. Terminology

This memo makes use of the terms defined in [RFC5440], [RFC8231], [RFC8283] and [RFC8402].

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. Use cases

3.1. PCE-based Central Control

A PCE-based central controller (PCECC) can simplify the processing of a distributed control plane by integrating with elements of SDN. Thus, the LSP/SR path can be calculated/setup/initiated and the label/SID 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.

3.1.1. PCECC for MPLS/SR-MPLS

[I-D.ietf-pce-pcep-extension-for-pce-controller] describes a mode where 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. For this to work, the PCE-based controller will take responsibility for managing some part of the MPLS label space for each router that it controls as described in section 3.1.2. of [RFC8283]. A mechanism for a PCC to inform the PCE of such a label space to control is needed within PCEP.

[I-D.ietf-pce-segment-routing] specifies extensions to PCEP that allow a stateful PCE to compute, update or initiate SR-TE paths. [I-D.zhao-pce-pcep-extension-pce-controller-sr] describes the mechanism for PCECC to allocate and provision the node/prefix/adjacency label (SID) via PCEP. To make such allocation, PCE needs to be aware of the label space from Segment Routing Global Block (SRGB) or Segment Routing Local Block (SRLB) [RFC8402] of the node that it controls. A mechanism for a PCC to inform the PCE of such label space to control is needed within PCEP. The full SRGB/SRLB of a node could be learned via existing IGP or BGP-LS mechanism.

3.1.2. PCECC for SRv6

[I-D.dhody-pce-pcep-extension-pce-controller-srv6] describes the mechanism for PCECC to allocate and provision the SRv6 SID via PCEP. An SRv6 SID is represented as LOC:FUNCT ([I-D.filsfils-spring-srv6-network-programming]) where LOC is the L most significant bits and FUNCT is the 128-L least significant bits. The FUNCT part of the SID is an opaque identification of a local function bound to the SID. To make such allocation, PCE needs to be aware of the Function ID space (FUNCT part) of the node that it controls. A mechanism for a PCC to inform the PCE of such a Function ID space to control is needed within PCEP.

3.2. Binding SID Allocation

The headend of an SR Policy binds a Binding SID to its policy [I-D.ietf-spring-segment-routing-policy]. The instantiation of which may involve a list of SIDs. Currently Binding SID are allocated by the node, but there is an inherent advantage in the Binding SID to be allocated by a PCE to allow SR policies to be dynamically created, updated according to the network status and operations. This is described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. Therefore, a PCE needs to obtain the authority and control to allocate Binding SID actively from the PCC's label space as described in above use case.

4. Overview

During PCEP Initialization Phase, Open messages are exchanged between PCCs and PCEs. The OPEN object may also contain a set of TLVs used to convey capabilities in the Open message. The term 'ID' in this document, could be a MPLS label, SRv6 Function ID or any other future ID space for PCE to control and allocate from. A PCC can include a corresponding ID-CONTROL-SPACE TLVs in the OPEN Object to inform the corresponding ID space information that it wants the PCE to control. This TLV MUST NOT be included by the PCE and MUST be ignored on

receipt by a PCC. This is an optional TLV, the PCE could be aware of the ID space from some other means outside of PCEP.

For delegating multiple types of ID space, multiple TLVs corresponding to each ID type MUST be included in an Open message. The ID type can be MPLS label or other type of ID. The following ID-CONTROL-SPACE TLV is defined in this document -

- o LABEL-CONTROL-SPACE TLV - for MPLS Labels (including for SR-MPLS)
- o FUNCTION-ID-CONTROL-SPACE TLV - for SRv6 SID Function ID

The procedure of ID space control to PCE is shown below:

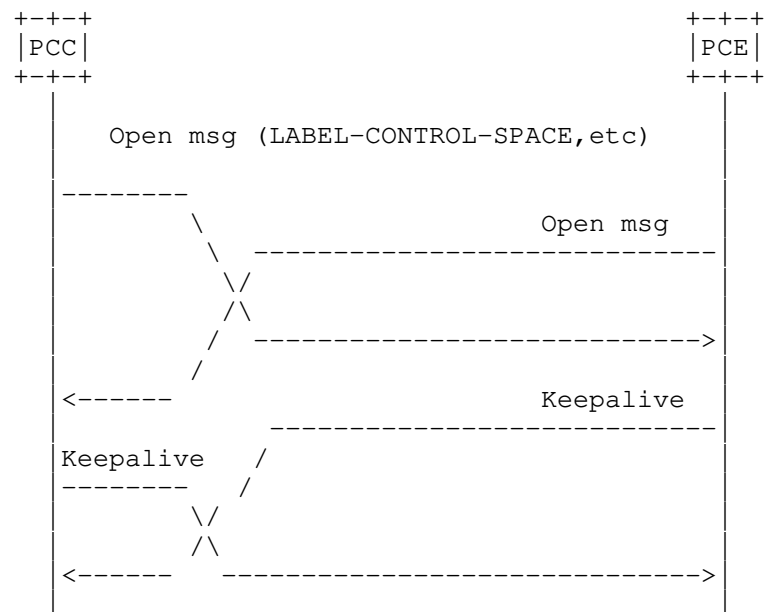


Figure 1: ID space control to PCE

If the ID space control procedure is successful, the PCE will return a KeepAlive message to the PCC. If there is any error in processing the corresponding TLV, an Error (PCErr) message will be sent to the PCC with Error-Type=1 (PCEP session establishment failure) and Error-value=TBD (ID space control failure).

After this process, a stateful PCE can learn the PCE controlled ID spaces of a node (PCC) under its control. A PCE can then allocate IDs within the control ID space. For example, a PCE can actively

allocate labels and download forwarding instructions for the PCECC LSP as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. A PCE can also allocate labels from the PCE controlled portion of the SRGB/SRLB for PCECC-SR [I-D.zhao-pce-pcep-extension-pce-controller-sr]. The full SRGB/SRLB of a node could be learned via existing IGP or BGP-LS mechanism.

5. Objects

5.1. Open Object

For advertising the PCE controlled ID space to a PCE, this document defines several TLVs within the OPEN object.

5.1.1. LABEL-CONTROL-SPACE TLV

For a PCC to inform the label space under the PCE control, this document defines a new LABEL-CONTROL-SPACE TLV.

The LABEL-CONTROL-SPACE TLV is an optional TLV in the OPEN object, and its format is shown in the following figure:

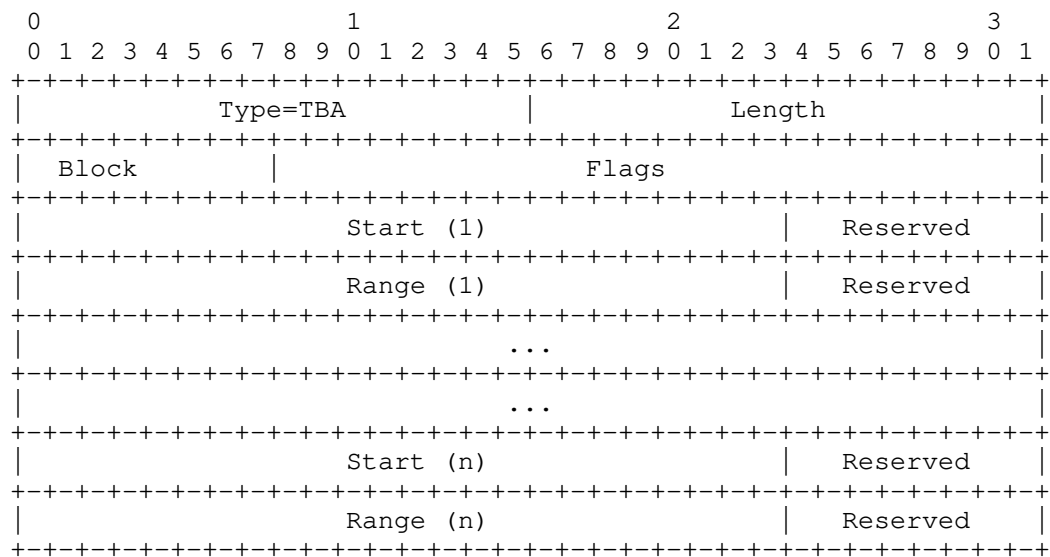


Figure 2: LABEL-CONTROL-SPACE TLV

The type (16 bits) of the TLV is TBA. The length field (16 bits) and has a variable value.

Block(8 bits): the number of ID blocks. The range of a block is described by a start field and a range field.

Flags (24 bits): No flag is currently defined. The unassigned bits of Flags field MUST be set to 0 on transmission and MUST be ignored on receipt.

Start(i) (24 bits): indicates the beginning of the label block i.

Range(i) (24 bits): indicates the range of the label block i.

Reserved: SHOULD be set to 0 on transmission and MUST be ignored on reception.

LABEL-CONTROL-SPACE TLV SHOULD be included only once in a Open Message. On receipt, only the first instance is processed and others MUST be ignored.

A stateful PCE can actively allocate labels and download forwarding instructions for the PCECC LSP as described in [I-D.ietf-pce-pcep-extension-for-pce-controller]. A PCE can also allocate labels from SRGB/SRLB for PCECC-SR [I-D.zhao-pce-pcep-extension-pce-controller-sr]. The Binding Segments can also be selected for the PCE controlled space [I-D.ietf-pce-pcep-extension-for-pce-controller].

5.1.2. FUNCT-ID-CONTROL-SPACE TLV

For a PCC to inform the SRv6 SID Function ID space under the PCE control, this document defines a new FUNCT-ID-CONTROL-SPACE TLV.

The FUNCT-ID-CONTROL-SPACE TLV is an optional TLV for use in the OPEN object, and its format is shown in the following figure:



Figure 3: FUNCT-ID-CONTROL-SPACE TLV

The type (16 bits) of the TLV is TBA. The length field (16 bits) and has a variable value.

Block(8 bits): the number of ID blocks. The range of a block is described by a start field and a range field.

Flags (24 bits): Following flags are currently defined

- o L-flag: Locator flag, set when the locator information is included in this TLV. If L-flag is unset, Loc Size and variable Locator field MUST NOT be included in this TLV, and the ID spaces are applicable to all Locators.

The unassigned bits of Flags field MUST be set to 0 on transmission and MUST be ignored on receipt.

Start(i) (128 bits): indicates the beginning of the Function ID block i.

Range(i) (128 bits): indicates the range of the Function ID block i.

Loc size(8 bits): indicates the bit length of a Locator. Appears only when the L-flag is set.

Locator (variable length): the value of a Locator. The Function ID spaces specified in this TLV are associated with this locator.

As per [RFC5440], the value portion of the PCEP TLV needs to be 4-bytes aligned, so a FUNCT-ID-CONTROL-SPACE TLV is padded with trailing zeros to a 4-byte boundary.

Multiple FUNCT-ID-CONTROL-SPACE TLVs MAY be included in a OPEN object to specify Function ID space specific to each locator.

A stateful PCE can actively allocate SRv6 SID and download forwarding instructions for the PCECC SRv6 path as described in [I-D.dhody-pce-pcep-extension-pce-controller-srv6].

Note that SRv6 SID allocation involves LOC:FUNCT; the LOC is assumed to be known at PCE and FUNCT is allocated from the PCE controlled Function ID block.

6. Other Considerations

In case of multiple PCEs, a PCC MAY decide to give control over different ID space to each instance of the PCE. In case a PCC includes the same ID space to multiple PCEs, the PCE SHOULD use synchronization mechanism (such as [I-D.litkowski-pce-state-sync]) to avoid allocating the same ID.

The PCE would allocate ID from the PCE controlled ID space. The PCC would not allocate ID by itself from this space as long as it has an active PCEP session to a PCE to which it has given control over the ID space.

Note that if there is any change in the ID space, the PCC MUST bring the session down and re-establish the session with new TLVs. During state synchronization the PCE would need to consider the new ID space into consideration and SHOULD re-establish the LSP/SR-paths if needed.

The PCC can regain control of the ID space by closing the PCEP session and require new session without ID space TLVs specified in this document.

7. IANA Considerations

TBA.

8. Security Considerations

TBA.

9. Acknowledgements

TBD.

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>>.
- [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>>.

- [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>>.

10.2. Informative References

- [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>>.
- [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-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-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-01 (work in progress), February 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-04 (work in progress), February 2019.
- [I-D.litkowski-pce-state-sync]
Litkowski, S., Sivabalan, S., and D. Dhody, "Inter Stateful Path Computation Element communication procedures", draft-litkowski-pce-state-sync-04 (work in progress), October 2018.

[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-02 (work in progress), October 2018.

[I-D.dhody-pce-pcep-extension-pce-controller-srv6]

Negi, M., Li, Z., and X. Geng, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for SRv6", draft-dhody-pce-pcep-extension-pce-controller-srv6-01 (work in progress), February 2019.

[I-D.filsfils-spring-srv6-network-programming]

Filsfils, C., Camarillo, P., Leddy, J., daniel.voyer@bell.ca, d., Matsushima, S., and Z. Li, "SRv6 Network Programming", draft-filsfils-spring-srv6-network-programming-07 (work in progress), February 2019.

Appendix A. Contributors

Dhruv Dhody

Huawei Technologies

Divyashree Techno Park, Whitefield

Bangalore, Karnataka 560066

India

EMail: dhruv.ietf@gmail.com

Authors' Addresses

Cheng Li

Huawei Technologies

Huawei Campus, No. 156 Beiqing Rd.

Beijing 100095

China

EMail: chengli13@huawei.com

Mach(Guoyi) Chen

Huawei Technologies

Huawei Campus, No. 156 Beiqing Rd.

Beijing 100095

China

EMail: Mach.chen@huawei.com

Jie Dong

Huawei Technologies

Huawei Campus, No. 156 Beiqing Rd.

Beijing 100095

China

EMail: jie.dong@huawei.com

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

EMail: lizhenbin@huawei.com

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

EMail: wangaj.bri@chinatelecom.cn

Chao Zhou
Cisco System
San Jose
USA

EMail: chao.zhou@cisco.com

Network Working Group
Internet-Draft
Intended status: Experimental
Expires: 22 September 2022

C. Li
H. Shi
Huawei Technologies
A. Wang
China Telecom
W. Cheng
China Mobile
C. Zhou
HPE
21 March 2022

PCE Controlled ID Space
draft-li-pce-controlled-id-space-11

Abstract

The Path Computation Element Communication Protocol (PCEP) provides a mechanisms for the Path Computation Elements (PCEs) to perform path computations in response to Path Computation Clients (PCCs) requests. The Stateful PCE extensions allow stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Label Switched Paths (LSPs) using PCEP. Furthermore, PCE can be used for computing paths in the SR networks.

Stateful PCE provide active control of MPLS-TE LSPs via PCEP, for a model where the PCC delegates control over one or more locally configured LSPs to the PCE. Further, stateful PCE could also create and remove PCE-initiated LSPs by itself. A PCE-based Central Controller (PCECC) simplify the processing of a distributed control plane by integrating with elements of Software-Defined Networking (SDN).

In some use cases, such as PCECC or Binding Segment Identifier (SID) for Segment Routing (SR), there are requirements for a stateful PCE to make allocation of labels, SIDs, etc. These use cases require PCE aware of various identifier spaces from where to make allocations on behalf of a PCC. This document describes a mechanism for a PCC to inform the PCE of the identifier space set aside for the PCE control via PCEP. The identifier could be an MPLS label, a SID or any other to-be-defined identifier that can be allocated by a PCE.

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 22 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	4
2.1. Requirements Language	4
3. Use cases	4
3.1. PCE-based Central Control	4
3.1.1. PCECC for MPLS/SR-MPLS	4
3.1.2. PCECC for SRv6	5
3.2. Binding SID Allocation	5
4. Overview	6
5. Objects	7
5.1. Open Object	7
5.1.1. LABEL-CONTROL-SPACE TLV	7
5.1.2. FUNCT-ID-CONTROL-SPACE TLV	9
6. Other Considerations	11
7. IANA Considerations	11
7.1. PCEP TLV Type Indicators	11
7.2. LABEL-CONTROL-SPACE TLV's Flag field	11
7.3. FUNCT-ID-CONTROL-SPACE TLV's Flag field	12
8. Security Considerations	12
9. Acknowledgements	13

10. References	13
10.1. Normative References	13
10.2. Informative References	14
Appendix A. Contributors	16
Authors' Addresses	17

1. Introduction

[RFC5440] defines the stateless Path Computation Element Communication Protocol (PCEP) for the Path Computation Elements (PCEs) to perform path computation in response to Path Computation Clients (PCCs) requests. For supporting stateful operations, [RFC8231] specifies a set of extensions to PCEP to enable stateful control of LSPs within and across PCEP sessions in compliance with [RFC4657]. Furthermore, [RFC8281] describes the setup, maintenance, and teardown of PCE-initiated LSPs under the stateful PCE model, without the need for local configuration on the PCC, thus allowing for a dynamic network that is centrally controlled and deployed.

[RFC8283] introduces the architecture for PCE as a central controller, it examines the motivations and applicability for PCEP as a control protocol in this environment, and introduces the implications for the protocol. Also, [RFC9050] specifies the procedures and PCEP extensions for using the PCE as a Central Controller (PCECC), where LSPs are calculated/set up/initiated and label forwarding entries are downloaded through extending PCEP. However, the document assumes that label range to be used by a PCE is known and set on both PCEP peers. This extension adds the capability to advertise the label range via a PCEP extension.

Similarly, [RFC9050] specifies the procedures and PCEP 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. However, the document assumes that label range to be used by a PCE is known and set on both PCEP peers. This extension adds the capability to advertise the range (from SRGB or SRLB of the node) via a PCEP extension.

In addition, [I-D.dhody-pce-pcep-extension-pce-controller-srv6] specifies the procedures and PCEP extensions of PCECC for SRv6. An SRv6 SID is represented as LOC:FUNCT ([RFC8986]) where LOC is the L most significant bits and FUNCT is the 128-L least significant bits. The FUNCT part of the SID is an opaque identification of a local function bound to the SID. This extension adds the capability to advertise the range of Function ID (FUNCT part) via a PCEP extension.

Once the PCC/node has given control over an ID space (for example labels), the PCC/node MUST NOT allocate the ID from this ID space. For example, a PCC/node MUST NOT use this labels from the PCE controlled label space to make allocation for VPN Prefix distributed via BGP or labels used for LDP/RSVP-TE signalling. This is done to make sure that the PCE control over ID space does not conflict with the existing node allocation.

The use case are described in Section 3. The ID space range information can be advertised via the TLVs in the Open message. The detailed procedures are described in Section 4, and the TLV format is specified in Section 5.

2. Terminology

This memo makes use of the terms defined in [RFC5440], [RFC8231], [RFC8283] and [RFC8402].

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. Use cases

3.1. PCE-based Central Control

A PCE-based Central Controller (PCECC) can simplify the processing of a distributed control plane by integrating with elements of SDN. Thus, the LSP/SR path can be calculated/set up/initiated and the label/SID 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.

3.1.1. PCECC for MPLS/SR-MPLS

[RFC9050] describes a mode where 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. For this to work, the PCE-based controller will take responsibility for managing some part of the MPLS label space for each router that it controls as described in section 3.1.2. of [RFC8283]. A mechanism for a PCC to inform the PCE of such a label space to control is

needed within PCEP.

[RFC8664] specifies extensions to PCEP that allow a stateful PCE to compute, update or initiate SR-TE paths. [RFC9050] describes the mechanism for PCECC to allocate and distribute the node/prefix/adjacency label (SID) via PCEP. To make such allocation, PCE needs to be aware of the label space from Segment Routing Global Block (SRGB) or Segment Routing Local Block (SRLB) [RFC8402] of the node that it can control. A mechanism for a PCC to inform the PCE of such label space to control is needed within PCEP. The full SRGB/SRLB of a node could be learned via existing IGP or BGP-LS mechanism.

3.1.2. PCECC for SRv6

[I-D.dhody-pce-pcep-extension-pce-controller-srv6] describes the mechanism for PCECC to allocate and provision the SRv6 SID via PCEP. An SRv6 SID is represented as LOC:FUNCT ([RFC8986]) where LOC is the L most significant bits and FUNCT is the 128-L least significant bits. The FUNCT part of the SID is an opaque identification of a local function bound to the SID. To make such allocation, PCE needs to be aware of the Function ID space (FUNCT part) of the node that it controls. A mechanism for a PCC to inform the PCE of such a Function ID space to control is needed within PCEP.

3.2. Binding SID Allocation

The headend of an SR Policy binds a Binding SID (BSID) [I-D.ietf-pce-binding-label-sid] to its policy [I-D.ietf-spring-segment-routing-policy]. The instantiation of which may involve a list of SIDs. The Binding SID can be allocated by the node as described in [I-D.ietf-pce-binding-label-sid], but there is an inherent advantage in the Binding SID to be allocated by a PCE to allow SR policies to be dynamically created, updated according to the network status and operations. This is described in [RFC9050]. Therefore, a PCE needs to obtain the authority and control to allocate Binding SID actively from the PCC's label space as described in above use case.

This is applicable for both SR-MPLS and SRv6 BSID.

4. Overview

During PCEP Initialization Phase, Open messages are exchanged between the PCCs and the PCEs. The OPEN object may also contain a set of TLVs used to convey the capabilities in the Open message. The term 'ID' in this document, could be a MPLS label, SRv6 Function ID or any other future ID space for PCE to control and allocate from. A PCC can include a corresponding ID-CONTROL-SPACE TLVs in the OPEN Object to inform the corresponding ID space information that it wants the PCE to control. This TLV MUST NOT be included by the PCE and MUST be ignored on receipt by a PCC. This is an optional TLV, the PCE could be aware of the ID space from some other means outside of PCEP.

For delegating multiple types of ID space, multiple TLVs corresponding to each ID type MUST be included in an Open message. The ID type can be MPLS label or other type of ID. The following ID-CONTROL-SPACE TLV is defined in this document -

- * LABEL-CONTROL-SPACE TLV - for MPLS Labels (including for SR-MPLS)
- * FUNCTION-ID-CONTROL-SPACE TLV - for SRv6 SID Function ID

The procedure of ID space control to PCE is shown below:

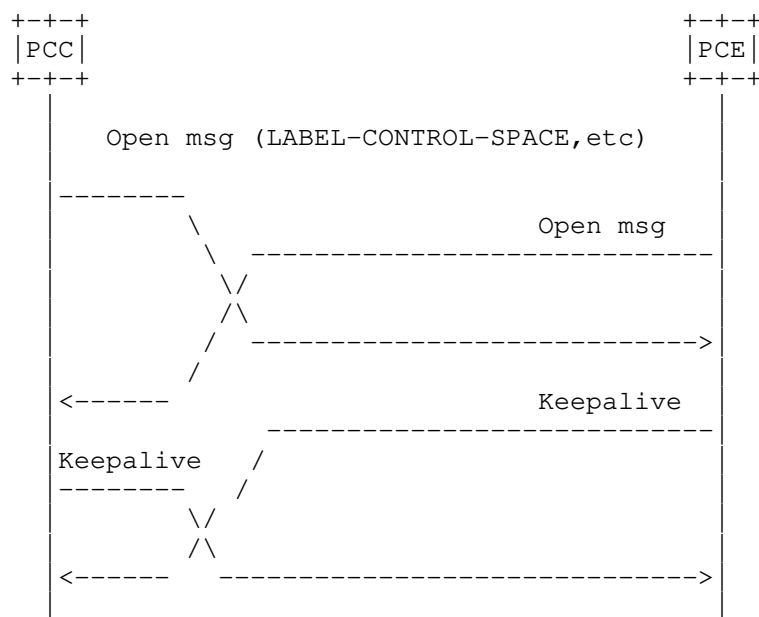


Figure 1: ID space control to PCE

If the ID space control procedure is successful, the PCE will return a KeepAlive message to the PCC. If there is any error in processing the corresponding TLV, an Error (PCErr) message will be sent to the PCC with Error-Type=1 (PCEP session establishment failure) and Error-value=TBD (ID space control failure).

After this process, a stateful PCE can learn the PCE-controlled ID spaces of a node (PCC) under its control. A PCE can then allocate IDs within the controlled-ID space. For example, a PCE can actively allocate labels and download forwarding instructions for the PCECC LSP as described in [RFC9050]. A PCE can also allocate labels from the PCE controlled portion of the SRGB/SRLB for PCECC-SR [RFC9050]. The full SRGB/SRLB of a node could be learned via existing IGP or BGP-LS mechanism.

The procedure for handling the FUNCTION-ID-CONTROL-SPACE TLV is same as above.

5. Objects

5.1. Open Object

For advertising the PCE-controlled ID space to a PCE, this document defines several TLVs within the OPEN object.

5.1.1. LABEL-CONTROL-SPACE TLV

For a PCC to inform the label space under the PCE control, this document defines a new LABEL-CONTROL-SPACE TLV.

The LABEL-CONTROL-SPACE TLV is an optional TLV in the OPEN object, and its format is shown in the following figure:

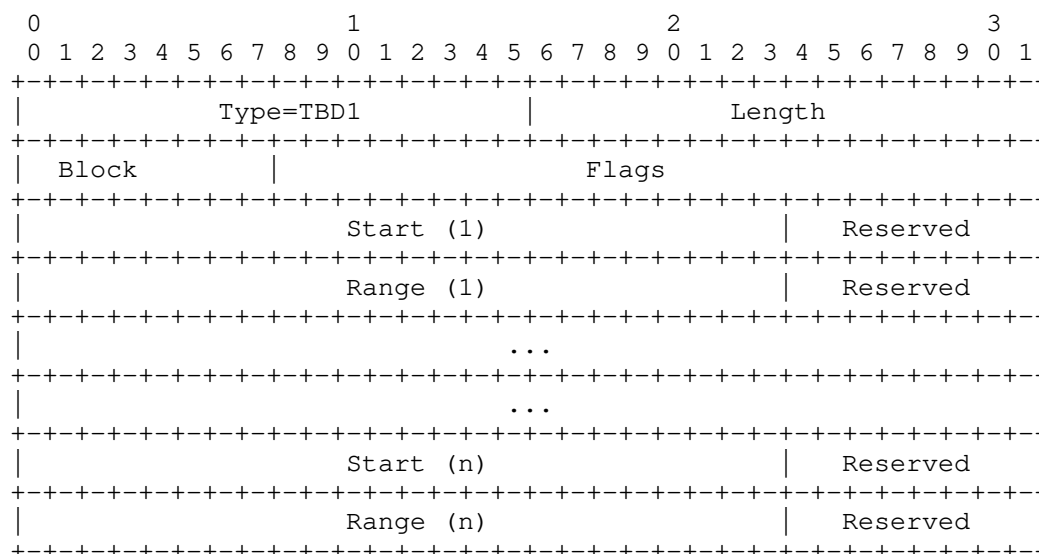


Figure 2: LABEL-CONTROL-SPACE TLV

The type (16 bits) of the TLV is TBD1. The length field (16 bits) and has a variable value.

Block(8 bits): the number of ID blocks. The range of a block is described by a start field and a range field.

Flags (24 bits): No flag is currently defined. The unassigned bits of Flags field MUST be set to 0 on transmission and MUST be ignored on receipt.

Start(i) (24 bits): indicates the beginning of the label block i.

Range(i) (24 bits): indicates the range of the label block i.

Reserved: MUST be set to 0 on transmission and MUST be ignored on reception.

LABEL-CONTROL-SPACE TLV SHOULD be included only once in a Open Message. On receipt, only the first instance is processed and others MUST be ignored.

A stateful PCE can actively allocate labels and download forwarding instructions for the PCECC LSP as described in [RFC9050]. A PCE can also allocate labels from SRGB/SRLB for PCECC-SR [I-D.ietf-pce-pcep-extension-pce-controller-sr]. The Binding Segments can also be selected for the PCE controlled space [RFC9050].

5.1.2. FUNCT-ID-CONTROL-SPACE TLV

For a PCC to inform the SRv6 SID Function ID space under the PCE control, this document defines a new FUNCT-ID-CONTROL-SPACE TLV.

The FUNCT-ID-CONTROL-SPACE TLV is an optional TLV for use in the OPEN object, and its format is shown in the following figure:

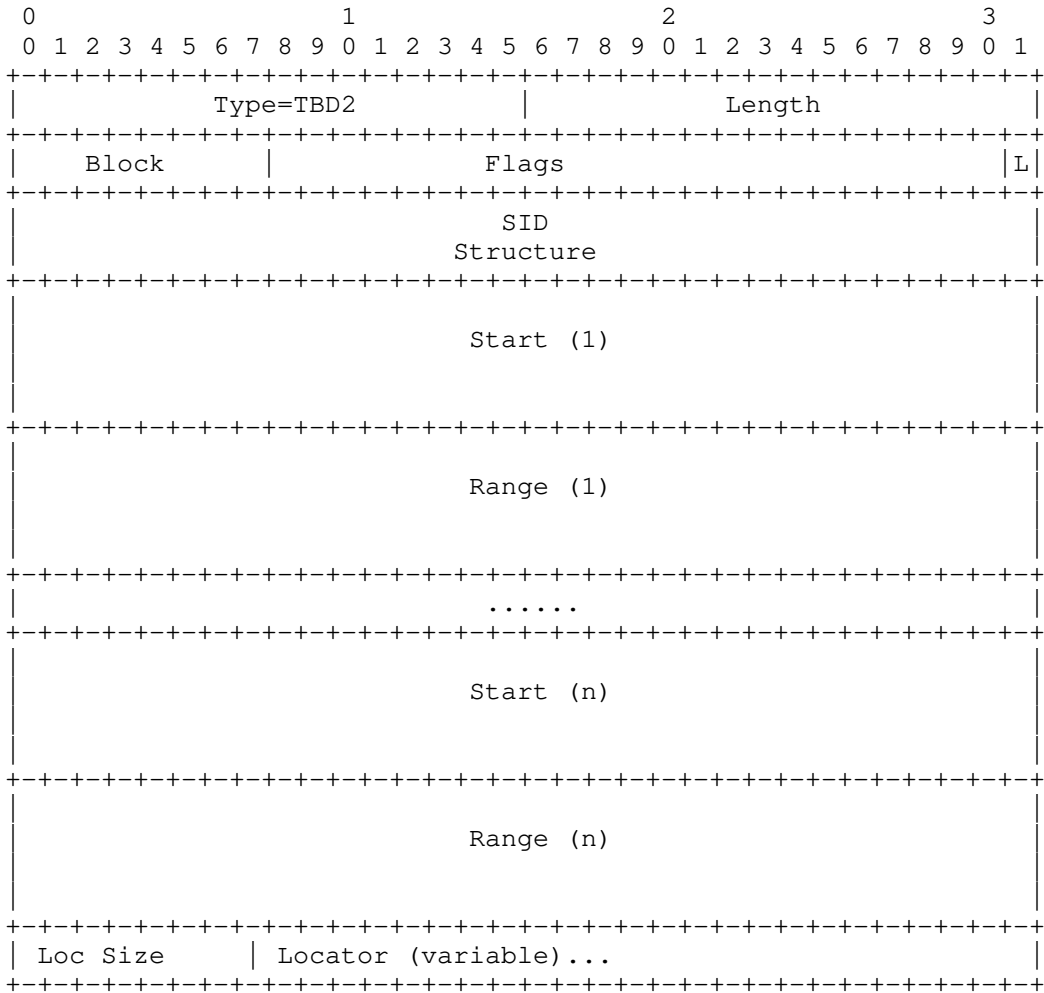


Figure 3: FUNCT-ID-CONTROL-SPACE TLV

The type (16 bits) of the TLV is TBD2. The length field (16 bits) and has a variable value.

Block(8 bits): the number of ID blocks. The range of a block is described by a start field and a range field.

Flags (24 bits): Following flags are currently defined

- * L-flag: Locator flag, set when the locator information is included in this TLV. If L-flag is unset, Loc Size and variable Locator field MUST NOT be included in this TLV, and the ID spaces are applicable to all Locators.

The unassigned bits of Flags field MUST be set to 0 on transmission and MUST be ignored on receipt.

SID Structure: 64-bit field formatted as per "SID Structure" in [I-D.ietf-pce-segment-routing-ipv6].

Start(i) (128 bits): indicates the beginning of the Function ID block i.

Range(i) (128 bits): indicates the range of the Function ID block i.

Loc size(8 bits): indicates the bit length of a Locator. Appears only when the L-flag is set.

Locator (variable length): the value of a Locator. The Function ID spaces specified in this TLV are associated with this locator.

As per [RFC5440], the value portion of the PCEP TLV needs to be 4-bytes aligned, so a FUNCT-ID-CONTROL-SPACE TLV is padded with trailing zeros to a 4-byte boundary.

Multiple FUNCT-ID-CONTROL-SPACE TLVs MAY be included in a OPEN object to specify Function ID space specific to each locator.

A stateful PCE can actively allocate SRv6 SID and download SIDs for the PCECC-SRv6 as described in [I-D.dhody-pce-pcep-extension-pce-controller-srv6].

Note that SRv6 SID allocation involves LOC:FUNCT; the LOC is assumed to be known at PCE and FUNCT is allocated from the PCE controlled Function ID block.

6. Other Considerations

In case of multiple PCEs, a PCC MAY decide to give control over different ID space to each instance of the PCE. In case a PCC includes the same ID space to multiple PCEs, the PCE MUST use synchronization mechanism (such as [I-D.ietf-pce-state-sync]) to avoid allocating the same ID.

The PCE would allocate ID from the PCE controlled ID space. The PCC would not allocate ID by itself from this space as long as it has an active PCEP session to a PCE to which it has given control over the ID space.

Note that if there is any change in the ID space, the PCC MUST bring the session down and re-establish the session with new TLVs. During state synchronization the PCE would need to consider the new ID space into consideration and SHOULD re-establish the LSP/SR-paths if needed.

The PCC can regain control of the ID space by closing the PCEP session and require new session without ID space TLVs specified in this document.

7. IANA Considerations

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" registry. This document requests IANA actions to allocate code points for the protocol elements defined in this document.

7.1. PCEP TLV Type Indicators

IANA maintains a subregistry called "PCEP TLV Type Indicators". IANA is requested to make an assignment from this subregistry as follows:

Value	Meaning	Reference
TBD1	LABEL-CONTROL-SPACE TLV	[This.I-D]
TBD2	FUNCT-ID-CONTROL-SPACE TLV	[This.I-D]

7.2. LABEL-CONTROL-SPACE TLV's Flag field

This document defines the LABEL-CONTROL-SPACE TLV and requests that IANA to create a new sub-registry to manage the value of the LABEL-CONTROL-SPACE TLV's 24-bits Flag field. New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

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

Currently, there is no allocation in this registry.

Bit	Name	Reference
0-23	Unassigned	[This.I-D]

7.3. FUNCT-ID-CONTROL-SPACE TLV's Flag field

This document defines the FUNCT-ID-CONTROL-SPACE TLV and requests that IANA to create a new sub-registry to manage the value of the FUNCT-ID-CONTROL-SPACE TLV's 24-bits Flag field. New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

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

Currently, there is no allocation in this registry.

Bit	Name	Reference
23	L-Bit	[This.I-D]
0-22	Unassigned	[This.I-D]

8. Security Considerations

The security considerations described in [RFC9050], [I-D.ietf-pce-pcep-extension-pce-controller-sr], and [I-D.dhody-pce-pcep-extension-pce-controller-srv6] and 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. Acknowledgements

TBD.

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

- [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-ipv6]
Li, C., Negl, M., 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-ipv6-12, 6 March 2022, <<https://www.ietf.org/internet-drafts/draft-ietf-pce-segment-routing-ipv6-12.txt>>.

10.2. Informative References

- [RFC4657] Ash, J., Ed. and J.L. 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>>.
- [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>>.
- [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>>.
- [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>>.

- [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://www.ietf.org/archive/id/draft-ietf-pce-pcep-extension-pce-controller-sr-04.txt>>.
- [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://www.ietf.org/archive/id/draft-ietf-pce-state-sync-01.txt>>.
- [I-D.ietf-spring-segment-routing-policy] Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", Work in Progress, Internet-Draft, draft-ietf-spring-segment-routing-policy-21, 19 March 2022, <<https://www.ietf.org/archive/id/draft-ietf-spring-segment-routing-policy-21.txt>>.
- [I-D.dhody-pce-pcep-extension-pce-controller-srv6] Li, Z., Peng, S., Geng, X., and M. Negi, "Path Computation Element Communication Protocol (PCEP) Procedures and Extensions for Using the PCE as a Central Controller (PCECC) for SRv6 SID Allocation and Distribution.", Work in Progress, Internet-Draft, draft-dhody-pce-pcep-extension-pce-controller-srv6-08, 6 March 2022, <<https://www.ietf.org/internet-drafts/draft-dhody-pce-pcep-extension-pce-controller-srv6-08.txt>>.

[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-pce-binding-label-sid] Sivabalan, S., Filsfils, C., Tantsura, J., Previdi, S., and C. L. (editor), "Carrying Binding Label/Segment Identifier (SID) in PCE-based Networks.", Work in Progress, Internet-Draft, draft-ietf-pce-binding-label-sid-14, 3 March 2022, <<https://www.ietf.org/archive/id/draft-ietf-pce-binding-label-sid-14.txt>>.

Appendix A. Contributors

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

EEmail: dhruv.ietf@gmail.com

Mach Chen
Huawei Technologies
China

EEmail: Mach.chen@huawei.com

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

EEmail: lizhenbin@huawei.com

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

EEmail: jie.dong@huawei.com

Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing
100095
China
Email: c.l@huawei.com

Hang Shi
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing
100095
China
Email: shihang9@huawei.com

Aijun Wang
China Telecom
Beiqijia Town,
Beijing
Changping District, 102209
China
Email: wangaj3@chinatelecom.cn

Weiqiang Cheng
China Mobile
Email: chengweiqiang@chinamobile.com

Chao Zhou
HPE
Email: chaozhou_us@yahoo.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 12, 2019

C. Li
M. Chen
Huawei Technologies
W. Cheng
China Mobile
Z. Li
J. Dong
Huawei Technologies
R. Gandhi
Cisco Systems, Inc.
March 11, 2019

PCEP Extensions for Associated Bidirectional Segment Routing (SR) Paths
draft-li-pce-sr-bidir-path-04

Abstract

The Path Computation Element Communication Protocol (PCEP) provides mechanisms for Path Computation Elements (PCEs) to perform path computations in response to Path Computation Clients (PCCs) requests.

The Stateful PCE extensions allow stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Label Switched Paths (LSPs) using PCEP. Furthermore, PCEP can be used for computing paths in Segment Routing (SR) TE networks.

This document defines PCEP extensions for grouping two reverse unidirectional SR Paths into an Associated Bidirectional SR Path when using a Stateful PCE for both PCE-Initiated and PCC-Initiated LSPs as well as when using a Stateless PCE.

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 9, 2019.

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
2. Terminology	4
2.1. Requirements Language	4
3. PCEP Extension for Bidirectional SR Path	4
3.1. Double-sided Bidirectional SR Path Association Group Object	5
4. Bidirectional Flag	6
5. Procedures for Associated Bidirectional SR Path Computation	6
5.1. PCE Initiated Associated Bidirectional SR Paths	7
5.2. PCC Initiated Associated Bidirectional SR Paths	8
5.3. Error Handling	9
6. IANA Considerations	10
6.1. Association Type	10
6.2. PCEP Errors	10
7. Security Considerations	10
8. Contributors	11
9. Acknowledgments	11
10. References	12
10.1. Normative References	12
10.2. Informative References	13
Authors' Addresses	14

1. Introduction

Segment routing (SR) [RFC8402] leverages the source routing and tunneling paradigms. SR supports to steer packets into an explicit forwarding path at the ingress node.

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables the communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalized MPLS (GMPLS) Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of TE LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP State Synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions. The model of operation where LSPs are initiated from the PCE is described in [RFC8281].

[I-D.ietf-pce-segment-routing] specifies extensions to the Path Computation Element Protocol (PCEP) [RFC5440] for SR networks, that allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request, report or delegate SR Paths.

[I-D.ietf-pce-segment-routing-ipv6] extend PCEP to support SR for IPv6 data plane.

[I-D.ietf-pce-association-group] introduces a generic mechanism to create a grouping of LSPs which can then be used to define associations between a set of LSPs and/or a set of attributes, for example primary and secondary LSP associations, and is equally applicable to the active and passive modes of a Stateful PCE [RFC8231] or a stateless PCE [RFC5440].

Currently, SR networks only support unidirectional paths. However, bidirectional SR Paths are required in some networks, for example, in mobile backhaul transport networks. The requirement of bidirectional SR Path is specified in [I-D.ietf-spring-mpls-path-segment].

[I-D.ietf-pce-association-bidir] defines PCEP extensions for grouping two reverse unidirectional MPLS TE LSPs into an Associated Bidirectional LSP when using a Stateful PCE for both PCE-Initiated and PCC-Initiated LSPs as well as when using a Stateless PCE.

This document extends the bidirectional association to segment routing by specifying PCEP extensions for grouping two reverse unidirectional SR Paths into a bidirectional SR Path.

[I-D.ietf-pce-association-bidir] specifies the Double-sided Bidirectional LSP Association procedure, where the PCE creates the association and provisions at both endpoints, the RSVP-TE does the signaling to the egress the status of the forward LSP and the ingress about the reverse LSP. Thus, the both endpoints learn the reverse LSPs forming the bidirectional LSP association. In case of SR, to support the bidirectional path use-case, this is done using the PCEP protocol. This is done so that both endpoints are aware of the the unidirectional SR Path, as well as the status and other SR path related information.

[I-D.li-pce-sr-path-segment] defines a procedure for Path Segment Identifier (PSID) in PCEP for SR using PATH-SEGMENT TLV. The PSID can be a Path Segment Identifier in SR-MPLS [I-D.ietf-spring-mpls-path-segment], or a Path Segment Identifier in SRv6 [I-D.li-spring-srv6-path-segment]. The PSID can be used for an associated bidirectional SR Path for identifying the SR Path.

2. Terminology

This document makes use of the terms defined in [I-D.ietf-pce-segment-routing]. The reader is assumed to be familiar with the terminology defined in [RFC5440], [RFC8231], [RFC8281], [I-D.ietf-pce-association-group] and [I-D.ietf-pce-association-bidir].

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. PCEP Extension for Bidirectional SR Path

As per [I-D.ietf-pce-association-group], LSPs are associated by adding them to a common association group.

[I-D.ietf-pce-association-bidir] specifies PCEP extensions for grouping two reverse unidirectional MPLS-TE LSPs into an Associated Bidirectional LSP for both single-sided and double-sided initiation cases by defining two new Bidirectional LSP Association Groups.

This document extends the procedure for associated bidirectional SR Paths by defining a new bidirectional association group (Double-sided Bidirectional SR Path Association Group). The document further

describes the mechanism for associating two unidirectional SR Paths into a bidirectional SR Path. [I-D.li-pce-sr-path-segment] defines a procedure for communicating Path Segment in PCEP for SR using PATH-SEGMENT TLV. The bidirectional SR Path can also use the PATH-SEGMENT TLV.

Note that an association group is defined in this document to define procedures specific to SR Paths (and the procedures are different than the RSVP-TE bidirectional association groups defined in [I-D.ietf-pce-association-bidir]).

3.1. Double-sided Bidirectional SR Path Association Group Object

As defined in [I-D.ietf-pce-association-bidir], two LSPs are associated as a bidirectional MPLS-TE LSP by a common bidirectional LSP association group. For associating two SR paths, this document defines a new association group called 'Double-sided Bidirectional SR Path Association Group' as follows:

- o Association Type (TBD1 to be assigned by IANA) = Double-sided Bidirectional SR Path Association Group

Similar to other bidirectional associations, this Association Type is operator-configured in nature and statically created by the operator on the PCEP peers. The paths belonging to this association is conveyed via PCEP messages to the PCEP peer. Operator-configured Association Range TLV [I-D.ietf-pce-association-group] MUST NOT be sent for these Association Types, and MUST be ignored, so that the entire range of association ID can be used for them. The handling of the Association ID, Association Source, optional Global Association Source and optional Extended Association ID in this association are set in the same way as [I-D.ietf-pce-association-bidir].

A member of the 'Double-sided Bidirectional SR Path Association Group' can take the role of a forward or reverse SR Path and follow the similar rules defined in [I-D.ietf-pce-association-bidir] for LSPs.

- o An SR Path (forward or reverse) can not be part of more than one 'Double-sided Bidirectional SR Path Association Group'.
- o The endpoints of the SR Paths in this associations cannot be different.

For describing the SR Paths in this association group, such as direction and co-routed information, this association group reuses the Bidirectional LSP Association Group TLV defined in [I-D.ietf-pce-association-bidir]. All fields and processing rules

are as per [I-D.ietf-pce-association-bidir].

4. Bidirectional Flag

As defined in [RFC5440], the B-flag in RP object MUST be set when the PCC specifies that the path computation request relates to a bidirectional TE LSP. In this document, the B-flag also MUST be set when the PCC specifies that the path computation request relates to a bidirectional SR Path. When a stateful PCE initiates or updates a bidirectional SR Paths including LSPs and SR paths, the B-flag in SRP object [I-D.ietf-pce-pcep-stateful-pce-gmpls] MAY be set as well.

5. Procedures for Associated Bidirectional SR Path Computation

Two unidirectional SR Paths can be associated by the association group object as specified in [I-D.ietf-pce-association-group]. A bidirectional LSP association group object is defined in [I-D.ietf-pce-association-bidir] (for MPLS-TE). This document extends these association mechanisms for bidirectional SR Paths. Two SR Paths can be associated together by using the Bidirectional SR Path Association Group defined in this document for PCEP messages. The PATH-SEGMENT TLV [I-D.li-pce-sr-path-segment] SHOULD also be included in the LSP object for these SR Paths to support required use-cases.

For bidirectional SR Paths, there is a need to know the reverse direction SR paths. The PCE SHOULD inform the reverse SR Paths to the ingress PCCs and vice versa. To achieve this, a PCInitiate message for the reverse SR Path is sent to the ingress PCC and a PCInitiate message for the forward SR Path is sent to the egress PCC (with the same association group). These PCInitiate message MUST NOT trigger initiation of SR Paths. The reverse direction SR Path can be used for several use-cases, such as directed BFD [I-D.ietf-mpls-bfd-directed].

For a bidirectional LSP computation when using both direction LSPs on a node, the same LSP would need to be identified using 2 different PLSP-IDs based on the PCEP session to the ingress or the egress. In other words, the LSP will have a PLSP-ID A at the ingress node while it will have the PLSP-ID B at the egress node. The PCE will maintain the two PLSP-IDs for the same LSP. For instance, an ingress PCC requests a bidirectional SR Path computation, and the PCE computes a forward LSP1 with PLSP-ID say 100. The reverse LSP2 from the egress to the ingress with PLSP-ID say 200 is allocated by the egress PCC. Since the PLSP-ID space is independent at each PCC, the PLSP-ID allocated by the egress PCC can not be used for the LSP at the

ingress PCC (PLSP-ID conflict may occur). Hence, the PCE needs to allocate a PLSP-ID for LSP2 from the ingress PCC's PLSP-ID space , say 101. Similarly for LSP1, it has PLSP-ID 100 at the ingress, and may have say PLSP-ID 201 at the egress node.

5.1. PCE Initiated Associated Bidirectional SR Paths

As specified in [I-D.ietf-pce-association-group], Bidirectional SR Path Association Group can be created by a Stateful PCE.

- o Stateful PCE can create and update the forward and reverse SR Paths independently for a 'Double-sided Bidirectional SR Path Association Group'.
- o Stateful PCE can establish and remove the association relationship on a per SR Path basis.
- o Stateful PCE can create and update the SR Path and the association on a PCC via PCInitiate and PCUpd messages, respectively, using the procedures described in [I-D.ietf-pce-association-group].
- o The PATH-SEGMENT TLV SHOULD be included for each SR Path in the LSP object.
- o The reverse direction SR Path (LSP2(R) at node S, LSP1(R) at node D) SHOULD be informed by PCE via PCInitiate message with the matching association group.

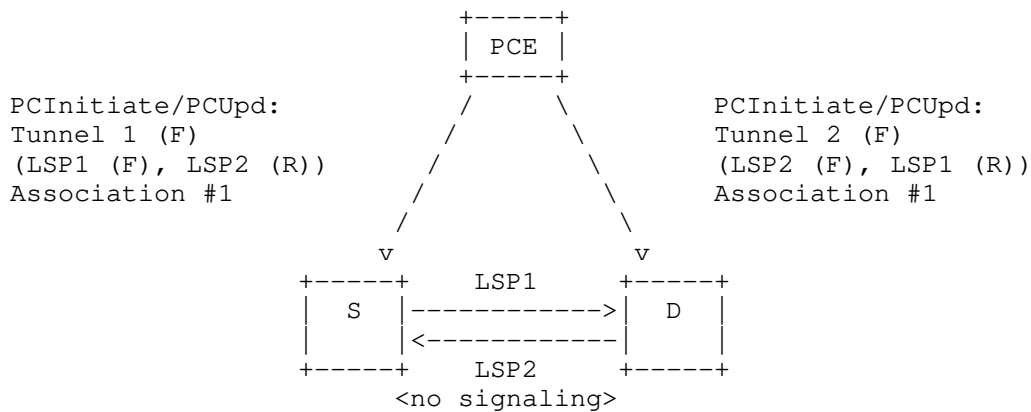


Figure 1: PCE-Initiated Double-sided Bidirectional SR Path with Forward and Reverse Direction SR Paths

5.2. PCC Initiated Associated Bidirectional SR Paths

As specified in [I-D.ietf-pce-association-group], Bidirectional SR Path Association Group can also be created by a PCC.

- o PCC can create and update the forward and reverse SR Paths independently for a 'Double-sided Bidirectional SR Path Association Group'.
- o PCC can establish and remove the association relationship on a per SR Path basis.
- o PCC MUST report the change in the association group of an SR Path to PCE(s) via PCRpt message.
- o PCC can report the forward and reverse SR Paths independently to PCE(s) via PCRpt message.
- o PCC can delegate the forward and reverse SR Paths independently to a Stateful PCE, where PCE would control the SR Paths.
- o Stateful PCE can update the SR Paths in the 'Double-sided Bidirectional SR Path Association Group' via PCUpd message, using the procedures described in [I-D.ietf-pce-association-group].
- o The PATH-SEGMENT TLV MUST be handled as defined in [I-D.li-pce-sr-path-segment].
- o The reverse direction SR Path (LSP2(R) at node S, LSP1(R) at node D) SHOULD be informed by PCE via PCInitiate message with the matching association group.

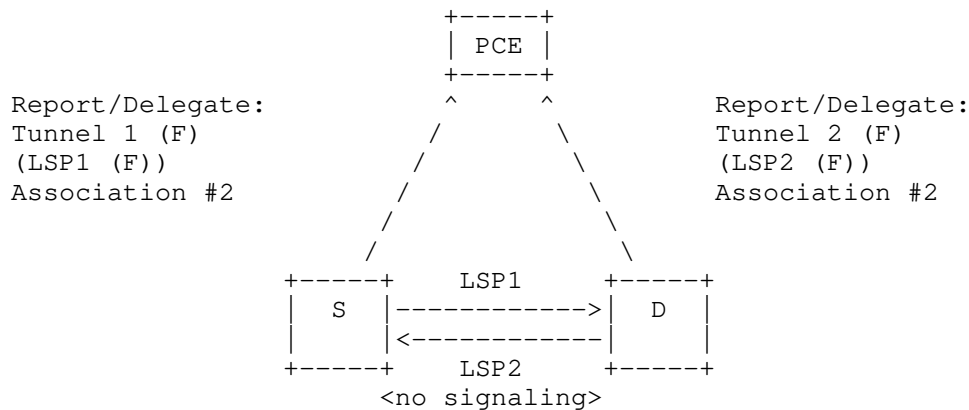


Figure 2a: Step 1: PCC-Initiated Double-sided Bidirectional SR Path with Forward Direction SR Paths

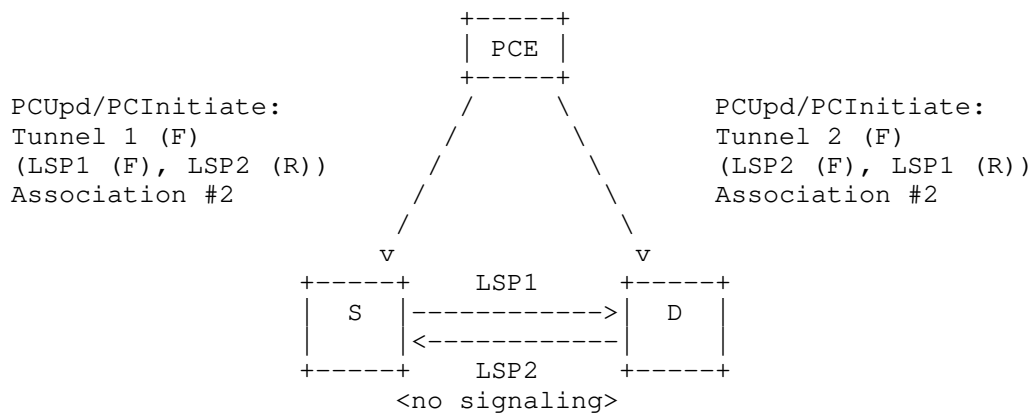


Figure 2b: Step 2: PCE-Upd/Initiated Double-sided Bidirectional Path Along with Reverse Direction SR Paths

5.3. Error Handling

The error handling as described in section 5.5 of [I-D.ietf-pce-association-bidir] continue to apply.

The PCEP Path Setup Type (PST) MUST be set to 'TE Path is Setup using Segment Routing' [I-D.ietf-pce-segment-routing] for the LSP belonging to the 'Double-sided Bidirectional SR Path Association Group'. In case a PCEP speaker receives a different PST value for this association group, it MUST send a PCErr message with Error-Type = 29

(Early allocation by IANA) (Association Error) and Error-Value = TBD2 (Bidirectional LSP Association - Path Setup Type Mismatch).

6. IANA Considerations

6.1. Association Type

This document defines a new Association Type for the Association Object defined [I-D.ietf-pce-association-group]. IANA is requested to make the assignment of a value for the sub-registry "ASSOCIATION Type Field" (to be created in [I-D.ietf-pce-association-group]), as follows:

Value	Name	Reference
TBD1	Double-sided Bidirectional SR Path Association Group	This document

6.2. PCEP Errors

This document defines new Error value for Error Type 29 (Association Error). IANA is requested to allocate new Error value within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry, as follows:

Error Type	Description	Reference
29	Association Error	
	Error value: TBD2 Bidirectional LSP Association - Path Setup Type Mismatch	This document

7. Security Considerations

The security considerations described in [RFC5440], [RFC8231], [RFC8281], and [I-D.ietf-pce-segment-routing] apply to the extensions defined in this document as well.

A new Association Type for the Association Object, 'Double-sided Associated Bidirectional SR Path Association Group' is introduced in this document. Additional security considerations related to LSP associations due to a malicious PCEP speaker is described in [I-D.ietf-pce-association-group] and apply to this Association Type. Hence, securing the PCEP session using Transport Layer Security (TLS)

[RFC8253] is recommended.

8. Contributors

The following people have substantially contributed to this document:

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

Email: dhruv.ietf@gmail.com

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan, Hubei 430223
China

Email: xiong.quan@zte.com.cn

9. Acknowledgments

Many thanks to Marina Fitzgeer for detailed review and comments.

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>>.
- [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-association-group] Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "PCEP Extensions for Establishing Relationships Between Sets of LSPs", draft-ietf-pce-association-group-08 (work in progress), March 2019.
- [I-D.ietf-pce-association-bidir] Gandhi, R., Barth, C., and B. Wen, "PCEP Extensions for Associated Bidirectional Label Switched Paths (LSPs)", draft-ietf-pce-association-bidir (work in progress).
- [I-D.ietf-pce-pcep-stateful-pce-gmpls] Lee, Y., Zhang, F., Casellas, R., Dios, O., and Z. Ali, "Path Computation Element (PCE) Protocol Extensions for Stateful PCE Usage in GMPLS-controlled Networks", draft-ietf-pce-pcep-stateful-pce-gmpls-10 (work in progress), March 2019.

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

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

[I-D.ietf-spring-mpls-path-segment]

Cheng, W., Li, H., Chen, M., Gandhi, R., and R. Zigler, "Path Segment in MPLS Based Segment Routing Network", draft-ietf-spring-mpls-path-segment (work in progress).

[I-D.li-spring-srv6-path-segment]

Li, C., Chen, M., Dhody, D., Li, Z., Dong, J., and R. Gandhi, "Path Segment for SRv6 (Segment Routing in IPv6)", draft-li-spring-srv6-path-segment-00 (work in progress), October 2018.

10.2. Informative References

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

[I-D.ietf-mpls-bfd-directed]

Mirsky, G., Tantsura, J., Varlashkin, I., and M. Chen,
"Bidirectional Forwarding Detection (BFD) Directed Return
Path", draft-ietf-mpls-bfd-directed-10 (work in progress),
September 2018.

Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: chengli13@huawei.com

Mach(Guoyi) Chen
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: Mach.chen@huawei.com

Weiqiang Cheng
China Mobile
China

Email: chengweiqiang@chinamobile.com

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

Email: lizhenbin@huawei.com

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: jie.dong@huawei.com

Rakesh Gandhi
Cisco Systems, Inc.
Canada

Email: rgandhi@cisco.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 9, 2019

C. Li
M. Chen
Huawei Technologies
W. Cheng
China Mobile
J. Dong
Z. Li
Huawei Technologies
R. Gandhi
Cisco Systems, Inc.
March 08, 2019

Path Computation Element Communication Protocol (PCEP) Extension for
Path Segment in Segment Routing (SR)
draft-li-pce-sr-path-segment-04

Abstract

The Path Computation Element (PCE) provides path computation functions in support of traffic engineering in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

The Source Packet Routing in Networking (SPRING) architecture describes how Segment Routing (SR) can be used to steer packets through an IPv6 or MPLS network using the source routing paradigm. A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a Path Computation Element (PCE).

Path identification is needed for several use cases such as performance measurement in Segment Routing (SR) network. This document specifies extensions to the Path Computation Element Protocol (PCEP) to support requesting, replying, reporting and updating the Path Segment ID (Path SID) between PCEP speakers.

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 9, 2019.

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
2. Terminology	4
2.1. Requirements Language	4
3. Overview of Path Segment Extensions in PCEP	4
4. Objects and TLVs	5
4.1. The OPEN Object	5
4.1.1. The SR PCE Capability sub-TLV	5
4.1.2. The SRv6 PCE Capability sub-TLV	6
4.1.3. PCECC-CAPABILITY sub-TLV	6
4.2. LSP Object	7
4.2.1. Path Segment TLV	7
4.3. FEC Object	9
4.4. CCI Object	10
5. Operations	11
5.1. PCC Allocated Path Segment	11
5.1.1. Egress PCC Allocated Path Segment	11
5.2. PCE Allocated Path Segment	15
5.2.1. PCE Controlled Label Spaces Advertisement	15
5.2.2. Ingress PCC request Path Segment to PCE	15
5.2.3. PCE allocated Path Segment on its own	17
6. Dataplane Considerations	17
7. IANA Considerations	18
7.1. SR PCE Capability Flags	18
7.2. SRv6 PCE Capability Flags	18
7.3. New LSP Flag Registry	18
7.4. New PCEP TLV	19

7.4.1. Path Segment TLV	19
7.5. New CCI Flag Registry	19
7.6. New FEC Type Registry	20
7.7. PCEP Error Type and Value	20
8. Security Considerations	20
9. Acknowledgments	20
10. Contributors	20
11. References	21
11.1. Normative References	21
11.2. Informative References	23
Authors' Addresses	23

1. Introduction

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables the communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalized MPLS (GMPLS) Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of TE LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP State Synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions. The model of operation where LSPs are initiated from the PCE is described in [RFC8281].

[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) [RFC8402] leverages the source routing and tunneling paradigms and supports steering packets into an explicit forwarding path at the ingress node.

An SR path needs to be identified in some use cases such as performance measurement. For identifying an SR path, [I-D.cheng-spring-mpls-path-segment] introduces a new segment that is referred to as Path Segment.

[I-D.ietf-pce-segment-routing] specifies extensions to the Path Computation Element Protocol (PCEP) [RFC5440] for SR networks, that allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request, report or delegate SR paths.

[I-D.negi-pce-segment-routing-ipv6] extend PCEP to support SR paths for IPv6 data plane.

[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 specifies a mechanism to carry the SR path identification information in PCEP messages [RFC5440] [RFC8231] [RFC8281]. The SR path identifier can be a Path Segment in SR-MPLS [I-D.cheng-spring-mpls-path-segment], or a Path Segment in SRv6 [I-D.li-spring-srv6-path-segment] or other IDs that can identify an SR path. This document also extends the PCECC-SR mechanism to inform the Path Segment to the egress PCC.

2. Terminology

This memo makes use of the terms defined in [RFC4655], [I-D.ietf-pce-segment-routing], and [RFC8402].

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. Overview of Path Segment Extensions in PCEP

This document specifies a mechanism of encoding (and allocating) Path Segment in PCEP extensions. For supporting Path Segment in PCEP, several TLVs and flags are defined. The formats of the objects and TLVs are described in Section 4. The procedures of Path Segment allocation are described in Section 5.

There are various modes of operations, such as -

- o The Path Segment can be allocated by Egress PCC. The PCE should request the Path Segment from Egress PCC.
- o The PCE can allocate a Path Segment on its own accord and inform the ingress/egress PCC, useful for PCE-initiated LSPs.

- o Ingress PCC can also request PCE to allocate the Path Segment, in this case, the PCE would either allocate and inform the assigned Path Segment to the ingress/egress PCC using PCEP messages, or first request egress PCC for Path Segment and then inform it to the ingress PCC.

The path information to the ingress PCC and PCE is exchanged via an extension to [I-D.ietf-pce-segment-routing] and [I-D.negi-pce-segment-routing-ipv6]. The Path Segment information to the egress PCC can be informed via an extension to the PCECC-SR procedures [I-D.zhao-pce-pcep-extension-pce-controller-sr].

For the PCE to allocate a Path Segment, the PCE SHOULD be aware of the MPLS label space from the PCCs. This is done via mechanism as described in [I-D.li-pce-controlled-id-space]. Otherwise, the PCE should request the egress PCC for Path Segment allocation.

4. Objects and TLVs

4.1. The OPEN Object

4.1.1. The SR PCE Capability sub-TLV

[I-D.ietf-pce-segment-routing] defined a new Path Setup Type (PST) and SR-PCE-CAPABILITY sub-TLV for SR. PCEP speakers use this sub-TLV to exchange information about their SR capability. The TLV defines a Flags field that includes one bit (L-flag) to indicate Local Significance [I-D.ietf-pce-segment-routing].

This document adds an additional flag for Path Segment allocation, as follows -

P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SR path identification (Path Segment, as per [I-D.cheng-spring-mpls-path-segment]).

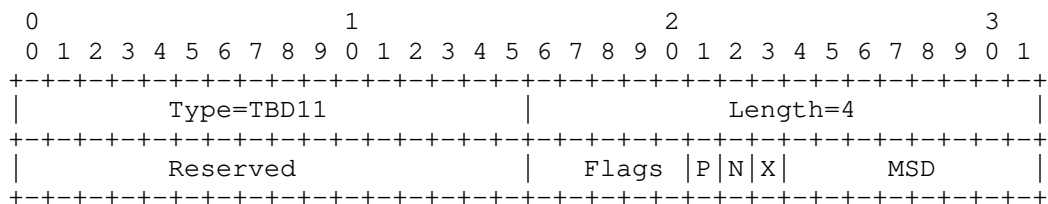


Figure 1: P-flag in SR-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and can be removed at the time of publication.

4.1.2. The SRv6 PCE Capability sub-TLV

[I-D.negi-pce-segment-routing-ipv6] defined a new Path Setup Type (PST) and SRv6-PCE-CAPABILITY sub-TLV for SRv6. PCEP speakers use this sub-TLV to exchange information about their SRv6 capability. The TLV includes a Flags field and one bit (L-flag) was allocated in [I-D.negi-pce-segment-routing-ipv6].

This document adds an additional flag for Path Segment allocation, as follows -

P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SRv6 path identification. (Path Segment, as per [I-D.li-spring-srv6-path-segment]).

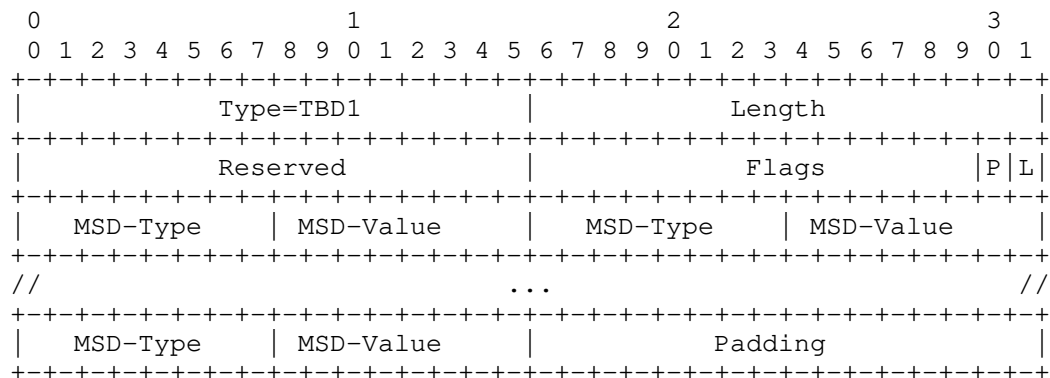


Figure 2: P-flag in SRv6-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and can be removed at the time of publication.

4.1.3. PCECC-CAPABILITY sub-TLV

Along with the SR sub-TLVs, the PCECC Capability as per [I-D.zhao-pce-pcep-extension-pce-controller-sr] should be advertised if the PCE allocates the Path Segment and acts as a Central Controller that manages the Label space.

The PCECC Capability should also be advertised on the egress PCEP session, along with the SR sub-TLVs. This is needed to ensure that

the PCE can use the PCECC objects/mechanism to request/inform the egress PCC of the Path Segment as described in this document.

4.2. LSP Object

The LSP Object is defined in Section 7.3 of [RFC8231]. This document adds the following flags to the LSP Object:

P (PCE Allocation bit): If the bit is set to 1, it indicates that the PCC requests PCE to allocate resource for this LSP. With the resource TLV, a PCE can understand what kind of resource should be allocated, such as Path Segment and Binding Segment. A PCC would set this bit to 1 and include a PATH-SEGMENT TLV in the LSP object to request for allocation of Path Segment by the PCE in the PCReq or PCRpt message. A PCE would also set this bit to 1 and include a PATH-SEGMENT TLV to indicate that the Path Segment is allocated by PCE and encoded in the PCRep, PCUpd or PCInitiate message. Further, a PCE would set this bit to 0 and include a PATH-SEGMENT TLV in the LSP object to indicate that the Path Segment should be allocated by the PCC as described in Section 5.1.1.

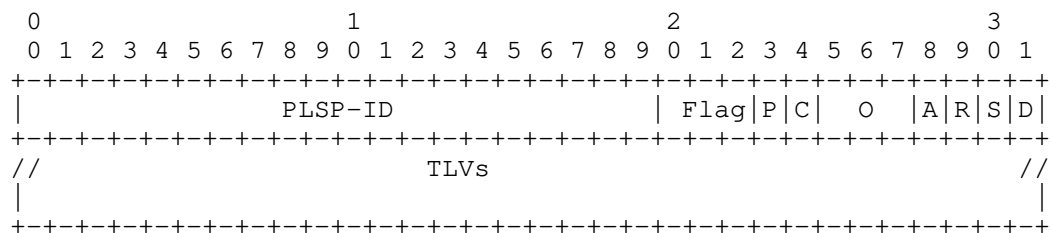


Figure 3: P-flag in LSP Object

The figure is included for the ease of the reader and can be removed at the time of publication.

4.2.1. Path Segment TLV

The PATH-SEGMENT TLV is an optional TLV for use in the LSP Object for Path Segment allocation. The type of this TLV is to be allocated by IANA (TBA4). The format is shown below.

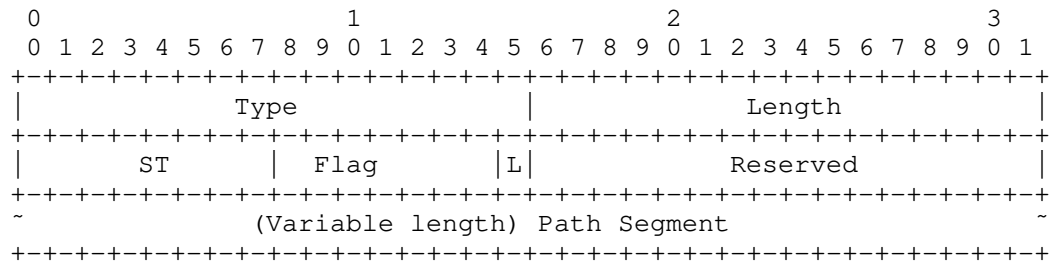


Figure 4: The PATH-SEGMENT TLV Format

The type (16-bit) of the TLV is TBA4 (to be allocated by IANA). The length (16-bit) has a fixed value of 8 octets. The value contains the following fields:

ST (The Segment type - 8 bits): The ST field specifies the type of the Path Segment field, which carries a Path Segment corresponding to the SR path.

- * 0: MPLS Path Segment, which is an MPLS label as defined in [I-D.cheng-spring-mpls-path-segment]. The PST type MUST be set to SR (MPLS).
- * 1: SRv6 Path Segment, which is a 128 bit IPv6 address as defined in [I-D.li-spring-srv6-path-segment]. The PST type MUST be set to SRv6.

Flags (8 bits): Two flags are currently defined:

- * L-Bit (Local/Global - 1 bit): If set, then the Path Segment carried by the PATH-SEGMENT TLV has local significance. If not set, then the Path Segment carried by this TLV has global significance (i.e. Path Segment is global within an SR domain).
- * The unassigned bits MUST be set to 0 and MUST be ignored at receipt.

Reserved (16 bits): MUST be set to 0 and MUST be ignored at receipt.

Path Segment: The Path Segment of an SR path. The Path Segment type is indicated by the ST field. When the ST is 0, it is a MPLS Path Segment [I-D.cheng-spring-mpls-path-segment] in the MPLS label format. When the ST field is 1, it is a 128-bit SRv6 Path Segment as defined in [I-D.li-spring-srv6-path-segment].

In general, only one instance of PATH-SEGMENT TLV will be included in LSP object. If more than one PATH-SEGMENT TLV is included, the first one is processed and others MUST be ignored. Multiple Path Segment allocation for use cases like alternate-making will be considered in future version of this draft.

When the Path Segment allocation is enable, a PATH-SEGMENT TLV MUST be included in the LSP object.

If the label space is maintained by PCC itself, and the Path Segment is allocated by Egress PCC, then the PCE should request the Path Segment from Egress PCC as described in Section 5.1.1. In this case, the PCE should send a PCUpdate or PCInitiate message to the egress PCC to request the Path Segment. The P-flag in LSP should be unset in this case.

If a PCEP node does not recognize the PATH-SEGMENT TLV, it would behave in accordance with [RFC5440] and ignore the TLV. If a PCEP node recognizes the TLV but does not support the TLV, it MUST send PCError with Error-Type = 2 (Capability not supported).

4.3. FEC Object

The FEC Object [I-D.zhao-pce-pcep-extension-pce-controller-sr] is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message for the PCECC-SR operations. The PCE MUST inform the Path Identification information to the Egress PCC. To do this, this document extends the procedures of [I-D.zhao-pce-pcep-extension-pce-controller-sr] by defining a new FEC object type for Path.

FEC Object-Type is TBA6 'Path'.

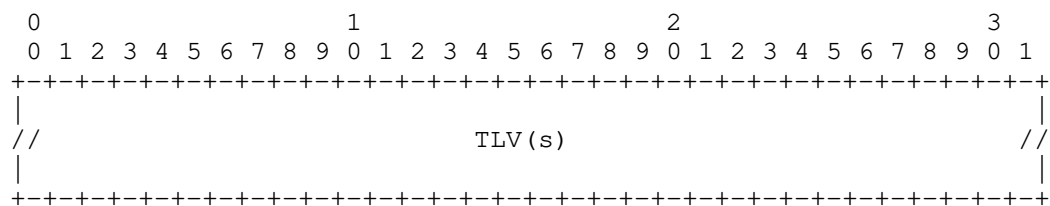


Figure 5: The path FEC object Format

One or more following TLV(s) are allowed in the 'path' FEC object -

- o SYMBOLIC-PATH-NAME TLV: As defined in [RFC8231], it is a human-readable string that identifies an LSP in the network.

- o LSP-IDENTIFIERS TLVs: As defined in [RFC8231], it is optional for SR, but could be used to encode the source, destination and other identification information for the path.
- o SPEAKER-ENTITY-ID TLV: As defined in [RFC8232], a unique identifier for the PCEP speaker, it is used to identify the Ingress PCC.

Either SYMBOLIC-PATH-NAME TLV or LSP-IDENTIFIERS TLV MUST be included. SPEAKER-ENTITY-ID TLV is optional. Only one instance of each TLV is processed, if more than one TLV of each type is included, the first one is processed and others MUST be ignored.

4.4. 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]. Further [I-D.zhao-pce-pcep-extension-pce-controller-sr] defined a CCI object type for SR.

The Path Segment information is encoded directly in the CCI SR object. The Path Segment TLV as described in the Section 4.2.1, MUST also be included in the CCI SR object as the TLV (as it includes additional information regarding the Path Segment identifier).

This document adds the following flags to the CCI Object:

- o C (PCC Allocation bit): 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 space. A PCC would set this bit to indicate that it has allocated the CC-ID and report it to the PCE.

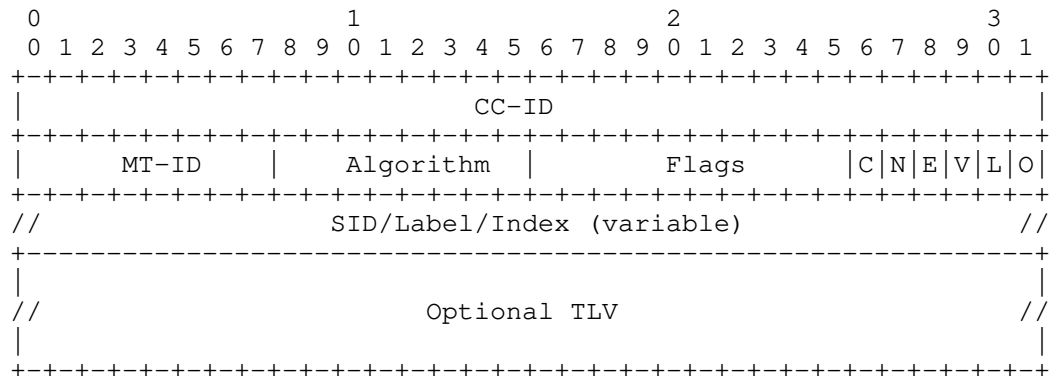


Figure 6: The CCI object for SR

(Editor's Note - An update is planned for [I-D.zhao-pce-pcep-extension-pce-controller-sr] in the next revision detailing this procedure, and the above text might move there.)

5. Operations

The Path Segment allocation and encoding is as per the stateful PCE operations for segment routing. The procedures are as per the corresponding extensions defined in [I-D.ietf-pce-segment-routing] and [I-D.negi-pce-segment-routing-ipv6] (which are further based on [RFC8231] and [RFC8281]). The additional operations for Path Segment are defined in this section.

To notify (or request) the Path Segment to the Egress PCC, the procedures are as per the PCECC-SR [I-D.zhao-pce-pcep-extension-pce-controller-sr] (which is based on [I-D.ietf-pce-pcep-extension-for-pce-controller]). The additional operations are defined in this section.

5.1. PCC Allocated Path Segment

5.1.1. Egress PCC Allocated Path Segment

As defined in [I-D.cheng-spring-mpls-path-segment], a Path Segment can be allocated by the egress PCC. In this case, the label space may be maintained on the PCC itself.

On receiving a stateful path computation request with Path Segment allocation request from an ingress PCC, or by initiating or updating an LSP with Path Segment actively, a PCE can request the egress PCC to allocate a Path Segment. This is needed if the PCE does not

control the Path Segment allocation for the egress PCC or the label space is maintained by the egress PCC itself.

The mechanism of Path Segment request and reply may be achieved by using PCInitiate and PCUpd message as described in this section.

5.1.1.1. Using CCI and FEC objects (PCECC)

The PCE can request the egress to allocate the Path Segment using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. The C flag in the CCI object is set to 1 and the CC-ID is set to a special value of 0x0000 to indicate that the allocation needs to be done by the PCC. The PATH-SEGMENT TLV is also included in CCI object along with the FEC object identifying the SR-Path. The egress PCC would allocate the Path Segment and would report to the PCE using the PCRpt message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr] with the allocated Path Segment in the CC-ID field as well as in the PATH-SEGMENT TLV.

(Editor's Note - An update is planned for [I-D.zhao-pce-pcep-extension-pce-controller-sr] in the next revision detailing this procedure)

If the value of CC-ID/Path Segment is 0 and the C flag is set, it indicates that the PCE is requesting a Path Segment for this LSP. If the CC-ID/Path Segment is set to a value 'n' and the C flag is set in the CCI object, it indicates that the PCE requests a specific value 'n' of Path Segment. If the Path Segment is allocated successfully, the egress PCC should report the Path Segment via PCRpt message with the CCI object along with the PATH-SEGMENT TLV. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment in CCI object is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

Once the PCE receives the PCRpt message with the CCI object, it can obtain the Path Segment information from the egress PCC and then update the path with Path Segment or reply to the ingress PCC, the path information with Path Segment.

If the SR-Path is setup the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, as described in [I-D.ietf-pce-segment-routing], a PCErr message will be sent back to the PCE. The PCE MUST request the withdraw of the Path Segment allocation by sending a PCInitiate message to remove the central

controller instruction as per [I-D.zhao-pce-pcep-extension-pce-controller-sr]. When the LSP is deleted or the Path Segment is removed, the PCE should synchronize with the egress PCC.

If the egress PCC wishes to withdraw or modify a previously reported Path Segment value, it MUST send a PCRpt message without any PATH-SEGMENT TLV or with the PATH-SEGMENT TLV containing the new Path Segment respectively in the CCI object. The PCE would further trigger the removal of the central controller instruction as per [I-D.zhao-pce-pcep-extension-pce-controller-sr].

If a PCE wishes to modify a previously requested Path Segment value, it MUST send a new PCInitiate message with an allocation request CC-ID/PATH-SEGMENT TLV containing the new Path Segment value and C flag is set. The PCE should trigger the removal of the older Path Segment next as per [I-D.zhao-pce-pcep-extension-pce-controller-sr].

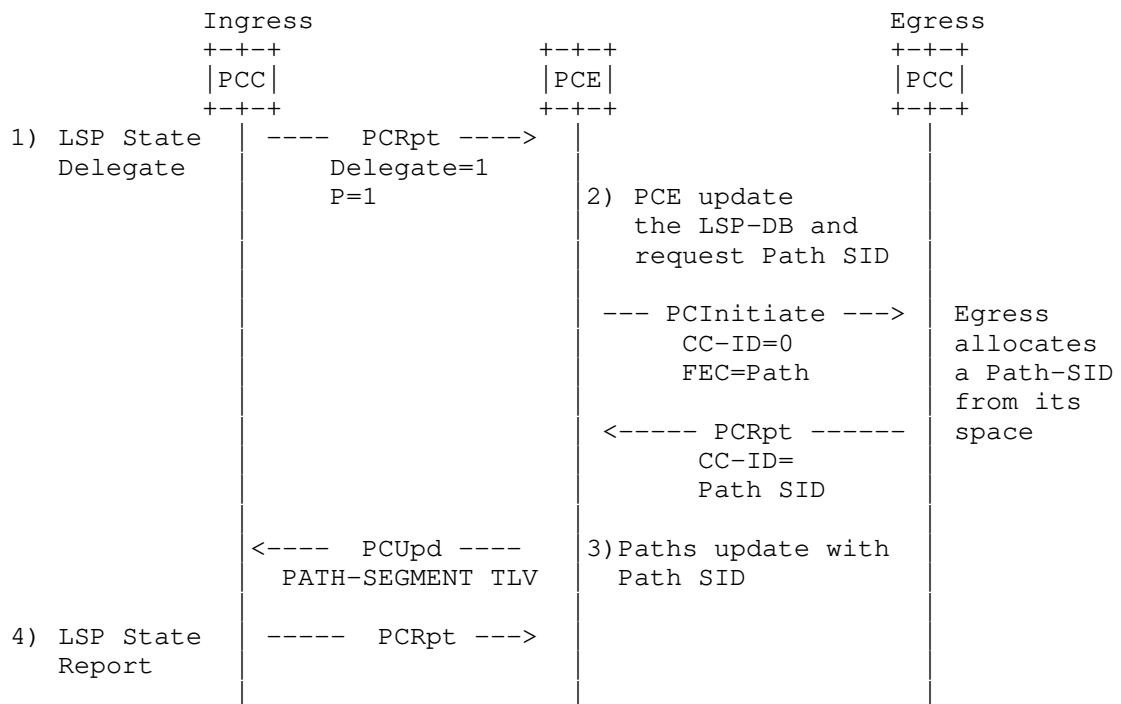


Figure 7: Egress PCC Allocated Path Segment

5.1.1.2. Using LSP objects (PCEP-SR)

The PATH-SEGMENT TLV MUST be included in an LSP object in the PCInitiate message sent from the PCE to the egress to request path identification allocation by the egress PCC. The P flag in LSP object MUST be set to 0. This PCInitiate message to egress PCC would be the similar to the one sent to ingress PCC as per [I-D.ietf-pce-segment-routing], but the egress PCC would only allocate the Path Segment and would not trigger the initiation/update operation.

If the value of Path Segment is 0x0 it indicates that the PCE is requesting a Path Segment for this LSP. If the Path Segment is set to a value 'n' and the P flag is unset in the LSP object, it indicates that the PCE requests a specific value 'n' of Path Segment. If the Path Segment is allocated successfully, the egress PCC should report the Path Segment via PCRpt message with PATH-SEGMENT TLV in LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

Once the PCE receives the PCRpt message, it can obtain the Path Segment information from the egress PCC and then update the path with Path Segment or reply to the ingress PCC, the path information with Path Segment.

If the SR-Path is setup, the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, as described in [I-D.ietf-pce-segment-routing], an PCErr message will be sent back to the PCE. The PCE MUST request the withdraw of the Path Segment allocation by sending a PCUpd message to remove the LSP and associated Path Segment by setting the R flag in the SRP object. When the LSP is deleted or the Path Segment is removed, the PCE should send a PCUpd message to synchronize with the egress PCC.

If the egress PCC wishes to withdraw or modify a previously reported Path Segment value, it MUST send a PCRpt message without any PATH-SEGMENT TLV or with the PATH-SEGMENT TLV containing the new Path Segment respectively.

If a PCE wishes to modify a previously requested Path Segment value, it MUST send a PCUpd message with PATH-SEGMENT TLV containing the new Path Segment value and P flag in LSP object would be unset. Absence of the PATH-SEGMENT TLV in PCUpd message means that the PCE wishes to withdraw the Path Segment.

If a PCC receives a valid Path Segment value from a PCE which is different than the current Path Segment, it MUST try to allocate the new value. If the new Path Segment is successfully allocated, the PCC MUST report the new value to the PCE. Otherwise, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

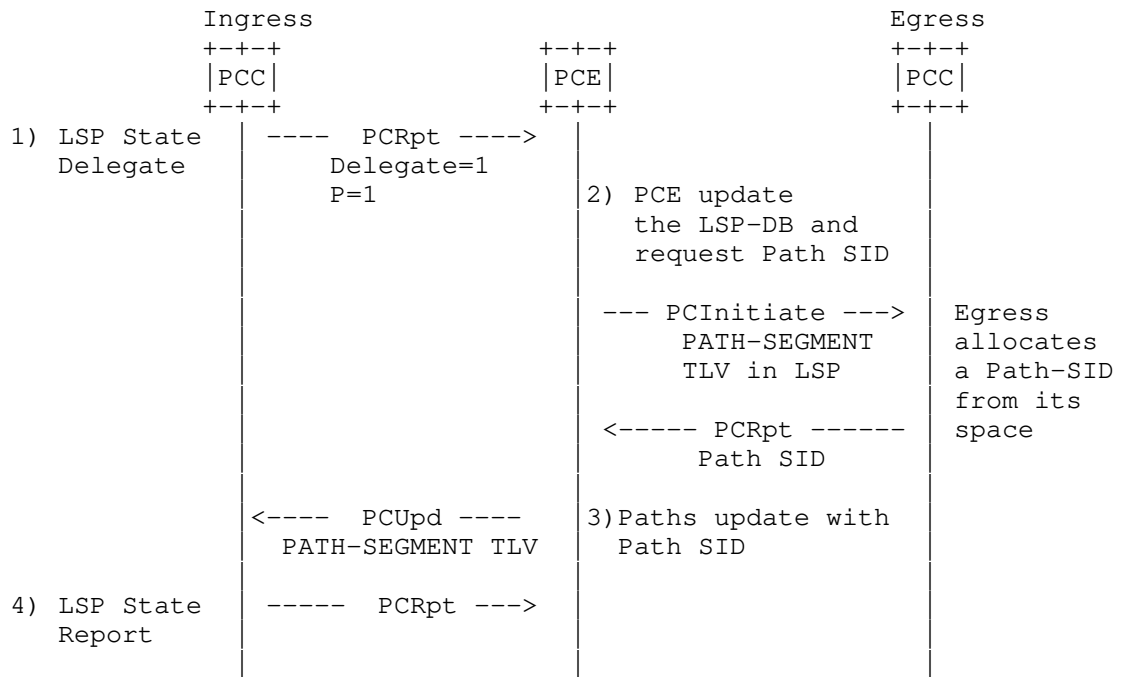


Figure 8: Egress PCC Allocated Path Segment

5.2. PCE Allocated Path Segment

5.2.1. PCE Controlled Label Spaces Advertisement

For allocating the Path Segments to SR paths by the PCEs, the PCE controlled label space MUST be known at PCEs via configurations or any other mechanism. The PCE controlled label spaces MAY be advertised as described in [I-D.li-pce-controlled-id-space].

5.2.2. Ingress PCC request Path Segment to PCE

The ingress PCC could request the Path Segment to be allocated by the PCE via PCRpt message as per [RFC8231]. The delegate flag (D-flag)

MUST also be set for this LSP. Also, the P-flag in the LSP object MUST be set.

A PATH-SEGMENT TLV MUST be included in the LSP object. If the value of Path Segment is 0x0, it indicates that the Ingress PCC is requesting a Path Segment for this LSP. If the Path Segment is set to a value 'n', it indicates that the ingress PCC requests a specific value 'n' of Path Segment.

If the Path Segment is allocated successfully, the PCE would further respond to Ingress PCC with PCUpd message as per [RFC8231] and MUST include the PATH-SEGMENT TLV in a LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

The active PCE would allocate the Path Segment as per the PATH-SEGMENT flags and in case PATH-SEGMENT is not included, the PCE MUST act based on the local policy.

The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

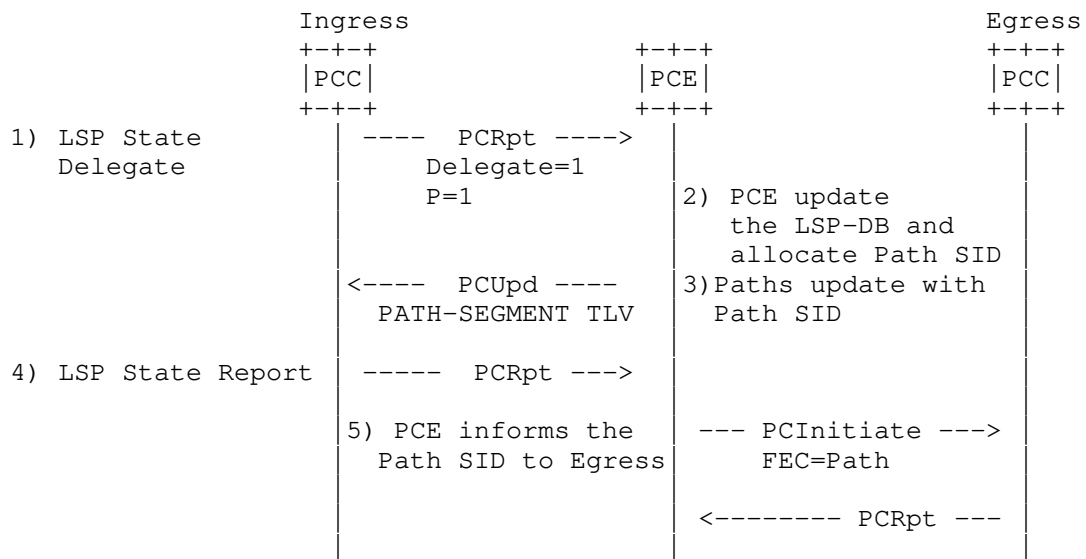


Figure 9: Ingress PCC request Path Segment to PCE

5.2.3. PCE allocated Path Segment on its own

The PCE could allocate the Path Segment on its own for a PCE-Initiated (or delegated LSP). The allocated Path Segment needs to be informed to the Ingress and Egress PCC. The PCE would use the PCInitiate message [RFC8281] or PCUpd message [RFC8231] towards the Ingress PCC and MUST include the PATH-SEGMENT TLV in the LSP object. The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

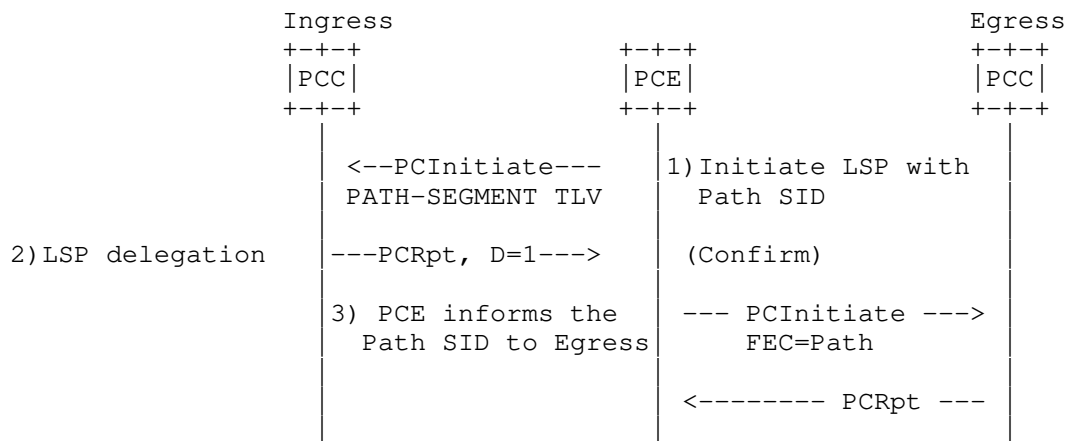


Figure 10: PCE allocated Path Segment on its own

6. Dataplane Considerations

As described in [I-D.cheng-spring-mpls-path-segment], in an SR-MPLS network, when a packet is transmitted along an SR path, the labels in the MPLS label stack will be swapped or popped. So that no label or only the last label may be left in the MPLS label stack when the packet reaches the egress node. Thus, the egress node cannot determine from which SR path the packet comes. For this reason, it introduces the Path Segment.

Apart from allocation and encoding of the Path Segment (described in this document) for the LSP, it would also be included in the SID/Label stack of the LSP (usually for processing by the egress). To support this, the Path Segment MAY also be a part of SR-ERO as prepared by the PCE as per [I-D.ietf-pce-segment-routing]. The PCC MAY also include the Path Segment while preparing the label stack based on the local policy and use-case.

It is important that the PCE learns the Maximum SID Depth (MSD) that can be imposed at each node/link of a given SR path to ensure that the SID stack depth does not exceed the number of SIDs the node is capable of imposing. As a new type of segment, Path Segment will be inserted in the SID list just like other SIDs. Thus, the PCE needs to consider the affect of Path Segment when computing a LSP with Path Segment allocation.

7. IANA Considerations

7.1. SR PCE Capability Flags

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 the following allocation in the aforementioned registry.

Bit	Description	Reference
TBA1	Path Segment Allocation is supported(P)	This document

7.2. SRv6 PCE Capability Flags

SRv6 PCE Capability TLV is defined in defined in [I-D.negi-pce-segment-routing-ipv6], and the registry to manage the Flag field of the SRv6 PCE Capability Flags is requested in [I-D.negi-pce-segment-routing-ipv6]. IANA is requested to make the following allocation in the aforementioned registry.

Bit	Description	Reference
TBA2	Path Segment Allocation is supported(P)	This document

7.3. 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 has allocated a new bit in the "LSP Object Flag Field" subregistry, as follows:

Bit	Description	Reference
TBA3	Request for Path Segment Allocation(P)	This document

7.4. New PCEP TLV

IANA is requested to add the assignment of a new allocation in the existing "PCEP TLV Type Indicators" subregistry as follows:

Value	Description	Reference
TBA4	PATH-SEGMENT TLV	This document

7.4.1. Path Segment TLV

This document requests that a new subregistry named "PATH-SEGMENT TLV Segment Type (ST) Field" to be created to manage the value of the ST field in the PATH-SEGMENT TLV.

Value	Description	Reference
0	MPLS Path Segment (MPLS label)	This document
1	SRv6 Path Segment (IPv6 address)	This document

Further, this document also requests that a new subregistry named "PATH-SEGMENT TLV Flag Field" to be created to manage the Flag field in the PATH-SEGMENT TLV. New values are 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

Bit	Description	Reference
7	Local Signification (L)	This document

7.5. New CCI Flag Registry

CCI object is defined in defined in [I-D.ietf-pce-pcep-extension-for-pce-controller], further [I-D.zhao-pce-pcep-extension-pce-controller-sr] defined a CCI object type for SR. and the subregistry to manage the Flag field of the CCI object for SR is requested in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. IANA is requested to make the following allocation in the aforementioned subregistry.

Bit	Description	Reference
TBA5	PCC is requested to allocate resource(C)	This document

7.6. New FEC Type Registry

A new PCEP object called FEC is defined in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. IANA is requested to allocate a new Object-Type for FEC object in the "PCEP Objects" subregistry.

Value	Description	Reference
TBA6	SR path	This document

7.7. PCEP Error Type and Value

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
TBA7	Path SID failure: Error-value = 1 Invalid SID	This document
	Error-value = 2 Unable to allocate Path SID	

8. Security Considerations

TBA

9. Acknowledgments

10. Contributors

The following people have substantially contributed to this document:

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

Email: dhruv.ietf@gmail.com

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan, Hubei 430223
China

Email: xiong.quan@zte.com.cn

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

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>>.
- [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>>.
- [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>>.
- [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]
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.negi-pce-segment-routing-ipv6]
Negi, M., Li, C., Sivabalan, S., and P. Kaladharan, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", draft-negi-pce-segment-routing-ipv6-04 (work in progress), February 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-04 (work in progress), February 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-01 (work in progress), February 2019.

[I-D.li-spring-srv6-path-segment]

Li, C., Chen, M., Dhody, D., Li, Z., Dong, J., and R. Gandhi, "Path Segment for SRv6 (Segment Routing in IPv6)", draft-li-spring-srv6-path-segment-00 (work in progress), October 2018.

11.2. 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>>.

[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>>.

[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.li-pce-controlled-id-space]

Li, C., Chen, M., Dong, J., Li, Z., Wang, A., and C. Zhou, "PCE Controlled ID Space", draft-li-pce-controlled-id-space-02 (work in progress), March 2019.

[I-D.cheng-spring-mpls-path-segment]

Cheng, W., Wang, L., Li, H., Chen, M., Gandhi, R., Zigler, R., and S. Zhan, "Path Segment in MPLS Based Segment Routing Network", draft-cheng-spring-mpls-path-segment-03 (work in progress), October 2018.

Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: chengli13@huawei.com

Mach(Guoyi) Chen
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: Mach.chen@huawei.com

Weiqiang Cheng
China Mobile
China

Email: chengweiqiang@chinamobile.com

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: jie.dong@huawei.com

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

Email: lizhenbin@huawei.com

Rakesh Gandhi
Cisco Systems, Inc.
Canada

Email: rgandhi@cisco.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 20, 2020

C. Li
M. Chen
Huawei Technologies
W. Cheng
China Mobile
J. Dong
Z. Li
Huawei Technologies
R. Gandhi
Cisco Systems, Inc.
Q. Xiong
ZTE Corporation
August 19, 2019

Path Computation Element Communication Protocol (PCEP) Extension for
Path Segment in Segment Routing (SR)
draft-li-pce-sr-path-segment-08

Abstract

The Path Computation Element (PCE) provides path computation functions in support of traffic engineering in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

The Source Packet Routing in Networking (SPRING) architecture describes how Segment Routing (SR) can be used to steer packets through an IPv6 or MPLS network using the source routing paradigm. A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a Path Computation Element (PCE).

Path identification is needed for several use cases such as performance measurement in Segment Routing (SR) network. This document specifies extensions to the Path Computation Element Communication Protocol (PCEP) to support requesting, replying, reporting and updating the Path Segment ID (Path SID) between PCEP speakers.

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 20, 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
2. Terminology	4
2.1. Requirements Language	4
3. Overview of Path Segment Extensions in PCEP	4
4. Objects and TLVs	5
4.1. The OPEN Object	5
4.1.1. The SR PCE Capability sub-TLV	5
4.1.2. PCECC-CAPABILITY sub-TLV	6
4.2. LSP Object	6
4.2.1. Path Segment TLV	7
4.3. FEC Object	8
4.4. CCI Object	9
5. Operations	10
5.1. Stateful PCE Operation	10
5.1.1. Ingress PCC-Initiated Path Segment Allocation	10
5.1.2. PCE Initiated Path Segment Allocation	12
5.2. PCECC Based Operation	13
5.2.1. PCE Controlled Label Spaces Advertisement	13
5.2.2. PCECC based Path Segment Allocation	13
6. Dataplane Considerations	15
7. Implementation Status	16
7.1. Huawei's Commercial Delivery	16
7.2. ZTE's Commercial Delivery	17
8. IANA Considerations	17

8.1.	SR PCE Capability Flags	17
8.2.	New LSP Flag Registry	17
8.3.	New PCEP TLV	17
8.3.1.	Path Segment TLV	18
8.4.	New FEC Type Registry	18
8.5.	PCEP Error Type and Value	19
9.	Security Considerations	19
10.	Manageability Considerations	19
10.1.	Control of Function and Policy	19
10.2.	Information and Data Models	20
10.3.	Liveness Detection and Monitoring	20
10.4.	Verify Correct Operations	20
10.5.	Requirements On Other Protocols	20
10.6.	Impact On Network Operations	20
11.	Acknowledgments	20
12.	References	20
12.1.	Normative References	20
12.2.	Informative References	22
Appendix A.	Contributors	23
Appendix B.	SRv6 extensions	23
B.1.	The SRv6 PCE Capability sub-TLV	24
B.2.	SRv6 PCE Capability Flags	24
B.3.	Path Segment TLV	25
Authors'	Addresses	25

1. Introduction

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables the communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalized MPLS (GMPLS) Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of TE LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP State Synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions. The model of operation where LSPs are initiated from the PCE is described in [RFC8281].

[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) [RFC8402] leverages the source routing and tunneling paradigms and supports steering packets into an explicit forwarding path at the ingress node.

An SR path needs to be identified in some use cases such as performance measurement. For identifying an SR path, [I-D.ietf-spring-mpls-path-segment] introduces a new segment that is referred to as Path Segment.

[I-D.ietf-pce-segment-routing] specifies extensions to the Path Computation Element Protocol (PCEP) [RFC5440] for SR networks, that allow a stateful PCE to compute and initiate SR-TE paths, as well as a PCC to request, report or delegate SR paths.

[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 specifies a mechanism to carry the SR path identification information in PCEP messages [RFC5440] [RFC8231] [RFC8281]. The SR path identifier can be a Path Segment in SR-MPLS [I-D.ietf-spring-mpls-path-segment], or other IDs that can identify an SR path. This document also extends the PCECC-SR mechanism to inform the Path Segment to the egress PCC.

2. Terminology

This memo makes use of the terms defined in [RFC4655], [I-D.ietf-pce-segment-routing], and [RFC8402].

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. Overview of Path Segment Extensions in PCEP

This document specifies a mechanism of allocating Path Segment and extends PCEP to encode it in PCEP messages. For supporting Path Segment in PCEP, several TLVs and flags are defined. The formats of

the objects and TLVs are described in Section 4. The procedures of Path Segment allocation are described in Section 5.

There are various modes of operations, such as -

- o The Path Segment can be allocated by Egress PCC. The PCE should request the Path Segment from Egress PCC.
- o The PCE can allocate a Path Segment on its own accord and inform the ingress/egress PCC, useful for PCE-initiated LSPs.
- o Ingress PCC can also request PCE to allocate the Path Segment, in this case, the PCE would either allocate and inform the assigned Path Segment to the ingress/egress PCC using PCEP messages, or first request egress PCC for Path Segment and then inform it to the ingress PCC.

The path information to the ingress PCC and PCE is exchanged via an extension to [I-D.ietf-pce-segment-routing] and [I-D.ietf-pce-segment-routing-ipv6]. The Path Segment information (for SR-MPLS) to the egress PCC can be informed via an extension to the PCECC-SR procedures [I-D.zhao-pce-pcep-extension-pce-controller-sr].

For the PCE to allocate a Path Segment on its own, the PCE needs to be aware of the MPLS label space from the PCCs. This is done via mechanism as described in [I-D.li-pce-controlled-id-space]. Otherwise, the PCE should request the egress PCC for Path Segment allocation.

4. Objects and TLVs

4.1. The OPEN Object

4.1.1. The SR PCE Capability sub-TLV

[I-D.ietf-pce-segment-routing] defined a new Path Setup Type (PST) and SR-PCE-CAPABILITY sub-TLV for SR-MPLS. PCEP speakers use this sub-TLV to exchange information about their SR capability. The TLV defines a Flags field [I-D.ietf-pce-segment-routing].

This document adds an additional flag for Path Segment allocation, as follows -

- o P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SR path identification (Path Segment, as per [I-D.ietf-spring-mpls-path-segment]).

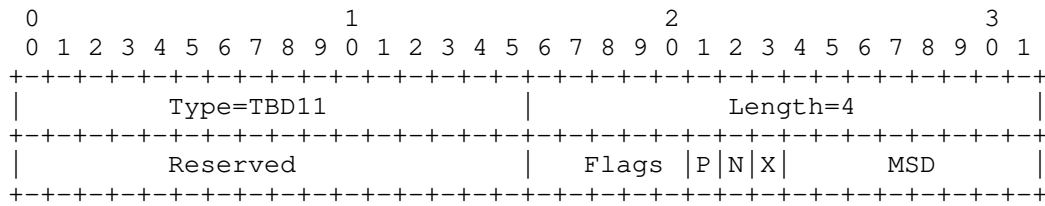


Figure 1: P-flag in SR-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and will be removed at the time of publication.

4.1.2. PCECC-CAPABILITY sub-TLV

Along with the SR sub-TLVs, the PCECC Capability as per [I-D.zhao-pce-pcep-extension-pce-controller-sr] should be advertised if the PCE allocates the Path Segment and acts as a Central Controller that manages the Label space.

The PCECC Capability should be advertised on the egress PCEP session, along with the SR sub-TLVs. This is needed to ensure that the PCE can use the PCECC objects/mechanism to request/inform the egress PCC of the Path Segment as described in Section 5.2.

4.2. LSP Object

The LSP Object is defined in Section 7.3 of [RFC8231]. This document adds a flag in the LSP Object:

- o P (PCE Allocation bit): If the bit is set to 1, it indicates that the PCC requests PCE to make allocations for this LSP. The TLV in LSP object identifies what should be allocated, such as Path Segment or Binding Segment. A PCC would set this bit to 1 and include a PATH-SEGMENT TLV in the LSP object to request for allocation of Path Segment by the PCE in the PCEP message. A PCE would also set this bit to 1 and include a PATH-SEGMENT TLV to indicate that the Path Segment is allocated by PCE and encoded in the PCEP message towards PCC. Further, a PCE would set this bit to 0 and include a PATH-SEGMENT TLV in the LSP object to indicate that the Path Segment should be allocated by the PCC as described in Section 5.1.1.

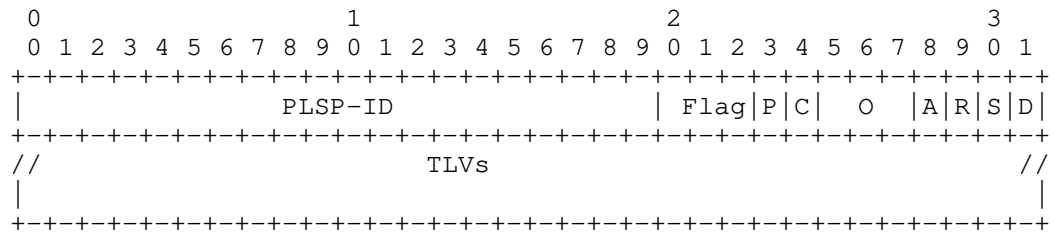


Figure 2: P-flag in LSP Object

The figure is included for the ease of the reader and will be removed at the time of publication.

4.2.1. Path Segment TLV

The PATH-SEGMENT TLV is an optional TLV for use in the LSP Object for Path Segment allocation. The type of this TLV is to be allocated by IANA (TBA4). The format is as shown below.

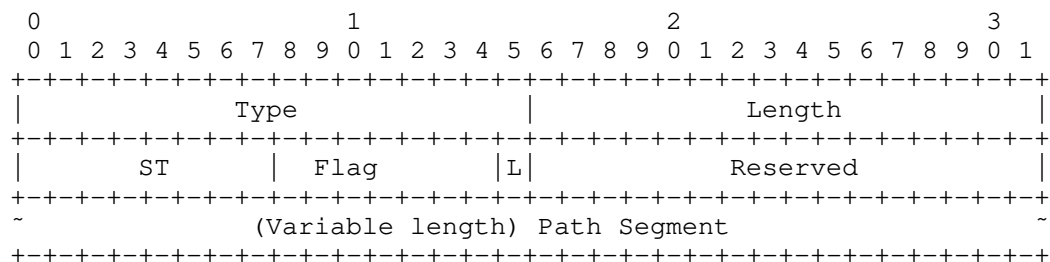


Figure 3: The PATH-SEGMENT TLV Format

The type (16-bit) of the TLV is TBA4 (to be allocated by IANA). The length (16-bit) has a variable length. The value contains the following fields:

- o ST (The Segment type - 8 bits): The ST field specifies the type of the Path Segment field, which carries a Path Segment corresponding to the SR path.
 - * 0: MPLS Path Segment, which is an MPLS label as defined in [I-D.ietf-spring-mpls-path-segment]. The PST type MUST be set to SR (MPLS).
 - * 1-255: Reserved for future use.
- o Flags (8 bits): One flag is currently defined:

- * L-Bit (Local/Global - 1 bit): If set, then the Path Segment carried by the PATH-SEGMENT TLV has local significance. If not set, then the Path Segment carried by this TLV has global significance (i.e. Path Segment is global within an SR domain).
- * The unassigned bits MUST be set to 0 and MUST be ignored at receipt.
- o Reserved (16 bits): MUST be set to 0 and MUST be ignored at receipt.
- o Path Segment: The Path Segment of an SR path. The Path Segment type is indicated by the ST field. When the ST is 0, it is a MPLS Path Segment [I-D.ietf-spring-mpls-path-segment] in the MPLS label format.

In general, only one instance of PATH-SEGMENT TLV will be included in LSP object. If more than one PATH-SEGMENT TLV is included, the first one is processed and others MUST be ignored. Multiple Path Segment allocation for use cases like alternate-making will be considered in future version of this draft.

When the Path Segment allocation is enabled, a PATH-SEGMENT TLV MUST be included in the LSP object.

If the label space is maintained by PCC itself, and the Path Segment is allocated by Egress PCC, then the PCE should request the Path Segment from Egress PCC as described in Section 5.1.1. In this case, the PCE should send a PCUpdate or PCInitiate message to the egress PCC to request the Path Segment. The P-flag in LSP should be unset in this case.

If a PCEP node does not recognize the PATH-SEGMENT TLV, it would behave in accordance with [RFC5440] and ignore the TLV. If a PCEP node recognizes the TLV but does not support the TLV, it MUST send PCErr with Error-Type = 2 (Capability not supported).

4.3. FEC Object

The FEC Object [I-D.zhao-pce-pcep-extension-pce-controller-sr] is used to specify the FEC information and carried within PCInitiate or PCRpt message for the PCECC-SR operations. The PCE MUST inform the Path Identification information to the Egress PCC. To do this, this document extends the procedures of [I-D.zhao-pce-pcep-extension-pce-controller-sr] by defining a new FEC object type for Path.

FEC Object-Type is TBA6 'Path'.

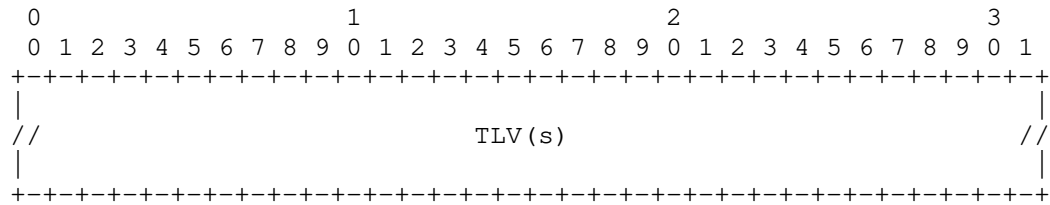


Figure 4: The path FEC object Format

One or more following TLV(s) are allowed in the 'path' FEC object -

- o SYMBOLIC-PATH-NAME TLV: As defined in [RFC8231], it is a human-readable string that identifies an LSP in the network.
- o LSP-IDENTIFIERS TLVs: As defined in [RFC8231], it is optional for SR, but could be used to encode the source, destination and other identification information for the path.
- o SPEAKER-ENTITY-ID TLV: As defined in [RFC8232], a unique identifier for the PCEP speaker, it is used to identify the Ingress PCC.

Either SYMBOLIC-PATH-NAME TLV or LSP-IDENTIFIERS TLV MUST be included. SPEAKER-ENTITY-ID TLV is optional. Only one instance of each TLV is processed, if more than one TLV of each type is included, the first one is processed and others MUST be ignored.

4.4. 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]. Further [I-D.zhao-pce-pcep-extension-pce-controller-sr] defined a CCI object type for SR.

The Path Segment information is encoded directly in the CCI SR object. The Path Segment TLV as described in the Section 4.2.1, MUST also be included in the CCI SR object as the TLV (as it includes additional information regarding the Path Segment identifier). The C flag in CCI object is used to indicate if the allocation needs to be done by the PCC.

5. Operations

The Path Segment allocation and encoding is as per the Stateful PCE operations for segment routing. The procedures are as per the corresponding extensions defined in [I-D.ietf-pce-segment-routing] and [I-D.ietf-pce-segment-routing-ipv6] (which are further based on [RFC8231] and [RFC8281]). The additional operations for Path Segment are defined in this section.

To notify (or request) the Path Segment to the Egress PCC, the procedures are as per the PCECC-SR [I-D.zhao-pce-pcep-extension-pce-controller-sr] (which is based on [I-D.ietf-pce-pcep-extension-for-pce-controller]). The additional operations are defined in this section.

5.1. Stateful PCE Operation

As defined in [I-D.ietf-spring-mpls-path-segment], a Path Segment can be allocated by the egress PCC. In this case, the label space is maintained on the PCC itself.

This section describes the mechanism of Path Segment allocation by using PCInitiate and PCUpd message in Stateful PCE model.

5.1.1. Ingress PCC-Initiated Path Segment Allocation

The ingress PCC could request the Path Segment to be allocated by the PCE via PCRpt message. The delegate flag (D-flag) MUST also be set for this LSP. Also, the P-flag in the LSP object MUST be set.

On receiving a delegation request with Path Segment allocation request from an ingress PCC, a stateful PCE requests the egress PCC to allocate a Path Segment.

The PATH-SEGMENT TLV MUST be included in an LSP object in the PCInitiate message sent from the PCE to the egress to request Path Segment allocation by the egress PCC. The P flag in LSP object MUST be set to 0. This PCInitiate message to egress PCC would be the similar to the one sent to ingress PCC as per [I-D.ietf-pce-segment-routing], but the egress PCC would only allocate the Path Segment and would not trigger the LSP initiation operation (as it would be the egress for this LSP).

If the value of Path Segment is 0x0, it indicates that the PCE is requesting a Path Segment for this LSP. If the Path Segment is set to a value 'n' and the P flag is unset in the LSP object, it indicates that the PCE requests a specific value 'n' of Path Segment. If the Path Segment is allocated successfully, the egress PCC reports

the Path Segment via PCRpt message with PATH-SEGMENT TLV in LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

Once the PCE receives the PCRpt message, it can obtain the Path Segment information from the egress PCC and then update the path with Path Segment by sending PCUpd message to the ingress PCC.

If the Path Segment is updated successfully, the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, an PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID") will be sent back to the PCE. The PCE MUST roll back the Path Segment value to the previous value (if any) by sending a PCUpd message to synchronize with the egress PCC.

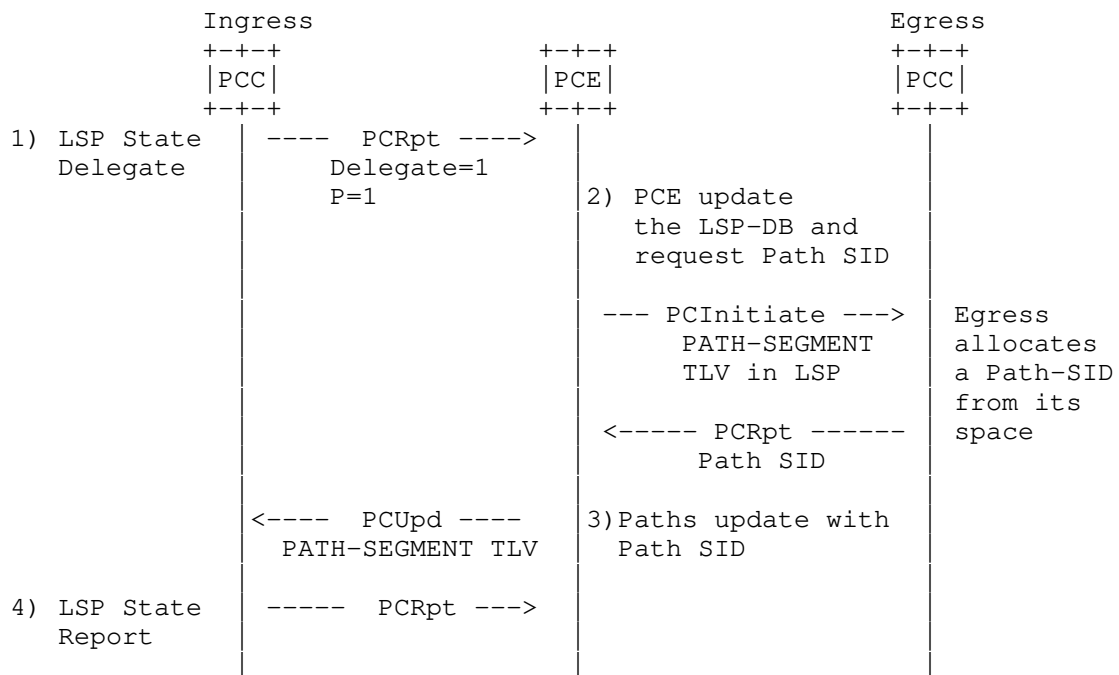


Figure 5: Ingress PCC-Initiated Path Segment Allocation

If the ingress PCC wishes to withdraw or modify a previously reported Path Segment value, it MUST send a PCRpt message without any PATH-

SEGMENT TLV or with the PATH-SEGMENT TLV containing the new Path Segment respectively. In this case, the PCE should synchronize with egress PCC via PCUpd message.

The Path Segment MUST be withdrawn when the corresponding LSP is removed. When the LSP is deleted, the PCE MUST request the egress PCC to withdraw the LSP and associated Path Segment via PCInitiate message with the R flag is set in the SRP object.

If an egress PCC receives a valid Path Segment value from a PCE which is different than the current Path Segment, it MUST try to allocate the new value. If the new Path Segment is successfully allocated, the egress PCC MUST report the new value to the PCE. Otherwise, it MUST send a PCErr message with Error-Type = TBA7 ("Path label/SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

5.1.1.2. PCE Initiated Path Segment Allocation

A stateful PCE also can initiate or update an LSP with Path Segment actively via requesting the egress PCC to allocate a Path Segment.

If a PCE wishes to modify a previously requested Path Segment value or allocate a Path Segment for an PCE-Initiated LSP, it MUST request the egress PCC to allocate a new value by sending a PCUpd message to the egress PCC with PATH-SEGMENT TLV containing the new Path Segment value. Also, the P flag in LSP object is unset. Absence of the PATH-SEGMENT TLV in PCUpd message means that the PCE wishes to withdraw the Path Segment.

The mechanism of requesting Path Segment is as per Section 5.1.1.

Once the PCE receives the PCRpt message, it can obtain the Path Segment information from the egress PCC and then update or initiate an LSP with Path Segment.

If the SR-Path is setup, the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, as described in [I-D.ietf-pce-segment-routing], an PCErr message will be sent back to the PCE. The PCE MUST request the egress PCC to withdraw the LSP and associated Path Segment via PCInitiate message with the R flag is set in the SRP object.

If the Path Segment is updated successfully, the ingress PCC will acknowledge with a PCRpt message to the PCE. In case of error, an PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID") will be sent back to the PCE. The PCE MUST

roll back the Path Segment value to the previous value (if any) by sending a PCUpd message to synchronize with the egress PCC.

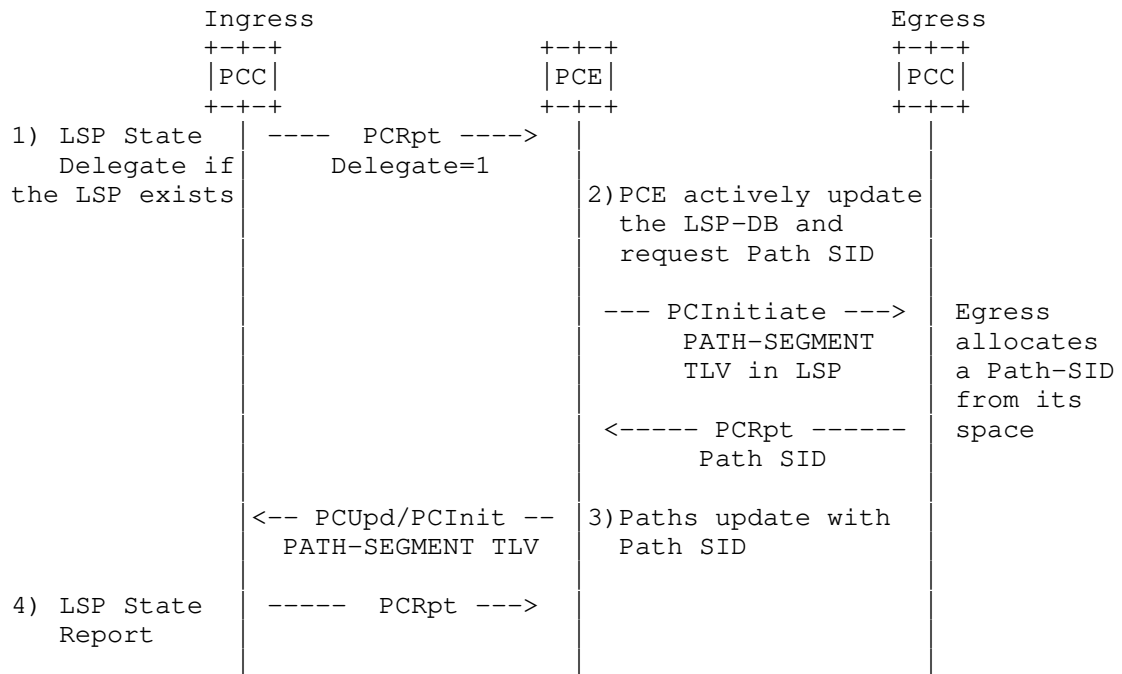


Figure 6: Stateful PCE-Initiated Path Segment Allocation

5.2. PCECC Based Operation

5.2.1. PCE Controlled Label Spaces Advertisement

For allocating the Path Segments to SR paths by the PCEs, the PCE controlled label space MUST be known at PCEs via configurations or any other mechanisms. The PCE controlled label spaces MAY be advertised as described in [I-D.li-pce-controlled-id-space].

5.2.2. PCECC based Path Segment Allocation

5.2.2.1. PCECC-Initiated

The PCE could allocate the Path Segment on its own for a PCE-Initiated (or delegated LSP). The allocated Path Segment needs to be informed to the Ingress and Egress PCC. The PCE would use the PCInitiate message [RFC8281] or PCUpd message [RFC8231] towards the Ingress PCC and MUST include the PATH-SEGMENT TLV in the LSP object.

The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

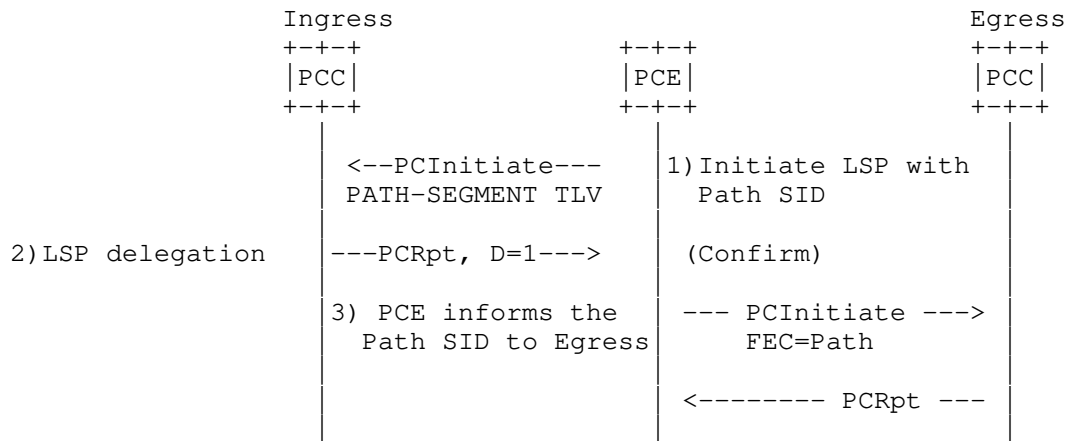


Figure 7: PCE allocated Path Segment on its own

5.2.2.2. Ingress PCC-Initiated PCECC

The ingress PCC could request the Path Segment to be allocated by the PCE via PCRpt message as per [RFC8231]. The delegate flag (D-flag) MUST also be set for this LSP. Also, the P-flag in the LSP object MUST be set.

A PATH-SEGMENT TLV MUST be included in the LSP object. If the value of Path Segment is 0x0, it indicates that the Ingress PCC is requesting a Path Segment for this LSP. If the Path Segment is set to a value 'n', it indicates that the ingress PCC requests a specific value 'n' of Path Segment.

If the Path Segment is allocated successfully, the PCE would further respond to Ingress PCC with PCUpd message as per [RFC8231] and MUST include the PATH-SEGMENT TLV in a LSP object. Else, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 1 ("Invalid SID"). If the value of Path Segment is valid, but the PCC is unable to allocate the Path Segment, it MUST send a PCErr message with Error-Type = TBA7 ("Path SID failure") and Error Value = 2 ("Unable to allocate the specified label/SID").

The active PCE would allocate the Path Segment as per the PATH-SEGMENT flags and in case PATH-SEGMENT is not included, the PCE MUST act based on the local policy.

The PCE would further inform the egress PCC about the Path Segment allocated by the PCE using the PCInitiate message as described in [I-D.zhao-pce-pcep-extension-pce-controller-sr].

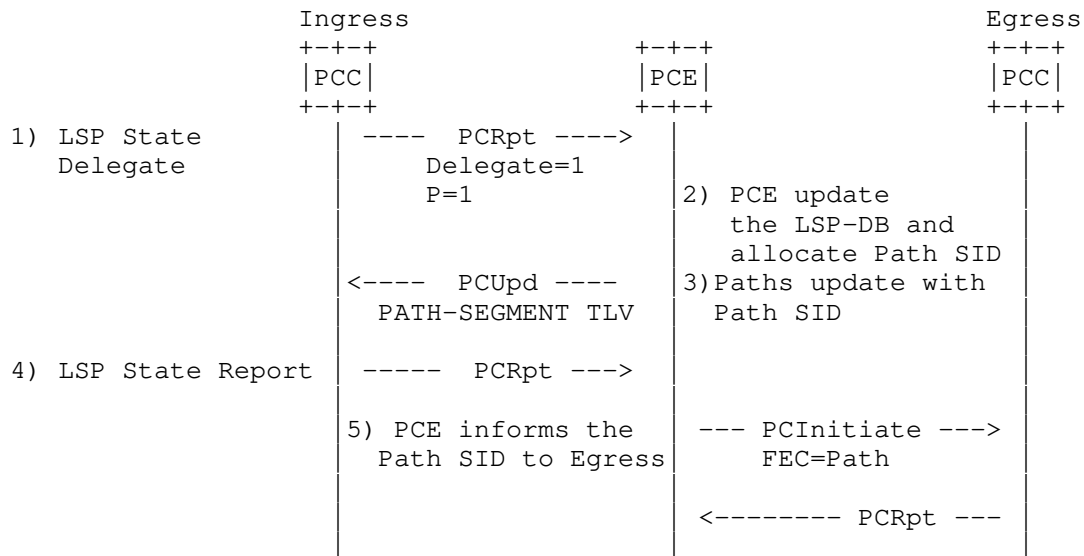


Figure 8: Ingress PCC request Path Segment to PCE

6. Dataplane Considerations

As described in [I-D.ietf-spring-mpls-path-segment], in an SR-MPLS network, when a packet is transmitted along an SR path, the labels in the MPLS label stack will be swapped or popped. So that no label or only the last label may be left in the MPLS label stack when the packet reaches the egress node. Thus, the egress node cannot determine from which SR path the packet comes. For this reason, it introduces the Path Segment.

Apart from allocation and encoding of the Path Segment (described in this document) for the LSP, it would also be included in the SID/Label stack of the LSP (usually for processing by the egress). To support this, the Path Segment MAY also be a part of SR-ERO as prepared by the PCE as per [I-D.ietf-pce-segment-routing]. The PCC MAY also include the Path Segment while preparing the label stack based on the local policy and use-case.

It is important that the PCE learns the Maximum SID Depth (MSD) that can be imposed at each node/link of a given SR path to ensure that the SID stack depth does not exceed the number of SIDs the node is

capable of imposing. As a new type of segment, Path Segment will be inserted in the SID list just like other SIDs. Thus, the PCE needs to consider the affect of Path Segment when computing a LSP with Path Segment allocation.

7. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to [RFC7942].

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".

7.1. Huawei's Commercial Delivery

The feature is developing based on Huawei VRP8.

- o Organization: Huawei
- o Implementation: Huawei's Commercial Delivery implementation based on VRP8.
- o Description: The implementation is under development and follows the mechanism as defined in section-5.1.1.
- o Maturity Level: Product
- o Contact: tanren@huawei.com

7.2. ZTE's Commercial Delivery

- o Organization: ZTE
- o Implementation: ZTE's Commercial Delivery implementation based on Rosng v8.
- o Description: The implementation is under development and follows the mechanism as defined in section-5.1.1.
- o Maturity Level: Product
- o Contact: zhan.shuangping@zte.com.cn

8. IANA Considerations

8.1. SR PCE Capability Flags

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 the following allocation in the "SR Capability Flag Field" sub-registry.

Bit	Description	Reference
TBA1	Path Segment Allocation is supported(P)	This document

8.2. 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 has allocated a new bit in the "LSP Object Flag Field" sub-registry, as follows:

Bit	Description	Reference
TBA3	Request for Path Segment Allocation(P)	This document

8.3. New PCEP TLV

IANA is requested to add the assignment of a new allocation in the existing "PCEP TLV Type Indicators" sub-registry as follows:

Value	Description	Reference
TBA4	PATH-SEGMENT TLV	This document

8.3.1. Path Segment TLV

This document requests that a new sub-registry named "PATH-SEGMENT TLV Segment Type (ST) Field" to be created to manage the value of the ST field in the PATH-SEGMENT TLV.

Value	Description	Reference
0	MPLS Path Segment (MPLS label)	This document
1-255	Reserved for future use	This document

Further, this document also requests that a new sub-registry named "PATH-SEGMENT TLV Flag Field" to be created to manage the Flag field in the PATH-SEGMENT TLV. New values are 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

Bit	Description	Reference
7	Local Signification(L)	This document

8.4. New FEC Type Registry

A new PCEP object called FEC is defined in [I-D.zhao-pce-pcep-extension-pce-controller-sr]. IANA is requested to allocate a new Object-Type for FEC object in the "PCEP Objects" sub-registry.

Value	Description	Reference
TBA6	Path	This document

8.5. PCEP Error Type and Value

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

Error-Type	Meaning	Reference
TBA7	Path SID failure: Error-value = 1 Invalid SID Error-value = 2 Unable to allocate Path SID	This document

9. Security Considerations

The security considerations described in [RFC5440], [RFC8231], [RFC8281] and [I-D.ietf-pce-segment-routing] are applicable to this specification. No additional security measure is required.

As described [I-D.ietf-pce-segment-routing] and [I-D.ietf-pce-pcep-extension-for-pce-controller], SR allows a network controller to instantiate and control paths in the network. A rogue PCE can manipulate Path SID allocations to have impact based on the usage of Path SID such as accounting, bi-directional etc.

Thus, 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

All manageability requirements and considerations listed in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing] apply to PCEP protocol extensions defined in this document. In addition, requirements and considerations listed in this section apply.

10.1. Control of Function and Policy

A PCEP implementation SHOULD allow the operator to configure the policy based on which it allocates the Path SID. This includes the Path SID scope.

10.2. Information and Data Models

The PCEP YANG module is defined in [I-D.ietf-pce-pcep-yang]. In future, this YANG module should be extended or augmented to provide the following additional information relating to Path SID.

An implementation SHOULD allow the operator to view the Path SID allocated to the LSP as well as Path SID as part of the computed SID list for the SR path.

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], [RFC8231], and [I-D.ietf-pce-segment-routing] .

10.5. Requirements On Other Protocols

Mechanisms defined in this document do not imply any new requirements on other protocols.

10.6. Impact On Network Operations

Mechanisms defined in [RFC5440], [RFC8231], and [I-D.ietf-pce-segment-routing] also apply to PCEP extensions defined in this document. Further, the mechanism described in this document can help the operator to request control of the LSPs at a particular PCE.

11. Acknowledgments

TBA

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>>.
- [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>>.
- [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>>.
- [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>>.
- [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]
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-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.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.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.li-spring-srv6-path-segment]
Li, C., Cheng, W., Chen, M., Dhody, D., Li, Z., Dong, J.,
and R. Gandhi, "Path Segment for SRv6 (Segment Routing in
IPv6)", draft-li-spring-srv6-path-segment-03 (work in
progress), August 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>>.
- [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>>.
- [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>>.

[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>>.

[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.ietf-spring-mpls-path-segment]

Cheng, W., Li, H., Chen, M., Gandhi, R., and R. Zigler, "Path Segment in MPLS Based Segment Routing Network", draft-ietf-spring-mpls-path-segment-00 (work in progress), March 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. Contributors

The following people have substantially contributed to this document:

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

Email: dhruv.ietf@gmail.com

Zafar Ali
Cisco Systems, Inc.

Email: zali@cisco.com

Appendix B. SRv6 extensions

This section would be rolled into the document once the SPRING WG adopts SRv6 path segment.

B.1. The SRv6 PCE Capability sub-TLV

[I-D.ietf-pce-segment-routing-ipv6] defined a new Path Setup Type (PST) and SRv6-PCE-CAPABILITY sub-TLV for SRv6. PCEP speakers use this sub-TLV to exchange information about their SRv6 capability. The TLV includes a Flags field and one bit (L-flag) was allocated in [I-D.ietf-pce-segment-routing-ipv6].

This document adds an additional flag for Path Segment allocation, as follows -

- o P (Path Segment Identification bit): A PCEP speaker sets this flag to 1 to indicate that it has the capability to encode SRv6 path identification. (Path Segment, as per [I-D.li-spring-srv6-path-segment]).

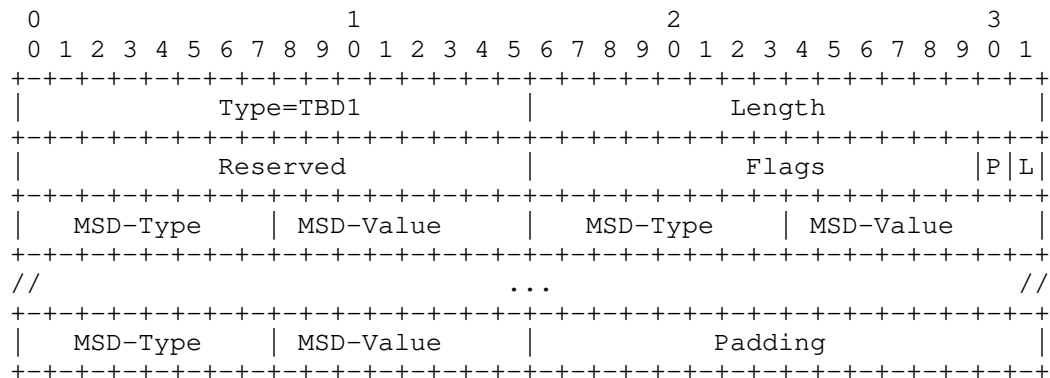


Figure 9: P-flag in SRv6-PCE-CAPABILITY TLV

The figure is included for the ease of the reader and can be removed at the time of publication.

B.2. SRv6 PCE Capability Flags

SRv6 PCE Capability TLV is defined in [I-D.ietf-pce-segment-routing-ipv6], and the registry to manage the Flag field of the SRv6 PCE Capability Flags is requested in [I-D.ietf-pce-segment-routing-ipv6]. IANA is requested to make the following allocation in the aforementioned registry.

Bit	Description	Reference
TBA2	Path Segment Allocation is supported(P)	This document

B.3. Path Segment TLV

A new assignment should be done to the "PATH-SEGMENT TLV Segment Type (ST) Field" sub-registry for SRv6.

Value	Description	Reference
1	SRv6 Path Segment (IPv6 addr)	This document
2-255	Reserved for future use	This document

Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: chengli13@huawei.com

Mach(Guoyi) Chen
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: Mach.chen@huawei.com

Weiqiang Cheng
China Mobile
China

Email: chengweiqiang@chinamobile.com

Jie Dong
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing 100095
China

Email: jie.dong@huawei.com

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

Email: lizhenbin@huawei.com

Rakesh Gandhi
Cisco Systems, Inc.
Canada

Email: rgandhi@cisco.com

Quan Xiong
ZTE Corporation
China

Email: xiong.quan@zte.com.cn

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 9, 2019

Q. Zhao
Z. Li
M. Negi
Huawei Technologies
C. Zhou
Cisco Systems
February 5, 2019

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) of SR-LSPs
draft-zhao-pcep-extension-pce-controller-sr-04

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 August 9, 2019.

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.	Security Considerations	20
9.	Manageability Considerations	21
9.1.	Control of Function and Policy	21
9.2.	Information and Data Models	21
9.3.	Liveness Detection and Monitoring	21
9.4.	Verify Correct Operations	21
9.5.	Requirements On Other Protocols	21
9.6.	Impact On Network Operations	21
10.	IANA Considerations	21
10.1.	PCECC-CAPABILITY TLV	21
10.2.	New Path Setup Type Registry	22
10.3.	PCEP Object	22
10.4.	PCEP-Error Object	22
11.	Acknowledgments	22
12.	References	23
12.1.	Normative References	23
12.2.	Informative References	24
Appendix A.	Contributor Addresses	27
Authors' Addresses	28

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.

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.

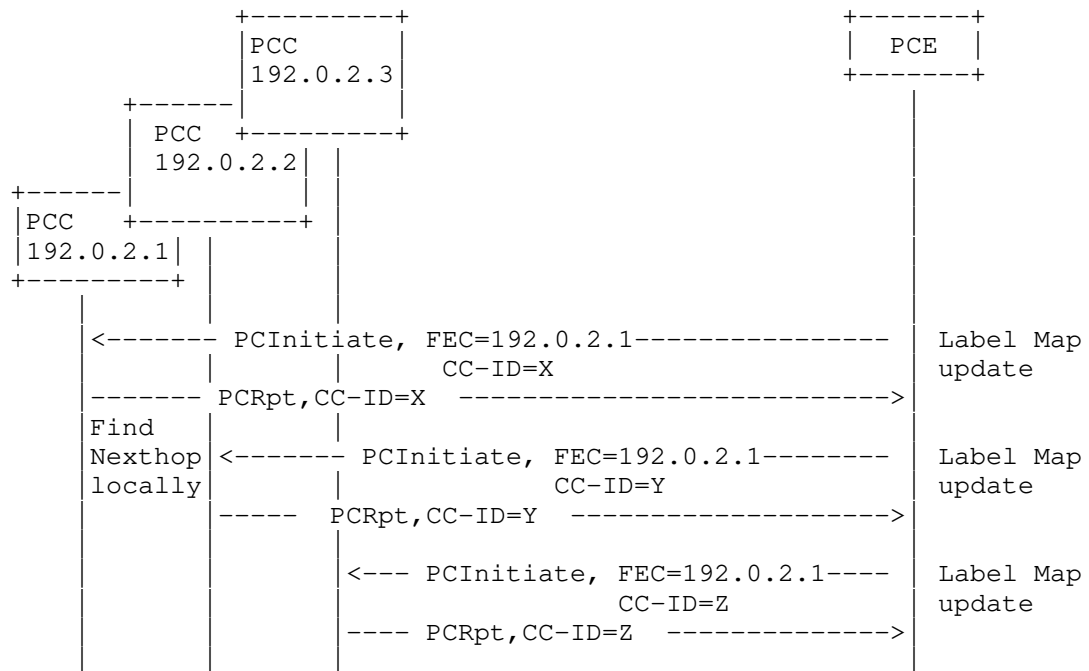
5.5.1.1. PCECC SR Node/Prefix SID allocation

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.

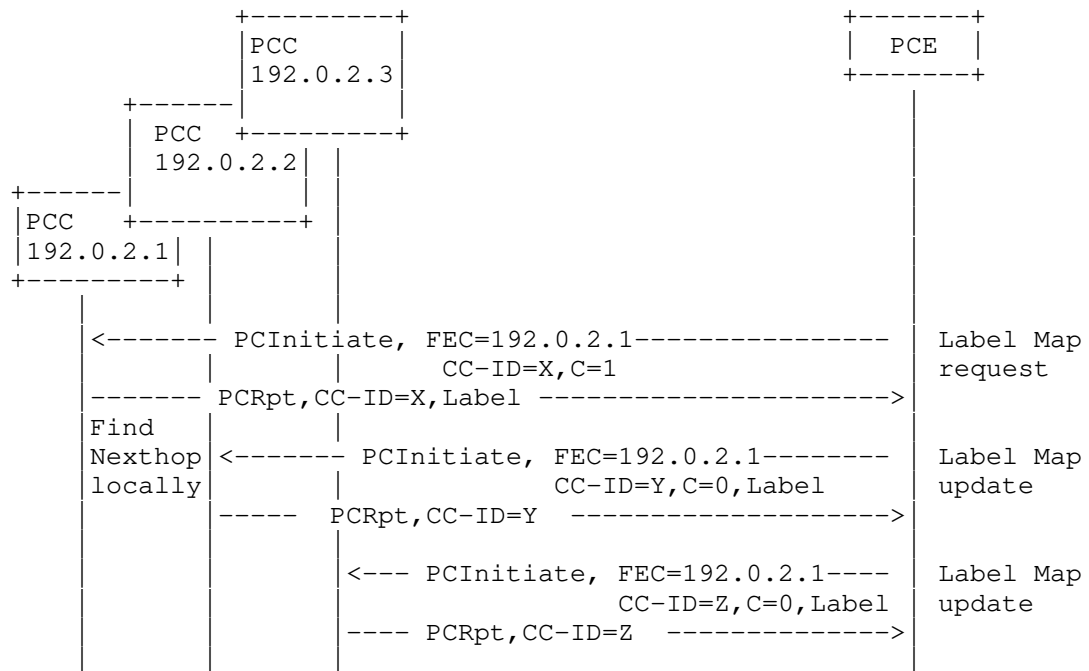


The forwarding behavior and the end result is similar to IGP based "Node-SID" in SR. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node.

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 -

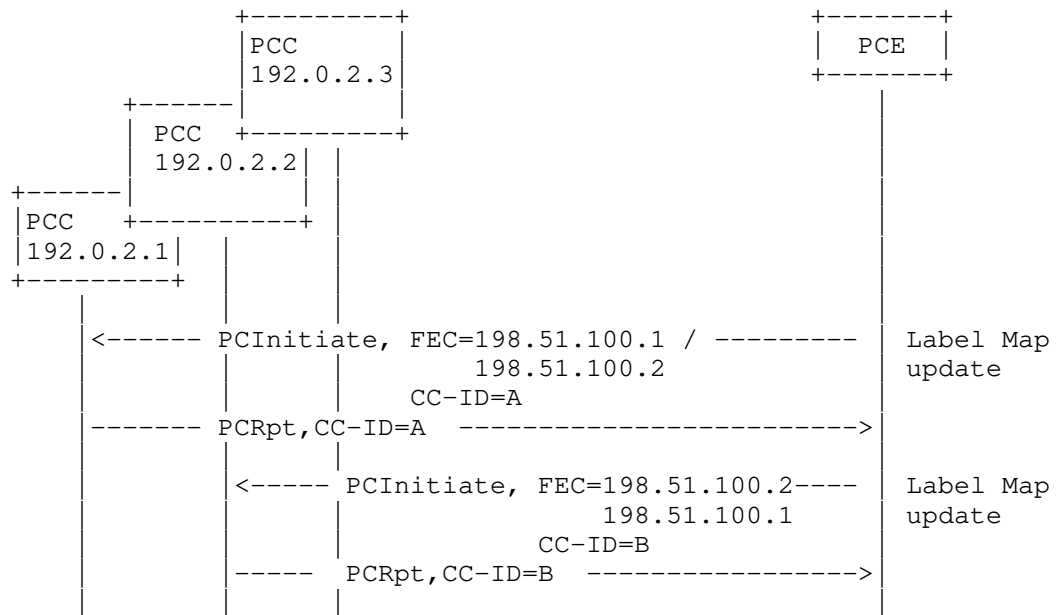


It should be noted that in this example, the request is made to the node 192.0.2.1 with C bit set in the CCI object to indicate that the allocation needs to be done by this PCC and it responds with the allocated label/SID to the PCE. The PCE would further inform the other PCCs in the network about the 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.



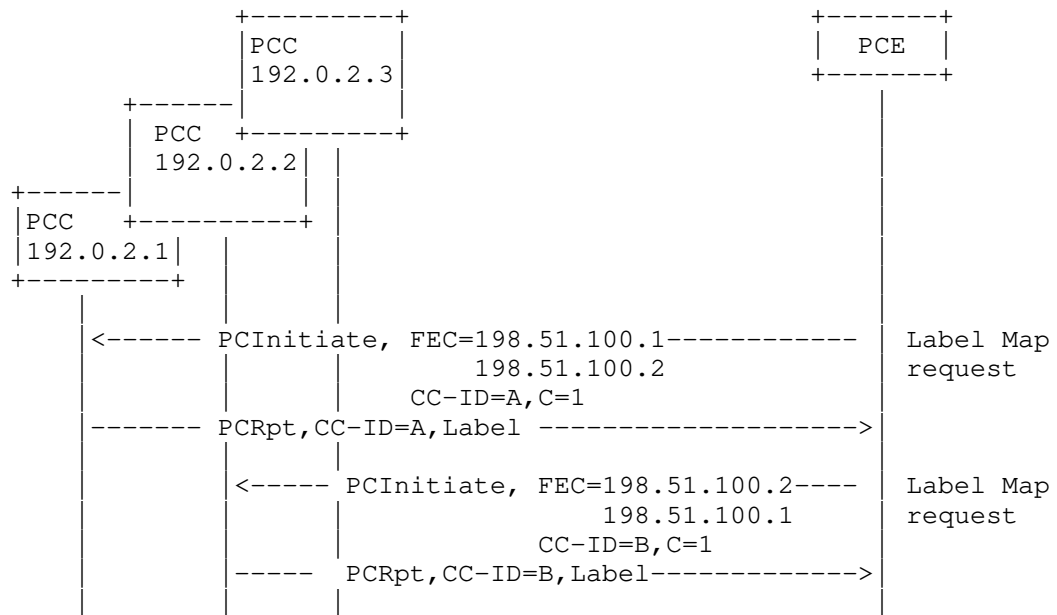
The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SR.

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

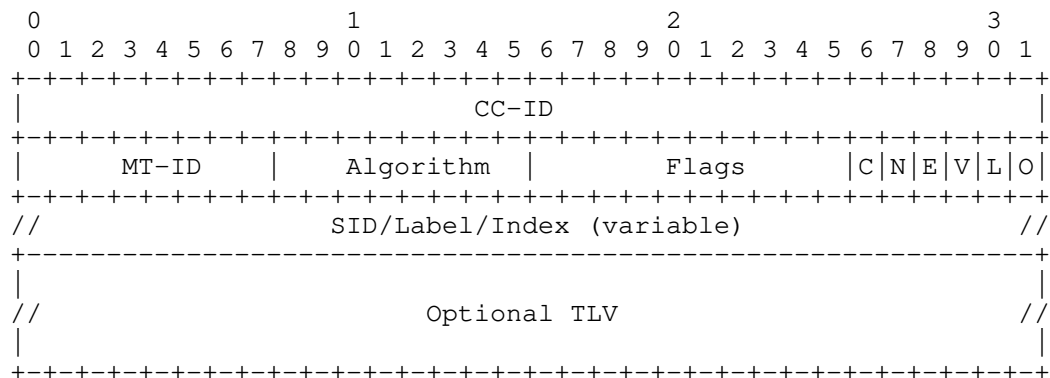
The PATH-SETUP-TYPE TLV is defined in [RFC8408]. PST = TBD is used when Path is setup via PCECC SR mode.

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                           |
|-----|

```

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.

Binding ID: TBD

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.

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 PCLabelUpd messages 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
31	S((PCECC-SR-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	Traffic engineering path is setup using PCECC-SR mode	This document

10.3. PCEP Object

IANA is requested to allocate new registry for FEC PCEP object.

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

10.4. 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.	
	Error-value = TBD :	FEC object missing
19	Invalid operation.	
	Error-value = TBD :	SR capability was not advertised

11. Acknowledgments

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

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>>.
- [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>>.

- [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>>.
- [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>>.

12.2. 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-02 (work in progress), October 2018.
- [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-09 (work in progress), October 2018.
- [I-D.ietf-pce-pcep-extension-for-pce-controller]
Zhao, Q., Li, Z., Dhody, D., Karunanithi, S., Farrel, A., 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-00 (work in progress), November 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-14 (work in progress), October 2018.

- [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-22 (work in progress), December 2018.
- [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., and D. Dhody, "Inter Stateful Path Computation Element communication procedures", draft-litkowski-pce-state-sync-04 (work in progress), October 2018.
- [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-12 (work in progress), December 2018.
- [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-18 (work in progress), December 2018.
- [I-D.sivabalan-pce-binding-label-sid]
Sivabalan, S., Filsfils, C., Tantsura, J., Hardwick, J., Previdi, S., and D. Dhody, "Carrying Binding Label/Segment-ID in PCE-based Networks.", draft-sivabalan-pce-binding-label-sid-05 (work in progress), October 2018.

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

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: 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: 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=TB4 (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.

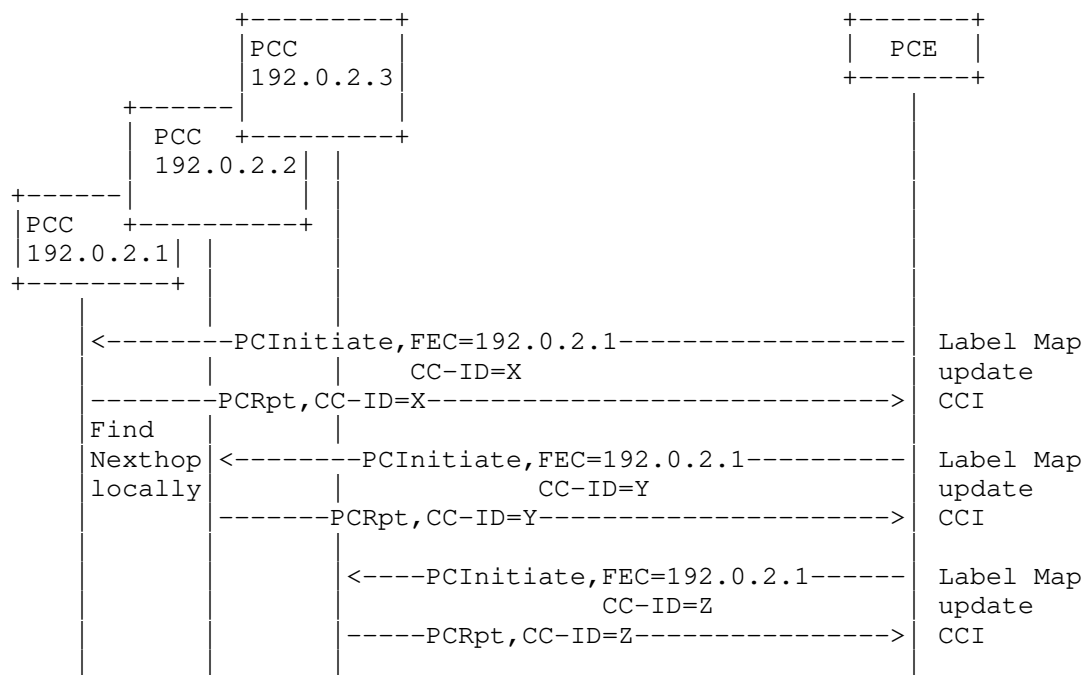
5.5.1.1. PCECC SR Node/Prefix SID allocation

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.

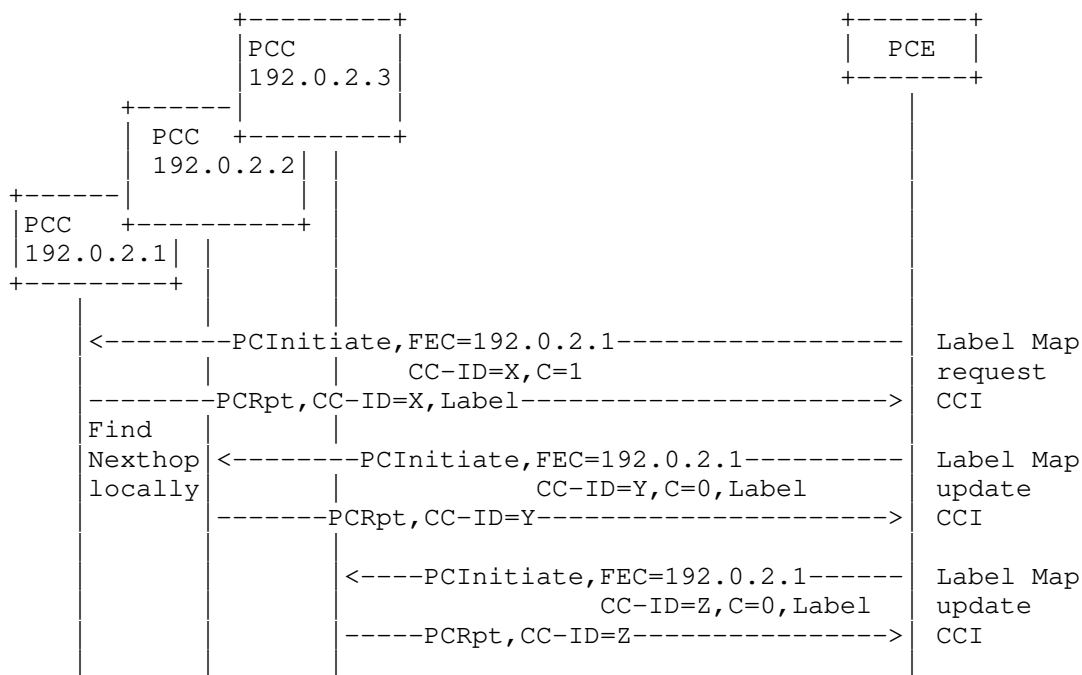


The forwarding behavior and the end result is similar to IGP based "Node-SID" in SR. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node as per [RFC8402].

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.

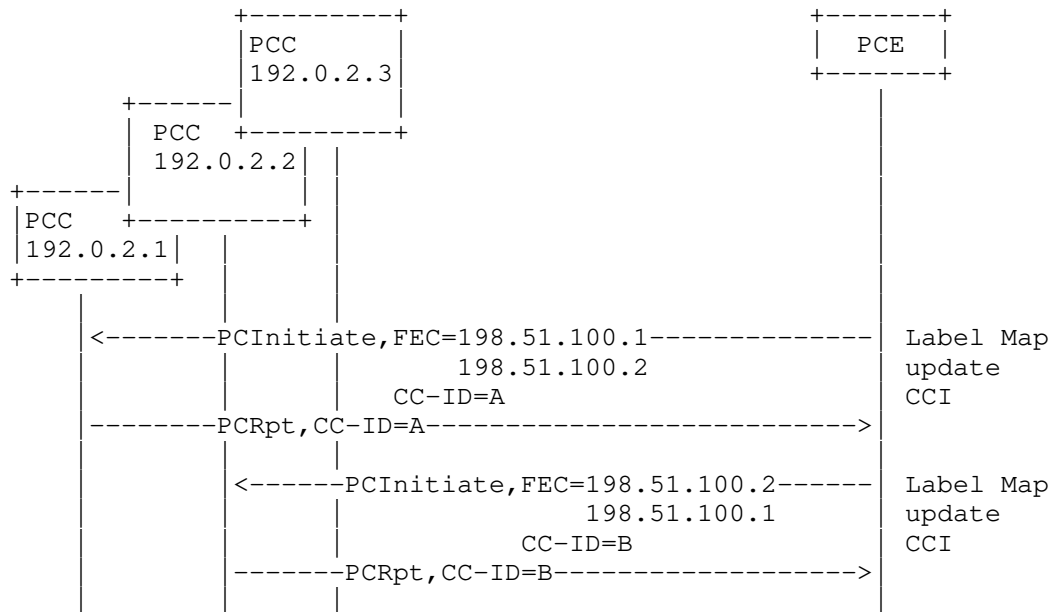
In the case where the label/SID allocation is made by the PCC itself (see Section 5.5.1.6), the PCE could request an allocation to be made by the PCC, and where the PCC would send a PCRpt with the allocated label/SID encoded in the CC-ID object as shown in Figure 2.



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.



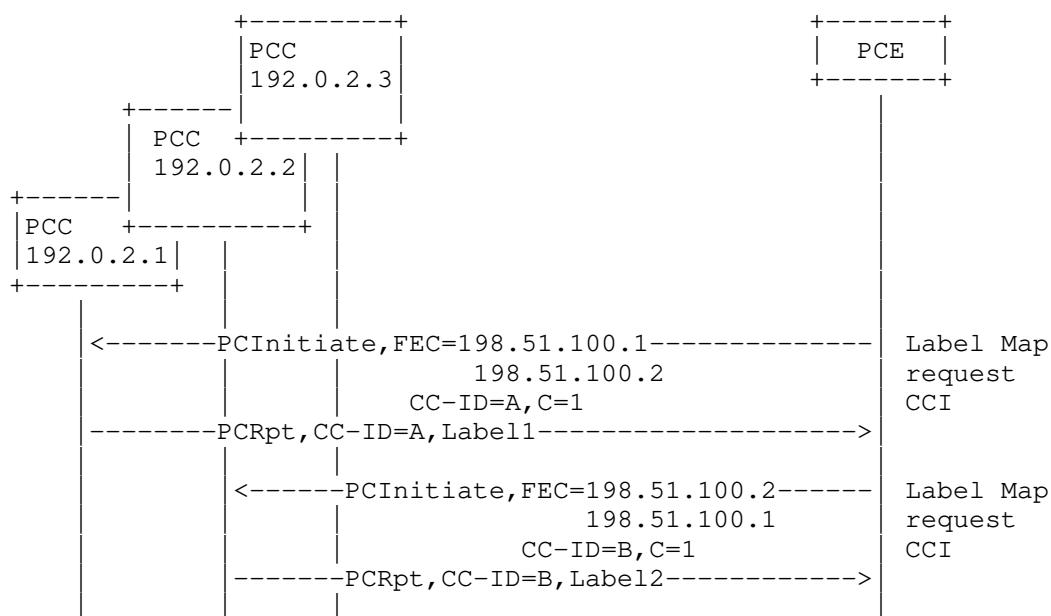
The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SR.

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).

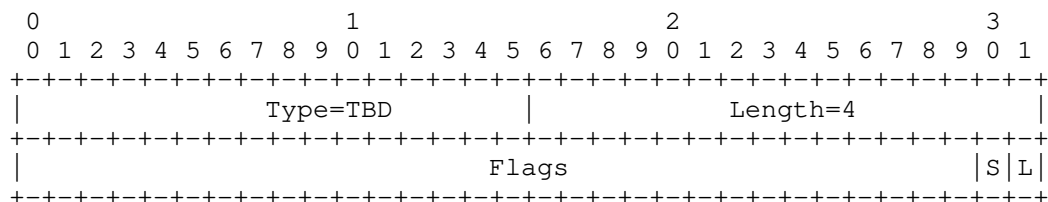
7. PCEP Objects

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.

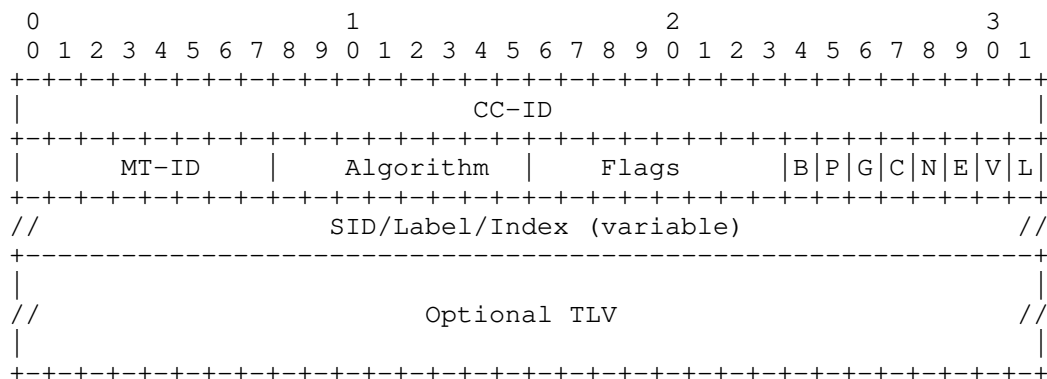
7.2. SR-TE Path Setup

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'.

```

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 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 : Invalid operation.	FEC object missing
	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] Filssils, 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., Filssils, 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., Filssils, 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., Filssils, 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