

PCE Working Group
Internet-Draft
Intended status: Experimental
Expires: 6 September 2022

D. Dhody
S. Peng
Huawei Technologies
Y. Lee
Samsung Electronics
D. Ceccarelli
Ericsson
A. Wang
China Telecom
G. Mishra
Verizon Inc.
S. Sivabalan
Ciena Corporation
5 March 2022

PCEP extensions for Distribution of Link-State and TE Information
draft-dhodylee-pce-pcep-ls-23

Abstract

In order to compute and provide optimal paths, a Path Computation Elements (PCEs) require an accurate and timely Traffic Engineering Database (TED). Traditionally, this TED has been obtained from a link state (LS) routing protocol supporting the traffic engineering extensions.

This document extends the Path Computation Element Communication Protocol (PCEP) with Link-State and TE Information as an experimental extension.

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 6 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	4
1.1. Scope	5
2. Terminology	6
3. Applicability	6
4. Requirements for PCEP extensions	7
5. New Functions to distribute link-state (and TE) via PCEP . .	8
6. Overview of Extensions to PCEP	9
6.1. New Messages	9
6.2. Capability Advertisement	9
6.3. Initial Link-State (and TE) Synchronization	10
6.3.1. Optimizations for LS Synchronization	12
6.4. LS Report	12
7. Transport	12
8. PCEP Messages	13
8.1. LS Report Message	13
8.2. The PCErr Message	13
9. Objects and TLV	14
9.1. TLV Format	14
9.2. Open Object	14
9.2.1. LS Capability TLV	14
9.3. LS Object	15
9.3.1. Routing Universe TLV	17
9.3.2. Route Distinguisher TLV	18
9.3.3. Virtual Network TLV	18
9.3.4. Local Node Descriptors TLV	18

9.3.5. Remote Node Descriptors TLV	19
9.3.6. Node Descriptors Sub-TLVs	20
9.3.7. Link Descriptors TLV	21
9.3.8. Prefix Descriptors TLV	21
9.3.9. PCEP-LS Attributes	22
9.3.9.1. Node Attributes TLV	22
9.3.9.2. Link Attributes TLV	22
9.3.9.3. Prefix Attributes TLV	23
9.3.10. Removal of an Attribute	23
10. Other Considerations	24
10.1. Inter-AS Links	24
11. Security Considerations	24
12. Manageability Considerations	24
12.1. Control of Function and Policy	24
12.2. Information and Data Models	25
12.3. Liveness Detection and Monitoring	25
12.4. Verify Correct Operations	25
12.5. Requirements On Other Protocols	26
12.6. Impact On Network Operations	26
13. IANA Considerations	26
13.1. PCEP Messages	26
13.2. PCEP Objects	26
13.3. LS Object	26
13.4. PCEP-Error Object	27
13.5. PCEP TLV Type Indicators	28
13.6. PCEP-LS Sub-TLV Type Indicators	28
14. TLV Code Points Summary	29
15. Implementation Status	30
15.1. Hierarchical Transport PCE controllers	30
15.2. ONOS-based Controller (MDSC and PNC)	31
16. Acknowledgments	31
17. References	31
17.1. Normative References	31
17.2. Informative References	32
Appendix A. Examples	35
A.1. All Nodes	35
A.2. Designated Node	37
A.3. Between PCEs	37
Appendix B. Contributor Addresses	38
Authors' Addresses	39

1. Introduction

In Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS), a Traffic Engineering Database (TED) is used in computing paths for connection-oriented packet services and for circuits. The TED contains all relevant information that a Path Computation Element (PCE) needs to perform its computations. It is important that the TED be 'complete and accurate' each time the PCE performs a path computation.

In MPLS and GMPLS, interior gateway routing protocols (Interior Gateway Protocol (IGPs)) have been used to create and maintain a copy of the TED at each node running the IGP. One of the benefits of the PCE architecture [RFC4655] is the use of computationally more sophisticated path computation algorithms and the realization that these may need enhanced processing power (not necessarily available at each node).

Section 4.3 of [RFC4655] describes the potential load of the TED on a network node and proposes an architecture where the TED is maintained by the PCE rather than the network nodes. However, it does not describe how a PCE would obtain the information needed to populate its TED. PCE may construct its TED by participating in the IGP ([RFC3630] and [RFC5305] for MPLS-TE; [RFC4203] and [RFC5307] for GMPLS). An alternative mechanism is offered by BGP-LS [I-D.ietf-idr-rfc7752bis] .

[RFC8231] describes a set of extensions to PCEP to provide stateful control. A stateful PCE has access to not only the information carried by the network's IGP, but also the set of active paths and their reserved resources for its computations. Path Computation Client (PCC) can delegate the rights to modify the LSP parameters to an Active Stateful PCE. This requires PCE to quickly be updated on any changes in the topology/TED, so that PCE can meet the need for updating LSPs effectively and in a timely manner. The fastest way for a PCE to be updated on TED changes is via a direct session with each network node and with an incremental update from each network node with only the attributes that gets modified.

[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. This model requires timely topology and TED update at the PCE.

[RFC5440] describes the specifications for the Path Computation Element Communication Protocol (PCEP). PCEP specifies the communication between a PCC and a PCE, or between two PCEs based on the PCE architecture [RFC4655].

This document describes a mechanism by which link-state and TE information can be collected from networks and shared with PCE using the PCEP itself. This is achieved using a new PCEP message format. The mechanism is applicable to physical and virtual links as well as further subjected to various policies.

A network node maintains one or more databases for storing link-state and TE information about nodes and links in any given area. Link attributes stored in these databases include: local/remote IP addresses, local/remote interface identifiers, link metric, and TE metric, link bandwidth, reservable bandwidth, per CoS class reservation state, preemption, and Shared Risk Link Groups (SRLG). The node's PCEP process can retrieve topology from these databases and distribute it to a PCE, either directly or via another PCEP Speaker, using the encoding specified in this document.

Further [RFC6805] describes Hierarchical-PCE architecture, where a parent PCE maintains a domain topology map. To build this domain topology map, the child PCE can carry the border nodes and inter-domain link information to the parent PCE using the mechanism described in this document. Further as described in [RFC8637], the child PCE can also transport abstract Link-State and TE information from child PCE to a Parent PCE using the mechanism described in this document to build an abstract topology at the parent PCE.

[RFC8231] describe LSP state synchronization between PCCs and PCEs in case of stateful PCE. This document does not make any change to the LSP state synchronization process. The mechanism described in this document are on top of the existing LSP state synchronization.

1.1. Scope

The procedures described in this document are experimental. The experiment is intended to enable research for the usage of PCEP to populate the Link-State and TE Information from a PCC to the PCE. For this purpose, this document specifies new PCEP message and object/TLVs.

The new message introduced by this document will not be understood by legacy implementations. On receiving the message, a legacy implementation will behave according to the rules for a unknown message as per [RFC5440]. It is assumed that this experiment will be conducted only when both the PCE and PCC form part of the experiment.

It is possible that a PCC or PCE can operate with peers, some of which form part of the experiment and some that do not. In this case, the capability exchange required before using this extension would take care of the mismatch. A PCEP speaker that offers this feature to its peer that does not support or does not wish to support the feature will not receive indication of support in the Open message, and so is expected to not use the feature. Thus this experimentation would not clash with or cause harm to existing deployments. Further since a PCEP speaker would use the new message only after capability exchange, there is no danger of this experimentation "escaping" to the wider Internet. A PCEP speaker that receives the new message that is part of the feature when use of the feature has not been agreed, will send an error message as described in Section 6.9 of [RFC5440]. A PCEP speaker that receives the new object that is part of the feature when use of the feature has not been agreed, will send an error message as described in Section 7.2 of [RFC5440].

The experiment will end three years after the RFC is published. At that point, the RFC authors will attempt to determine how widely this has been implemented and deployed. When the results of implementation and deployment are available, this document (or part there of) will be updated and refined, and then it could be moved from Experimental to Standards Track.

2. Terminology

The terminology is as per [RFC4655] and [RFC5440].

3. Applicability

The mechanism specified in this draft is applicable to deployments:

- * Where there is no IGP or BGP-LS running in the network.
- * Where there is no IGP or BGP-LS running at the PCE to learn link-state and TE information.
- * Where there is IGP or BGP-LS running but with a need for a faster and direct TE and link-state population and convergence at the PCE.
 - A PCE may receive partial information (say basic TE, link-state) from IGP and other information (optical and impairment) from PCEP.
 - A PCE may receive an incremental update (as opposed to the full (entire) information of the node/link).

- A PCE may receive full information from both existing mechanisms (IGP or BGP-LS) and PCEP.
- * Where there is a need for transporting (abstract) Link-State and TE information from child PCE to a Parent PCE in H-PCE [RFC6805]; as well as for Provisioning Network Controller (PNC) to Multi-Domain Service Coordinator (MDSC) in Abstraction and Control of TE Networks (ACTN) [RFC8453].
- * Where there is an existing PCEP session between all the nodes and the PCE-based central controller (PCECC) [RFC8283], and the operator would like to use PCEP as direct southbound interface to all the nodes in the network. This enables the operator to use PCEP as a single direct protocol between the controller and all the nodes in the network. In this mode, all nodes send only the local information.

Based on the local policy and deployment scenario, a PCC chooses to send only local information or both local and remote learned information. How a PCE manages the link-state (and TE) information is implementation specific and thus out of the scope of this document.

The prefix information in PCEP-LS can also help in determining the domain of the tunnel destination in the H-PCE (and ACTN) scenario. Section 4.5 of [RFC6805] describe various mechanisms and procedures that might be used, PCEP-LS provides a simple mechanism to exchange this information within PCEP.

[RFC8453] defines three types of topology abstraction - (1) Native/White Topology; (2) Black Topology; and (3) Grey Topology. Based on the local policy, the PNC (or child PCE) would share the domain topology to the MDSC (or Parent PCE) based on the abstraction type. The protocol extensions defined in this document can carry any type of topology abstraction.

4. Requirements for PCEP extensions

Following key requirements associated with link-state (and TE) distribution are identified for PCEP:

1. The PCEP speaker supporting this draft MUST have a mechanism to advertise the Link-State (and TE) distribution capability.

2. PCC supporting this draft MUST have the capability to report the link-state (and TE) information to the PCE. This MUST include self originated (local) information and MAY also allow remote information learned via routing protocols. PCC MUST be capable to do the initial bulk sync at the time of session initialization as well as any changes there after.
 3. A PCE MAY learn link-state (and TE) from PCEP as well as from existing mechanisms like IGP/BGP-LS. PCEP extensions MUST have a mechanism to correlate the information learned via other means. There MUST NOT be any changes to the existing link-state (and TE) population mechanism via IGP/BGP-LS. PCEP extension SHOULD keep the properties in a protocol (IGP or BGP-LS) neutral way, such that an implementation need not know about any OSPF or IS-IS or BGP-LS protocol specifics.
 4. It SHOULD be possible to encode only the changes in link-state (and TE) properties (after the initial sync) in PCEP messages. This leads to faster convergence.
 5. The same mechanism SHOULD be used for both MPLS TE as well as GMPLS, optical, and impairment aware properties.
 6. The same mechanism SHOULD be used for PCE to PCE Link-state (and TE) synchronization.
5. New Functions to distribute link-state (and TE) via PCEP

Several new functions are required in PCEP to support distribution of link-state (and TE) information. A function can be initiated either from a PCC towards a PCE (C-E) or from a PCE towards a PCC (E-C). The new functions are:

- * Capability advertisement (E-C,C-E): both the PCC and the PCE MUST announce during PCEP session establishment that they support PCEP extensions for distribution of link-state (and TE) information defined in this document.
- * Link-State (and TE) synchronization (C-E): after the session between the PCC and a PCE is initialized, the PCE must learn Link-State (and TE) information before it can perform path computations. In the case of stateful PCE it is RECOMMENDED that this operation be done before LSP state synchronization.
- * Link-State (and TE) Report (C-E): a PCC sends an LS (and TE) report to a PCE whenever the Link-State and TE information changes.

6. Overview of Extensions to PCEP

6.1. New Messages

In this document, we define a new PCEP message called LS Report (LSRpt), a PCEP message sent by a PCC to a PCE to report link-state (and TE) information. Each LS Report in an LSRpt message can contain the node or link properties. A unique PCEP specific LS identifier (LS-ID) is also carried in the message to identify a node or link and that remains constant for the lifetime of a PCEP session. This identifier on its own is sufficient when no IGP or BGP-LS running in the network for PCE to learn link-state (and TE) information. In case PCE learns some information from PCEP and some from the existing mechanism, the PCC SHOULD include the mapping of IGP or BGP-LS identifier to map the information populated via PCEP with IGP/BGP-LS. See Section 8.1 for details.

6.2. Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of LS (and TE) distribution via PCEP extensions. A PCEP Speaker includes the "LS Capability" TLV, described in Section 9.2.1, in the OPEN Object to advertise its support for PCEP-LS extensions. The presence of the LS Capability TLV in PCC's OPEN Object indicates that the PCC is willing to send LS Reports with local link-state (and TE) information. The presence of the LS Capability TLV in PCE's Open message indicates that the PCE is interested in receiving LS Reports with local link-state (and TE) information.

The PCEP extensions for LS (and TE) distribution MUST NOT be used if one or both PCEP Speakers have not included the LS Capability TLV in their respective OPEN message. If the PCE that supports the extensions of this draft but did not advertise this capability, then upon receipt of an LSRpt message from the PCC, it SHOULD generate a PCErr with error-type 19 (Invalid Operation), error-value TBD1 (Attempted LS Report if LS capability was not advertised) and it will terminate the PCEP session.

The LS reports sent by PCC MAY carry the remote link-state (and TE) information learned via existing means like IGP and BGP-LS only if both PCEP Speakers set the R (remote) Flag in the "LS Capability" TLV to 'Remote Allowed (R Flag = 1)'. If this is not the case and LS reports carry remote link-state (and TE) information, then a PCErr with error-type 19 (Invalid Operation) and error-value TBD1 (Attempted LS Report if LS remote capability was not advertised) and it will terminate the PCEP session.

6.3. Initial Link-State (and TE) Synchronization

The purpose of LS Synchronization is to provide a checkpoint-in-time state replica of a PCC's link-state (and TE) database in a PCE. State Synchronization is performed immediately after the Initialization phase (see [RFC5440]). In case of stateful PCE ([RFC8231]) it is RECOMMENDED that the LS synchronization should be done before LSP state synchronization.

During LS Synchronization, a PCC first takes a snapshot of the state of its database, then sends the snapshot to a PCE in a sequence of LS Reports. Each LS Report sent during LS Synchronization has the SYNC Flag in the LS Object set to 1. The end of synchronization marker is an LSRpt message with the SYNC Flag set to 0 for an LS Object with LS-ID equal to the reserved value 0. If the PCC has no link-state to synchronize, it will only send the end of synchronization marker.

Either the PCE or the PCC MAY terminate the session using the PCEP session termination procedures during the synchronization phase. If the session is terminated, the PCE MUST clean up the state it received from this PCC. The session re-establishment MUST be re-attempted per the procedures defined in [RFC5440], including the use of a back-off timer.

If the PCC encounters a problem which prevents it from completing the LS synchronization, it MUST send a PCErr message with error-type TBD2 (LS Synchronization Error) and error-value 2 (indicating an internal PCC error) to the PCE and terminate the session.

The PCE does not send positive acknowledgments for properly received LS synchronization messages. It MUST respond with a PCErr message with error-type TBD2 (LS Synchronization Error) and error-value 1 (indicating an error in processing the LSRpt) if it encounters a problem with the LS Report it received from the PCC and it MUST terminate the session.

The LS reports can carry local as well as remote link-state (and TE) information depending on the R flag in LS capability TLV.

The successful LS Synchronization sequence is shown in Figure 1.

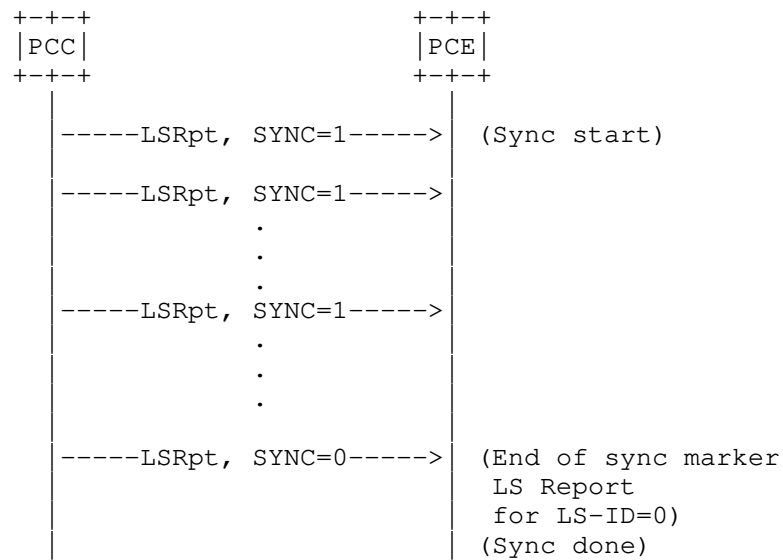


Figure 1: Successful LS synchronization

The sequence where the PCE fails during the LS Synchronization phase is shown in Figure 2.

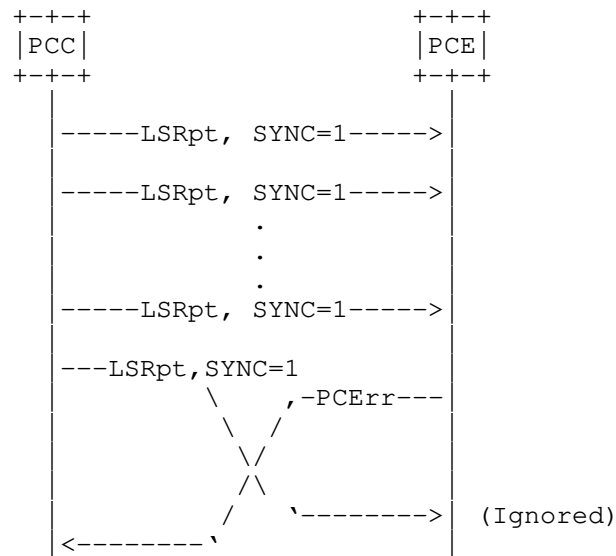


Figure 2: Failed LS synchronization (PCE failure)

The sequence where the PCC fails during the LS Synchronization phase is shown in Figure 3.

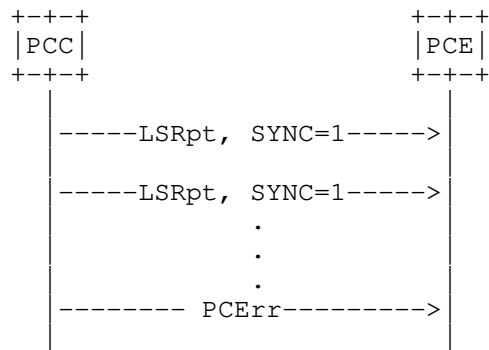


Figure 3: Failed LS synchronization (PCC failure)

6.3.1. Optimizations for LS Synchronization

These optimizations are described in [I-D.kondreddy-pce-pcep-ls-sync-optimizations].

6.4. LS Report

The PCC MUST report any changes in the link-state (and TE) information to the PCE by sending an LS Report carried on an LSRpt message to the PCE. Each node and Link would be uniquely identified by a PCEP LS identifier (LS-ID). The LS reports may carry local as well as remote link-state (and TE) information depending on the R flag in LS capability TLV. It MAY also include the mapping of IGP or BGP-LS identifier to map the information populated via PCEP with IGP/BGP-LS identifiers.

More details about the LSRpt message are in Section 8.1.

7. Transport

A permanent PCEP session (section 4.2.8 of [RFC5440]) MUST be established between a PCE and PCC supporting link-state (and TE) distribution via PCEP. In the case of session failure, session re-establishment is re-attempted as per the procedures defined in [RFC5440].

8. 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.

8.1. LS Report Message

A PCEP LS Report message (also referred to as LSRpt message) is a PCEP message sent by a PCC to a PCE to report the link-state (and TE) information. An LSRpt message can carry more than one LS Reports (LS object). The Message-Type field of the PCEP common header for the LSRpt message is set to [TBD3].

The format of the LSRpt message is as follows:

```
<LSRpt Message> ::= <Common Header>
                        <ls-report-list>
```

Where:

```
<ls-report-list> ::= <LS>[<ls-report-list>]
```

The LS object is a mandatory object which carries LS information of a node/prefix or a link. Each LS object has a unique LS-ID as described in Section 9.3. If the LS object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=[TBD4] (LS object missing).

A PCE may choose to implement a limit on the LS information a single PCC can populate. If an LSRpt is received that causes the PCE to exceed this limit, it MUST send a PCErr message with error-type 19 (invalid operation) and error-value 4 (indicating resource limit exceeded) in response to the LSRpt message triggering this condition and SHOULD terminate the session.

8.2. The PCErr Message

If a PCEP speaker has advertised the LS capability on the PCEP session, the PCErr message MAY include the LS object. If the error reported is the result of an LS report, then the LS-ID number MUST be the one from the LSRpt that triggered the error.

The format of a PCErr message from [RFC5440] is extended as follows:

```

<PCErr Message> ::= <Common Header>
                    ( <error-obj-list> [<Open>] ) | <error>
                    [<error-list>]

<error-obj-list> ::= <PCEP-ERROR> [<error-obj-list>]

<error> ::= [<request-id-list> | <ls-id-list>]
            <error-obj-list>

<request-id-list> ::= <RP> [<request-id-list>]

<ls-id-list> ::= <LS> [<ls-id-list>]

<error-list> ::= <error> [<error-list>]

```

9. Objects and TLV

The PCEP objects defined in this document are compliant with the PCEP object format defined in [RFC5440]. The P flag and the I flag of the PCEP objects defined in this document MUST always be set to 0 on transmission and MUST be ignored on receipt since these flags are exclusively related to path computation requests.

9.1. TLV Format

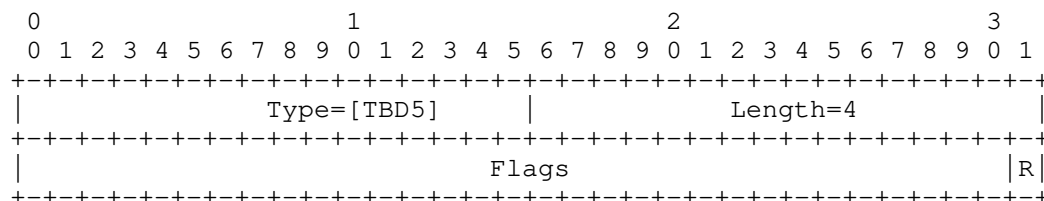
The TLV and the sub-TLV format (and padding) in this document, is as per section 7.1 of [RFC5440].

9.2. Open Object

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

9.2.1. LS Capability TLV

The LS-CAPABILITY TLV is an optional TLV for use in the OPEN Object for link-state (and TE) distribution via PCEP capability advertisement. Its format is shown in the following figure:



The type of the TLV is [TBD5] and it has a fixed length of 4 octets.

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

- * R (remote allowed - 1 bit): if set to 1 by a PCC, the R Flag indicates that the PCC allows reporting of remote LS information learned via other means like IGP and BGP-LS; if set to 1 by a PCE, the R Flag indicates that the PCE is capable of receiving remote LS information (from the PCC point of view). The R Flag must be advertised by both PCC and PCE for LSRpt messages to report remote as well as local LS information on a PCEP session. The TLVs related to IGP/BGP-LS identifier MUST be encoded when both PCEP speakers have the R Flag set.

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

Advertisement of the LS capability implies support of local link-state (and TE) distribution, as well as the objects, TLVs and procedures defined in this document.

9.3. LS Object

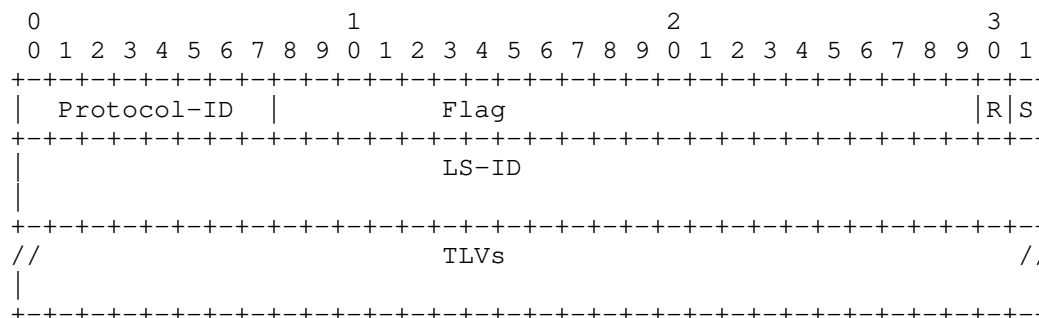
The LS (link-state) object MUST be carried within LSRpt messages and MAY be carried within PCErr messages. The LS object contains a set of fields used to specify the target node or link. It also contains a flag indicating to a PCE that the LS synchronization is in progress. The TLVs used with the LS object correlate with the IGP/BGP-LS encodings.

LS Object-Class is TBD6.

Four Object-Type values are defined for the LS object so far:

- * LS Node: LS Object-Type is 1.
- * LS Link: LS Object-Type is 2.
- * LS IPv4 Topology Prefix: LS Object-Type is 3.
- * LS IPv6 Topology Prefix: LS Object-Type is 4.

The format of all types of LS object is as follows:



Protocol-ID (8-bit): The field provides the source information. The protocol could be an IGP, BGP-LS, or an abstraction algorithm. In case PCC only provides local information of the PCC, it MUST use Protocol-ID as Direct. The following values are defined (some of the initial values are the same as [I-D.ietf-idr-rfc7752bis]):

Protocol-ID	Source protocol
1	IS-IS Level 1
2	IS-IS Level 2
3	OSPFv2
4	Direct
5	Static configuration
6	OSPFv3
7	BGP
8	RSVP-TE
9	Segment Routing
10	PCEP
11	Abstraction

Flags (24-bit):

- * S (SYNC - 1 bit): the S Flag MUST be set to 1 on each LSRpt sent from a PCC during LS Synchronization. The S Flag MUST be set to 0 in other LSRpt messages sent from the PCC.
- * R (Remove - 1 bit): On LSRpt messages, the R Flag indicates that the node/link/prefix has been removed from the PCC and the PCE SHOULD remove from its database. Upon receiving an LS Report with the R Flag set to 1, the PCE SHOULD remove all state for the node/link/prefix identified by the LS Identifiers from its database.

LS-ID(64-bit): A PCEP-specific identifier for the node, link, or prefix information. A PCC creates a unique LS-ID for each node/link/prefix that is constant for the lifetime of a PCEP session. The PCC will advertise the same LS-ID on all PCEP sessions it maintains at a given time. All subsequent PCEP messages then address the node/link/prefix by the LS-ID. The values of 0 and 0xFFFFFFFFFFFFFFFF are reserved.

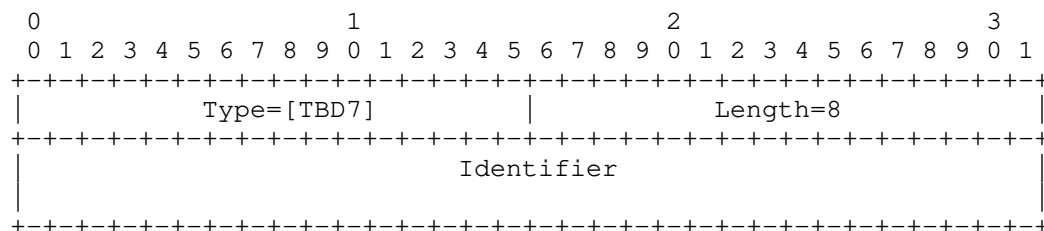
Unassigned bits are considered reserved. They MUST be set to 0 on transmission and MUST be ignored on receipt.

TLVs that may be included in the LS Object are described in the following sections.

9.3.1. Routing Universe TLV

In the case of remote link-state (and TE) population when existing IGP/BGP-LS are also used, OSPF and IS-IS may run multiple routing protocol instances over the same link as described in [I-D.ietf-idr-rfc7752bis]. See [RFC8202] and [RFC6549] for more information. These instances define an independent "routing universe". The 64-bit 'Identifier' field is used to identify the "routing universe" where the LS object belongs. The LS objects representing IGP objects (nodes or links or prefix) from the same routing universe MUST have the same 'Identifier' value; LS objects with different 'Identifier' values MUST be considered to be from different routing universes.

The format of the optional ROUTING-UNIVERSE TLV is shown in the following figure:



The below table lists the 'Identifier' values that are defined as well-known in this draft (same as [I-D.ietf-idr-rfc7752bis]).

Identifier	Routing Universe
0	Default Layer 3 Routing topology

If this TLV is not present the default value 0 is assumed.

9.3.2. Route Distinguisher TLV

To allow identification of VPN link, node, and prefix information in PCEP-LS, a Route Distinguisher (RD) [RFC4364] is used. The LS objects from the same VPN MUST have the same RD; LS objects with different RD values MUST be considered to be from different VPNs.

The ROUTE-DISTINGUISHER TLV is defined in [RFC9168] as a Flow Specification TLVs with a separate registry. This document also adds the ROUTE-DISTINGUISHER TLV with TBD15 in the PCEP TLV registry to be used inside the LS object.

9.3.3. Virtual Network TLV

To realize ACTN, the MDSC needs to build a multi-domain topology. This topology is best served if this is an abstracted view of the underlying network resources of each domain. It is also important to provide a customer view of the network slice for each customer. There is a need to control the level of abstraction based on the deployment scenario and business relationship between the controllers.

Virtual service coordination function in ACTN incorporates customer service-related knowledge into the virtual network operations in order to seamlessly operate virtual networks while meeting customer's service requirements. [I-D.ietf-teas-actn-requirements] describes various VN operations initiated by a customer/application. In this context, there is a need for associating the abstracted link-state and TE topology with a VN "construct" to facilitate VN operations in PCE architecture.

VIRTUAL-NETWORK-TLV as per [I-D.ietf-pce-vn-association] can be included in LS object to identify the link, node, and prefix information belongs to a particular VN.

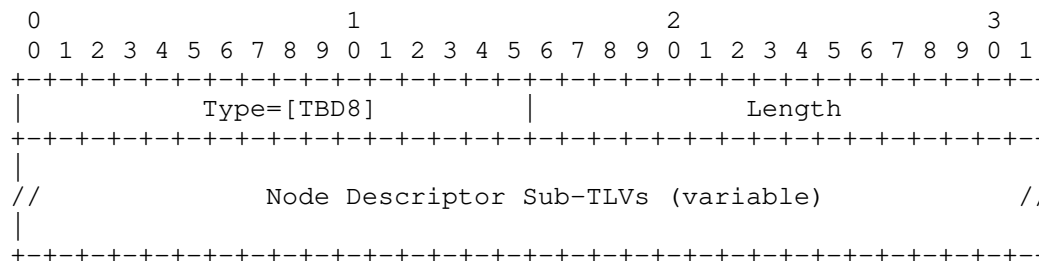
9.3.4. Local Node Descriptors TLV

As described in [I-D.ietf-idr-rfc7752bis], each link is anchored by a pair of Router-IDs that are used by the underlying IGP, namely, 48-bit ISO System-ID for IS-IS and 32-bit Router-ID for OSPFv2 and OSPFv3. In case of additional auxiliary Router-IDs used for TE, these MUST also be included in the link attribute TLV (see Section 9.3.9.2).

It is desirable that the Router-ID assignments inside the Node Descriptors TLV are globally unique. Some considerations for globally unique Node/Link/Prefix identifiers are described in [I-D.ietf-idr-rfc7752bis].

The Local Node Descriptors TLV contains Node Descriptors for the node anchoring the local end of the link. This TLV MUST be included in the LS Report when during a given PCEP session a node/link/prefix is first reported to a PCE. A PCC sends to a PCE the first LS Report either during State Synchronization, or when a new node/link/prefix is learned at the PCC. The value contains one or more Node Descriptor Sub-TLVs, which allows the specification of a flexible key for any given node/link/prefix information such that the global uniqueness of the node/link/prefix is ensured.

This TLV is applicable for all LS Object-Type.

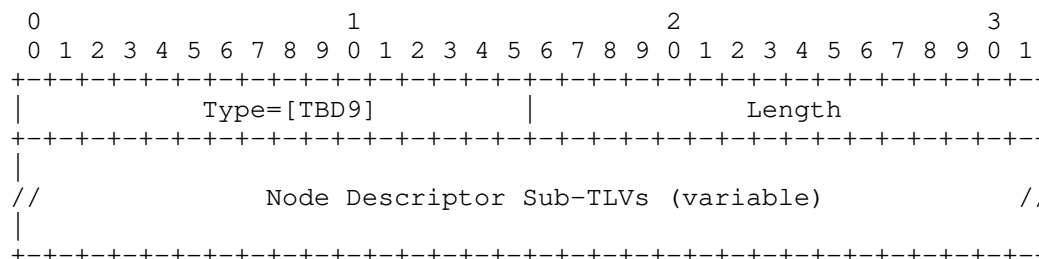


The value contains one or more Node Descriptor Sub-TLVs defined in Section 9.3.6.

9.3.5. Remote Node Descriptors TLV

The Remote Node Descriptors contain Node Descriptors for the node anchoring the remote end of the link. This TLV MUST be included in the LS Report when during a given PCEP session a link is first reported to a PCE. A PCC sends to a PCE the first LS Report either during State Synchronization, or when a new link is learned at the PCC. The length of this TLV is variable. The value contains one or more Node Descriptor Sub-TLVs defined in Section 9.3.6.

This TLV is applicable for LS Link Object-Type.



9.3.6. Node Descriptors Sub-TLVs

The Node Descriptors TLV (Local and Remote) carries one or more Node Descriptor Sub-TLV follows the format of all PCEP TLVs as defined in [RFC5440], however, the Type values are selected from a new PCEP-LS sub-TLV IANA registry (see Section 13.6).

Type values are chosen so that there can be commonality with BGP-LS [I-D.ietf-idr-rfc7752bis]. This is possible because the "BGP-LS Node Descriptor, Link Descriptor, Prefix Descriptor, and Attribute TLVs" registry marks 0-255 as reserved. Thus the space of the sub-TLV values for the Type field can be partitioned as shown below -

Range	
0	Reserved - must not be allocated.
1 .. 255	New PCEP sub-TLV allocated according to the registry defined in this document.
256 .. 65535	Per BGP registry defined by [I-D.ietf-idr-rfc7752bis]. Not to be allocated in this registry.

All Node Descriptors TLVs defined for BGP-LS can then be used with PCEP-LS as well. One new PCEP sub-TLVs for Node Descriptor are defined in this document.

Sub-TLV	Description	Length	Value defined in
1	SPEAKER-ENTITY-ID	Variable	[RFC8232]

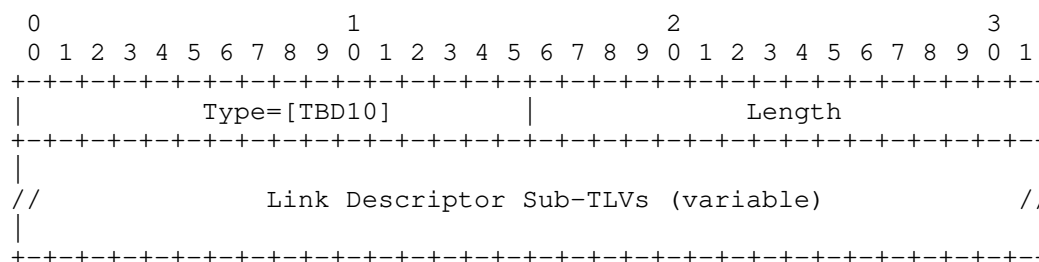
A new sub-TLV type (1) is allocated for SPEAKER-ENTITY-ID sub-TLV. The length and value fields are as per [RFC8232].

9.3.7. Link Descriptors TLV

The Link Descriptors TLV contains Link Descriptors for each link. This TLV MUST be included in the LS Report when during a given PCEP session a link is first reported to a PCE. A PCC sends to a PCE the first LS Report either during State Synchronization, or when a new link is learned at the PCC. The length of this TLV is variable. The value contains one or more Link Descriptor Sub-TLVs.

The 'Link descriptor' TLVs uniquely identify a link among multiple parallel links between a pair of anchor routers similar to [I-D.ietf-idr-rfc7752bis].

This TLV is applicable for LS Link Object-Type.



All Link Descriptors TLVs defined for BGP-LS can then be used with PCEP-LS as well. No new PCEP sub-TLVs for Link Descriptor are defined in this document.

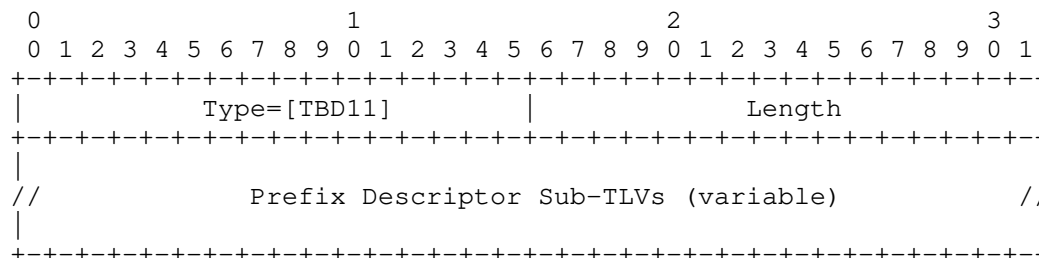
The format and semantics of the 'value' fields in most 'Link Descriptor' sub-TLVs correspond to the format and semantics of value fields in IS-IS Extended IS Reachability sub-TLVs, defined in [RFC5305], [RFC5307] and [RFC6119]. Although the encodings for 'Link Descriptor' TLVs were originally defined for IS-IS, the TLVs can carry data sourced either by IS-IS or OSPF or direct.

The information about a link present in the LSA/LSP originated by the local node of the link determines the set of sub-TLVs in the Link Descriptor of the link as described in [I-D.ietf-idr-rfc7752bis].

9.3.8. Prefix Descriptors TLV

The Prefix Descriptors TLV contains Prefix Descriptors that uniquely identify an IPv4 or IPv6 Prefix originated by a Node. This TLV MUST be included in the LS Report when during a given PCEP session a prefix is first reported to a PCE. A PCC sends to a PCE the first LS Report either during State Synchronization, or when a new prefix is learned at the PCC. The length of this TLV is variable.

This TLV is applicable for LS Prefix Object-Types for both IPv4 and IPv6.

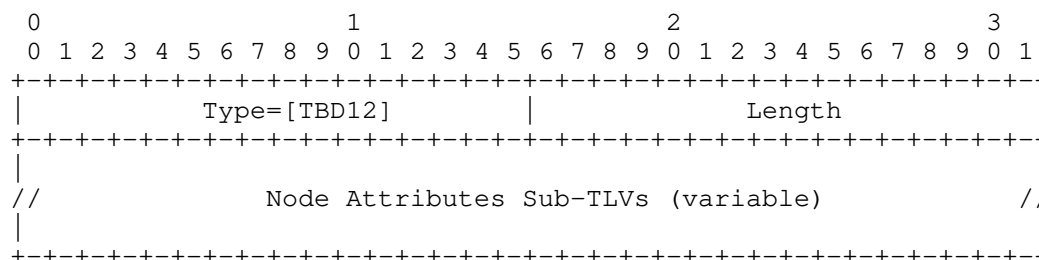


All Prefix Descriptors TLVs defined for BGP-LS can then be used with PCEP-LS as well. No new PCEP sub-TLVs for Prefix Descriptor are defined in this document.

9.3.9. PCEP-LS Attributes

9.3.9.1. Node Attributes TLV

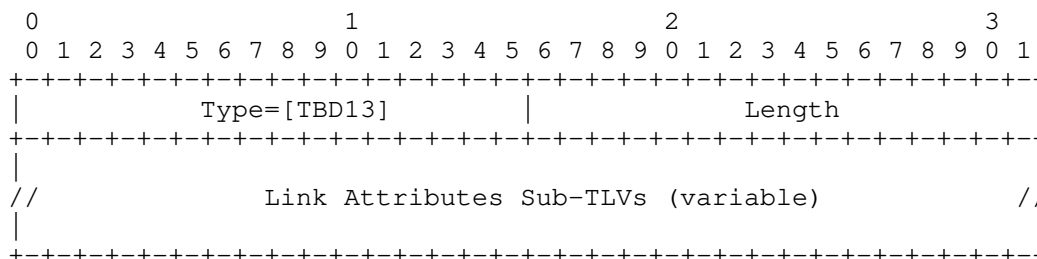
This is an optional attribute that is used to carry node attributes. This TLV is applicable for LS Node Object-Type.



All Node Attributes TLVs defined for BGP-LS can then be used with PCEP-LS as well. No new PCEP sub-TLVs for Node Attributes are defined in this document.

9.3.9.2. Link Attributes TLV

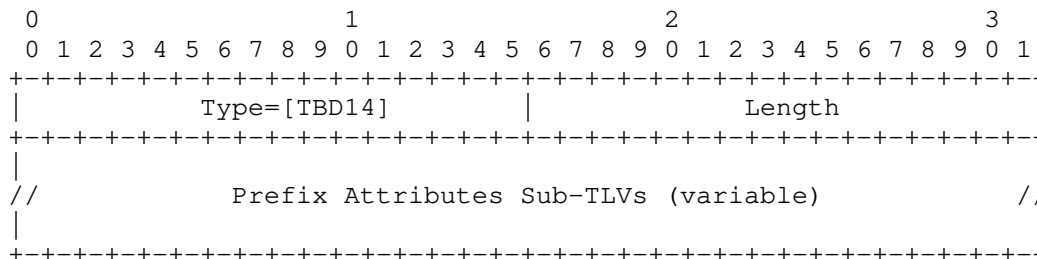
This TLV is applicable for LS Link Object-Type. The format and semantics of the 'value' fields in some 'Link Attribute' sub-TLVs correspond to the format and semantics of the 'value' fields in IS-IS Extended IS Reachability sub-TLVs, defined in [RFC5305], [RFC5307] and [I-D.ietf-idr-rfc7752bis]. Although the encodings for 'Link Attribute' TLVs were originally defined for IS-IS, the TLVs can carry data sourced either by IS-IS or OSPF or direct.



All Link Attributes TLVs defined for BGP-LS can then be used with PCEP-LS as well. No new PCEP sub-TLVs for Link Attributes are defined in this document.

9.3.9.3. Prefix Attributes TLV

This TLV is applicable for LS Prefix Object-Types for both IPv4 and IPv6. Prefixes are learned from the IGP (IS-IS or OSPF) or BGP topology with a set of IGP attributes (such as metric, route tags, etc.). This section describes the different attributes related to the IPv4/IPv6 prefixes. Prefix Attributes TLVs SHOULD be encoded in the LS Prefix Object.



All Prefix Attributes TLVs defined for BGP-LS can then be used with PCEP-LS as well. No new PCEP sub-TLVs for Prefix Attributes are defined in this document.

9.3.10. Removal of an Attribute

One of the key objectives of PCEP-LS is to encode and carry only the impacted attributes of a Node, a Link, or a Prefix. To accommodate this requirement, in case of a removal of an attribute, the sub-TLV MUST be included with no 'value' field and length=0 to indicate that the attribute is removed. On receiving a sub-TLV with zero length, the receiver removes the attribute from the database. An absence of a sub-TLV that was included earlier MUST be interpreted as no change.

10. Other Considerations

10.1. Inter-AS Links

The main source of LS (and TE) information is the IGP, which is not active on inter-AS links. In some cases, the IGP may have information of inter-AS links ([RFC5392], [RFC5316]). In other cases, an implementation SHOULD provide a means to inject inter-AS links into PCEP. The exact mechanism used to provision the inter-AS links is outside the scope of this document.

11. Security Considerations

This document extends PCEP for LS (and TE) distribution including a new LSRpt message with a new object and TLVs. Procedures and protocol extensions defined in this document do not effect the overall PCEP security model. See [RFC5440], [RFC8253]. Tampering with the LSRpt message may have an effect on path computations at PCE. It also provides adversaries an opportunity to eavesdrop and learn sensitive information and plan sophisticated attacks on the network infrastructure. The PCE implementation SHOULD provide mechanisms to prevent strains created by network flaps and amount of LS (and TE) information. Thus it is suggested that any mechanism used for securing the transmission of other PCEP message be applied here as well. As a general precaution, it is RECOMMENDED that these PCEP extensions only are activated on authenticated and encrypted sessions belonging to the same administrative authority.

Further, as stated in [RFC6952], PCEP implementations SHOULD support the TCP-AO [RFC5925] and not use TCP MD5 because of TCP MD5's known vulnerabilities and weaknesses. PCEP also support Transport Layer Security (TLS) [RFC8253] as per the recommendations and best current practices in [RFC7525].

12. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440] apply to PCEP protocol extensions defined in this document. In addition, requirements, and considerations listed in this section apply.

12.1. Control of Function and Policy

A PCE or PCC implementation MUST allow configuring the PCEP-LS capabilities as described in this document.

A PCC implementation SHOULD allow configuration to suggest if remote information learned via routing protocols should be reported or not.

An implementation SHOULD allow the operator to specify the maximum number of LS data to be reported.

An implementation SHOULD also allow the operator to create abstracted topologies that are reported to the peers and create different abstractions for different peers.

An implementation SHOULD allow the operator to configure a 64-bit identifier for Routing Universe TLV.

12.2. Information and Data Models

An implementation SHOULD allow the operator to view the LS capabilities advertised by each peer. To serve this purpose, the PCEP YANG module [I-D.ietf-pce-pcep-yang] can be extended to include advertised capabilities.

An implementation SHOULD also provide the statistics:

- * Total number of LSRpt sent/received, as well as per neighbor
- * Number of errors received for LSRpt, per neighbor
- * Total number of locally originated Link-State Information

These statistics should be recorded as absolute counts since system or session start time. An implementation MAY also enhance this information by recording peak per-second counts in each case.

An operator SHOULD define an import policy to limit inbound LSRpt to "drop all LSRpt from a particular peer" as well provide means to limit inbound LSRpts.

12.3. 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].

12.4. Verify Correct Operations

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

12.5. Requirements On Other Protocols

Mechanisms defined in this document do not imply any new requirements on other protocols.

12.6. Impact On Network Operations

Mechanisms defined in this document do not have any impact on network operations in addition to those already listed in [RFC5440].

13. IANA Considerations

This document requests IANA actions to allocate code points for the protocol elements defined in this document.

13.1. PCEP Messages

IANA created a registry for "PCEP Messages". Each PCEP message has a message type value. This document defines a new PCEP message value.

Value	Meaning	Reference
TBD3	LSRpt	[This I-D]

13.2. PCEP Objects

This document defines the following new PCEP Object-classes and Object-values:

Object-Class Value	Name	Reference
TBD6	LS Object	[This I-D]
	Object-Type=1	
	(LS Node)	
	Object-Type=2	
	(LS Link)	
	Object-Type=3	
	(LS IPv4 Prefix)	
	Object-Type=4	
	(LS IPv6 Prefix)	

13.3. LS Object

This document requests that a new sub-registry, named "LS Object Protocol-ID Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field of the LSP object. New values are to be assigned by Standards Action [RFC8126].

Value	Meaning	Reference
0	Reserved	[This I-D]
1	IS-IS Level 1	[This I-D]
2	IS-IS Level 2	[This I-D]
3	OSPFv2	[This I-D]
4	Direct	[This I-D]
5	Static configuration	[This I-D]
6	OSPFv3	[This I-D]
7	BGP	[This I-D]
8	RSVP-TE	[This I-D]
9	Segment Routing	[This I-D]
10	PCEP	[This I-D]
11	Abstraction	[This I-D]
12-255	Unassigned	

Further, this document also requests that a new sub-registry, named "LS Object Flag Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field of the LSP object. 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

The following values are defined in this document:

Bit	Description	Reference
0-21	Unassigned	
22	R (Remove bit)	[This I-D]
23	S (Sync bit)	[This I-D]

13.4. PCEP-Error Object

IANA is requested to make the following allocation in the "PCEP-ERROR Object Error Types and Values" registry.

Error-Type	Meaning	Reference
6	Mandatory Object missing Error-Value=TBD4 (LS object missing)	[RFC5440] [This I-D]
19	Invalid Operation Error-Value=TBD1 (Attempted LS Report if LS remote capability was not advertised)	[RFC8231] [This I-D]
TBD2	LS Synchronization Error Error-Value=1 (An error in processing the LSRpt) Error-Value=2 (An internal PCC error)	[This I-D]

13.5. PCEP TLV Type Indicators

This document defines the following new PCEP TLVs.

Value	Meaning	Reference
TBD5	LS-CAPABILITY TLV	[This I-D]
TBD7	ROUTING-UNIVERSE TLV	[This I-D]
TBD15	ROUTE-DISTINGUISHER TLV	[This I-D]
TBD8	Local Node Descriptors TLV	[This I-D]
TBD9	Remote Node Descriptors TLV	[This I-D]
TBD10	Link Descriptors TLV	[This I-D]
TBD11	Prefix Descriptors TLV	[This I-D]
TBD12	Node Attributes TLV	[This I-D]
TBD13	Link Attributes TLV	[This I-D]
TBD14	Prefix Attributes TLV	[This I-D]

13.6. PCEP-LS Sub-TLV Type Indicators

This document specifies the PCEP-LS Sub-TLVs. IANA is requested to create an "PCEP-LS Sub-TLV Types" sub-registry for the sub-TLVs carried in the PCEP-LS TLV (Local and Remote Node Descriptors TLV, Link Descriptors TLV, Prefix Descriptors TLV, Node Attributes TLV, Link Attributes TLV and Prefix Attributes TLV).

Allocations from this registry are to be made according to the following assignment policies [RFC8126]:

Range	Assignment policy
0	Reserved - must not be allocated.
1 .. 251	Specification Required
252 .. 255	Experimental Use
256 .. 65535	Reserved - must not be allocated. Usage mirrors the BGP-LS TLV registry [I-D.ietf-idr-rfc7752bis]

IANA is requested to pre-populate this registry with values defined in this document as follows, taking the new values from the range 1 to 251:

Value	Meaning
1	SPEAKER-ENTITY-ID

14. TLV Code Points Summary

This section contains the global table of all TLVs in LS object defined in this document.

TLV	Description	Ref TLV	Value defined in:
TBD7	Routing Universe	--	Sec 9.2.1
TBD15	Route Distinguisher	--	Sec 9.2.2
*	Virtual Network	--	[ietf-pce-vn-association]
TBD8	Local Node Descriptors	256	[I-D.ietf-idr-rfc7752bis] /3.2.1.2
TBD9	Remote Node Descriptors	257	[I-D.ietf-idr-rfc7752bis] /3.2.1.3
TBD10	Link Descriptors	--	Sec 9.2.8
TBD11	Prefix Descriptors	--	Sec 9.2.9
TBD12	Node Attributes	--	Sec 9.2.10.1
TBD13	Link Attributes	--	Sec 9.2.10.2
TBD14	Prefix Attributes	--	Sec 9.2.10.3

* this TLV is defined in a different PCEP document

Figure 4: TLV Table

15. Implementation Status

The PCEP-LS protocol extensions as described in this I-D were implemented and tested for a variety of applications. Apart from the below implementation, there exist other experimental implementations done for optical networks.

15.1. Hierarchical Transport PCE controllers

The PCEP-LS has been implemented as part of IETF97 Hackathon and Bits-N-Bites demonstration. The use-case demonstrated was DCI use-case of ACTN architecture in which to show the following scenarios:

- connectivity services on the ACTN based recursive hierarchical SDN/PCE platform that has the three-tier level SDN controllers (two-tier level MDSC and PNC) on the top of the PTN systems managed by EMS.
- Integration test of two tier-level MDSC: The SBI of the low level MDSC is the YANG based Korean national standards and the one of the high-level MDSC the PCEP-LS based ACTN protocols.

- Performance test of three types of SDN controller based recovery schemes including protection, reactive, and proactive restoration. PCEP-LS protocol was used to demonstrate a quick report of failed network components.

15.2. ONOS-based Controller (MDSC and PNC)

Huawei (PNC, MDSC) and SKT (MDSC) implemented PCEP-LS during Hackathon and IETF97 Bits-N-Bites demonstration. The demonstration was ONOS-based ACTN architecture in which to show the following capabilities:

Both packet PNC and optical PNC (with optical PCEP-LS extensions) implemented PCEP-LS on its SBI as well as its NBI (towards MDSC).

SKT orchestrator (acting as MDSC) also supported PCEP-LS (as well as RestConf) towards packet and optical PNCs on its SBI.

Further description can be found at ONOS-PCEP (<https://wiki.onosproject.org/display/ONOS/PCEP+Protocol>) and the code at ONOS-PCEP-GITHUB (<https://github.com/opennetworkinglab/onos/tree/master/protocols/pcep>).

16. Acknowledgments

This document borrows some of the structure and text from the [I-D.ietf-idr-rfc7752bis].

Thanks to Eric Wu, Venugopal Kondreddy, Mahendra Singh Negi, Avantika, and Zhengbin Li for the reviews.

Thanks to Ramon Casellas for his comments and suggestions based on his implementation experience.

17. References

17.1. Normative References

[I-D.ietf-idr-rfc7752bis]
Talaulikar, K., "Distribution of Link-State and Traffic Engineering Information Using BGP", Work in Progress, Internet-Draft, draft-ietf-idr-rfc7752bis-10, 10 November 2021, <<https://datatracker.ietf.org/doc/html/draft-ietf-idr-rfc7752bis-10>>.

- [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>>.
- [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>>.
- [RFC6119] Harrison, J., Berger, J., and M. Bartlett, "IPv6 Traffic Engineering in IS-IS", RFC 6119, DOI 10.17487/RFC6119, February 2011, <<https://www.rfc-editor.org/info/rfc6119>>.
- [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>>.
- [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>>.

17.2. Informative References

- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V. P., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-yang-18, 25 January 2022, <<https://datatracker.ietf.org/doc/html/draft-ietf-pce-pcep-yang-18>>.
- [I-D.ietf-pce-vn-association]
Lee, Y., Zheng, H., and D. Ceccarelli, "Path Computation Element communication Protocol (PCEP) extensions for Establishing Relationships between sets of LSPs and Virtual Networks", Work in Progress, Internet-Draft,

draft-ietf-pce-vn-association-05, 15 October 2021,
<<https://datatracker.ietf.org/doc/html/draft-ietf-pce-vn-association-05>>.

[I-D.ietf-teas-actn-requirements]

Lee, Y., Ceccarelli, D., Miyasaka, T., Shin, J. Y., and K. Lee, "Requirements for Abstraction and Control of TE Networks", Work in Progress, Internet-Draft, draft-ietf-teas-actn-requirements-09, 2 March 2018,
<<https://datatracker.ietf.org/doc/html/draft-ietf-teas-actn-requirements-09>>.

[I-D.kondreddy-pce-pcep-ls-sync-optimizations]

Kondreddy, V. R. and M. S. Negi, "Optimizations of PCEP Link-State (LS) Synchronization Procedures", Work in Progress, Internet-Draft, draft-kondreddy-pce-pcep-ls-sync-optimizations-00, 9 October 2015,
<<https://datatracker.ietf.org/doc/html/draft-kondreddy-pce-pcep-ls-sync-optimizations-00>>.

[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>>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February 2006, <<https://www.rfc-editor.org/info/rfc4364>>.

[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>>.

[RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, DOI 10.17487/RFC5316, December 2008, <<https://www.rfc-editor.org/info/rfc5316>>.

[RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, DOI 10.17487/RFC5392, January 2009, <<https://www.rfc-editor.org/info/rfc5392>>.

- [RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <<https://www.rfc-editor.org/info/rfc5925>>.
- [RFC6549] Lindem, A., Roy, A., and S. Mirtorabi, "OSPFv2 Multi-Instance Extensions", RFC 6549, DOI 10.17487/RFC6549, March 2012, <<https://www.rfc-editor.org/info/rfc6549>>.
- [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>>.
- [RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, DOI 10.17487/RFC6952, May 2013, <<https://www.rfc-editor.org/info/rfc6952>>.
- [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>>.
- [RFC8202] Ginsberg, L., Previdi, S., and W. Henderickx, "IS-IS Multi-Instance", RFC 8202, DOI 10.17487/RFC8202, June 2017, <<https://www.rfc-editor.org/info/rfc8202>>.
- [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>>.

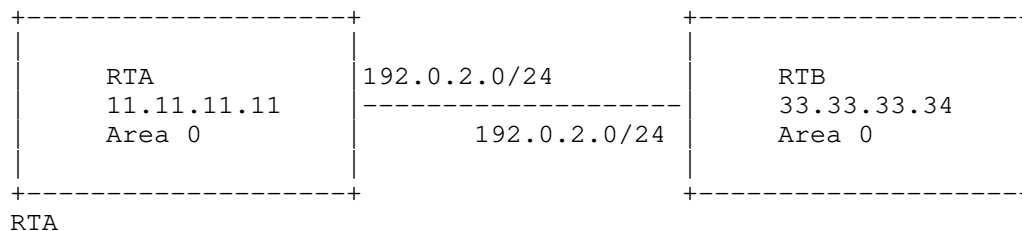
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8283] Farrel, A., 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>>.
- [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>>.
- [RFC8637] Dhody, D., Lee, Y., and D. Ceccarelli, "Applicability of the Path Computation Element (PCE) to the Abstraction and Control of TE Networks (ACTN)", RFC 8637, DOI 10.17487/RFC8637, July 2019, <<https://www.rfc-editor.org/info/rfc8637>>.
- [RFC9168] Dhody, D., Farrel, A., and Z. Li, "Path Computation Element Communication Protocol (PCEP) Extension for Flow Specification", RFC 9168, DOI 10.17487/RFC9168, January 2022, <<https://www.rfc-editor.org/info/rfc9168>>.

Appendix A. Examples

These examples are for illustration purposes only to show how the new PCEP-LS message could be encoded. They are not meant to be an exhaustive list of all possible use cases and combinations.

A.1. All Nodes

Each node (PCC) in the network chooses to provide its own local node and link information, and in this way PCE can build the full link-state and TE information.



LS Node

TLV - Local Node Descriptors
 Sub-TLV - 514: OSPF Area-ID: 0.0.0.0
 Sub-TLV - 515: IGP Router-ID: 11.11.11.11
TLV - Node Attributes TLV
 Sub-TLV(s)

LS Link

TLV - Local Node Descriptors
 Sub-TLV - 514: OSPF Area-ID: 0.0.0.0
 Sub-TLV - 515: IGP Router-ID: 11.11.11.11
TLV - Remote Node Descriptors
 Sub-TLV - 514: OSPF Area-ID: 0.0.0.0
 Sub-TLV - 515: IGP Router-ID: 22.22.22.22
TLV - Link Descriptors
 Sub-TLV - 259: IPv4 interface: 192.0.2.1
 Sub-TLV - 260: IPv4 neighbor: 192.0.2.2
TLV - Link Attributes TLV
 Sub-TLV(s)

RTB

LS Node

TLV - Local Node Descriptors
 Sub-TLV - 514: OSPF Area-ID: 0.0.0.0
 Sub-TLV - 515: IGP Router-ID: 22.22.22.22
TLV - Node Attributes TLV
 Sub-TLV(s)

LS Link

TLV - Local Node Descriptors
 Sub-TLV - 514: OSPF Area-ID: 0.0.0.0
 Sub-TLV - 515: IGP Router-ID: 22.22.22.22
TLV - Remote Node Descriptors
 Sub-TLV - 514: OSPF Area-ID: 0.0.0.0
 Sub-TLV - 515: IGP Router-ID: 11.11.11.11
TLV - Link Descriptors
 Sub-TLV - 259: IPv4 interface: 192.0.2.2
 Sub-TLV - 260: IPv4 neighbor: 192.0.2.1
TLV - Link Attributes TLV
 Sub-TLV(s)

A similar example with IPv6 address (say 2001:db8::1 and 2001:db8::2) for the links could be imagined with all other information as same and just IPv6 interface and neighbor TLVs.

A.2. Designated Node

A designated node(s) in the network will provide its own local node as well as all learned remote information, and in this way PCE can build the full link-state and TE information.

As described in Appendix A.1, the same LS Node and Link objects will be generated with a difference that it would be a designated router say RTA that generate all this information.

A.3. Between PCEs

As per Hierarchical-PCE [RFC6805], Parent PCE builds an abstract domain topology map with each domain as an abstract node and inter-domain links as an abstract link. Each child PCE may provide this information to the parent PCE. Considering the example in figure 1 of [RFC6805], following LS object will be generated:

PCE1

LS Node

TLV - Local Node Descriptors

Sub-TLV - 512: Autonomous System: 100 (Domain 1)

Sub-TLV - 515: IGP Router-ID: 11.11.11.11 (abstract)

LS Link

TLV - Local Node Descriptors

Sub-TLV - 512: Autonomous System: 100

Sub-TLV - 515: IGP Router-ID: 11.11.11.11 (abstract)

TLV - Remote Node Descriptors

Sub-TLV - 512: Autonomous System: 200 (Domain 2)

Sub-TLV - 515: IGP Router-ID: 22.22.22.22 (abstract)

TLV - Link Descriptors

Sub-TLV - 259: IPv4 interface: 192.0.2.1

Sub-TLV - 260: IPv4 neighbor: 192.0.2.2

TLV - Link Attributes TLV

Sub-TLV(s)

LS Link

TLV - Local Node Descriptors

Sub-TLV - 512: Autonomous System: 100

Sub-TLV - 515: IGP Router-ID: 11.11.11.11 (abstract)

TLV - Remote Node Descriptors

Sub-TLV - 512: Autonomous System: 200

Sub-TLV - 515: IGP Router-ID: 22.22.22.22 (abstract)

TLV - Link Descriptors

Sub-TLV - 259: IPv4 interface: 198.51.100.1

Sub-TLV - 260: IPv4 neighbor: 198.51.100.2

TLV - Link Attributes TLV
Sub-TLV(s)

LS Link

TLV - Local Node Descriptors
Sub-TLV - 512: Autonomous System: 100
Sub-TLV - 515: IGP Router-ID: 11.11.11.11 (abstract)
TLV - Remote Node Descriptors
Sub-TLV - 512: Autonomous System: 400 (Domain 4)
Sub-TLV - 515: IGP Router-ID: 44.44.44.44 (abstract)
TLV - Link Descriptors
Sub-TLV - 259: IPv4 interface: 203.0.113.1
Sub-TLV - 260: IPv4 neighbor: 203.0.113.2
TLV - Link Attributes TLV
Sub-TLV(s)

* similar information will be generated by other PCE
to help form the abstract domain topology.

Further the exact border nodes and abstract internal path between the border nodes may also be transported to the Parent PCE to enable ACTN as described in [RFC8637] using the similar LS node and link objects encodings.

Appendix B. Contributor Addresses

Udayasree Palle

Email: udayasreereddy@gmail.com

Sergio Belotti
Nokia

Email: sergio.belotti@nokia.com

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

Email: satishk@huawei.com

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

Email: c.l@huawei.com

Authors' Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore 560066
Karnataka
India
Email: dhruv.ietf@gmail.com

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

Young Lee
Samsung Electronics
Seoul
South Korea
Email: younglee.tx@gmail.com

Daniele Ceccarelli
Ericsson
Torshamnsgatan, 48
Stockholm
Sweden
Email: daniele.ceccarelli@ericsson.com

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing
102209
China
Email: wangaijun@tsinghua.org.cn

Gyan Mishra
Verizon Inc.
Email: gyan.s.mishra@verizon.com

Siva Sivabalan
Ciena Corporation
Email: ssivabal@ciena.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: February 2, 2020

I. Minei
Google, Inc.
E. Crabbe
Individual Contributor
S. Sivabalan
Cisco Systems, Inc.
H. Ananthakrishnan
Netflix
D. Dhody
Huawei
Y. Tanaka
NTT Communications Corporation
August 1, 2019

Path Computation Element Communication Protocol (PCEP) Extensions for
Establishing Relationships Between Sets of Label Switched Paths (LSPs)
draft-ietf-pce-association-group-10

Abstract

This document introduces a generic mechanism to create a grouping of Label Switched Paths (LSPs) in the context of a Path Computation Element (PCE). This grouping can then be used to define associations between sets of LSPs or between a set of LSPs and a set of attributes (such as configuration parameters or behaviours), and is equally applicable to the stateful PCE (active and passive modes) as well as the stateless PCE.

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 February 2, 2020.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<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. Architectural Overview	4
3.1. Motivation	4
3.2. Relationship with the SVEC List	5
3.3. Operation Overview	5
3.4. Operator-configured Association Range	6
4. Discovery of Supported Association Types	7
4.1. ASSOC-Type-List TLV	7
4.1.1. Procedure	8
5. Operator-configured Association Range TLV	8
5.1. Procedure	10
6. ASSOCIATION Object	11
6.1. Object Definition	11
6.1.1. Global Association Source TLV	13
6.1.2. Extended Association ID TLV	13
6.1.3. Association Source	14
6.1.4. Unique Identification for an Association Group	14
6.2. Relationship with the RSVP ASSOCIATION	15
6.3. Object Encoding in PCEP messages	15
6.3.1. Stateful PCEP messages	15
6.3.2. Request Message	17
6.3.3. Reply Message	18
6.4. Processing Rules	19

7.	IANA Considerations	21
7.1.	PCEP Object	21
7.2.	PCEP TLV	22
7.3.	Association Flags	22
7.4.	Association Type	23
7.5.	PCEP-Error Object	23
8.	Security Considerations	24
9.	Manageability Considerations	24
9.1.	Control of Function and Policy	24
9.2.	Information and Data Models	24
9.3.	Liveness Detection and Monitoring	25
9.4.	Verify Correct Operations	25
9.5.	Requirements on Other Protocols	25
9.6.	Impact on Network Operations	25
10.	Acknowledgments	25
11.	Contributors	25
12.	References	26
12.1.	Normative References	26
12.2.	Informative References	27
	Appendix A. Example for Operator-configured Association Range .	28
	Authors' Addresses	29

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].

[RFC4872] defines the RSVP ASSOCIATION object, which was defined in the context of GMPLS-controlled Label Switched Paths (LSPs) to be used to associate recovery LSPs with the LSP they are protecting. This object also has broader applicability as a mechanism to associate RSVP state and [RFC6780] described how the ASSOCIATION object can be more generally applied by defining the Extended ASSOCIATION Object.

This document introduces a generic mechanism to create a grouping of LSPs. This grouping can then be used to define associations between sets of LSPs or between a set of LSPs and a set of attributes (such as configuration parameters or behaviours), and is equally applicable to stateful PCE (active and passive modes) and stateless PCE. The associations could be created dynamically and conveyed to a PCEP peer within PCEP, or it could be configured manually by an operator on the PCEP peers. Refer Section 3.3 for more details.

2. Terminology

This document uses the following terms defined in [RFC5440]: PCC, PCE, PCEP Peer, Path Computation Request (PCReq), Path Computation Reply (PCRep), and PCEP Error (PCErr).

This document uses the following terms defined in [RFC8051]: Stateful PCE, Active Stateful PCE, Passive Stateful PCE, and Delegation.

This document uses the following terms defined in [RFC8231]: LSP State Report (PCRpt), LSP Update Request (PCUpd), and State Timeout Interval.

This document uses the following terms defined in [RFC8281]: PCE-initiated LSP, and LSP Initiate Request (PCInitiate).

3. Architectural Overview

3.1. Motivation

Stateful PCE provides the ability to update existing LSPs and to instantiate new ones. There are various situations where several LSPs need to share common information. E.g., to support for PCE-controlled make-before-break, an association between the original and the re-optimized path is desired. Similarly, for end-to-end protection, the association between working and protection LSPs is required (see [I-D.ietf-pce-stateful-path-protection]). For diverse paths, an association between a group of LSPs could be used (see [I-D.ietf-pce-association-diversity]). Another use for the LSP grouping is for applying a common set of configuration parameters or behaviours to a set of LSPs.

For a stateless PCE, it might be useful to associate a path computation request to an association group, thus enabling it to associate a common set of policy, configuration parameters or behaviours with the request.

Some associations could be created dynamically, such as association between the working and protection LSPs of a tunnel, whereas some

associations could be created by the operator manually, such as policy-based association, where the LSP could join an operator-configured existing association.

Rather than creating separate mechanisms for each use case, this document defines a generic mechanism that can be reused as needed.

3.2. Relationship with the SVEC List

Note that, [RFC5440] defines a mechanism for the synchronization of a set of path computation requests by using the SVEC (Synchronization VECTOR) object, that specifies the list of synchronized requests that can either be dependent or independent. The SVEC object identifies the relationship between the set of path computation requests, identified by 'Request-ID-number' in RP (Request Parameters) object. [RFC6007] further clarifies the use of the SVEC list for synchronized path computations when computing dependent requests as well as describes a number of usage scenarios for SVEC lists within single-domain and multi-domain environments.

The motivation behind the association group defined in this document and the SVEC object are quite different, though some use cases may overlap. PCEP extensions that define a new association type should clarify the relationship between the SVEC object and the association type, if any.

3.3. Operation Overview

LSPs are associated with other LSPs with which they interact by adding them to a common association group. Association groups as defined in this document can be applied to LSPs originating at the same head end or different head ends.

Some associations could be created dynamically by a PCEP speaker and the associations (along with the set of LSPs) are conveyed to a PCEP peer. Whereas some associations are configured by the operator on the PCEP peers involved beforehand, a PCEP speaker then could ask for a LSP to join the operator-configured association. Usage of dynamic and configured association is usually dependent on the type of the association.

For the operator-configured association, the association parameters such as the association identifier, association type, as well as the association source IP address, are manually configured by the operator. In case of dynamic association, the association parameters such as the association identifier, are allocated dynamically by the PCEP speaker, the association source is set as local PCEP speaker

address, unless local policy dictates otherwise, in which case association source is set based on the local policy.

The dynamically created association can be reported to the PCEP peer via the PCEP messages as per the stateful extensions. When the operator-configured association is known to the PCEP peer beforehand, a PCEP peer could ask for a LSP to join the operator-configured association via the stateful PCEP messages.

The associations are properties of the LSP and thus could be stored in the LSP state database. The dynamic association exists as long as the LSP state exists. In case of PCEP session termination, the LSP state clean-up MUST also take care of associations.

Multiple types of associations can exist, each with their own association identifier space. The definition of the different association types and their behaviours is outside the scope of this document. The establishment and removal of the association relationship can be done on a per LSP basis. An LSP may join multiple association groups, of different or of the same association type.

3.4. Operator-configured Association Range

Some association types are dynamic, some are operator-configured and some could be both. For the association types that could be both dynamic and operator-configured and use the same association source, it is necessary to distinguish a range of association identifiers that are marked for operator-configured associations to avoid any association identifier clash within the scope of the association source. This document assumes that these two ranges are configured.

A range of association identifiers for each Association type (and Association Source) are kept for the operator-configured associations. Dynamic associations MUST NOT use the association identifier from this range.

This range as set at the PCEP speaker (PCC or PCE, as an association source) needs to be communicated to a PCEP peer in the Open Message. A new TLV is defined in this specification for this purpose (Section 5). See Appendix A for an example.

Association identifier range for sources other than the PCEP speaker (for example an NMS system) is not communicated in PCEP and the procedure for operator-configured association range setting is outside the scope of this document.

4. Discovery of Supported Association Types

This section defines PCEP extensions so as to support the capability advertisement of the association types supported by a PCEP speaker.

A new PCEP ASSOC-Type-List (Association Types list) TLV is defined. The PCEP ASSOC-Type-List TLV is carried within an OPEN object. This way, during PCEP session-setup phase, a PCEP speaker can advertise to a PCEP peer the list of supported Association types.

4.1. ASSOC-Type-List TLV

The PCEP ASSOC-Type-List TLV is OPTIONAL. It MAY be carried within an OPEN object sent by a PCEP speaker in an Open message to a PCEP peer so as to indicate the list of supported Association types.

The PCEP ASSOC-Type-List TLV format is compliant with the PCEP TLV format defined in [RFC5440]. That is, the TLV is composed of 2 octets for the type, 2 octets specifying the TLV length, and a Value field. The Length field defines the length of the value portion in octets. The TLV is padded to 4-octet alignment, and padding is not included in the Length field (e.g., a 3-octet value would have a length of three, but the total size of the TLV would be eight octets).

The PCEP ASSOC-Type-List TLV has the following format:

TYPE: TBD
 LENGTH: $N * 2$ (where N is the number of association types)
 VALUE: list of 2-byte association type code points, identifying the association types supported by the sender of the Open message.

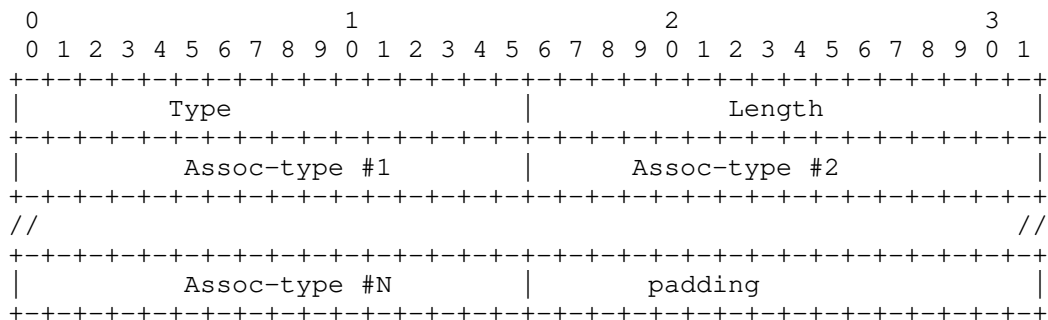


Figure 1: The ASSOC-Type-List TLV format

Assoc-type (2 bytes): Association type code point identifier. IANA manages the "ASSOCIATION Type Field" code point registry (see Section 7.4).

4.1.1. Procedure

An ASSOC-Type-List TLV within an OPEN object in an Open message is included by a PCEP speaker in order to advertise a set of one or more supported association types. The ASSOC-Type-List TLV MUST NOT appear more than once in an OPEN object. If it appears more than once, the PCEP session MUST be rejected with error type 1 and error value 1 (PCEP session establishment failure / Reception of an invalid Open message). As specified in [RFC5440], a PCEP peer that does not recognize the ASSOC-Type-List TLV will silently ignore it.

Association type (to be defined in other documents) can specify if the association type advertisement is mandatory for it. Thus, the ASSOC-Type-List TLV MUST be included if at least one mandatory association type needs to be advertised and the ASSOC-Type-List TLV MAY be included otherwise. For an association type that specifies that the advertisement is mandatory, a missing Assoc-type in the ASSOC-Type-List TLV (or missing ASSOC-Type-List TLV) is to be interpreted as the association type is not supported by the PCEP speaker.

The absence of the ASSOC-Type-List TLV in an OPEN object MUST be interpreted as an absence of information on the list of supported Association types (rather than the Association type is not supported). In this case, the PCEP speaker could still use the ASSOCIATION object: if the peer does not support the association, it will react as per the procedure described in Section 6.4.

In case the use of the ASSOC-Type-List TLV is triggered by support for a mandatory association type, then it is RECOMMENDED that the PCEP implementation include all supported Association types (including optional) to ease the operations of the PCEP peer.

5. Operator-configured Association Range TLV

This section defines PCEP extension to support the advertisement of the Operator-configured Association Range used for an Association type by the PCEP speaker (as an Association source).

A new PCEP OP-CONF-ASSOC-RANGE (Operator-configured Association Range) TLV is defined. The PCEP OP-CONF-ASSOC-RANGE TLV is carried within an OPEN object. This way, during PCEP session-setup phase, a PCEP speaker can advertise to a PCEP peer the Operator-configured Association Range for an association type.

The PCEP OP-CONF-ASSOC-RANGE TLV is OPTIONAL. It MAY be carried within an OPEN object sent by a PCEP speaker in an Open message to a PCEP peer. The OP-CONF-ASSOC-RANGE TLV format is compliant with the PCEP TLV format defined in [RFC5440]. That is, the TLV is composed of 2 bytes for the type, 2 bytes specifying the TLV length, and a Value field. The Length field defines the length of the value portion in bytes.

The PCEP OP-CONF-ASSOC-RANGE TLV has the following format:

TYPE: 29 (Early allocation by IANA)

LENGTH: $N * 8$ (where N is the number of association types)

VALUE:

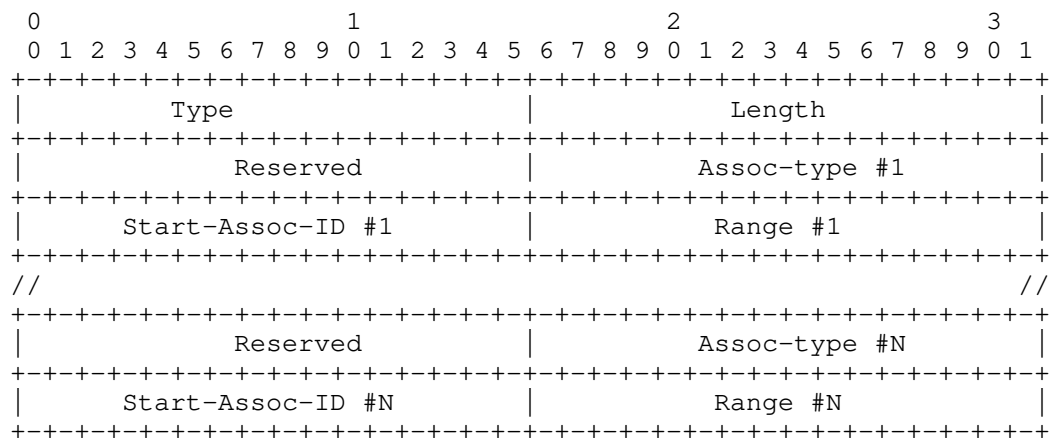


Figure 2: The OP-CONF-ASSOC-RANGE TLV format

The Value portion includes the following fields, repeated for each association type:

Reserved (2 bytes): This field MUST be set to 0 on transmission and MUST be ignored on receipt.

Assoc-type (2 bytes): The association type (Section 7.4). The association types are defined in separate documents.

Start-Assoc-ID (2 bytes): The start association identifier for the Operator-configured Association Range for the particular association type. The values 0 and 0xffff MUST NOT be used and on receipt of these values in the TLV, the session is rejected with error message sent (see Section 5.1).

Range (2 bytes): The number of associations marked for the Operator-configured Associations. The Range MUST be greater than 0, and it MUST be such that (Start-Assoc-ID + Range) do not cross the association identifier range of 0xffff. In case this condition is not satisfied, the session is rejected with error message sent (see Section 5.1).

5.1. Procedure

A PCEP speaker MAY include an OP-CONF-ASSOC-RANGE TLV within an OPEN object in an Open message sent to a PCEP peer in order to advertise the Operator-configured Association Range for an association type. The OP-CONF-ASSOC-RANGE TLV MUST NOT appear more than once in an OPEN object. If it appears more than once, the PCEP session MUST be rejected with error type 1 and error value 1 (PCEP session establishment failure / Reception of an invalid Open message).

As specified in [RFC5440], a PCEP peer that does not recognize the OP-CONF-ASSOC-RANGE TLV will silently ignore it.

The Operator-configured Association Range SHOULD be included for each association type that could be both dynamic and operator-configured. For association types that are only dynamic or only operator-configured, this TLV MAY be skipped, in which case the full range of association identifier is considered dynamic or operator-configured respectively. Each association type (that are defined in separate documents) can specify the default value for the operator-configured association range for their respective association type.

The absence of the OP-CONF-ASSOC-RANGE TLV in an OPEN object MUST be interpreted as an absence of explicit Operator-configured Association Range at the PCEP peer. In this case, the default behavior as per each association type applies. If the association source is not a PCEP speaker, the default value for the operator-configured association range is used for the association source.

If the Assoc-type is not recognized or supported by the PCEP speaker, it MUST ignore that respective Start-Assoc-ID and Range. If the Assoc-type is recognized/supported but the Start-Assoc-ID or Range are set incorrectly, the PCEP session MUST be rejected with error type 1 and error value 1 (PCEP session establishment failure / Reception of an invalid Open message). The incorrect range include the case when the (Start-Assoc-ID + Range) crosses the association identifier range of 0xffff.

A given Assoc-type MAY appear more than once in the OP-CONF-ASSOC-RANGE TLV in the case of a non-contiguous Operator-configured Association Range. The PCEP speaker originating this TLV MUST NOT

carry overlapping ranges for an association type. If a PCEP peer receives overlapping ranges for an association type, it MUST consider the Open message malformed and MUST reject the PCEP session with error type 1 and error value 1 (PCEP session establishment failure / Reception of an invalid Open message).

There may be cases where an operator-configured association was configured with association parameters (such as association identifier, association type and association source) at the local PCEP speaker, and later the PCEP session gets established with the association source and a new operator-configured range is learned during session establishment. At this time, the local PCEP speaker MUST remove any associations that are not in the new operator-configured range (by disassociating any LSPs that are part of it (and notifying this change to the PCEP peer)). If a PCEP speaker receives an association for an operator-configured association and the association identifier is not in the operator-configured association range for the association type and association source, it MUST generate an error (as described in Section 6.4).

6. ASSOCIATION Object

6.1. Object Definition

Association groups and their memberships are defined using a new ASSOCIATION object.

ASSOCIATION Object-Class is 40 (Early allocation by IANA).

ASSOCIATION Object-Type is 1 for IPv4 and its format is shown in Figure 3:

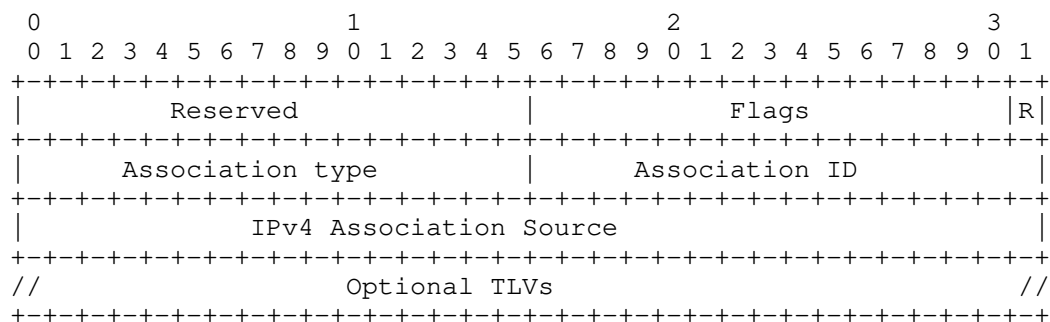


Figure 3: The IPv4 ASSOCIATION Object format

ASSOCIATION Object-Type is 2 for IPv6 and its format is shown in Figure 4:

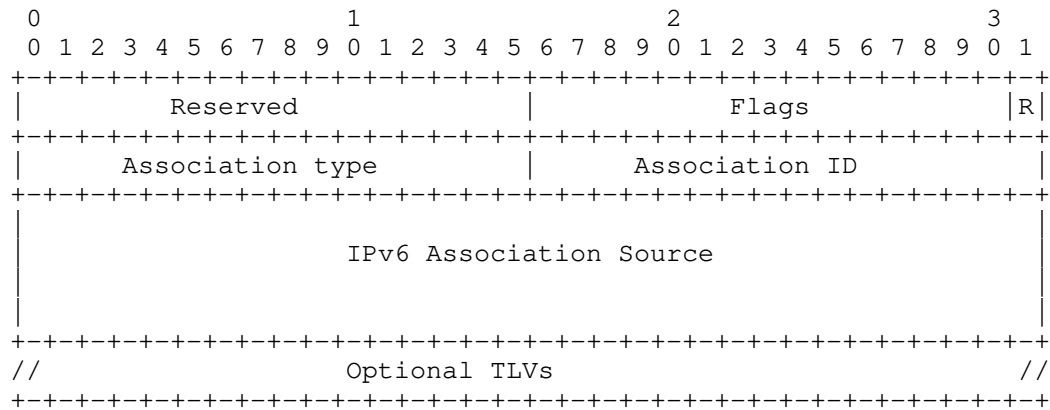


Figure 4: The IPv6 ASSOCIATION Object format

Reserved (2-byte): MUST be set to 0 and ignored upon receipt.

Flags (2-byte): The following flags are currently defined:

R (Removal - 1 bit): when set, the requesting PCEP peer requires the removal of an LSP from the association group. When unset, the PCEP peer indicates that the LSP is added or retained as part of the association group. This flag is used for the ASSOCIATION object in the PCRpt and the PCUpd message, the flag is ignored in other PCEP messages.

The unassigned flags MUST be set to zero on transmission and MUST be ignored on receipt.

Association type (2-byte): the association type (Section 7.4). The association types are defined in separate documents.

Association ID (2-byte): the identifier of the association group. When combined with other association parameters, such as Association Type and Association Source, this value uniquely identifies an association group. The values 0xffff and 0x0 are reserved. The value 0xffff is used to indicate all association groups and could be used with R flag to indicate removal for all associations for the LSP within the scope of association type and association source.

Association Source: Contains a valid IPv4 address (4 bytes) if the ASSOCIATION Object-Type is 1 or a valid IPv6 address (16 bytes) if the ASSOCIATION Object-Type is 2. The address provides scoping for the Association ID. See Section 6.1.3 for details.

Optional TLVs: The optional TLVs follow the PCEP TLV format of [RFC5440]. This document defines two optional TLVs. Other documents can define more TLVs in future.

6.1.1. Global Association Source TLV

The Global Association Source TLV is an optional TLV for use in the Association Object. The meaning and the usage of Global Association Source is as per Section 4 of [RFC6780].

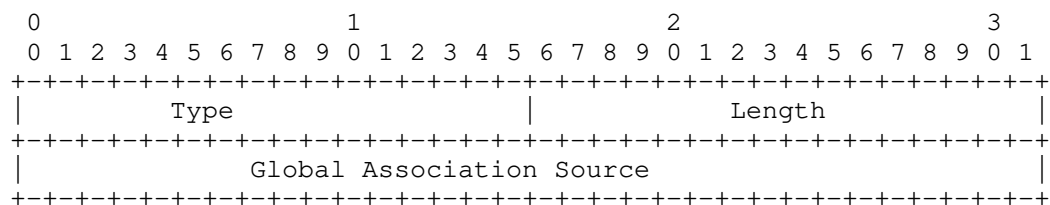


Figure 5: The Global Association Source TLV format

Type: 30 (Early allocation by IANA).

Length: Fixed value of 4 bytes.

Global Association Source: as defined in Section 4 of [RFC6780].

6.1.2. Extended Association ID TLV

The Extended Association ID TLV is an optional TLV for use in the Association Object. The meaning and the usage of Extended Association ID is as per Section 4 of [RFC6780].

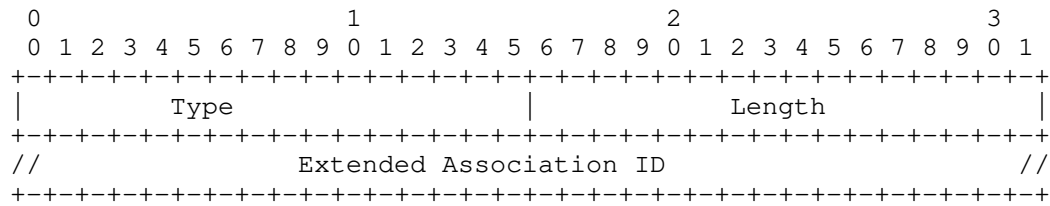


Figure 6: The Extended Association ID TLV format

Type: 31 (Early allocation by IANA).

Length: variable.

Extended Association ID: as defined in Section 4 of [RFC6780].

6.1.3. Association Source

The Association Source field in the ASSOCIATION object is set to a valid IP address to identify the node that originates the association. In case of dynamic associations, the association source is usually set as the local PCEP speaker address, unless local policy dictates otherwise, in which case association source is set based on the local policy. In case of PCE redundancy, local policy could set the source as a virtual IP address which identifies all instances of the PCE. In case of operator-configured association, the association source is manually configured and it could be set as one of the PCEP speakers, Network Management System (NMS), or any other valid IP address that scopes the association identifier for the association type.

6.1.4. Unique Identification for an 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. In this document, all these fields are collectively called 'association parameters'. Note that the ASSOCIATION object MAY include other optional TLVs (not defined in this document) based on the association types, that provides 'information' related to the association type, this document uses the term 'association information' for it.

6.2. Relationship with the RSVP ASSOCIATION

The format of PCEP ASSOCIATION Object defined in this document is aligned with the RSVP ASSOCIATION object ([RFC6780]). Various Association types related to RSVP association are defined in [RFC4872], [RFC4873], and [RFC7551]. The PCEP extensions that define new association types, should clarify how the PCEP associations would work with RSVP associations and vice-versa.

6.3. Object Encoding in PCEP messages

Message formats in this document are expressed using Reduced BNF (RBNF) as used in [RFC5440] and defined in [RFC5511].

6.3.1. Stateful PCEP messages

The ASSOCIATION Object MAY be carried in the Path Computation Update (PCUpd), Path Computation Report (PCRpt) and Path Computation Initiate (PCInitiate) messages.

When carried in PCRpt message, it is used to report the association group membership pertaining to a LSP to a stateful PCE. The PCRpt message are used for both initial state synchronization operations (Section 5.6 of [RFC8231]) as well as whenever the state of the LSP changes. If the LSP belongs to an association group, then the associations MUST be included during the state synchronization operations.

The PCRpt message can also be used to remove an LSP from one or more association groups by setting the R flag to 1 in the ASSOCIATION object.

When an LSP is first reported to the PCE, the PCRpt message MUST include all the association groups that it belongs to. Any subsequent report message SHOULD include only the associations that are being modified or removed.

The PCRpt message is defined in [RFC8231] and updated as below:

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

Where:

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

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

Where:

```
<path> ::= <intended-path>
           [<actual-attribute-list> <actual-path>]
           <intended-attribute-list>
```

```
<association-list> ::= <ASSOCIATION> [<association-list>]
```

When an LSP is delegated to a stateful PCE, the stateful PCE can create a new association group for this LSP, or associate it with one or more existing association groups. This is done by including the ASSOCIATION Object in a PCUpd message. A stateful PCE can also remove a delegated LSP from one or more association groups by setting the R flag to 1 in the ASSOCIATION object.

The PCUpd message SHOULD include the association groups that are being modified or removed, there is no need to include associations that remains unchanged.

The PCUpd message is defined in [RFC8231] and updated as below:

```
<PCUpd Message> ::= <Common Header>
                        <update-request-list>
```

Where:

```
<update-request-list> ::= <update-request> [<update-request-list>]
```

```
<update-request> ::= <SRP>
                    <LSP>
                    [<association-list>]
                    <path>
```

Where:

```
<path> ::= <intended-path> <intended-attribute-list>
```

```
<association-list> ::= <ASSOCIATION> [<association-list>]
```


Unless a PCEP speaker wants to delete an association from an LSP or make changes to the association, it does not need to carry the ASSOCIATION object in future stateful messages.

A PCE initiating a new LSP can also include the association groups that this LSP belongs to. This is done by including the ASSOCIATION Object in a PCInitiate message. The PCInitiate message MUST include all the association groups that it belongs to. The PCInitiate message is defined in [RFC8281] and updated as below:

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

Where:

```
<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>
                                       [<association-list>]
                                       [<attribute-list>]
```

Where:

```
<association-list> ::= <ASSOCIATION> [<association-list>]
```

6.3.2. Request Message

In case of passive (stateful or stateless) PCE, the ASSOCIATION Object is OPTIONAL and MAY be carried in the Path Computation Request (PCReq) message.

When carried in a PCReq message, the ASSOCIATION Object is used to associate the path computation request to an association group. The association (and the other LSPs) should be known to the PCE beforehand. These could be operator-configured or dynamically learned before via stateful PCEP messages. The R flag in ASSOCIATION object within PCReq message MUST be set to 0 while sending and ignored on receipt.

The PCReq message is defined in [RFC5440] and updated in [RFC8231] , it is further updated below for association:

```
<PCReq Message> ::= <Common Header>
                    [<svec-list>]
                    <request-list>
```

Where:

```
<svec-list> ::= <SVEC> [<svec-list>]
<request-list> ::= <request> [<request-list>]
```

```
<request> ::= <RP>
              <END-POINTS>
              [<LSP>]
              [<LSPA>]
              [<BANDWIDTH>]
              [<metric-list>]
              [<association-list>]
              [<RRO> [<BANDWIDTH>]]
              [<IRO>]
              [<LOAD-BALANCING>]
```

Where:

```
<association-list> ::= <ASSOCIATION> [<association-list>]
```

Note that the LSP object MAY be present for the passive stateful PCE mode.

6.3.3. Reply Message

In case of passive (stateful or stateless) PCE, the ASSOCIATION Object is OPTIONAL and MAY be carried in the Path Computation Reply (PCRep) message with the NO-PATH object. The ASSOCIATION object in PCRep message indicates the association group that cause the PCE to fail to find a path.

The PCRep message is defined in [RFC5440] and updated in [RFC8231] , it is further updated below for association:

```
<PCRep Message> ::= <Common Header>
                        <response-list>
```

Where:

```
<response-list> ::= <response> [<response-list>]
```

```
<response> ::= <RP>
                [<LSP>]
                [<NO-PATH>]
                [<association-list>]
                [<attribute-list>]
                [<path-list>]
```

Where:

```
<association-list> ::= <ASSOCIATION> [<association-list>]
```

Note that the LSP object MAY be present for the passive stateful PCE mode.

6.4. Processing Rules

Association groups can be operator-configured on the necessary PCEP speakers and the PCEP speakers can join the existing association groups. In addition, a PCC or a PCE can create association groups dynamically and the PCEP speaker can also report the associations to its peer via PCEP messages. The operator-configured associations are created via configurations (where all association parameters are manually set) and exist until explicitly removed via configurations. The PCEP speaker can add LSPs to these configured associations and carry this via stateful PCEP messages. The dynamic associations are created dynamically by the PCEP speaker (where all association parameters are populated dynamically). The association group is attached to the LSP state, and the association group exists till there is at least one LSP as part of the association. As described in Section 6.1.4, the association parameters are the combination of Association type, Association ID and Association Source as well as optional global source and extended association identifier, that uniquely identifies an association group. The information related to the association types encoded via the TLVs of a particular association type (not described in this document) are the association information (Section 6.1.4).

If a PCEP speaker does not recognize the ASSOCIATION object in the stateful message, it will return a PCErr message with Error-Type "Unknown Object" as described in [RFC5440]. In case of PCReq message, the PCE would react based on the P flag as per [RFC5440]. If a PCEP speaker understands the ASSOCIATION object but does not support the 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". If any association parameters are invalid in the ASSOCIATION object, the PCEP speaker would consider this as malformed object and handle it as malformed message [RFC5440]. On receiving a PCEP message with ASSOCIATION, if a PCEP speaker finds that too many LSPs belong to the association group, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 2 "Too many LSPs in the association group". If a PCEP speaker cannot handle a new association, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 3 "Too many association groups". These numbers MAY be set by operator or decided based on a local policy.

If a PCE peer is unwilling or unable to process the ASSOCIATION object in the stateful message, it MUST return a PCErr message with the Error-Type "Not supported object" and follow the relevant procedures described in [RFC5440]. In case of PCReq message, the PCE would react based on the P flag as per [RFC5440]. On receiving a PCEP message with ASSOCIATION, if a PCEP speaker could not add the LSP to the association group for any reason, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 7 "Cannot join the association group".

If a PCEP speaker receives an ASSOCIATION object for an operator-configured association and the association identifier is not in the operator-configured association range for the Association type and Association Source, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 8 "Association identifier not in range".

If a PCEP speaker receives ASSOCIATION in PCReq message, and the association is not known (association is not configured, or created dynamically, or learned from a PCEP peer), it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 4 "Association unknown".

If the association information (related to the association group as a whole) received from the peer does not match with the local operator-configured information, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 5 "Operator-configured association information mismatch". On receiving association information (related to the association group as a whole) that does not match with the association information previously received about the same association from a peer, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 6 "Association information mismatch". Note that information related to each LSP within the

association as part of the association information TLVs could be different.

If a PCEP speaker receives an ASSOCIATION object with the R bit set for removal, and the association group (identified by association parameters) is not known, it MUST return a PCErr message with Error-Type 26 (Early allocation by IANA) "Association Error" and Error-Value 4 "Association unknown".

The dynamic associations are cleared along with the LSP state information as per the [RFC8231]. When a PCEP session is terminated, after expiry of State Timeout Interval at PCC, the LSP state associated with that PCEP session is reverted to operator-defined default parameters or behaviours. Same procedure is also followed for the association groups. On session termination at the PCE, when the LSP state reported by PCC is cleared, the association groups are also cleared. When there are no LSPs in an association group, the association is considered to be empty and thus deleted.

In case the LSP is delegated to another PCE on session failure, the associations (and association information) set by the PCE remains intact, unless updated by the new PCE that takes over.

Upon LSP delegation revocation, the PCC MAY clear the association created by the PCE, but in order to avoid traffic loss, it SHOULD perform this in a make-before-break fashion (same as [RFC8231]).

7. IANA Considerations

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" registry at <<http://www.iana.org/assignments/pcep>>.

7.1. PCEP Object

The "PCEP Numbers" registry contains a subregistry "PCEP Objects". IANA is requested to confirm the early allocation of the following code point in the PCEP Objects registry.

Object-Class Value	Name	Reference
40 (Early allocation by IANA)	Association	[This.I-D]
	Object-Type	
	0: Reserved	
	1: IPv4	
	2: IPv6	

7.2. PCEP TLV

IANA is requested to confirm the early allocation of the following code point in the "PCEP TLV Type Indicators" registry.

Value	Meaning	Reference
29 (Early allocation by IANA)	Operator-configured Association Range	[This.I-D]
30 (Early allocation by IANA)	Global Association Source	[This.I-D]
31 (Early allocation by IANA)	Extended Association ID	[This.I-D]

IANA is requested to fix the meaning for value 31 in the above registry to 'Extended Association ID', it is currently mentioned as 'Extended Association Id'.

IANA is also requested to make a new assignment for the existing "PCEP TLV Type Indicators" registry as follows:

Value	Meaning	Reference
TBD	ASSOC-Type-List	[This.I-D]

7.3. Association Flags

This document requests IANA to create a subregistry of the "PCEP Numbers" for the bits carried in the Flags field of the ASSOCIATION object. The subregistry is called "ASSOCIATION Flags Field". 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
15	R (Removal)	[This.I-D]

7.4. Association Type

This document requests IANA to create a subregistry of the "PCEP Numbers" for the Association Type field of the the ASSOCIATION object. The subregistry is called "ASSOCIATION Type Field". New values are to be assigned by Standards Action [RFC8126]. Each value should be tracked with the following qualities:

- o Type
- o Name
- o Reference

There are no association types specified in this document, future documents should request the assignment of association types from this subregistry.

7.5. PCEP-Error Object

IANA is requested to confirm the early allocation of the following code points within the "PCEP-ERROR Object Error Types and Values" sub-registry of the "PCEP Numbers" registry, as follows:

Error-Type	Meaning
26	Association Error [This.I-D]
(early	Error-value=0:
alloc by	Unassigned
IANA)	Error-value=1:
	Association type is not supported
	Error-value=2:
	Too many LSPs in the association group
	Error-value=3:
	Too many association groups
	Error-value=4:
	Association unknown
	Error-value=5:
	Operator-configured association
	information mismatch
	Error-value=6:
	Association information mismatch
	Error-value=7:
	Cannot join the association group
	Error-value=8:
	Association identifier not in range

8. Security Considerations

The security considerations described in [RFC8231] and [RFC5440] apply to the extensions described in this document as well. Additional considerations related to a malicious PCEP speaker are introduced, as associations could be spoofed and could be used as an attack vector. An attacker could attempt to create too many associations in an attempt to load the PCEP peer. The PCEP peer responds with PCErr as described in Section 6.4. An attacker could impact LSP operations by creating bogus associations. Further, association groups 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.

Much of the information carried in the ASSOCIATION object, as per this document is not extra sensitive. It often reflects information that can also be derived from the LSP Database, but association provides a much easier grouping of related LSPs and messages. Implementations and operator can and should use indirect values in ASSOCIATION as a way to hide any sensitive business relationships.

9. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440] and [RFC8231] apply to PCEP protocol extensions defined in this document. In addition, requirements and considerations listed in this section apply.

9.1. Control of Function and Policy

A PCE or PCC implementation MUST allow operator-configured associations and SHOULD allow setting of the operator-configured association range (Section 3.4) as described in this document.

9.2. 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 association groups.

An implementation SHOULD allow the operator to view the associations configured or created dynamically. Further implementation SHOULD allow to view associations reported by each peer, and the current set of LSPs in the association.

It might also be useful to find out how many associations for each association type currently exist and to know how many free association identifiers are available for a particular association type and source.

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

Mechanisms defined in this document do not imply any new requirements on other protocols.

9.6. Impact on Network Operations

Mechanisms defined in [RFC5440] and [RFC8231] also apply to PCEP extensions defined in this document.

10. Acknowledgments

We would like to thank Yuji Kamite and Joshua George for their contributions to this document. Also thanks to Venugopal Reddy, Cyril Margaria, Rakesh Gandhi, Mike Koldychev, and Adrian Farrel for their useful comments.

We would like to thank Julien Meuric for shepherding this document and providing comments with text suggestions.

Thanks to Stig Venaas for the RTGDIR review comments.

Thanks to Alvaro Retana, Mirja Kuhlewind, Martin Vigoureux, Barry Leiba, Eric Vyncke, Suresh Krishnan, and Benjamin Kaduk for the IESG comments.

11. Contributors

Stephane Litkowski
Orange

Email: stephane.litkowski@orange.com

Xian Zhang
Huawei Technologies
F3-1-B RnD Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129
China

Email: zhang.xian@huawei.com

Mustapha Aissaoui
Nokia

Email: mustapha.aissaoui@nokia.com

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>>.
- [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>>.
- [RFC6780] Berger, L., Le Faucheur, F., and A. Narayanan, "RSVP ASSOCIATION Object Extensions", RFC 6780, DOI 10.17487/RFC6780, October 2012, <<https://www.rfc-editor.org/info/rfc6780>>.

- [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>>.
- [RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a Stateful Path Computation Element (PCE)", RFC 8051, DOI 10.17487/RFC8051, January 2017, <<https://www.rfc-editor.org/info/rfc8051>>.
- [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>>.

12.2. 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>>.
- [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, DOI 10.17487/RFC4872, May 2007, <<https://www.rfc-editor.org/info/rfc4872>>.
- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, DOI 10.17487/RFC4873, May 2007, <<https://www.rfc-editor.org/info/rfc4873>>.

- [RFC6007] Nishioka, I. and D. King, "Use of the Synchronization VECtor (SVEC) List for Synchronized Dependent Path Computations", RFC 6007, DOI 10.17487/RFC6007, September 2010, <<https://www.rfc-editor.org/info/rfc6007>>.
- [RFC7551] Zhang, F., Ed., Jing, R., and R. Gandhi, Ed., "RSVP-TE Extensions for Associated Bidirectional Label Switched Paths (LSPs)", RFC 7551, DOI 10.17487/RFC7551, May 2015, <<https://www.rfc-editor.org/info/rfc7551>>.
- [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.ietf-pce-pcep-yang] Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-12 (work in progress), July 2019.
- [I-D.ietf-pce-stateful-path-protection] Ananthakrishnan, H., Sivabalan, S., Barth, C., Minei, I., and M. Negi, "PCEP Extensions for Associating Working and Protection LSPs with Stateful PCE", draft-ietf-pce-stateful-path-protection-07 (work in progress), June 2019.
- [I-D.ietf-pce-association-diversity] Litkowski, S., Sivabalan, S., Barth, C., and M. Negi, "Path Computation Element Communication Protocol (PCEP) Extension for LSP Diversity Constraint Signaling", draft-ietf-pce-association-diversity-08 (work in progress), July 2019.

Appendix A. Example for Operator-configured Association Range

Consider an association type T1 (which allows both dynamic and operator-configured association with a default range of <0x1000, 0xffff>). Consider that, because of need of the network, the PCE needs to create more dynamic associations and would like to change the association range to <0xbffe, 0xffff> instead. During PCEP session establishment the PCE would advertise the new range, the PCC could skip advertising as the default values are used. If a PCC is creating a dynamic association (with PCC as association source) it needs to pick a free association identifier for type T1 in the range of <0x1, 0x0fff> whereas if a PCE is creating a dynamic association (with PCE as association source) it needs to pick a free association

identifier from the range <0x1, 0xbffd>. Similarly if an operator-configured association is manually configured with the PCC as association source, it should be from the range <0x1000, 0xffff> whereas if the PCE is association source, it should be from <0xbffe, 0xffff>. In case the association source is not a PCEP peer (for example an NMS system), then the default range of <0x1000, 0xffff> is considered.

Authors' Addresses

Ina Minei
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: inaminei@google.com

Edward Crabbe
Individual Contributor

Email: edward.crabbe@gmail.com

Siva Sivabalan
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: msiva@cisco.com

Hariharan Ananthakrishnan
Netflix

Email: hari@netflix.com

Dhruv Dhody
Huawei
Divyashree Techno Park, Whitefield
Bangalore, KA 560066
India

Email: dhruv.ietf@gmail.com

Yosuke Tanaka
NTT Communications Corporation
Granpark Tower 3-4-1 Shibaura, Minato-ku
Tokyo 108-8118
Japan

Email: yosuke.tanaka@ntt.com

PCE Working Group
Internet Draft
Intended status: Standard Track
Expires: August 2018

Francesco Lazzeri
Daniele Ceccarelli
Ericsson
Young Lee
Dhruv Dhody
Huawei
February 26, 2018

Extensions to the Path Computation Element Protocol (PCEP) for residual
path bandwidth support

draft-lazzeri-pce-residual-bw-01

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 August 26, 2018.

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 (<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.

Abstract

The PCEP protocol has objective functions to optimize path attributes like the residual bandwidth. While this is enough for some applications, it's not possible to return the computed values of such attributes to the PCC, or put bounds on them.

This document describes extensions to the PCE Communication Protocol (PCEP) providing new path-related bandwidth metrics allowing a PCE to compute paths taking into account and returning to the PCC information about the remaining bandwidth along the computed paths.

Table of Contents

1. Requirements for managing the residual bandwidth as a metric...	4
2. New metrics definition.....	5
2.1. Link and Path Unreserved bandwidth.....	5
2.2. Link and Path Residual bandwidth.....	5
3. PCEP protocol extensions.....	6
4. Non-Understanding/Non-Support Residual Bandwidth.....	8
4.1. Mode of Operation.....	9
5. Procedures.....	10
5.1. Use cases.....	10
6. IANA considerations.....	11
6.1. METRIC types.....	11
6.2. New Error-Values.....	12
7. Security Considerations.....	12
8. References.....	13
8.1. Normative References.....	13
8.2. Informative References.....	14
9. Contributors.....	15
Authors' Addresses.....	15
Intellectual Property Statement.....	15
Disclaimer of Validity.....	16

Introduction

The objective of this document is to define an extension to the PCEP [RFC5440] providing information about the bandwidth still available for future reservations on a given path, that is the minimum unreserved bandwidth and the minimum residual bandwidth among all the links of that path.

This is not a new concept to PCEP. In [RFC5541] two objective functions are defined, called minimum load path (MLP) and maximum residual bandwidth path (MBP). Both of them allow to find paths with optimal value of bandwidth-related metrics, defined on a per-link basis, considering the links traversed by that path.

For example, the residual bandwidth of a path is defined as the minimum value of the residual bandwidth on each link in the path. Specifying that OF inside the SVEC object of a PCReq message, the PCE tries and finds the path with the maximum value of the path residual bandwidth.

Unfortunately, being an objective function, MBP can only be used to find a path that optimizes the residual bandwidth, but its value cannot be returned for a path computed with some other objectives (and also when MBP itself is used), or used as a bound.

The same applies to the unreserved bandwidth. The difference between residual and unreserved bandwidth is well described in [RFC7471]:

"The calculation of Residual Bandwidth is different than that of Unreserved Bandwidth [RFC3630]. Residual Bandwidth subtracts tunnel reservations from Maximum Bandwidth (i.e., the link capacity) [RFC3630] and provides an aggregated remainder across priorities. Unreserved Bandwidth, on the other hand, is subtracted from the Maximum Reservable Bandwidth (the bandwidth that can theoretically be reserved) and provides per priority remainders. Residual Bandwidth and Unreserved Bandwidth [RFC3630] can be used concurrently, and each has a separate use case (e.g., the former can be used for applications like Weighted ECMP while the latter can be used for call admission control)".

Having this information would allow a PCC to reuse a path resulting from a path computation to route additional LSPs without requesting new path computations (with the same end-points and constraints), until the maximum path unreserved bandwidth is taken (or a path deployment fails).

1. Requirements for managing the residual bandwidth as a metric

Path computation with optimization of the load or of the residual bandwidth has been defined as important objective functions in [RFC5541].

Managing the unreserved bandwidth (related to the load) and the residual bandwidth of a path as additional metrics, adds the capability to return their value, or putting a bound on their value. This is an added value in distributed PCE applications, like e.g. in ACTN architecture [ACTN-FW] and [PCE-APP]. The following associated key requirements are identified for PCEP:

1. A PCE supporting this draft MUST have the capability to compute end-to-end (E2E) paths with either unreserved bandwidth or with residual bandwidth constraints. It MUST also support the combination of these new constraints with existing constraints, like IGP metric, TE metric, hop limit, and network performance constraints as defined in [RFC5440] and [PCEP-SERV-AWARE].

2. A PCC MUST be able to specify either unreserved bandwidth or residual bandwidth constraints in a Path Computation Request (PCReq) message to be applied during the path computation.

3. A PCC MUST be able to request that a PCE optimizes a path using either unreserved bandwidth or residual bandwidth as objective metric.

4. A PCE that supports this specification is not required to provide unreserved bandwidth or residual bandwidth path computation to any PCC at any time.

Therefore, it MUST be possible for a PCE to reject a PCReq message with reason codes that indicate unreserved bandwidth or residual bandwidth is not supported. Furthermore, a PCE that does not support this specification will either ignore or reject such requests using pre-existing mechanisms, therefore the requests MUST be identifiable to legacy PCEs and rejections by legacy PCEs MUST be acceptable within this specification.

5. A PCE that supports this specification MUST be able to return unreserved or residual bandwidth information of the computed path in a Path Computation Reply (PCRep) message.

2. New metrics definition

2.1. Link and Path Unreserved bandwidth

The unreserved bandwidth of a link is the bandwidth available for future allocation on the link at a given priority, that is the difference between the Maximum Reservable Bandwidth of the link and total bandwidth used on that link by LSPs with priority equal or lower (higher value) than the specified priority. In order to define the path unreserved bandwidth, the following concepts and notation need to be introduced:

- o A network comprises of a set of N links $\{L_i, (i=1\dots N)\}$.
- o A path of a point to point (P2P) LSP is a list of K links $\{L_{pi}, (i=1\dots K)\}$.
- o The maximum reservable bandwidth of the link L_i , named R_i .
- o The bandwidth allocated to LSPs at priority p on the link L_i is the sum of the bandwidth of all the LSPs passing through the link L_i with priority $\geq p$, named $B_i(p)$.
- o The unreserved bandwidth at priority p of the link L_i is $U_i(p) = R_i - B_i(p)$

The path unreserved bandwidth at a given priority k is defined as the minimum value of the unreserved bandwidth at priority k among all the links along the P2P path. Specifically, extending on the above mentioned terminology:

- o Path unreserved bandwidth metric at priority is defined as:
$$PU(p) = \min \{U_i(p), (i=1\dots K)\}$$

2.2. Link and Path Residual bandwidth

The residual bandwidth of a link is the bandwidth physically left free for future allocation on the link. In order to define the path residual bandwidth, the following concepts and notation need to be introduced:

- o A network comprises of a set of N links $\{L_i, (i=1\dots N)\}$.

- o A path of a point to point (P2P) LSP is a list of K links $\{L_{pi}, (i=1...K)\}$
- o The maximum bandwidth of the link L_i , named B_i .
- o The sum of the bandwidth of all the LSPs passing through the link L_i , that is the bandwidth allocated on the link, named A_i .
- o The residual bandwidth of the link L_i is $r(i) = B_i - A_i$.

The path residual bandwidth is defined as the minimum value of the residual bandwidth among all the links along the P2P path. Specifically, extending on the above mentioned terminology:

- o Path residual bandwidth metric for the P2P path is defined as:
 $PB = \min \{r(L_{pi}), (i=1...K)\}$

3. PCEP protocol extensions

This section defines PCEP extensions to fulfill the requirements outlined in Section 2. The proposed solution is used to support path unreserved bandwidth and path residual bandwidth as additional metrics of the PCEP protocol. The METRIC object is defined in section 7.8 of [RFC5440], comprising metric-value, metric-type (T field) and a flags field comprising a number of bit-flags.

This document defines two new types for the METRIC object:

T = TBD1: Path Unreserved Bandwidth

When the T field is set to TBD1, the value of the metric-value field is set to the Path Unreserved Bandwidth for the traffic type and priority requested in the PCReq message.

The same format used by [RFC5440] for the BANDWIDTH object body is used here to represent the value of a path unreserved bandwidth bound or returned value, as shown in the following:

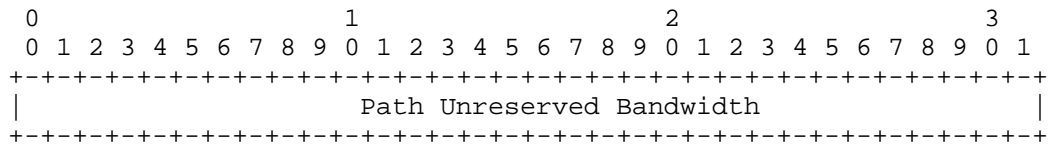


Figure 1: PATH UNRESERVED BANDWIDTH value format

Path Unreserved Bandwidth (32 bits): The path unreserved bandwidth is encoded in 32 bits in IEEE floating point format (see [IEEE.754.1985]), expressed in bytes per second.

The PATH UNRESERVED BANDWIDTH value has a fixed length of 4 bytes.

T = TBD2: Path Residual Bandwidth

When the T field is set to TBD2, the value of the metric-value field is set to the Path Residual Bandwidth for the traffic type requested in the PCReq message.

When the T field is set to TBD2, the value of the metric-value field is set to the Path Residual Bandwidth for the traffic type requested in the PCReq message.

The same format used by [RFC5440] for the BANDWIDTH object body is used here to represent the value of a path residual bandwidth bound or returned value, as shown in the following:

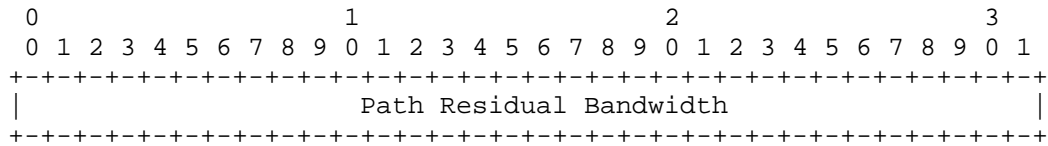


Figure 1: PATH RESIDUAL BANDWIDTH value format

Path Residual Bandwidth (32 bits): The path residual bandwidth is encoded in 32 bits in IEEE floating point format (see [IEEE.754.1985]), expressed in bytes per second.

The PATH RESIDUAL BANDWIDTH value has a fixed length of 4 bytes.

Editor NOTE: these definitions provide support only of PSC signal type. For other signal types (e.g. ODU, WDM) these fields can be filled with the number of unreserved or residual fixed containers (e.g. 3 ODU0) related to the type of traffic specified in the PCReq. This has to be discussed.

A PCC MAY use the path unreserved or residual bandwidth in a PCReq message to request a path meeting the end to end unreserved or residual bandwidth requirement. In this case, the B bit MUST be set to suggest a bound (a minimum) for the path residual bandwidth metric that must be guaranteed for the PCC to consider the computed path as acceptable. The path unreserved or residual bandwidth metrics must be greater than or equal to the value specified in the metric-value field.

The P bit MAY be set to specify the constraint as mandatory, or MAY be left cleared to specify the bound as optional.

A PCC can also use this metric to ask PCE to optimize (that is maximize) the path residual bandwidth during path computation. In this case, the B bit MUST be cleared.

A PCE MAY use the path residual bandwidth metric in a PCRep message along with a NO-PATH object in the case where the PCE cannot compute a path meeting this constraint.

A PCE can also use this metric to send the computed path residual bandwidth metric to the PCC.

4. Non-Understanding/Non-Support Residual Bandwidth

If a PCE receives a PCReq message containing a METRIC object with type PATH UNRESERVED BANDWIDTH or PATH RESIDUAL BANDWIDTH and the PCE does not understand or support those metric types, and the P bit is clear in the METRIC object header then the PCE SHOULD simply ignore the METRIC object as per the processing specified in [RFC5440].

If the PCE does not understand the new METRIC types, and the P bit is set in the METRIC object header, then the PCE MUST send a PCErr message containing a PCEP-ERROR Object with Error-Type = 4

(Not supported object) and Error-value = 4 (Unsupported parameter) [RFC5440][RFC5541].

If the PCE understands but does not support the new METRIC type, and the P bit is set in the METRIC object header, then the PCE MUST send a PCErr message containing a PCEP-ERROR Object with Error-Type = 4 (Not supported object) with Error-value = TBD3 (Unsupported path unreserved bandwidth constraint) or TBD4 (Unsupported path residual bandwidth constraint).

The path computation request MUST then be cancelled.

If the PCE understands the new METRIC type, but the local policy has been configured on the PCE to not allow network performance constraint, and the P bit is set in the METRIC object header, then the PCE MUST send a PCErr message containing a PCEP-ERROR Object with Error-Type = 5 (Policy violation) with Error-value = TBD5 (Not Allowed path unreserved bandwidth constraint) or TBD6 (Not Allowed path residual bandwidth constraint). The path computation request MUST then be cancelled.

4.1. Mode of Operation

As explained in [RFC5440], the METRIC object is optional and can be used for several purposes. In a PCReq message, a PCC MAY insert one or more METRIC objects:

- o To indicate the metric (path unreserved or path residual bandwidth) that MUST be optimized by the path computation algorithm.
- o To indicate a bound on the METRIC (path unreserved or path residual bandwidth) that MUST NOT be exceeded for the path to be considered as acceptable by the PCC.

In a PCRep message, the PCE MAY insert the METRIC object with an Explicit Route Object (ERO) so as to provide the METRIC (residual bandwidth) for the computed path.

The PCE MAY also insert the METRIC object with a NO-PATH object to indicate that the metric constraint could not be satisfied.

The path computation algorithmic aspects used by the PCE to optimize a path with respect to a specific metric are outside the scope of this document.

All the rules of processing the METRIC object as explained in [RFC5440] are applicable to the new metric types as well.

5. Procedures

The new metrics defined in this document don't add or change the procedures already defined for PCEP protocol in [RFC5440] and [RFC5541].

In particular, the existing objective function MBP is still usable as appropriate, being equivalent to the usage of the Path Residual Bandwidth metric with the B bit cleared.

The new metric can be used to define new procedures especially in the scope of SDN and ACTN, which are out of the scope of this document.

5.1. Use cases

The first use case is the application of the residual bandwidth to simplify the computation of an end-to-end path across a multi-domain network.

The ability of a hierarchy of PCEs to compute accurate end-to-end paths across multiple domains is recognized as an important requirement in many applications.

In particular, this is a key requirement for networks with a centralized path computation function (e.g. hierarchical PCE or SDN). In such scenarios, a hierarchy of PCEs is often implemented, where, as illustrated in [RFC6805], a parent H-PCE coordinates the operations of a set of child (domain) PCEs in order to compute end-to-end paths across the network.

An H-PCE (either stateful or stateless) can make the best of residual bandwidth metrics, using paths from erstwhile path computations to deploy multiple LSPs (having the same end-points and constraints) without additional requests, until either the remaining In a hierarchical architecture of PCEs, domain PCEs just know the topology of their domains, while the parent PCE has in general detailed information about the managed domains and the relevant inter-domain links, but not necessarily enough information about the internals of each domain, so that it's capable to compute accurately an end-to-end path.

The residual bandwidth information would also be beneficial for implementing abstractions of the domain topology, building the

abstract connectivity incrementally, based only on really used constraints, as soon as path computation results are returned. One of the key features of SDN is the support of network abstraction, that is, as described in [RFC7926], the capability of applying policy to a set of information about a network, in order to produce selective information that represents the potential ability to connect across the domain.

The process of abstraction produces a connectivity graph, which can be used by the parent PCE to compute an accurate path based on the abstracted topology. The main issue is that the connectivity graph can be huge, depending on the size of the domain topology and the number of end-points defined on the edge of the domain.

One way to provide similar information is to store the result of path computations requested to the child PCEs (performed by e.g. TE-tunnels "compute only") and try reusing them if possible to save further path computation iterations between parent and child PCEs. In any case a selection of path computation constraints has to be defined against the abstract topology in order to reduce the number of the abstract links or TE-tunnels exported by the connectivity graph, as it's impractical to compute or pre-compute all the constraints combinations. It's also very important to reduce the number of updates of such connectivity information to the parent PCE in order not to flood it with a continuous stream of updates.

6. IANA considerations

6.1. METRIC types

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" at <http://www.iana.org/assignments/pcep>. Within this registry IANA maintains one sub-registry for "METRIC object T field".

Two new metric types are defined in this document for the METRIC object (specified in [RFC5440]).

IANA is requested to make the following allocations:

Value	Description	Reference
TBD1	Path unreserved bandwidth metric	[This I.D.]

TBD2 Path residual bandwidth metric [This I.D.]

6.2. New Error-Values

IANA maintains a registry of Error-Types and Error-values for use in PCEP messages. This is maintained as the "PCEP-ERROR Object Error Types and Values" sub-registry of the "Path Computation Element Protocol (PCEP) Numbers" registry.

IANA is requested to make the following allocations:

Four new Error-values are defined for the Error-Type "Not supported object" (type 4) and "Policy violation" (type 5).

Error-Type	Meaning and error values	Reference
4	Not supported object Error-value=TBD3 Unsupported Path unreserved bandwidth constraint Error-value=TBD4 Unsupported Path residual bandwidth constraint	[This I.D.]
5	Policy violation Error-value=TBD5 Not allowed Path unreserved bandwidth constraint Error-value=TBD6 Not allowed Path residual bandwidth constraint	[This I.D.]

7. Security Considerations

This document defines new METRIC types, which do not add any new security concerns beyond those discussed in [RFC5440] and [RFC5541] in itself.

In some scenarios, path unreserved bandwidth and path residual bandwidth information could be considered sensitive and could be used to influence path computation and setup with adverse effect.

Snooping of PCEP messages with such data, or using PCEP messages for network reconnaissance, may give an attacker sensitive information about the capabilities of the network. Thus, such deployment should employ suitable PCEP security mechanisms like TCP Authentication Option (TCP-AO) [RFC5925] or [PCEPS].

The Transport Layer Security (TLS) based procedure in [PCEPS] is considered as a security enhancement and thus much better suited for the sensitive residual bandwidth information.

8. References

8.1. Normative References

- [RFC5440] Vasseur JP., Ed. and JL. Le Roux, Ed.,
"Path Computation Element (PCE) Communication
Protocol(PCEP)", RFC 5440,
DOI 10.17487/RFC5440, March 2009,
<<http://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,
<<http://www.rfc-editor.org/info/rfc5541>>.
- [RFC5925] Touch, J., Mankin, A., and R. Bonica,
"The TCP Authentication Option", RFC 5925,
DOI 10.17487/RFC5925, June 2010,
<<http://www.rfc-editor.org/info/rfc5925>>.
- [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,
<<http://www.rfc-editor.org/info/rfc7420>>.
- [PCEPS] Lopez, D.Lopez, D., Dios, O., Wu, W., and D. Dhody,
"Secure Transport for PCEP", draft-ietf-pce-pceps-11
(work in progress), January 2017.
- [IEEE.754.1985] IEEE, "Standard for Binary Floating-Point Arithmetic",
IEEE 754, August 1985

8.2. Informative References

- [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, November 2012,
<<http://www.rfc-editor.org/info/rfc6805>>.
- [RFC7471] Giacalone S., Ward D., Drake J., Atlas A. and S.
Previdi, "OSPF Traffic Engineering (TE) Metric
Extensions", RFC 7471,
DOI 10.17487/RFC7471, March 2015,
<<http://www.rfc-editor.org/info/rfc7471>>
- [RFC7926] Farrel, A. et al., "Problem Statement and Architecture
for Information Exchange Between Interconnected
Traffic Engineered Networks", RFC 7926, July 2016.
- [ACTN-FW] Ceccarelli, D. and Y. Lee, "Framework for Abstraction
and Control of Traffic Engineered Networks", draft-
ietf-teas-actn-framework-03 (work in progress),
February 2017.
- [PCE-APP] Dhody, D. Lee, Y. Ceccarelli, D. "Applicability of
Path Computation Element (PCE) for Abstraction and
Control of TE Networks (ACTN)" draft-dhody-pce-
applicability-actn-02

9. Contributors

Authors' Addresses

Francesco Lazzeri
Ericsson
Via Melen 77
Genova - Italy
Email: francesco.lazzeri@ericsson.com

Daniele Ceccarelli
Ericsson AB
Gronlandsgatan 21
Kista - Sweden
Email: daniele.ceccarelli@ericsson.com

Young Lee
Huawei Technologies
5340 Legacy Drive
Plano, TX 75023, USA
Phone: (469)277-5838
Email: leeyoung@huawei.com

Dhruv Dhody
Huawei Technologies,
Divyashree Technopark, Whitefield
Bangalore, India
Email: dhruv.ietf@gmail.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or

users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

PCE Working Group
Internet Draft
Intended status: Standard Track
Expires: April 23, 2019

Y. Lee (Editor)
H. Zheng
Huawei

R. Casellas
R. Vilalta
CTTC

D. Ceccarelli
F. Lazzeri
Ericsson

October 22, 2018

PCEP Extension for Flexible Grid Networks

draft-lee-pce-flexible-grid-03

Abstract

This document provides the Path Computation Element Communication Protocol (PCEP) extensions for the support of Routing and Spectrum Assignment (RSA) in Flexible Grid 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 December 22, 2018.

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 (<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. Terminology.....	3
2. Requirements Language.....	3
3. Introduction.....	3
4. Spectrum Assignment (SA) Object.....	4
4.1. Frequency-Slot Selection TLV.....	6
4.2. Frequency-slot Restriction Constraint TLV.....	8
4.2.1. Frequency-Slot Restriction Field.....	10
5. Encoding of a RSA Path Reply.....	10
5.1. Error Indicator.....	11
5.2. NO-PATH Indicator.....	12
6. Manageability Considerations.....	12
6.1. Control of Function and Policy.....	13
6.2. Information and Data Models.....	13
6.3. Verifying Correct Operation.....	13
6.4. Requirements on Other Protocols and Functional Components	13
6.5. Impact on Network Operation.....	14
7. Security Considerations.....	14
8. IANA Considerations.....	14
8.1. New PCEP Object.....	14
8.2. New PCEP TLV: Frequency Slot Selection TLV.....	15
8.3. New PCEP TLV: Frequency Slot Restriction Constraint TLV..	15
8.4. New PCEP TLV: Spectrum Allocation TLV.....	15
8.5. New No-Path Reasons.....	16
8.6. New Error-Types and Error-Values.....	16
9. References.....	17

9.1. Informative References.....	17
9.2. Normative References.....	18
10. Contributors.....	19
Authors' Addresses.....	20

1. Terminology

This document uses the terminology defined in [RFC4655], [RFC5440] and [RFC7698].

2. 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 [RFC2119].

3. Introduction

[RFC4655] defines a PCE based path computation architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is said to be any network component that makes such a request and may be, for instance, an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communications Protocol (PCEP) is the communication protocol used between a PCC and a PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

[PCEP-WSO] provides the PCEP extensions for the support of Routing and Wavelength Assignment (RWA) in Wavelength Switched Optical Networks (WSO) based on the requirements specified in [RFC6163] and [RFC7449].

[RFC7698] provides Framework and Requirements for GMPLS-Based Control of Flexi-Grid Dense Wavelength Division Multiplexing (DWDM)

Networks. To allow efficient allocation of optical spectral bandwidth for systems that have high bit-rates, the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) has extended its Recommendations G.694.1 and G.872 to include a new Dense Wavelength Division Multiplexing (DWDM) grid by defining a set of nominal central frequencies, channel spacings, and the concept of the "frequency slot". In such an environment, a data-plane connection is switched based on allocated, variable-sized frequency ranges within the optical spectrum, creating what is known as a flexible grid (flexi-grid).

This document provides PCEP extensions to support Routing and Spectrum Assignment (RSA) in in Spectrum Switched Optical Networks (SSON) [RFC7698].

Figure 2 shows one typical PCE based implementation, which is referred to as the Combined Routing and Spectrum Assignment (R&SA) [RFC7698]. With this architecture, the two processes of routing and spectrum assignment are accessed via a single PCE. This architecture is the base architecture from which the PCEP extensions are going to be specified in this document.

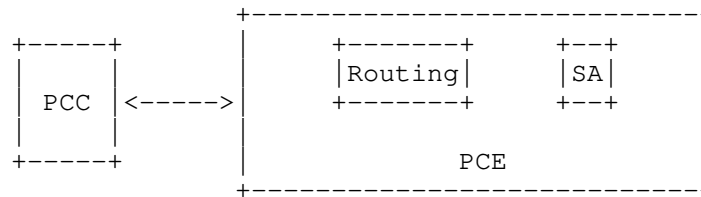


Figure 1 Combined Process (R&SA) architecture

4. Spectrum Assignment (SA) Object

Spectrum allocation can be performed by the PCE by different means:

- (a) By means of Explicit Label Control (ELC) where the PCE allocates which label to use for each interface/node along the path.
- (b) By means of a Label Set where the PCE provides a range of potential frequency slots to allocate by each node along the path. This document aligns with GMPLS extensions for PCEP [PCEP-GMPLS] for generic property such as label, label-set and label assignment

noting that frequency is a type of label. Frequency restrictions and constraints are also formulated in terms of labels per [RFC7579].

Option (b) allows distributed spectrum allocation (performed during signaling) to complete spectrum assignment.

Additionally, given a range of potential spectrums to allocate, the request SHOULD convey the heuristic / mechanism to the allocation.

The format of a PCReq message after incorporating the Spectrum Assignment (SA) object is as follows:

```
<PCReq Message> ::= <Common Header>
                        [<svec-list>]
                        <request-list>
```

Where:

```
<request-list> ::= <request> [<request-list>]
<request> ::= <RP>
                <GENERALIZED ENDPOINTS>
                [ <SA> ]
                [other optional objects...]
```

If the SA object is present in the request, it MUST be encoded after the ENDPOINTS object.

The format of the Spectrum Assignment (SA) object body 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								
Reserved										Flags										M																			
Frequency-Slot Selection TLV																																							
Frequency-Slot Restriction Constraint TLV																																							

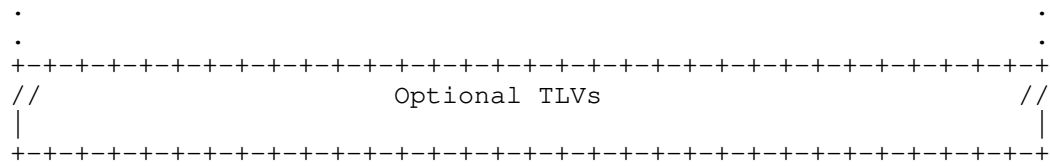


Figure 2 SA Object

- o Reserved (16 bits)
- o Flags (16 bits)

The following new flags SHOULD be set

- . M (Mode - 1 bit): M bit is used to indicate the mode of spectrum assignment. When M bit is set to 1, this indicates that the spectrum assigned by the PCE must be explicit. That is, the selected way to convey the allocated spectrum is by means of Explicit Label Control (ELC) [RFC4003] for each hop of a computed LSP. Otherwise, the spectrum assigned by the PCE needs not be explicit (i.e., it can be suggested in the form of label set objects in the corresponding response, to allow distributed SA. In such case, the PCE MUST return a Label Set Field as described in Section 2.6 of [RFC7579] in the response. See Section 5 of this document for the encoding discussion of a Label Set Field in a PCRep message.

4.1. Frequency-Slot Selection TLV

The Frequency-Slot Selection TLV is used to indicate the frequency-slot selection constraint in regard to the order of frequency-slot assignment to be returned by the PCE. This TLV is only applied when M bit is set in the SA Object specified in Section 3.1. This TLV MUST NOT be used when the M bit is cleared.

The Frequency-Slot Selection sub-TLV value field is defined as:

```

      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
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|S|  FSA Method  |                               Reserved          |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+

```

Where:

S (Symmetry, 1 bit): This flag is only meaningful when the request is for a bidirectional LSP (see [RFC5440]).

0 denotes requiring the same frequency-slot in both directions;
 1 denotes that different spectrums on both directions are allowed.

Frequency-Slot Assignment (FSA) Method (7 bits):

- 0: unspecified (any); This does not constrain the SA method used by a PCC. This value is implied when the Frequency-Slot Selection sub-TLV is absent.
- 1: First-Fit. All the feasible frequency slots are numbered (based on "n" parameter), and this SA method chooses the available frequency-slot with the lowest index (of "n" parameter).
- 2: Random. This SA method chooses an feasible frequency-slot ("n" parameter) randomly.
- 3-127: Unassigned.

The processing rules for this TLV are as follows:

If a PCE does not support the attribute(s), its behavior is specified below:

- S bit not supported: a PathErr MUST be generated with the Error Code "Routing Problem" (24) with error sub-code "Unsupported Frequency slot Selection Symmetry value" (TDB).
- FSA method not supported: a PathErr MUST be generated with the Error Code "Routing Problem" (24) with error sub-code

"Unsupported Frequency Slot Assignment value" (TDB).

A Frequency Slot Selection TLV can be constructed by a node and added to an ERO Hop Attributes subobject in order to be processed by downstream nodes (transit and egress). As defined in [RFC7570], the R bit reflects the LSP_REQUIRED_ATTRIBUTE and LSP_ATTRIBUTE semantic defined in [RFC5420], and it SHOULD be set accordingly.

Once a node properly parses the Spectrum Selection sub-TLV received in an ERO Hop Attributes subobject, the node use the indicated spectrum assignment method (at that hop) for the LSP. In addition, the node SHOULD report compliance by adding an RRO Hop Attributes subobject with the WSON Processing Hop Attribute TLV (and its sub-TLVs) that indicate the utilized method. Frequency-Slot Selection TLVs carried in an RRO Hop Attributes subobject are subject to [RFC7570] and standard RRO processing; see [RFC3209].

4.2. Frequency-slot Restriction Constraint TLV

For any request that contains a Frequency-slot assignment, the requester (PCC) MUST be able to specify a restriction on the frequency-slots to be used. This restriction is to be interpreted by the PCE as a constraint on the tuning ability of the origination laser transmitter or on any other maintenance related constraints.

The format of the Frequency-Slot Restriction Constraint TLV is as follows:

```
<Frequency-slot Restriction Constraint> ::=
    <Action> <Count> <Reserved>
    (<Link Identifiers> <Freq-slot Restriction>)...
```

Where

```
<Link Identifiers> ::= <Link Identifier> [<Link Identifiers>]
```

See Section 4.3.1 in [PCEP-WSON] for the encoding of the Link Identifiers Field.

The Frequency slot Restriction Constraint TLV type is TBD. This TLV MAY appear more than once to be able to specify multiple restrictions.

The TLV data is defined as follows:

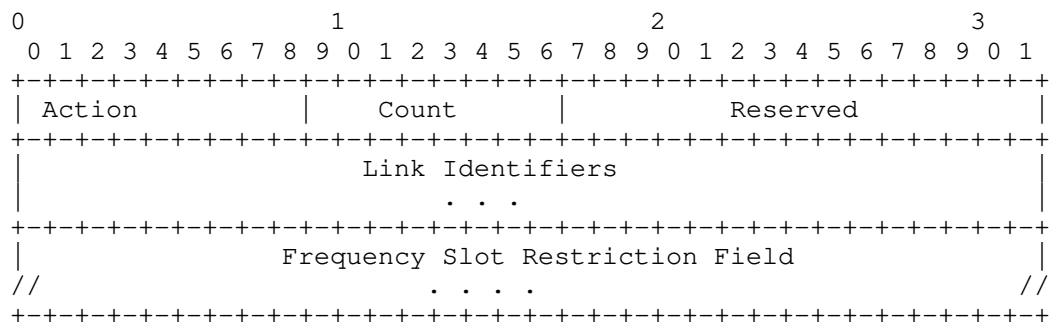


Figure 3 spectrum Restriction Constraint TLV Encoding

- o Action: 8 bits
 - . 0 - Inclusive List indicates that one or more link identifiers are included in the Link Set. Each identifies a separate link that is part of the set.
 - . 1 - Inclusive Range indicates that the Link Set defines a range of links. It contains two link identifiers. The first identifier indicates the start of the range (inclusive). The second identifier indicates the end of the range (inclusive). All links with numeric values between the bounds are considered to be part of the set. A value of zero in either position indicates that there is no bound on the corresponding portion of the range. Note that the Action field can be set to 0 when unnumbered link identifier is used.

- o Count: The number of the link identifiers (8 bits)

Note that a PCC MAY add a spectrum restriction that applies to all links by setting the Count field to zero and specifying just a set of spectrums.

Note that all link identifiers in the same list must be of the same type.

- o Reserved: Reserved for future use (16 bits)

- o Link Identifiers: Identifies each link ID for which restriction is applied. The length is dependent on the link format and the Count field. See Section 4.3.1 in [PCEP-WSN] for Link Identifier encoding and Section 3.3.1 for the Spectrum Restriction Field encoding, respectively.

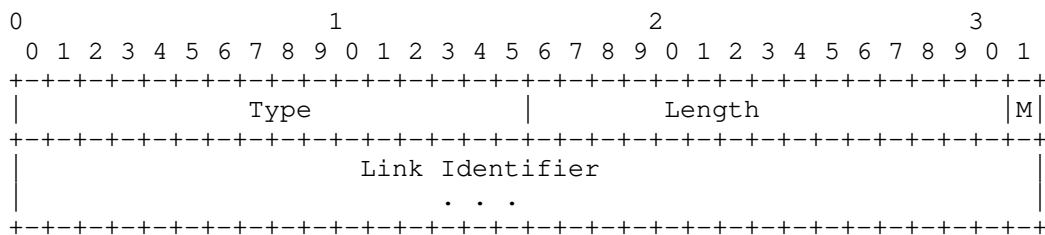
4.2.1. Frequency-Slot Restriction Field

The Frequency-Slot Restriction Field of the Frequency slot restriction TLV is encoded as defined in <https://tools.ietf.org/html/draft-ietf-ccamp-flexible-grid-ospf-ext-09#section-4.1.1>.

5. Encoding of a RSA Path Reply

This section provides the encoding of a RSA Path Reply for frequency slot allocation as discussed in Section 4. Spectrum Allocation TLV

The Spectrum Allocation TLV type is TBD, recommended value is TBD. The TLV data is defined as follows:



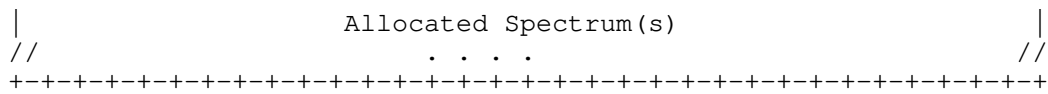


Figure 4 Spectrum Allocation TLV Encoding

- o Type (16 bits): The type of the TLV.
- o Length (15 bits): The length of the TLV including the Type and Length fields.
- o M (Mode): 1 bit
 - 0 indicates the allocation is under Explicit Label Control.
 - 1 indicates the allocation is expressed in Label Sets.

Note that all link identifiers in the same list must be of the same type.

- o **Link Identifier (variable):** Identifies the interface to which assignment spectrum(s) is applied. See Section 3.3 for Link Identifier encoding.
- o **Allocated Spectrum(s) (variable):** Indicates the allocated spectrum(s) to the link identifier. See Section 3.3.1 for encoding details.

This TLV is encoded as an attributes TLV, per [RFC5420], which is carried in the ERO LSP Attribute Subobjects per [RFC7570]. The type value of the Spectrum Restriction Constraint TLV is TBD by IANA.

5.1. Error Indicator

To indicate errors associated with the RSA request, a new Error Type (TDB) and subsequent error-values are defined as follows for inclusion in the PCEP-ERROR Object:

A new Error-Type (TDB) and subsequent error-values are defined as follows:

- . Error-Type=TBD; Error-value=1: if a PCE receives a RSA request and the PCE is not capable of processing the request due to insufficient memory, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=TDB) and an Error-value(Error-value=1). The PCE stops processing the request. The corresponding RSA request MUST be cancelled at the PCC.
- . Error-Type=TBD; Error-value=2: if a PCE receives a RSA request and the PCE is not capable of RSA computation, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=TDB) and an Error-value (Error-value=2). The PCE stops processing the request. The corresponding RSA computation MUST be cancelled at the PCC.

5.2. NO-PATH Indicator

To communicate the reason(s) for not being able to find RSA for the path request, the NO-PATH object can be used in the corresponding response. The format of the NO-PATH object body is defined in [RFC5440]. The object may contain a NO-PATH-VECTOR TLV to provide additional information about why a path computation has failed.

One new bit flag is defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object.

- . Bit TDB: When set, the PCE indicates no feasible route was found that meets all the constraints (e.g., spectrum restriction, etc.) associated with RSA.

6. Manageability Considerations

Manageability of SSON Routing and Spectrum Assignment (RSA) with PCE must address the following considerations:

6.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- . The ability to send a Flexi-Grid RSA request.

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- . The support for Flexi-Grid RSA .
- . A set of Flexi-Grid RSA specific policies (authorized sender, request rate limiter, etc).

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

6.2. Information and Data Models

Extensions to the PCEP YANG module may include to cover the Flexi-Grid RSA information introduced in this document. 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 section 8.3 of [RFC5440].

6.3. Verifying Correct Operation

Mechanisms defined in this document do not imply any new verification requirements in addition to those already listed in section 8.4 of [RFC5440]

6.4. Requirements on Other Protocols and Functional Components

The PCE Discovery mechanisms ([RFC5089] and [RFC5088]) may be used to advertise Flexi-Grid RSA path computation capabilities to PCCs.

This draft has requirements on other protocols (ERO objects, etc. which are under TEAS or CCAMP.)

6.5. Impact on Network Operation

Mechanisms defined in this document do not imply any new network operation requirements in addition to those already listed in section 8.6 of [RFC5440].

7. Security Considerations

This document has no requirement for a change to the security models within PCEP. However, the additional information distributed in order to address the RSA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

8. IANA Considerations

IANA maintains a registry of PCEP parameters. IANA has made allocations from the sub-registries as described in the following sections.

8.1. New PCEP Object

As described in Section 4.1, a new PCEP Object is defined to carry frequency-slot assignment related constraints. IANA is to allocate the following from "PCEP Objects" sub-registry (<http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-objects>):

Object Class Value	Name	Object Type	Reference

TDB	SA	1: Spectrum Assignment	[This.I-D]

8.2. New PCEP TLV: Frequency Slot Selection TLV

As described in Sections 4.2, a new PCEP TLV is defined to indicate spectrum selection constraints. IANA is to allocate this new TLV from the "PCEP TLV Type Indicators" subregistry (<http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-tlv-type-indicators>).

Value	Description	Reference
TBD	Spectrum Selection	[This.I-D]

8.3. New PCEP TLV: Frequency Slot Restriction Constraint TLV

As described in Section 4.3, a new PCEP TLV is defined to indicate wavelength restriction constraints. IANA is to allocate this new TLV from the "PCEP TLV Type Indicators" subregistry (<http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-tlv-type-indicators>).

Value	Description	Reference
TBD	Frequency Slot Restriction Constraint	[This.I-D]

8.4. New PCEP TLV: Spectrum Allocation TLV

As described in Section 5, a new PCEP TLV is defined to indicate the allocation of freq-slots(s) by the PCE in response to a request by the PCC. IANA is to allocate this new TLV from the "PCEP TLV Type Indicators" subregistry (<http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-tlv-type-indicators>).

Value	Description	Reference
TBD	Spectrum Allocation	[This.I-D]

8.5. New No-Path Reasons

As described in Section 4.3, a new bit flag are defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object. This flag, when set, indicates that no feasible route was found that meets all the RSA constraints (e.g., spectrum restriction, signal compatibility, etc.) associated with a RSA path computation request.

IANA is to allocate this new bit flag from the "PCEP NO-PATH-VECTOR TLV Flag Field" subregistry
(<http://www.iana.org/assignments/pcep/pcep.xhtml#no-path-vector-tlv>).

Bit	Description	Reference
TBD	No RSA constraints met	[This.I-D]

8.6. New Error-Types and Error-Values

As described in Section 5.1, new PCEP error codes are defined for WSON RWA errors. IANA is to allocate from the "PCEP-ERROR Object Error Types and Values" sub-registry
(<http://www.iana.org/assignments/pcep/pcep.xhtml#pcep-error-object>).

Error-Type	Meaning	Error-Value	Reference
TDB	Flexi-Grid RSA Error	1: Insufficient Memory	[This.I-D]
		2: RSA computation Not supported	[This.I-D]

9. References

9.1. Informative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2863] McCloghrie, K. and F. Kastenholtz, "The Interfaces Group MIB", RFC 2863, June 2000.
- [RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", RFC 4003, February 2005.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol", RFC 5440, March 2009.
- [RFC5088] Le Roux, JL, JP. Vasseur, Y. Ikejiri, and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery," RFC 5088, January 2008.
- [RFC5089] Le Roux, JL, JP. Vasseur, Y. Ikejiri, and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery," RFC 5089, January 2008.
- [RFC6163] Lee, Y. and Bernstein, G. (Editors), and W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, March 2011.
- [RFC6566] Y. Lee, G. Bernstein, D. Li, G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSN) with Impairments", RFC 6566, March 2012.
- [RFC7420] Koushik, A., E. Stephan, Q. Zhao, D. King, and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, December 2014.

- [RFC7446] Y. Lee, G. Bernstein. (Editors), "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", RFC 7446, February 2015.
- [RFC7449] Lee, Y., et. al., "PCEP Requirements for WSON Routing and Wavelength Assignment", RFC 7449, February 2015.

9.2. Normative References

- [PCEP-GMPLS] Margaria, et al., "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions, work in progress.
- [RFC5420] Farrel, A. "Encoding of Attributes for MPLS LSP Establishment Using Resource Reservation Protocol Traffic Engineering (RSVP-TE)", RFC5420, February 2009.
- [RFC5521] Oki, E, T. Takeda, and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, April 2009.
- [RFC6205] Tomohiro, O. and D. Li, "Generalized Labels for Lambda-Switching Capable Label Switching Routers", RFC 6205, January, 2011.
- [RFC7570] Margaria, et al., "Label Switched Path (LSP) Attribute in the Explicit Route Object (ERO)", RFC 7570, July 2015.
- [RFC7689] Bernstein et al, "Signaling Extensions for Wavelength Switched Optical Networks", RFC 7689, November 2015.
- [RFC7688] Y. Lee, and G. Bernstein, "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", RFC 7688, November 2015.
- [RFC7698] O. Gonzalez de Dios, R. Casellas, editors, "Framework and Requirements for GMPLS-Based Control of Flexi-Grid Dense Wavelength Division Multiplexing (DWDM) Networks", RFC 7698, November 2015.
- [RFC7581] Bernstein and Lee, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", RFC7581, June 2015.

[RFC7579] Bernstein and Lee, "General Network Element Constraint Encoding for GMPLS Controlled Networks", RFC 7579, June 2015.

[PCEP-WSON] Y. Lee (Ed.), and R. Casellas (Ed.), "PCEP Extension for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-rwa-ext, work in progress.

10. Contributors

Authors' Addresses

Young Lee, Editor
Huawei Technologies

Email: leeyoung@huawei.com

Haomian Zheng
Huawei Technologies

Email: zhenghaomian@huawei.com

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Email: ramon.casellas@cttc.es

Ricard Vilalta
CTTC
Email: ricard.vilalta@cttc.es

Daniele Ceccarelli
Ericsson AB
Gronlandsgatan 21
Kista - Stockholm
Email: daniele.ceccarelli@ericsson.com

Francesco Lazzeri
Ericsson
Via Melen 77
Genova - Italy
Email: francesco.lazzeri@ericsson.com

PCE Working Group
Internet Draft
Intended Status: Standard
Expires: June 27, 2018

Young Lee
Dhruv Dhody
Huawei

Daniele Ceccarelli
Ericsson

December 27, 2017

PCEP Extensions for Stitching LSPs in Hierarchical Stateful PCE
Model

draft-lee-pce-lsp-stitching-hpce-01.txt

Abstract

This document extends the Path Communication Element Communication Protocol (PCEP) to coordinate an end-to-end inter-domain tunnel setup over a multi-domain networks in the context of Hierarchical Stateful PCE environments. This document uses Stitching Label (SL) to stitch per-domain LSPs.

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 June 27, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

1. Introduction.....	2
2. Network Settings and concepts.....	4
2.1. Stateful H-PCE Stitching Procedure.....	6
2.2. Applicability to ACTN.....	10
3. Security Considerations.....	10
4. IANA Considerations.....	10
5. References.....	11
5.1. Normative References.....	11
5.2. Informative References.....	11
Appendix A. Contributor Addresses.....	14
Author's Addresses.....	14

1. Introduction

In Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS), a Traffic Engineering Database (TED) is used in computing paths for connection oriented packet services and for circuits. The TED contains all relevant information that a Path Computation Element (PCE) needs to perform its computations. It is important that the TED should be complete and accurate anytime so that the PCE can perform path computations.

In MPLS and GMPLS networks, Interior Gateway routing Protocols (IGPs) have been used to create and maintain a copy of the TED at

each node. One of the benefits of the PCE architecture [RFC4655] is the use of computationally more sophisticated path computation algorithms and the realization that these may need enhanced processing power not necessarily available at each node participating in an IGP.

[Stateful-PCE] describes a set of extensions to PCEP to provide stateful control. A stateful PCE has access to not only the information carried by the network's Interior Gateway Protocol (IGP), but also the set of active paths and their reserved resources for its computations. PCC can delegate the rights to modify the LSP parameters to an Active Stateful PCE.

[RFC6805] describes a Hierarchical PCE (H-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 the Hierarchical PCE (H-PCE) architecture [RFC6805], 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] 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 of [Stateful H-PCE] describe the per domain stitched LSP mode, where the individual per domain LSP are stitched together.

[PCE-CC] introduces the architecture for PCE as a central controller, and examines the motivations and applicability for PCEP as a southbound interface. Section 2.1.3 describes the approach with hierarchical controllers.

[BRPC-Stitch] describes how inter-domain labels over the inter-domain interfaces are determined in the multi-domain BRPC-based PCE environments. Further, the document introduces the concept of Stitching Label (SL) and Inter-domain Path Setup Type [PST]. This document also uses these concepts in the hierarchical Stateful PCE model.

This document extends the Path Communication Element Communication Protocol (PCEP) to coordinate an end-to-end tunnel for a virtual network over multi-domain networks in the context of Hierarchical Stateful PCE environments.

2. Network Settings and concepts

This section describes network settings for this draft. Figure 1 shows the context of Hierarchical Stateful PCE architecture where multi-domain LSP stitching is required for an end-to-end tunnel associated with a VN member.

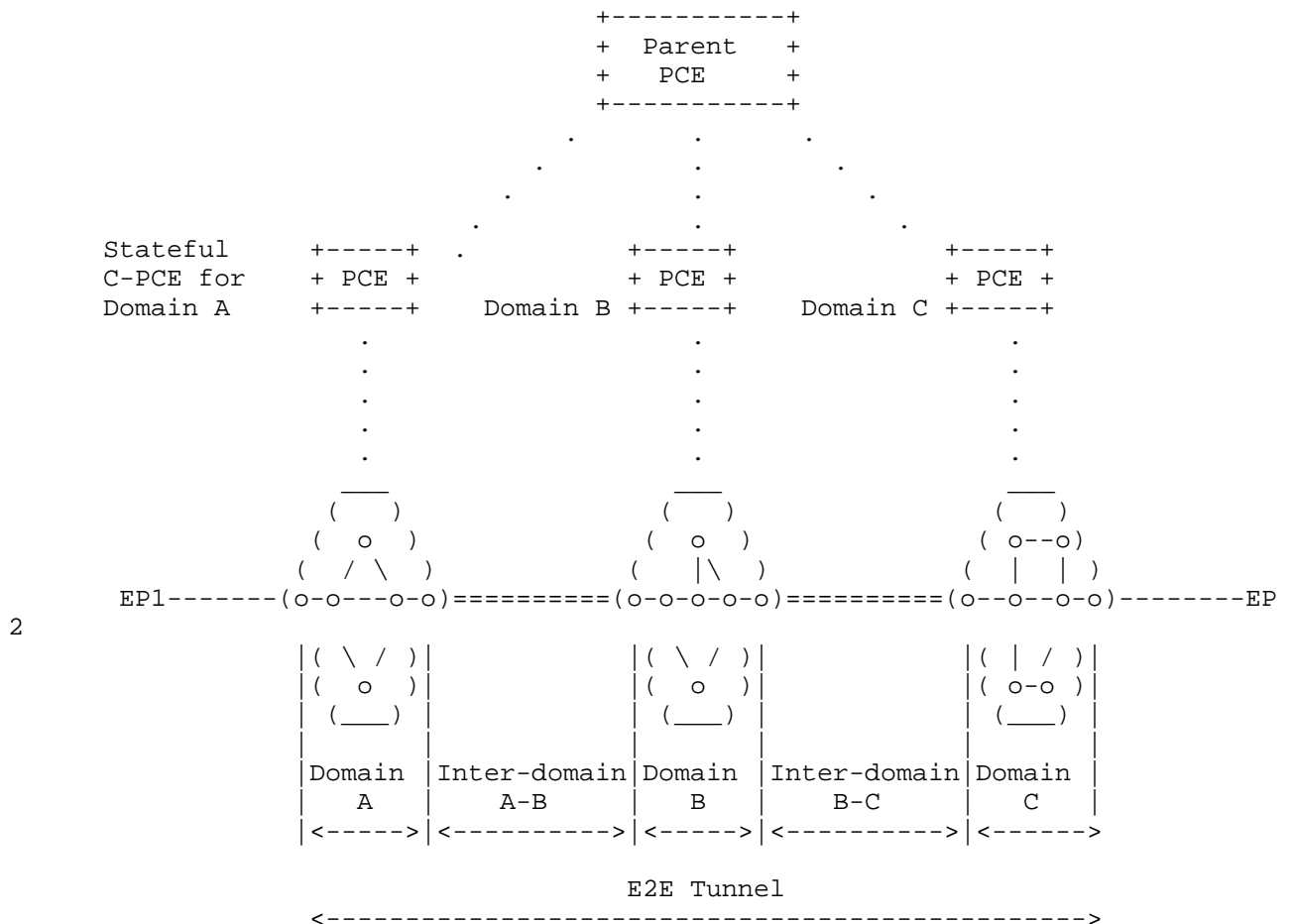


Figure 1: Multi-domain LSP stitching for an end-to-end tunnel

The draft provides PCE mechanisms to identify and isolate an end-to-end tunnel for a virtual network by concatenating a set of LSP/tunnel segments comprising an end-to-end tunnel. From Figure 1, there are a set of segments comprising an end-to-end tunnel: Per

Domain LSP A, Inter-domain Link A-B, Per Domain LSP B, Inter-domain Link B-C, and Per Domain C.

It is important to realize that this end-to-end tunnel for a virtual network should be identifiable from other tunnels in the networks so as to guarantee its performance objective associated with this particular tunnel. See Section 2.2 for ACTN applicability for detailed discussion on this aspect.

As per [BRPC-Stitch], Stitching Label (SL) is defined as a dedicated label that is used to stitch two tunnels (RSVP-TE tunnels or Segment Routing paths). This label is exchanged between exit BN(i) and entry BN(i+1) via PCEP. In case of H-PCE, the SL is conveyed from entry BN(i+1) to the child PCE(i+1) to the parent PCE, and then to child PCE(i) to the entry BN(i). The exit BN(i) learns the SL via the per-domain LSP setup technique (RSVP-TE, SR, PCECC etc).

[BRPC-Stitch] define new LSP setup types for BRPC mode, this document also uses the same LSP setup type for the Stateful H-PCE mode.

- o TBD1: Inter-Domain Traffic engineering end-to-end path is setup using H-PCE method. This new LSP-TYPE value MUST be set in a PCInitiate messages sends by a P-PCE (Parent PCE) to its C-PCE (child PCE) of transit and destination domains to initiate a new inter-domain LSP tunnel. In turn, the C-PCE MUST return a Stitching Label SL in the RRO of the PCRpt message to P-PCE.

- o TBD2: Inter-Domain Traffic engineering local path is setup using RSVP-TE. This new LSP-TYPE value MUST be set in the PCInitiate message sends by a C-PCE(i) requesting to a PCC of domain(i) to initiate a new local LSP tunnel(i) which is part of an inter-domain LSP tunnel. This LSP-TYPE value MUST be used by the C-PCE(i) only after receiving a PCInitiate message with an LSP-TYPE equal to TBD1 from a P-PCE. In turn, the PCC of domain(i) MUST return a Stitching Label SL in the RRO of the PCRpt message.

- o TBD3: Inter-Domain Traffic engineering local path is setup using Segment Routing (SR). This new LSP-TYPE value MUST be set in the PCInitiate message sends by a C-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 LSP-TYPE value MUST be used by the C-PCE(i) only after receiving a PCInitiate message with

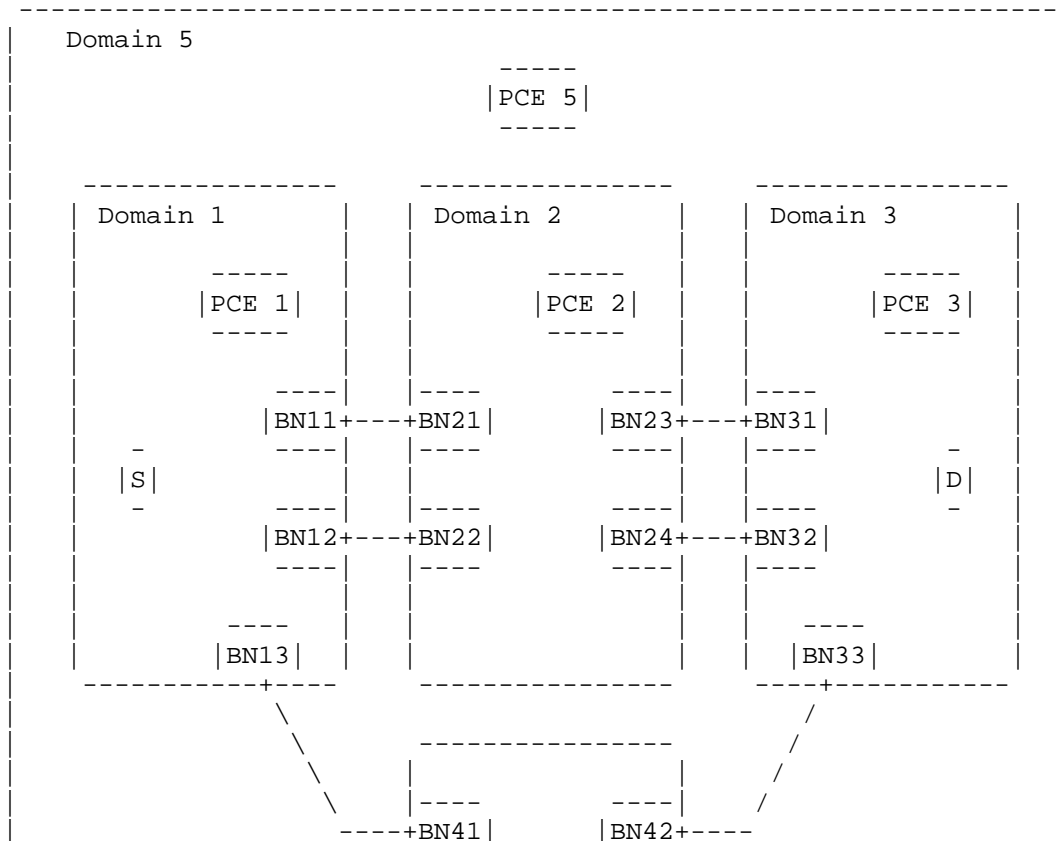
an LSP-TYPE equal to TBD1 from a P-PCE. In turn, the PCC MUST return a Stitching Label SL in the RRO of the PCRpt message.

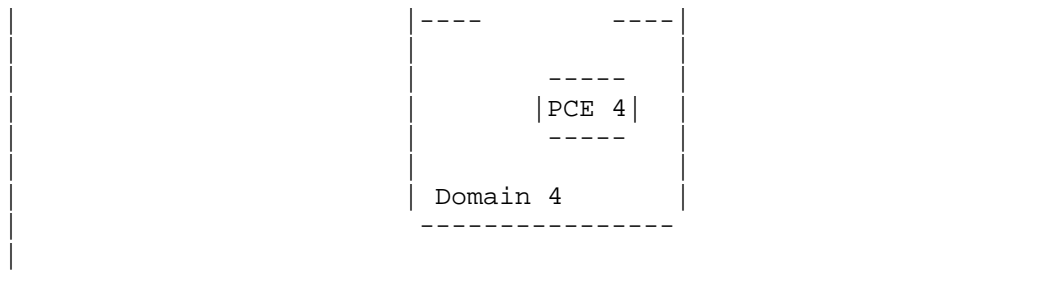
[Editor's Note - This draft authors plan to discuss with authors of [BRPC-Stitch] to simplify this, as any new path setup type like PCECC would require another path-setup type to be defined here.]

Thus, these LSP-TYPE value MUST be set in PCEP messages sends by a Parent PCE to child PCE as well as between child PCE and the PCCs when SL is used.

2.1. Stateful H-PCE Stitching Procedure

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





Section 3.3.1 of [Stateful H-PCE] describes the per-domain stitched LSP mode and list all the steps needed. To support SL based stitching, the steps are modified as follows -

Using the reference architecture described in Figure above:

- (1) 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 say -

- o S-BN41
- o BN41-BN33
- o BN33-D

For LSP (BN33-D)

- (2) 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 LSP-TYPE=TBD1.

- (3) The PCE3 further propagates the initiate message to BN33 with the ERO and LSP-TYPE=TBD2/TBD3 based on setup type.

- (4) BN33 initiates the setup of the LSP as per the path and reports to the PCE3 the LSP status ("GOING-UP").

- (5) The PCE3 further reports the status of the LSP to the P-PCE (PCE5).
- (6) 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).
- (7) 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 (SL33,BN33..D).

For LSP (BN41-BN33)

- (8) The P-PCE (PCE5) sends the initiate request to the child PCE (PCE4) via PCInitiate message for LSP (BN41-BN33) with ERO=(BN41..BN42,SL33,BN33) and LSP-TYPE=TBD1.
- (9) The PCE4 further propagates the initiate message to BN41 with the ERO and LSP-TYPE=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.
- (10) BN41 initiates the setup of the LSP as per the path and reports to the PCE4 the LSP status ("GOING-UP").
- (11) The PCE4 further reports the status of the LSP to the P-PCE (PCE5).
- (12) The node BN41 notifies the LSP state to PCE4 when the state is "UP" it also carry the stitching label (SL41) in RRO as (SL41,BN41..BN33).

(13) 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 (SL41,BN41..BN33).

For LSP (S-BN41)

(14) The P-PCE (PCE5) sends the initiate request to the child PCE (PCE1) via PCInitiate message for LSP (S-BN41) with ERO=(S..BN13,SL41,BN41).

(15) 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.

(16) S initiates the setup of the LSP as per the path and reports to the PCE1 the LSP status ("GOING-UP").

(17) The PCE1 further reports the status of the LSP to the P-PCE (PCE5).

(18) The node S notifies the LSP state to PCE1 when the state is "UP".

(19) 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 LFIB 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.

2.2. Applicability to ACTN

[ACTN] 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 [Stateful H-PCE] 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.

3. Security Considerations

Procedures and protocol extensions defined in this document do not effect the overall PCEP security model. See [RFC5440], [I-D.ietf-pce-pceps]. It is suggested that any mechanism used for securing the transmission of other PCEP message be applied here as well. As a general precaution, it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions belonging to the same administrative authority.

4. IANA Considerations

This document requests IANA actions to allocate code points for the protocol elements defined in this document.

5. References

5.1. Normative References

- [RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4674] Le Roux, J., Ed., "Requirements for Path Computation Element (PCE) Discovery", RFC 4674, October 2006.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5250] Berger, L., Bryskin, I., Zinin, A., and R. Coltun, "The OSPF Opaque LSA Option", RFC 5250, July 2008.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

5.2. Informative References

- [JMS] Java Message Service, Version 1.1, April 2002, Sun Microsystems.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [BGP-LS] Gredler, H., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and TE information using BGP", draft-ietf-idr-ls-distribution, work in progress.
- [S-PCE-GMPLS] X. Zhang, et. al, "Path Computation Element (PCE) Protocol Extensions for Stateful PCE Usage in GMPLS-controlled Networks", draft-ietf-pce-pcep-stateful-pce-gmpls, work in progress.
- [RFC7399] A. Farrel and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, October 2015.
- [RFC7449] Y. Lee, G. Bernstein, "Path Computation Element Communication Protocol (PCEP) Requirements for Wavelength Switched Optical Network (WSO) Routing and Wavelength Assignment", RFC 7449, February 2015.
- [RFC4456] Bates, T., Chen, E., and R. Chandra, "BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP)", RFC 4456, April 2006.
- [RFC6163] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163,
- [G.680] ITU-T Recommendation G.680, Physical transfer functions of optical network elements, July 2007.
- [ACTN-Frame] D. Ceccarelli, and Y. Lee (Editors), "Framework for Abstraction and Control of TE Networks", draft-ietf-teas-actn-framework, work in progress.
- [RFC6805] A. Farrel and D. King, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, November 2012.
- [PCEP-LS-Arch] Y. Lee, D. Dhody and D. Ceccarelli, "Architecture and Requirement for Distribution of Link-State and TE Information via PCEP", draft-leedhody-teas-pcep-ls, work in progress.

- [PCEP-LS] D. Dhody, Y. Lee and D. Ceccarelli "PCEP Extension for Distribution of Link-State and TE Information.", work in progress, September 21, 2015
- [Stateful-PCE] Crabbe, E., Minei, I., Medved, J., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce, work in progress.
- [PCE-Initiated] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", draft-ietf-pce-pce-initiated-lsp, work in progress.
- [Stateful H-PCE] D. Dhody, Y. Lee and D. Ceccarelli, "Hierarchical Stateful Path Computation Element (PCE)", draft-ietf-pce-stateful-hpce, work-in-progress.
- [FlexOSPF] X. Zhang, H. Zheng, R. Casellas, O. Gonzalez de Dios, D. Ceccarelli, "GMPLS OSPF Extensions in support of Flexi-grid DWDM networks", draft-ietf-ccamp-flexible-grid-ospf-ext-05, work in progress.
- [PST] Sivabalan, S., Medved, J., Minei, I., Crabbe, E., Varga, R., Tantsura, J., and J. Hardwick, "Conveying path setup type in PCEP messages", draft-ietf-pce-lsp-setup-type-03 work in progress.

Appendix A. Contributor Addresses

Author's Addresses

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

Email: leeyoung@huawei.com

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India
Email: dhruv.ietf@gmail.com

Daniele Ceccarelli
Ericsson
Torshamnsgatan, 48
Stockholm
Sweden

Email: daniele.ceccarelli@ericsson.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 2, 2018

D. Dhody, Ed.
Huawei Technologies
A. Farrel, Ed.
Juniper Networks
Z. Li
Huawei Technologies
December 29, 2017

PCEP Extension for Flow Specification
draft-li-pce-pcep-flowspec-03

Abstract

The Path Computation Element (PCE) is a functional component capable of selecting the paths through a traffic engineered network. These paths may be supplied in response to requests for computation, or may be unsolicited directions issued by the PCE to network elements. Both approaches use the PCE Communication Protocol (PCEP) to convey the details of the computed path.

Traffic flows may be categorized and described using "Flow Specifications". RFC 5575 defines the Flow Specification and describes how it may be distributed in BGP to allow specific traffic flows to be associated with routes.

This document specifies a set of extensions to PCEP to support dissemination of Flow Specifications. This allows a PCE to indicate what traffic should be placed on each path that it is aware of.

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 July 2, 2018.

Copyright Notice

Copyright (c) 2017 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. Procedures for PCE Use of Flow Specifications	4
3.1. Capability Advertisement	5
3.1.1. PCEP OPEN Message	5
3.1.2. IGP PCE Capabilities Advertisement	5
3.2. Dissemination Procedures	6
3.3. Flow Specification Synchronization	7
4. PCE FlowSpec Capability TLV	7
5. PCEP Flow Spec Object	8
6. Flow Filter TLV	9
7. Flow Specification TLVs	9
8. Detailed Procedures	12
8.1. Default Behavior and Backward Compatibility	13
8.2. Composite Flow Specifications	13
8.3. Modifying Flow Specifications	13
8.4. Multiple Flow Specifications	13
8.5. Adding and Removing Flow Specifications	14
8.6. VPN Identifiers	14
8.7. Priorities and Overlapping Flow Specifications	14
9. PCEP Messages	15
10. IANA Considerations	18
10.1. PCEP Objects	18
10.2. PCEP TLV Type Indicators	18

10.3. Flow Specification TLV Type Indicators	18
10.4. PCEP Error Codes	19
10.5. PCE Capability Flag	19
11. Security Considerations	20
12. Manageability Considerations	20
13. Acknowledgements	21
14. References	21
14.1. Normative References	21
14.2. Informative References	22
Appendix A. Contributors	23
Authors' Addresses	24

1. Introduction

[RFC4655] defines the Path Computation Element (PCE), a functional component capable of computing paths for use in traffic engineering networks. PCE was originally conceived for use in Multiprotocol Label Switching (MPLS) for Traffic Engineering (TE) networks to derive the routes of Label Switched Paths (LSPs). However, the scope of PCE was quickly extended to make it applicable to Generalized MPLS (GMPLS) networks, and more recent work has brought other traffic engineering technologies and planning applications into scope (for example, Segment Routing (SR) [I-D.ietf-pce-segment-routing]).

[RFC5440] describes the Path Computation Element Communication Protocol (PCEP). PCEP defines the communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, enabling computation of path for MPLS-TE LSPs.

Stateful PCE [RFC8231] specifies a set of extensions to PCEP to enable control of TE-LSPs by a PCE that retains state about the the LSPs provisioned in the network (a stateful PCE). [RFC8281] describes the setup, maintenance, and teardown of LSPs initiated by a stateful PCE without the need for local configuration on the PCC, thus allowing for a dynamic network that is centrally controlled. [RFC8283] introduces the architecture for PCE as a central controller and describes how PCE can be viewed as a component that performs computation to place 'flows' within the network and decide how these flows are routed.

Dissemination of traffic flow specifications (Flow Specifications) was introduced for BGP in [RFC5575]. A Flow Specification is comprised of traffic filtering rules and actions. The routers that receive a Flow Specification can classify received packets according to the traffic filtering rules and can direct packets based on the actions.

When a PCE is used to initiate tunnels (such as TE-LSPs or SR paths) using PCEP, it is important that the head end of the tunnels understands what traffic to place on each tunnel. The data flows intended for a tunnel can be described using Flow Specifications, and when PCEP is in use for tunnel initiation it makes sense for that same protocol to be used to distribute the Flow Specifications that describe what data is to flow on those tunnels.

This document specifies a set of extensions to PCEP to support dissemination of Flow Specifications. The extensions include the creation, update, and withdrawal of Flow Specifications via PCEP and can be applied to tunnels initiated by the PCE or to tunnels where control is delegated to the PCE by the PCC. Furthermore, a PCC requesting a new path can include Flow Specifications in the request to indicate the purpose of the tunnel allowing the PCE to factor this in during the path computation.

Flow Specifications are carried in TLVs within a new Flow Spec Object defined in this document. The flow filtering rules indicated by the Flow Specifications are mainly defined by BGP Flow Specifications.

2. Terminology

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

The following term from [RFC5575] is used frequently throughout this document:

Flow Specification (FlowSpec): A Flow Specification is an n-tuple consisting of several matching criteria that can be applied to IP traffic, including filters and actions. Each FlowSpec consists of a set of filters and a set of actions.

This document uses the terms "stateful PCE" and "active PCE" as advocated in [RFC7399].

3. Procedures for PCE Use of Flow Specifications

There are three elements of procedure:

- o A PCE and a PCC must be able to indicate whether or not they support the use of Flow Specifications.
- o A PCE or PCC must be able to include Flow Specifications in PCEP messages with clear understanding of the applicability of those Flow Specifications in each case including whether the use of such

information is mandatory, constrained, or optional, and how overlapping Flow Specifications will be resolved..

- o Flow Specification information/state must be synchronized between PCEP peers so that, on recovery, the peers have the same understanding of which Flow Specifications apply.

The following subsections describe these points.

3.1. Capability Advertisement

3.1.1. PCEP OPEN Message

During PCEP session establishment, a PCC or PCE that supports the procedures described in this document announces this fact by including the "PCE FlowSpec Capability" TLV (described in Section 4) in the OPEN Object carried in the PCEP Open message.

The presence of the PCE FlowSpec Capability TLV in the OPEN Object in a PCE's OPEN message indicates that the PCE can support distribute the FlowSpec to PCCs and can receive FlowSpecs in messages from the PCCs.

The presence of the PCE FlowSpec Capability TLV in the OPEN Object in a PCC's OPEN message indicates that the PCC supports the FlowSpec functionality described in this document.

If either one of a pair of PCEP peers does not indicate support of the functionality described in this document by not including the PCE FlowSpec Capability TLV in the OPEN Object in its OPEN message, then the other peer MUST NOT include a FlowSpec object in any PCEP message sent to the peer that does not support the procedures. If a FlowSpec object is received even though support has not been indicated, the receiver will respond with a PCerr message reporting the objects containing the FlowSpec as described in [RFC5440]: that is, it will use 'Unknown Object' if it does not support this specification, and 'Not supported object' if it supports this specification but has not chosen to support FlowSpec objects on this PCEP session.

3.1.2. IGP PCE Capabilities Advertisement

The ability to advertise support for PCEP and PCE features in IGP advertisements is provided for OSPF in [RFC5088] and for IS-IS in [RFC5089]. The mechanism uses the PCE Discovery TLV which has a PCE-CAP-FLAGS sub-TLV containing bit-flags each of which indicates support for a different feature.

This document defines a new PCE-CAP-FLAGS sub-TLV bit, the FlowSpec Capable flag (bit number TBD1). Setting the bit indicates that an advertising PCE supports the procedures defined in this document.

Note that while PCE FlowSpec Capability may be advertised during discovery, PCEP speakers that wish to use Flow Specification in PCEP MUST negotiate PCE FlowSpec Capability during PCEP session setup, as specified in Section 3.1.1. A PCC MAY initiate PCE FlowSpec Capability negotiation at PCEP session setup even if it did not receive any IGP PCE capability advertisement.

3.2. Dissemination Procedures

This section describes the procedures to support Flow Specifications in PCEP messages.

The primary purpose of distributing Flow Specification information is to allow a PCE to indicate to a PCC what traffic it should place on a path (such as an LSP or an SR path). This means that the Flow Specification may be included in:

- o PCInitiate messages so that an active PCE can indicate the traffic to place on a path at the time that the PCE instantiates the path.
- o PCUpd messages so that an active PCE can indicate or change the traffic to place on a path that has already been set up.
- o PCRpt messages so that a PCC can report the traffic that the PCC plans to place on the path.
- o PCReq messages so that a PCC can indicate what traffic it plans to place on a path at the time it requests the PCE to perform a computation in case that information aids the PCE in its work.
- o PCRep messages so that a PCE that has been asked to compute a path can suggest which traffic could be placed on a path that a PCC may be about to set up.
- o PCErr messages so that issues related to paths and the traffic they carry can be reported to the PCE by the PCC, and so that problems with other PCEP messages that carry Flow Specifications can be reported.

To carry Flow Specifications in PCEP messages, this document defines a new PCEP object called the PCEP Flow Spec Object. The object is OPTIONAL in the messages described above and MAY appear more than once in each message.

The PCEP Flow Spec Object carries zero or one Flow Filter TLV which describes a traffic flow.

The inclusion of multiple PCEP Flow Spec Objects allows multiple traffic flows to be placed on a single path.

Once a PCE and PCC have established that they can both support the use of Flow Specifications in PCEP messages, such information may be exchanged at any time for new or existing paths.

The application and prioritization of Flow Specifications is described in Section 8.7.

3.3. Flow Specification Synchronization

The Flow Specifications are carried along with the LSP State information as per [RFC8231] making the Flow Specifications part of the LSP database (LSP-DB). Thus, the synchronization of the Flow Specification information is done as part of LSP-DB synchronization. This may be achieved using normal state synchronization procedures as described in [RFC8231] or enhanced state synchronization procedures as defined in [RFC8232].

The approach selected will be implementation and deployment specific and will depend on issues such as how the databases are constructed and what level of synchronization support is needed.

4. PCE FlowSpec Capability TLV

The PCE-FLOWSPEC-CAPABILITY TLV is an optional TLV that can be carried in the OPEN Object [RFC5440] to exchange PCE FlowSpec capabilities of PCEP speakers.

The format of the PCE-FLOWSPEC-CAPABILITY TLV follows the format of all PCEP TLVs as defined in [RFC5440] and is shown in Figure 1.

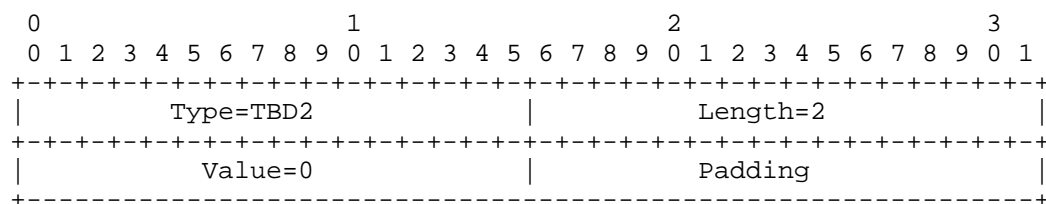


Figure 1: PCE-FLOWSPEC-CAPABILITY TLV format

The type of the PCE-FLOWSPEC-CAPABILITY TLV is TBD2 and it has a fixed length of 2 octets. The Value field is set to default value 0. The two bytes of padding MUST be set to zero and ignored on receipt.

The inclusion of this TLV in an OPEN object indicates that the sender can perform FlowSpec handling as defined in this document.

5. PCEP Flow Spec Object

The PCEP Flow Spec object defined in this document is compliant with the PCEP object format defined in [RFC5440]. It is OPTIONAL in the PCReq, PCRep, PCErr, PCInitiate, PCRpt, and PCUpd messages and MAY be present zero, one, or more times. Each instance of the object specifies a traffic flow.

The PCEP Flow Spec object carries a FlowSpec filter rule encoded in a TLV (as defined in Section 6).

The FLOW SPEC Object-Class is TBD3 (to be assigned by IANA).

The FLOW SPEC Object-Type is 1.

The format of the body of the PCEP Flow Spec object is shown in Figure 2

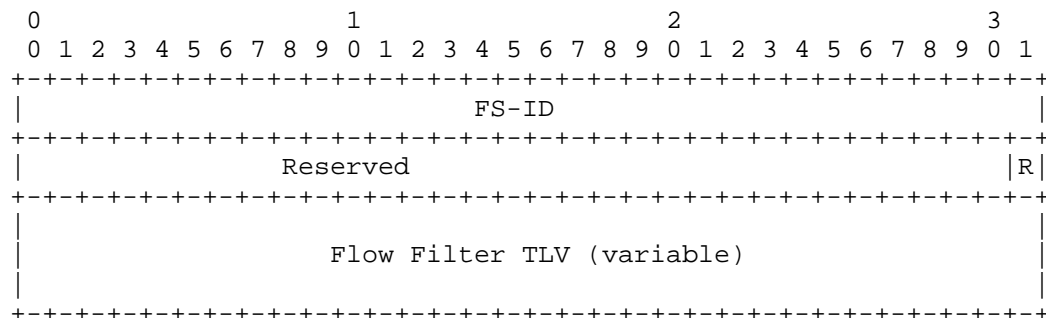


Figure 2: PCEP Flow Spec Object Body Format

FS-ID (32-bits): A PCEP-specific identifier for the FlowSpec information. A PCE creates an FS-ID for each FlowSpec, the value is unique within the scope of the PCE and is constant for the lifetime of a PCEP session. All subsequent PCEP messages can identify the FlowSpec using the FS-ID. The values 0 and 0xFFFFFFFF are reserved and MUST NOT be used.

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

R bit: The Remove bit is set when a PCEP Flow Spec Object is included in a PCEP message to indicate removal of the Flow Specification from the associated tunnel. If the bit is clear, the Flow Specification is being added or modified.

Flow Filter TLV (variable): One TLV MAY be included.

The Flow Filter TLV is OPTIONAL when the R bit is set. The TLV MUST be present when the R bit is clear. If the TLV is missing when the R bit is clear, the PCEP peer MUST respond with a PCErr message with error-type TBD8 (FlowSpec Error), error-value 2 (Malformed FlowSpec).

6. Flow Filter TLV

A new PCEP TLV is defined to convey Flow Specification filtering rules that specify what traffic is carried on a path. The TLV follows the format of all PCEP TLVs as defined in [RFC5440]. The Type field values come from the codepoint space for PCEP TLVs and has the value TBD4.

The Value field contains one or more sub-TLVs (the Flow Specification TLVs) as defined in Section 7. Only one Flow Filter TLV can be present and represents the complete definition of a Flow Specification for traffic to be placed on the tunnel indicated by the PCEP message in which the PCEP Flow Spec Object is carried. The set of Flow Specification TLVs in a single instance of a Flow Filter TLV are combined to indicate the specific Flow Specification.

Further Flow Specifications can be included in a PCEP message by including additional Flow Spec objects.

7. Flow Specification TLVs

Flow Filter TLV carries one or more Flow Specification sub-TLV. The Flow Specification TLV also follows the format of all PCEP TLVs as defined in [RFC5440], however, the Type values are selected from a separate IANA registry (see Section 10) rather than from the common PCEP TLV registry.

Type values are chosen so that there can be commonality with Flow Specifications defined for use with BGP. This is possible because the BGP Flow Spec encoding uses a single octet to encode the type where PCEP uses two octets. Thus the space of values for the Type field is partitioned as shown in Figure 3.

Range	
0	Reserved - must not be allocated.
1 .. 255	Per BGP registry defined by [RFC5575]. Not to be allocated in this registry.
256 .. 65535	New PCEP Flow Specs allocated according to the registry defined in this document.

Figure 3: Flow Specification TLV Type Ranges

The content of the Value field Flow in each TLV is specific to the type and describes the parameters of the Flow Specification. The definition of the format of many of these Value fields is inherited from BGP specifications as shown in Figure 4. Specifically, the inheritance is from [RFC5575] and [I-D.ietf-idr-flow-spec-v6], but may also be inherited from future BGP specifications.

When multiple Flow Specification TLVs are present in a single Flow Filter TLV they are combined to produce a more detailed description of a flow. For examples and rules about how this is achieved, see [RFC5575].

An implementation that receives a PCEP message carrying a Flow Specification TLV with a type value that it does not recognize or does not support MUST respond with a PCErr message with error-type TBD8 (FlowSpec Error), error-value 1 (Unsupported FlowSpec) and MUST NOT install the Flow Specification.

When used in other protocols (such as BGP) these Flow Specifications are also associated with actions to indicate how traffic matching the Flow Specification should be treated. In PCEP, however, the only action is to associate the traffic with a tunnel and to forward matching traffic on to that path, so no encoding of an action is needed.

Section 8.7 describes how overlapping Flow Specifications are prioritized and handled.

Type	Description	Value defined in
*	Destination IPv4 Prefix	[RFC5575]
*	Source IPv4 Prefix	[RFC5575]
*	IP Protocol	[RFC5575]
*	Port	[RFC5575]
*	Destination port	[RFC5575]
*	Source port	[RFC5575]
*	ICMP type	[RFC5575]
*	ICMP code	[RFC5575]
*	TCP flags	[RFC5575]
*	Packet length	[RFC5575]
*	DSCP	[RFC5575]
*	Fragment	[RFC5575]
*	Flow Label	[I-D.ietf-idr-flow-spec-v6]
*	Destination IPv6 Prefix	[I-D.ietf-idr-flow-spec-v6]
*	Source IPv6 Prefix	[I-D.ietf-idr-flow-spec-v6]
*	Next Header	[I-D.ietf-idr-flow-spec-v6]
TBD5	Route Distinguisher	[I-D.dhodylee-pce-pcep-ls]
TBD6	IPv4 Multicast Flow	[This.I-D]
TBD7	IPv6 Multicast Flow	[This.I-D]

* Indicates that the TLV Type value comes from the value used in BGP.

Figure 4: Table of Flow Specification TLV Types

All Flow Specification TLVs with Types in the range 1 to 255 have Values defined for use in BGP (for example in [RFC5575] and [I-D.ietf-idr-flow-spec-v6]) and are set using the BGP encoding, but without the type or length octets (the relevant information is in the Type and Length fields of the TLV). The Value field is padded with trailing zeros to achieve 4-byte alignment if necessary.

[I-D.dhodylee-pce-pcep-ls] defines a way to convey identification of a VPN in PCEP via a Route Distinguisher (RD) [RFC4364] and encoded in ROUTE-DISTINGUISHER TLV. A Flow Specification TLV with Type TBD5 carries a Value field matching that in the ROUTE-DISTINGUISHER TLV and is used to identify that other flow filter information (for example, an IPv4 destination prefix) is associated with a specific VPN identified by the RD. See Section 8.6 for further discussion of VPN identification.

Although it may be possible to describe a multicast Flow Specification from the combination of other Flow Specification TLVs with specific values, it is more convenient to use a dedicated Flow Specification TLV. Flow Specification TLVs with Type values TBD6 and TBD7 are used to identify a multicast flow for IPv4 and IPv6 respectively. The Value field is encoded as shown in Figure 5.

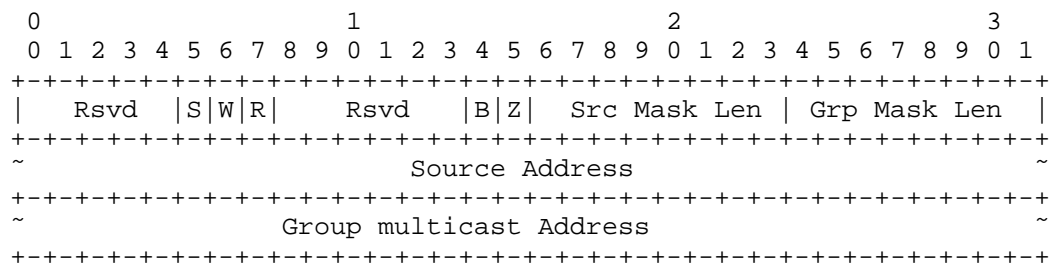


Figure 5: Multicast Flow Specification TLV Encoding

The fields of the two Multicast Flow Specification TLVs are as described in Section 4.9.1 of [RFC7761] noting that the two address fields are 32 bits for the IPv4 Multicast Flow and 128 bits for the IPv6 Multicast Flow. Reserved fields MUST be set to zero and ignored on receipt.

8. Detailed Procedures

This section outlines some specific detailed procedures for using the protocol extensions defined in this document.

8.1. Default Behavior and Backward Compatibility

The default behavior is that no Flow Specification is applied to a tunnel. That is, the default is that the Flow Spec object is not used as is the case in all systems before the implementation of this specification.

In this case it is a local matter (such as through configuration) how tunnel head ends are instructed what traffic to place on a tunnel.

[RFC5440]describes how receivers respond when they see unknown PCEP objects.

8.2. Composite Flow Specifications

Flow Specifications may be represented by a single Flow Specification TLV or may require a more complex description using multiple Flow Specification TLVs. For example, a flow indicated by a source-destination pair of IPv6 addresses would be described by the combination of Destination IPv6 Prefix and Source IPv6 Prefix Flow Specification TLVs.

8.3. Modifying Flow Specifications

A PCE may want to modify a Flow Specification associated with a tunnel, or a PCC may want to report a change to the Flow Specification it is using with a tunnel.

It is important that the specific Flow Specification is identified so that it is clear that this is a modification of an existing flow and not the addition of a new flow as described in Section 8.4. The FS-ID field of the PCEP Flow Spec Object is used to identify a specific Flow Specification.

When modifying a Flow Specification, all Flow Specification TLVs for the intended specification of the flow MUST be included in the PCEP Flow Spec Object and the FS-ID MUST be retained from the previous description of the flow.

8.4. Multiple Flow Specifications

It is possible that multiple flows will be place on a single tunnel. In some cases it is possible to to define these within a single PCEP Flow Spec Object: for example, two Destination IPv4 Prefix TLVs could be included to indicate that packets matching either prefix are acceptable. PCEP would consider this as a single Flow Specification identified by a single FS-ID.

In other scenarios the use of multiple Flow Specification TLVs would be confusing. For example, if flows from A to B and from C to D are to be included then using two Source IPv4 Prefix TLVs and two Destination IPv4 Prefix TLVs would be confusing (are flows from A to D included?). In these cases, each Flow Specification is carried in its own PCEP Flow Spec Object with multiple objects present on a single PCEP message. Use of separate objects also allows easier removal and modification of Flow Specifications.

8.5. Adding and Removing Flow Specifications

The Remove bit in the the PCEP Flow Spec Object is left clear when a Flow Specification is being added or modified.

To remove a Flow Specification, a PCEP Flow Spec Object is included with the FS-ID matching the one being removed, and the R bit set to indicate removal. In this case it is not necessary to include any Flow Specification TLVs.

If the R bit is set and Flow Specification TLVs are present an implementation MAY ignore them. If the implementation checks the Flow Specification TLVs against those recorded for the FS-ID of the Flow Specification being removed and finds a mismatch, the Flow Specification MUST still be removed and the implementation SHOULD record a local exception or log.

8.6. VPN Identifiers

VPN instances are identified in BGP using Route Distinguishers (RDs) [RFC4364]. These values are not normally considered to have any meaning outside of the network, and they are not encoded in data packets belonging to the VPNs. However, RDs provide a useful way of identifying VPN instances and are often manually or automatically assigned to VPNs as they are provisioned.

Thus the RD provides a useful way to indicate that traffic for a particular VPN should be placed on a given tunnel. The tunnel head end will need to interpret this Flow Specification not as a filter on the fields of data packets, but using the other mechanisms that it uses to identify VPN traffic. This could be based on the incoming port (for port-based VPNs) or may leverage knowledge of the VRF that is in use for the traffic.

8.7. Priorities and Overlapping Flow Specifications

TBD

An implementation that receives a PCEP message carrying a Flow Specification that it cannot resolve against other Flow Specifications already installed MUST respond with a PCErr message with error-type TBD8 (FlowSpec Error), error-value 3 (Unresolvable conflict) and MUST NOT install the Flow Specification.

9. PCEP Messages

The figures below use the notation defined in [RFC5511].

The FLOW SPEC Object is OPTIONAL and MAY be carried in the PCEP messages.

The PCInitiate message is defined in [RFC8281] and updated as below:

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

Where:

```
<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>]
                                       [<flowspec-list>]
```

Where:

```
<flowspec-list> ::= <FLOWSPEC> [<flowspec-list>]
```

The PCUpd message is defined in [RFC8231] and updated as below:

<PCUpd Message> ::= <Common Header>
 <update-request-list>

Where:

<update-request-list> ::= <update-request>
 [<update-request-list>]

<update-request> ::= <SRP>
 <LSP>
 <path>
 [<flowspec-list>]

Where:

<path> ::= <intended-path><intended-attribute-list>

<flowspec-list> ::= <FLOWSPEC> [<flowspec-list>]

The PCRpt message is defined in [RFC8231] and updated as below:

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

Where:

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

<state-report> ::= [<SRP>]
 <LSP>
 <path>
 [<flowspec-list>]

Where:

<path> ::= <intended-path>
 [<actual-attribute-list><actual-path>]
 <intended-attribute-list>

<flowspec-list> ::= <FLOWSPEC> [<flowspec-list>]

The PCReq message is defined in [RFC5440] and updated in [RFC8231],
 it is further updated below for flow specification:

```
<PCReq Message> ::= <Common Header>
                        [<svec-list>]
                        <request-list>
```

Where:

```
<svec-list> ::= <SVEC> [<svec-list>]

<request-list> ::= <request> [<request-list>]

<request> ::= <RP>
               <END-POINTS>
               [<LSP>]
               [<LSPA>]
               [<BANDWIDTH>]
               [<metric-list>]
               [<RRO> [<BANDWIDTH>]]
               [<IRO>]
               [<LOAD-BALANCING>]
               [<flowspec-list>]
```

Where:

```
<flowspec-list> ::= <FLOWSPEC> [<flowspec-list>]
```

The PCRep message is defined in [RFC5440] and updated in [RFC8231], it is further updated below for flow specification:

```
<PCRep Message> ::= <Common Header>
                     <response-list>
```

Where:

```
<response-list> ::= <response> [<response-list>]

<response> ::= <RP>
               [<LSP>]
               [<NO-PATH>]
               [<attribute-list>]
               [<path-list>]
               [<flowspec-list>]
```

Where:

```
<flowspec-list> ::= <FLOWSPEC> [<flowspec-list>]
```

10. 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.

10.1. PCEP Objects

Each PCEP object has an Object-Class and an Object-Type. IANA maintains a subregistry called "PCEP Objects". IANA is requested to make an assignment from this subregistry as follows:

Object-Class	Value Name	Object-Type	Reference
TBD3	FLOW SPEC	0 (Reserved)	[This.I-D]
		1	[This.I-D]

10.2. 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
TBD2	PCE-FLOWSPEC-CAPABILITY TLV	[This.I-D]
TBD4	FLOW FILTER TLV	[This.I-D]

10.3. Flow Specification TLV Type Indicators

IANA is requested to create a new subregistry call the PCEP Flow Specification TLV Type Indicators registry.

Allocations from this registry are to be made according to the following assignment policies [RFC8126]:

Range	Assignment policy
0	Reserved - must not be allocated.
1 .. 255	Reserved - must not be allocated. Usage mirrors the BGP FlowSpec registry [RFC5575].
258 .. 64506	Specification Required
64507 .. 65531	First Come First Served
65532 .. 65535	Experimental

IANA is requested to pre-populate this registry with values defined in this document as follows:

Value	Meaning
TBD5	Route Distinguisher
TBD6	IPv4 Multicast
TBD7	IPv6 Multicast

10.4. PCEP Error Codes

IANA maintains a subregistry called "PCEP-ERROR Object Error Types and Values". Entries in this subregistry are described by Error-Type and Error-value. IANA is requested to make the following assignment from this subregistry:

Error-Type	Meaning	Error-value	Reference
TBD8	FlowSpec error	0: Unassigned	[This.I-D]
		1: Unsupported FlowSpec	[This.I-D]
		2: Malformed FlowSpec	[This.I-D]
		3: Unresolvable conflict	[This.I-D]
		4-255: Unassigned	[This.I-D]

10.5. PCE Capability Flag

IANA maintains a subregistry called "Open Shortest Path First v2 (OSPFv2) Parameters" with a sub-registry called "Path Computation

Element (PCE) Capability Flags". IANA is requested to assign a new capability bit from this registry as follows:

Bit	Capability Description	Reference
TBD1	FlowSpec	[This.I-D]

11. Security Considerations

We may assume that a system that utilizes a remote PCE is subject to a number of vulnerabilities that could allow spurious LSPs or SR paths to be established or that could result in existing paths being modified or torn down. Such systems, therefore, apply security considerations as described in [RFC5440], [RFC6952], and [RFC8253].

The description of Flow Specifications associated with paths set up or controlled by a PCE add an further detail that could be attacked without tearing down LSPs or SR paths but causing traffic to be misrouted within the network. Therefore, the use of the security mechanisms for PCEP referenced above is important.

Visibility into the information carried in PCEP does not have direct privacy concerns for end-users' data, however, knowledge of how data is routed in a network may make that data more vulnerable. Of course, the ability to interfere with the way data s routed also makes the data more vulnerable. Furthermore, knowledge of the connected end-points (such as multicast receivers or VPN sites) is usually considered private customer information. Therefore, implementations or deployments concerned to protect privacy MUST apply the mechanisms described in the documents referenced above.

Experience with Flow Specifications in BGP systems indicates that they can become complex and that the overlap of Flow Specifications installed in different orders can lead to unexpected results. Although this is not directly a security issue per se, the confusion and unexpected forwarding behavior may be engineered or exploited by an attacker. Therefore, implementers and operators SHOULD pay careful attention to the Manageability Considerations described in Section 12.

12. Manageability Considerations

TBD

13. Acknowledgements

Thanks to Julian Lucek and Sudhir Cheruathur for useful discussions.

14. References

14.1. Normative References

- [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-08 (work in progress), June 2017.
- [I-D.ietf-idr-flow-spec-v6]
McPherson, D., Raszuk, R., Pithawala, B., akarch@cisco.com, a., and S. Hares, "Dissemination of Flow Specification Rules for IPv6", draft-ietf-idr-flow-spec-v6-09 (work in progress), November 2017.
- [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>>.
- [RFC5575] Marques, P., Sheth, N., Raszuk, R., Greene, B., Mauch, J., and D. McPherson, "Dissemination of Flow Specification Rules", RFC 5575, DOI 10.17487/RFC5575, August 2009, <<https://www.rfc-editor.org/info/rfc5575>>.
- [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>>.
- [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>>.

14.2. Informative References

- [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-11 (work in progress),
November 2017.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private
Networks (VPNs)", RFC 4364, DOI 10.17487/RFC4364, February
2006, <<https://www.rfc-editor.org/info/rfc4364>>.
- [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>>.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R.
Zhang, "OSPF Protocol Extensions for Path Computation
Element (PCE) Discovery", RFC 5088, DOI 10.17487/RFC5088,
January 2008, <<https://www.rfc-editor.org/info/rfc5088>>.
- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R.
Zhang, "IS-IS Protocol Extensions for Path Computation
Element (PCE) Discovery", RFC 5089, DOI 10.17487/RFC5089,
January 2008, <<https://www.rfc-editor.org/info/rfc5089>>.
- [RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of
BGP, LDP, PCEP, and MSDP Issues According to the Keying
and Authentication for Routing Protocols (KARP) Design
Guide", RFC 6952, DOI 10.17487/RFC6952, May 2013,
<<https://www.rfc-editor.org/info/rfc6952>>.
- [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>>.
- [RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I.,
Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent
Multicast - Sparse Mode (PIM-SM): Protocol Specification
(Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761, March
2016, <<https://www.rfc-editor.org/info/rfc7761>>.
- [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>>.

- [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>>.
- [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>>.

Appendix A. Contributors

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

Email: shankara@huawei.com

Qiandeng Liang
Huawei Technologies
101 Software Avenue,
Yuhuatai District
Nanjing
210012
China

Email: liangqiandeng@huawei.com

Cyril Margaria

Juniper Networks
200 Somerset Corporate Boulevard, Suite 4001
Bridgewater, NJ
08807
USA

Email: cmargaria@juniper.net

Colby Barth
Juniper Networks
200 Somerset Corporate Boulevard, Suite 4001
Bridgewater, NJ
08807
USA

Email: cbarth@juniper.net

Xia Chen
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China

Email: jescia.chenxia@huawei.com

Shunwan Zhuang
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China

Email: zhuangshunwan@huawei.com

Authors' Addresses

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

Email: dhruv.ietf@gmail.com

Adrian Farrel (editor)
Juniper Networks

Email: afarrel@juniper.net

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

Email: lizhenbin@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 5, 2018

U. Palle
D. Dhody
S. Karunanithi
Huawei Technologies
March 4, 2018

LABEL-DB Synchronization Procedures for a PCE as a central
controller(PCECC)
draft-palle-pce-controller-labeldb-sync-03

Abstract

PCE as a central controller specifies the procedures and PCEP protocol extensions where LSPs are calculated/setup/initiated and label forwarding entries are downloaded through a centralized PCE server to each network devices along the LSP path while leveraging the existing PCE technologies as much as possible.

Labels downloaded to forwarding entries requires a reliable synchronization mechanism between the path computation clients (PCCs) and the PCE. The basic mechanism for label database (LABEL-DB) synchronization is part of the PCE as a central controller specification. This document presents motivations for optimizations to the LABEL-DB synchronization and the corresponding PCEP procedures and extensions.

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, 2018.

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	3
1.1. Requirements Language	3
2. LABEL-DB Synchronization	3
3. Optimizations for LABEL-DB Synchronization	4
3.1. LABEL-DB Synchronization Avoidance Procedure	4
3.2. Incremental LABEL-DB Synchronization Procedure	8
4. PCEP Extensions	10
4.1. Extension of SRP object	10
4.2. Extension of PCECC Capability sub-TLV	10
4.3. New LABEL-DB-VERSION TLV	11
5. IANA Considerations	12
5.1. PCEP TLV Type Indicators	12
5.2. PCECC-CAPABILITY sub-TLV	12
6. Manageability Considerations	12
6.1. Control of Function and Policy	12
6.2. Information and Data Models	12
6.3. Liveness Detection and Monitoring	12
6.4. Verify Correct Operations	12
6.5. Requirements On Other Protocols	13
6.6. Impact On Network Operations	13
7. Security Considerations	13
8. Acknowledgements	13
9. References	13
9.1. Normative References	13
9.2. Informative References	14
Authors' Addresses	14

1. Introduction

[I-D.zhao-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller [RFC8283] and user cases where LSPs are calculated/setup/initiated/downloaded through extending the existing PCE architectures and PCEP.

[I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr] specifies reliable synchronization mechanism between the path computation clients (PCCs) and the PCECC.

This draft specify the optimizations for LABEL-DB synchronization and the corresponding PCEP procedures and extensions.

[Important Note - [I-D.zhao-pce-pcep-extension-for-pce-controller] defined new messages for label download and cleanup. The authors and WG also debated on the use of existing PCEP messages. The appendix of the document includes the details related to use of existing messages. This document is related to new messages as a new procedure for Label DB sync was defined. In case existing messages are used, some modifications needs to be made to the existing LSP-DB synchronization mechanism to also handle the label synchronization. These details are not present in the document at this stage.]

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. LABEL-DB Synchronization

PCECC MUST maintains the LABEL-DB for each PCEP session separately. The purpose of LABEL-DB synchronization is to make sure that the PCECC's view of LABEL-DB matches with the PCC's LABEL-DB. The LABEL-DB synchronization MUST be performed from PCECC to PCC immediately after the LSP state synchronization. [RFC8231] describes the basic mechanism for LSP state synchronization. [RFC8233] describes the optimizations for LSP state synchronization.

Full LABEL-DB synchronization performed from PCECC to PCC on Initial session UP or every session flap is described in [I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr].

Providing an Optimizations for LABEL-DB synchronization can result in significant savings in both control-plane data exchanges and the time it takes for the PCC to become fully operational.

Optimizations for LABEL-DB synchronization describes the need that both PCEP speakers support label database version capability and maintain label database version for each session. See Section 3 for detail procedures.

[Editor's Note: [I-D.zhao-pce-pcep-extension-for-pce-controller] defines new messages PCLabelUpd and PCLabelRpt. Questions were raised on the need for the new messages. Further appendix in [I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr] describes how the existing messages can be extended to add this functionality. WG needs to decide the final direction i.e. new specific messages are needed or existing PCEP messages can be extended. The optimization procedure would need to be modified based on the above decision.]

3. Optimizations for LABEL-DB Synchronization

This section add some of the optimization mechanisms for LABEL-DB synchronization. By default, the full LABEL-DB synchronization is performed.

3.1. LABEL-DB Synchronization Avoidance Procedure

The LABEL-DB synchronization MAY be skipped following a PCEP session restart if there is no change in the LABEL-DB of the session at PCECC, during the period prior to session re-initialization. To be able to make this determination, labels must be exchanged and maintained by both PCECC and PCC during normal operation. This is accomplished by keeping track of the changes to the label database, using a version tracking field called the Label Database Version Number.

The Label Database Version Number, carried in LABEL-DB-VERSION TLV (see Section 4.3), is owned by a PCECC and it MUST be incremented by 1 for each successive change in the PCECC's label database. The Label Database Version Number MUST start at 1 and may wrap around. Values 0 and 0xFFFFFFFFFFFFFFFF are reserved. If either of the two values are used during LABEL-DB synchronization, the PCC speaker receiving this node should send back a PCErr with Error-type TBD1 Error-value 3 'Received an invalid Label Database Version Number', and close the PCEP session. Operations that trigger a change to the Label database include an addition or deletion of labels that would trigger a label update to the PCC.

LABEL-DB synchronization avoidance is advertised on a PCEP session during session startup using the INCLUDE-LABEL-DB-VERSION (I) bit in the PCECC capability sub-TLV (see Section 4.2). The PCEP peer MAY include the SPEAKER-ENTITY-ID TLV described in [RFC8233] in the OPEN message to identify the peer in case of IP address change.

If both PCEP speakers set the I flag in the OPEN object's PCECC Capability sub-TLV to 1, the PCECC MUST include the LABEL-DB-VERSION TLV in each LABEL object of the PCLabelUpd message. If the LABEL-DB-VERSION TLV is missing in a PCLabelUpd message, the PCC will generate an error with Error-Type 6 (mandatory object missing) and Error-Value TBD2 'LABEL-DB-VERSION TLV missing' and close the session. If LABEL-DB synchronization avoidance has not been enabled on a PCEP session, the PCECC SHOULD NOT include the LABEL-DB-VERSION TLV in the LABEL Object and the PCC SHOULD ignore it were it to receive one.

If a PCC's label database survived the restart of a PCEP session, the PCC will include the LABEL-DB-VERSION TLV in its OPEN object, and the TLV will contain the last Label Database Version Number received on an Label Update from the PCECC in the previous PCEP session. If a PCECC's Label Database survived the restart of a PCEP session, the PCECC will include the LABEL-DB-VERSION TLV in its OPEN object and the TLV will contain the latest Label Database Version Number. If a PCEP speaker's label database did not survive the restart of a PCEP session, the PCEP speaker MUST NOT include the LABEL-DB-VERSION TLV in the OPEN object.

If both PCEP speakers include the LABEL-DB-VERSION TLV in the OPEN Object and the TLV values match, the PCECC MAY skip LABEL-DB synchronization. Otherwise, the PCECC MUST perform full LABEL-DB synchronization ([I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr]) or incremental LABEL-DB synchronization (see Section 3.2) to the PCC. In case, the PCECC attempts to skip LABEL-DB synchronization, by setting the SYNC Flag to 0 on the first Label Update from the PCECC, the PCC MUST send back a PCErr with Error-type TBD1 (Label Database Synchronization Error) and Error-value 4(Label Database Version mismatch), and close the PCEP session.

If LABEL-DB synchronization is required, then prior to completing the initialization phase, the PCC MUST mark any labels in the label database that were previously updated by the PCECC as stale. When the PCECC updates a label during LABEL-DB synchronization, if the label already exists in the label database, the PCC MUST update the label database and clear the stale marker from the label. When it has finished LABEL-DB synchronization, the PCECC MUST immediately send an end of synchronization marker. The end of synchronization marker is a Path Computation Label Update (PCLabelUpd) message with a

SRP object containing the SYNC flag set to 0 (see Section 4.1) and Label as 0 in the LABEL object. The LABEL-DB-VERSION TLV MUST be included in this PCLabelUpd message. On receiving this Label Update, the PCC MUST report all the labels in the label database that are still marked as stale to PCECC.

Note that a PCECC/PCC MAY force LABEL-DB synchronization by not including the LABEL-DB-VERSION TLV in its OPEN object.

Figure 1 shows an example sequence where the LABEL-DB synchronization is skipped.

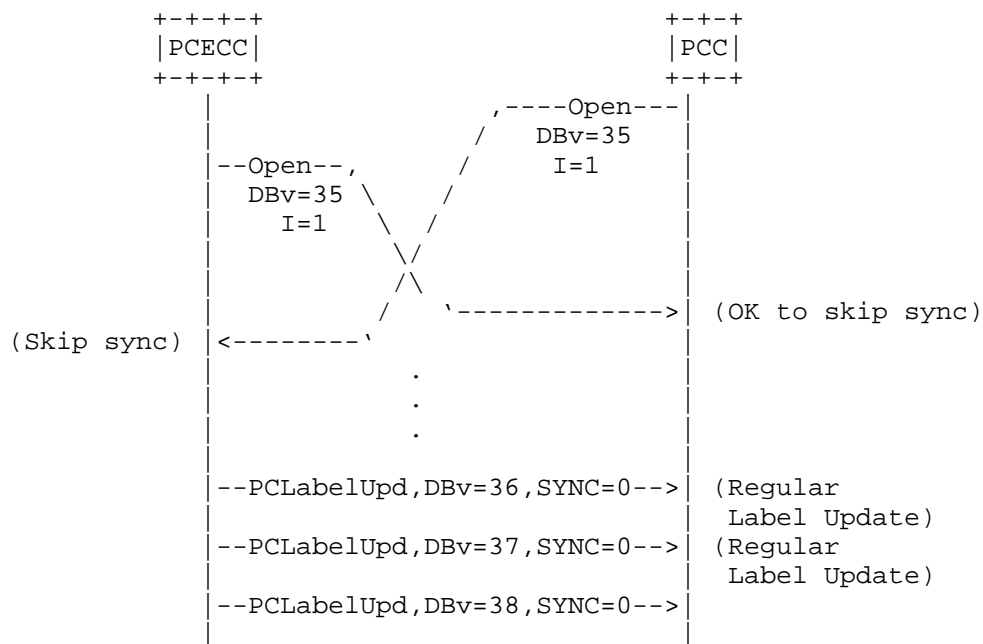


Figure 1: LABEL-DB synchronization Skipped

Figure 2 shows an example sequence where the LABEL-DB synchronization is performed due to label database version mismatch during the PCEP session setup. Note that the same LABEL-DB synchronization sequence would happen if either the PCC or the PCECC would not include the LABEL- DB-VERSION TLV in their respective Open messages.

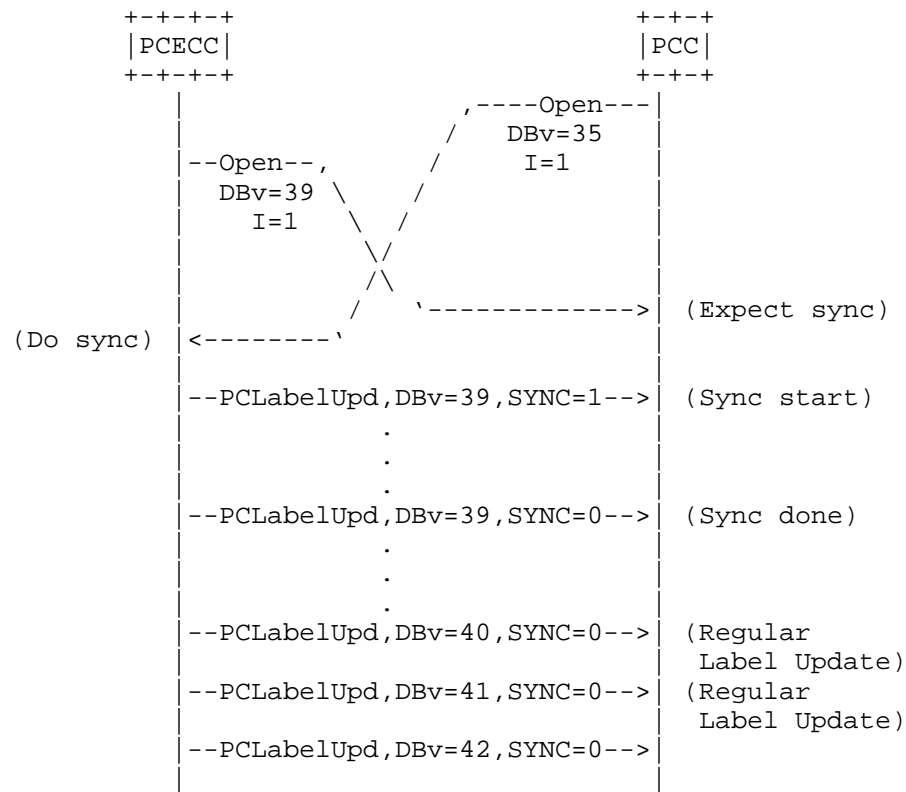


Figure 2: LABEL-DB synchronization Performed

Figure 3 shows an example sequence where the LABEL-DB synchronization is skipped, but because one or both PCEP speakers set the I Flag to 0, the PCECC does not send LABEL-DB-VERSION TLVs in subsequent PCLabelUpd messages to the PCC. If the current PCEP session restarts, the PCEP speakers will have to perform full LABEL-DB synchronization, since the PCC does not know the PCECC's latest Label Database Version Number information.

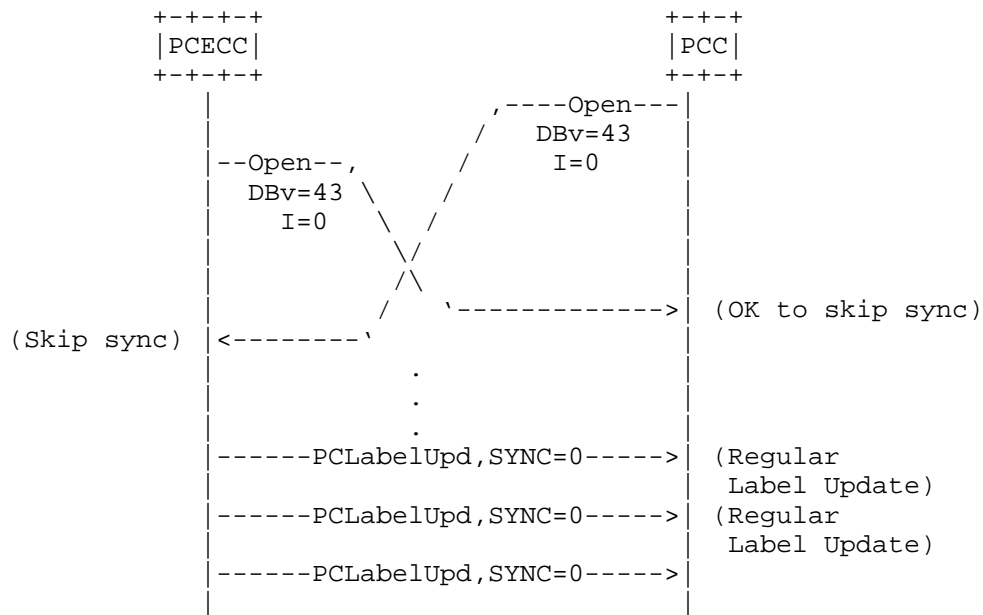


Figure 3: LABEL-DB Synchronization Skipped, no LABEL-DB-VERSION TLVs sent from PCECC

3.2. Incremental LABEL-DB Synchronization Procedure

If a PCC restarts and its label database survived, PCECC with mismatched Label Database Version Number will send all their Labels information (full LABEL-DB) to the PCC, even if only a small number of changes happened. It can take a long time and consume large communication channel bandwidth.

This section extends the idea to only synchronize the delta (changes) in case of Label Database Version Number of both PCEP peers is non-zero and mismatch.

If both PCEP speakers include the LABEL-DB-VERSION TLV in the OPEN object and the LABEL-DB-VERSION TLV values match, the PCECC MAY skip LABEL-DB synchronization. Otherwise, the PCECC MUST perform LABEL-DB synchronization. Incremental label database synchronization capability is advertised on a PCEP session during session startup using the DELTA-LABEL-SYNC-CAPABILITY (D) bit in the capabilities TLV (see Section 4.2). Instead of dumping full LABEL-DB to the PCC again, the PCECC synchronizes the delta (changes) as described in Figure 4 when D flag and I flag is set to 1 by both PCC and PCECC. Other combinations of D and I flags setting by PCC and PCECC result in full LABEL-DB synchronization procedure as described in

[I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr]. The PCECC MAY force a full LABEL-DB synchronization by setting the D flag to zero in the OPEN message.

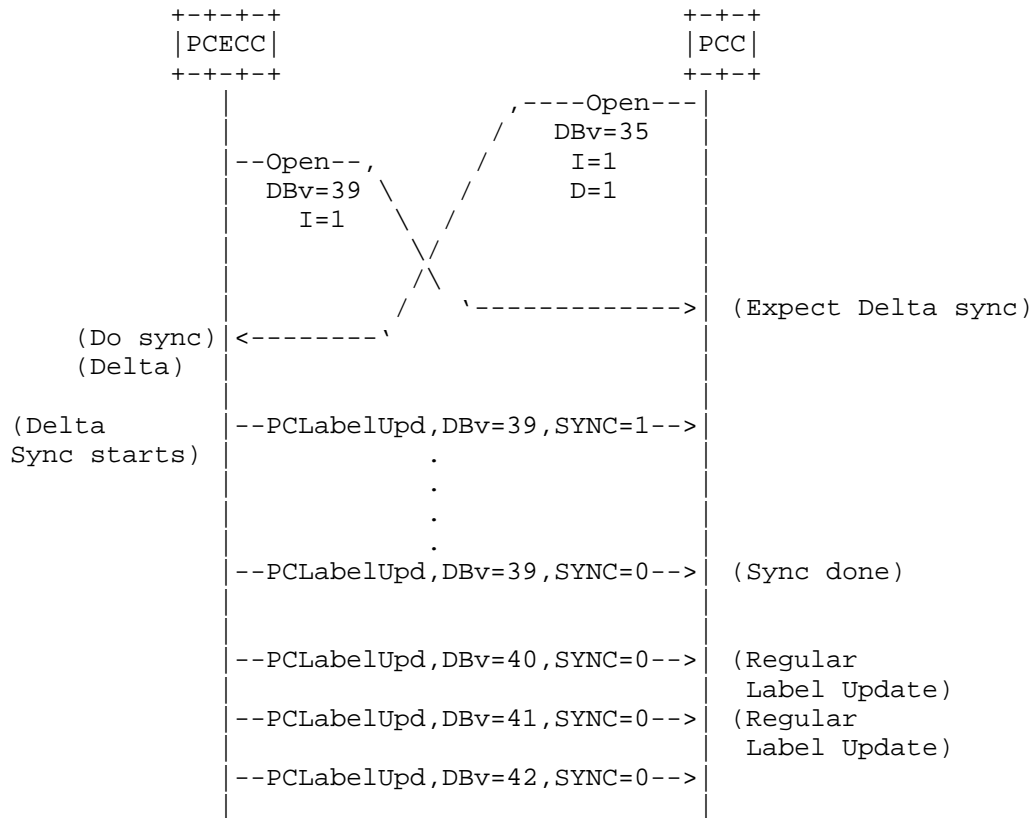


Figure 4: Incremental Synchronization Procedure

As per Section 3.1, the Label Database Version Number is incremented each time a change is made to the PCECC's label database. Each label is associated with the DB version at the time of its addition. This is needed to determine which label and what information needs to be synchronized in incremental LABEL-DB synchronization.

It is not necessary for a PCECC to store a complete history of label database change, but rather remember the labels (including label addition and deletion) that happened between the PCEP session(s) restart in order to carry out incremental LABEL-DB synchronization. After the synchronization procedure finishes, the PCECC can dump this

history information. In the example shown in Figure 4, the PCECC needs to store the label changes that happened between DB Version 35 to 39 and synchronizes these changes only when performing incremental label update. So a PCECC needs to remember at least the label changes that happened after an existing PCEP session with a PCC goes down to have any chance of doing incremental synchronization when the session is re-established.

If a PCECC finds out it does not have sufficient information to complete incremental synchronization after advertising incremental LABEL-DB synchronization capability, it MUST send a PCErr with Error-Type TBD1 and Error-Value 5 'A PCECC indicates to a PCC that it can not complete the LABEL-DB synchronization' and terminate the session. The PCECC SHOULD re-establish the session with the D bit set to 0 in the OPEN message.

The other procedures and error checks remain unchanged from the default LABEL-DB synchronization defined in [I-D.zhao-pce-pcep-extension-for-pce-controller] and [I-D.zhao-pce-pcep-extension-pce-controller-sr].

4. PCEP Extensions

4.1. Extension of SRP object

SRP object extension for SYNC flag to specify the LABEL-DB synchronization operation is defined in [I-D.zhao-pce-pcep-extension-for-pce-controller].

4.2. Extension of PCECC Capability sub-TLV

PCECC Capability sub-TLV is defined in [I-D.zhao-pce-pcep-extension-for-pce-controller]. This draft defines a new 'INCLUDE-LABEL-DB-VERSION' flag (I bit) to specify the label database version capability and 'DELTA-LABEL-SYNC-CAPABILITY' to specify the incremental label database synchronization capability.

The TLV format is as per [RFC5440]. The format of the PCECC Capability sub-TLV is shown Figure 5:



Figure 5: PCECC Capability sub-TLV

I (INCLUDE-LABEL-DB-VERSION - 1 bit): if set to 1 by both PCEP Speakers, the PCECC will include the LABEL-DB-VERSION TLV in each LABEL Object.

D (DELTA-LABEL-SYNC-CAPABILITY - 1 bit): if set to 1 by a PCEP speaker, it indicates that the PCEP speaker allows incremental (delta) LABEL-DB synchronization.

4.3. New LABEL-DB-VERSION TLV

The Label Database Version Number (LABEL-DB-VERSION) TLV is an optional TLV that MAY be included in the OPEN object and the LABEL object.

The TLV format is as per [RFC5440]. The format of the LABEL-DB-VERSION TLV is shown in the following figure:

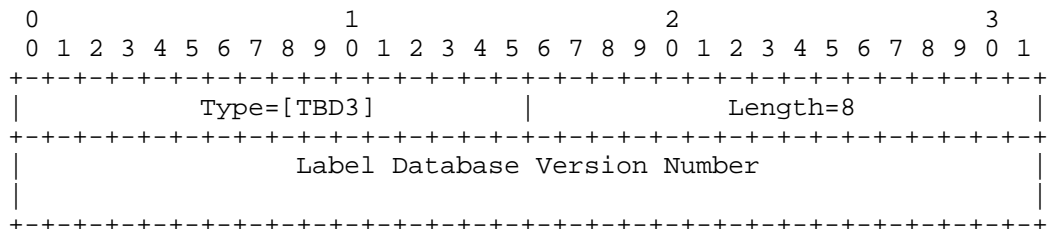


Figure 6: LABEL-DB-VERSION TLV format

The type of the TLV is [TBD3] and it has a fixed length of 8 octets. The value contains a 64-bit unsigned integer, representing the Label Database Version Number.

5. IANA Considerations

5.1. 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]	LABEL-DB-VERSION TLV	This document

5.2. PCECC-CAPABILITY sub-TLV

[I-D.zhao-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY sub-TLV and
[I-D.zhao-pce-pcep-extension-pce-controller-sr] extends this sub-TLV to add PCECC-SR-CAPABILITY.

Requests that IANA creates a registry to manage the value of the new PCECC-CAPABILITY sub-TLV's Flag field. IANA is requested to allocate a new bits in the PCECC-CAPABILITY sub-TLV Flag Field registry, as follows:

Bit	Description	Reference
TBD	I (INCLUDE-LABEL-DB-VERSION)	This document
TBD	D (DELTA-LABEL-SYNC-CAPABILITY)	This document

6. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440], [RFC8231] and [I-D.zhao-pce-pcep-extension-for-pce-controller] apply to PCEP protocol extensions defined in this document. In addition, requirements and considerations listed in this section apply.

6.1. Control of Function and Policy

6.2. Information and Data Models

6.3. Liveness Detection and Monitoring

6.4. Verify Correct Operations

6.5. Requirements On Other Protocols

6.6. Impact On Network Operations

7. Security Considerations

The security considerations listed in [RFC8231] and [I-D.zhao-pce-pcep-extension-for-pce-controller] apply to this document as well. Securing the PCEP session using Transport Layer Security (TLS) [RFC8253], as per the recommendations and best current practices in [RFC7525], is RECOMMENDED.

8. Acknowledgements

This document borrows some of the structure and text from [RFC8253], and would like to thanks the authors and contributors of the document.

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [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>>.
- [I-D.zhao-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-zhao-pce-pcep-extension-for-pce-controller-06 (work in progress), October 2017.

9.2. Informative References

- [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>>.
- [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>>.
- [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.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-01 (work in progress), October 2017.

Authors' Addresses

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

Email: udayasreereddy@gmail.com

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

PCE WG
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2019

Quan. Xiong
Fangwei. Hu
Ran. Chen
ZTE Corporation
October 22, 2018

PCE Multi-layer LSP Association
draft-xiong-pce-multilayer-lsp-association-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. [I-D.ietf-pce-association-group] proposed an association mechanism for a set of LSPs.

This document proposes a set of extensions to PCEP to associate a grouping of multi-layer LSPs. The extensions define a mechanism to create associations between upper-layer LSP and related lower-layer LSPs.

Status of This Memo

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

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

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

This Internet-Draft will expire on April 25, 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. Requirements Language	3
2.1. Terminology	3
3. Overview	3
3.1. Motivation	3
3.2. Operation Overview	4
4. Extensions to the PCEP	5
4.1. Association Type and Group	5
4.2. MULTI-LAYER-ASSOCIATION TLV	5
5. PCEP Procedure	6
5.1. Multi-Layer LSPs Associations Creation	6
5.2. Bandwidth Adjustment	6
5.3. TE Links Optimization	7
6. Security Considerations	7
7. IANA Considerations	7
7.1. Association Object Type	7
7.2. LAYER-ASSOCIATION TLV	7
8. Acknowledgements	8
9. References	8
9.1. Informative References	8
9.2. Normative References	8
Authors' Addresses	8

1. Introduction

[RFC5440] describes the Path Computation Element Protocol (PCEP) which is used between a Path Computation Element (PCE) and a Path Computation Client (PCC) (or other PCE) to enable computation of Multi-protocol Label Switching (MPLS) for Traffic Engineering Label Switched Path (TE LSP). [I-D.ietf-pce-association-group] proposed an association mechanism to create a grouping of LSPs in the context of a PCE.

This document proposes a set of extensions to PCEP to associate a grouping of multi-layer LSPs. The extensions define a mechanism to create associations between upper-layer LSP and related lower-layer LSPs.

2. 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 [RFC2119].

2.1. Terminology

The terminology is defined as [RFC5440], [I-D.ietf-pce-stateful-pce-app] and [I-D.ietf-pce-association-group].

3. Overview

3.1. Motivation

In GMPLS/MPLS networks, service provider network is divided into several service layers according to the requirements and customer network is the upper layer with the lower layers as the Forwarding Adjacency LSP (FA-LSP) as shown in Figure 1. The service connection is established with the set up of multi-layer LSPs.

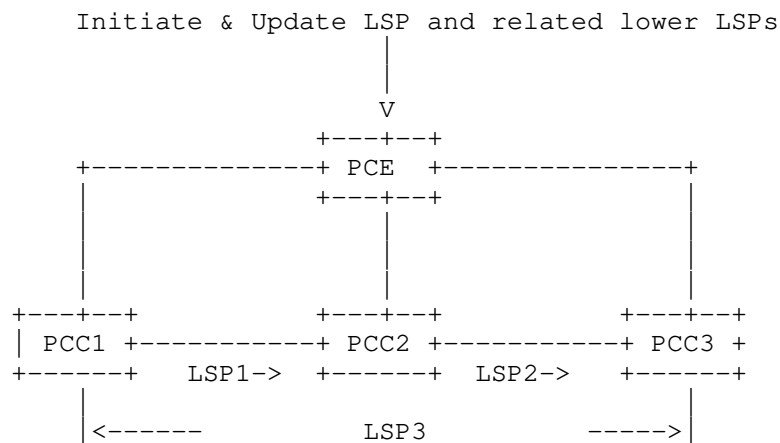


Figure 1 Usecase for multi-layer LSPs

As discussed in [I-D.ietf-pce-stateful-pce-app], it consists of a set of one or more TE LSPs in the lower layer which provides TE links to the upper layer in Multi-Layer Networks (MLN). The requirement is to control of the multi-layer LSPs and related TE links. The establishment or teardown of a lower layer LSP needs to take into consideration the state of existing LSPs or new LSP request in the upper layer.

As discussed in [I-D.ietf-pce-stateful-pce] , the stateful PCE MAY determine to optimize the link and path based on the lower layer of the LSP and its upper TE Link, and in the case of the failure of the lower level LSP, it MAY update the upper network LSP path according to the existing resources and the status of the LSP.

The stateful PCE provides the ability to update the LSP, in the process of bandwidth adjustment, it MAY be necessary to adjust the bandwidth of related lower layer LSPs, which provide the TE link for the upper layer LSP. The association of multi-layer LSPs can reduce the repeated operations and optimize the information interaction between PCC and PCE.

In overlayer multi-domain scenario, the lower-layer LSPs in each domain may be initiated by respective domain's PCE and stitched together to an association group with an end-to-end LSP as its upper-layer LSP.

In these cases, it is necessary to add multi-layer LSPs to an association group.

3.2. Operation Overview

[I-D.ietf-pce-association-group] introduces a generic mechanism to create a grouping of LSPs. This grouping can then be used to define associations between sets of LSPs or between a set of LSPs and a set of attributes.

In order to solve the problem of multi-layer LSP control in PCE network, this document proposes the association of the multi-layer LSPs. The upper LSP is associated with its related lower LSPs by adding them to a multi-layer association group.

One new optional Association Object type is defined carried in the Association object defined in [I-D.ietf-pce-association-group]. This document proposes a new association type called "Layer Association Type" and related TLV called "LAYER-ASSOCIATION TLV".

As defined in [I-D.ietf-pce-association-group], multi-layer LSPs associations could be created dynamically or configured by the operator when operator-configured association is needed.

The handling and policy of multi-layer LSPs Association is similar to the generic association and some processing rules as shown in session 4.2.

4. Extensions to the PCEP

4.1. Association Type and Group

[I-D.ietf-pce-association-group] introduces the ASSOCIATION object and this document proposes a new Association type for multi-layer LSPs association to associate multi-layer LSPs into one group for further operation. An association ID will be used to identify the group and a new Association Type is defined in this document, based on the generic Association object :

Association type = TBD1 ("Multi-Layer Association Type") for Multi-Layer Association Group (MLAG)

MLAG may carry optional TLVs including but not limited to :

MULTI-LAYER-ASSOCIATION-TLV: Used to identify the upper-layer LSP and lower-layer LSP in multi-layer information, described in Section 4.2.

As [I-D.ietf-pce-association-group] specified, the capability advertisement of the association types supported by a PCEP speaker is performed by defining a ASSOC-Type-List TLV to be carried within an OPEN object. The association type which defined in this document should be added in the list and be advertised between the PCEP speakers before the multi-layer association.

This Association-Type is operator-configured and created by the operator manually on the PCEP peers. The LSP belonging to this associations is conveyed via PCEP messages to the PCEP peer. Operator-configured Association Range SHOULD NOT be set for this association-type, and MUST be ignored, so that the full range of association identifier can be utilized.

4.2. MULTI-LAYER-ASSOCIATION TLV

This document proposes LAYER-ASSOCIATION TLV for the association of multi-layer LSPs. The TLV is optional. The format of the new Association TLV is shown in Figure 4:

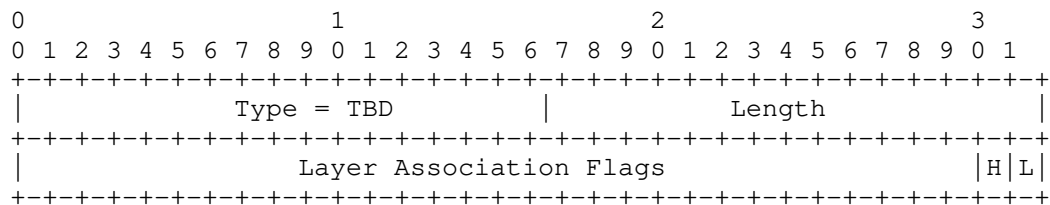


Figure 4: The LAYER-ASSOCIATION TLV format

The type of the TLV is [TBD] which indicates the LAYER ASSOCIATION TLV. The fields in the format are:

Length:16bits,the length of the TLV.

Layer Association Flags-H:1bit, indicates LSP of the upper layer when it is set.

Layer Association Flags-L:1bit, indicates LSP of the lower layer when it is set.

5. PCEP Procedure

Once a group of multilayer LSPs is created, the upper layer LSP is associated with its related lower layer LSPs. Association objects can be carried in PCReq, PCRpt, PCUpd, or PCInit messages.

5.1. Multi-Layer LSPs Associations Creation

As defined in [I-D.ietf-pce-association-group], association groups can be created by both PCC and PCE.

In stateless PCE, the association object with "Layer Association type" is carried in PCReq message from PCC to PCE, indicating that the LSP joins one existing multi-layer LSPs association group or create a new one. If the LSP is belong to upper layer then set the "H" bit in "LAYER-ASSOCIATION TLV", otherwise set the "L" bit when it is lower layer LSP.

In stateful PCE, PCE MAY create a new association group or associate a LSP to an existing association group carried in PCInit message after the LSP delegation from PCC to the PCE as discussed in [I-D.ietf-pce-pce-initiated-lsp]. In state synchronization process between PCC and PCE, PCC also need to report the existing multi-layer LSPs association groups to PCE. If the association group changes, PCC needs to report the relevant group changes to PCE through the PCRpt message.

5.2. Bandwidth Adjustment

The stateful PCE provides the ability to update the LSP, in the process of bandwidth adjustment, for example, enlarge the bandwidth of the upper layer LSP, it MUST be necessary to adjust the bandwidth of related lower layer LSPs, which provide the TE link for it.

Once the multi-layer LSPs associated in a group, the PCE MAY send the PCUpd message to the PCC with the association object to adjust the upper layer LSP. Once receiving the request, PCC will search the

relevant lower layer LSPs and adjust their bandwidth before the adjustment of the upper layer LSP.

5.3. TE Links Optimization

The stateful PCE MAY determine to optimize the link and path based on the lower layer of the LSP and its upper TE Link, and in the case of the failure of the lower level LSP, it MAY update the upper network LSP path and re-optimize resource usage across multi-layers.

When removing the upper layer LSP, PCC or PCE MAY release each of lower layer LSPs which associated in a group and re-use the resources for other upper layer LSP according to the existing resources and the status of the LSP.

6. Security Considerations

TBD

7. IANA Considerations

7.1. Association Object Type

This document defines a new association type in Association object which originally defined in [I-D.ietf-pce-association-group]. IANA is requested to make allocations from the registry, as follows:

Value	Name	Reference
TBD	Layer Association Type	[this document]

Table 1

7.2. LAYER-ASSOCIATION TLV

This document defines the following TLV in Association object which originally defined in [I-D.ietf-pce-association-group]. IANA is requested to make allocations from the registry, as follows:

Value	Name	Reference
TBD	LAYER-ASSOCIATION TLV	[this document]

Table 2

8. Acknowledgements

TBD.

9. References

9.1. Informative References

- [I-D.ietf-pce-stateful-pce-app]
Zhang, X. and I. Minei, "Applicability of a Stateful Path Computation Element (PCE)", draft-ietf-pce-stateful-pce-app-08 (work in progress), October 2016.

9.2. 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-06 (work in progress), June 2018.
- [I-D.ietf-pce-pce-initiated-lsp]
Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", draft-ietf-pce-pce-initiated-lsp-11 (work in progress), October 2017.
- [I-D.ietf-pce-stateful-pce]
Crabbe, E., Minei, I., Medved, J., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-21 (work in progress), June 2017.
- [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>>.

Authors' Addresses

Quan Xiong
ZTE Corporation
No.6 Huashi Park Rd
Wuhan, Hubei 430223
China

Phone: +86 27 83531060
Email: xiong.quan@zte.com.cn

Fangwei Hu
ZTE Corporation
No.889 Bibo Rd
Shanghai 201203
China

Phone: +86 21 68896273
Email: hu.fangwei@zte.com.cn

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

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 20, 2018

Q. Zhao
Z. Li
D. Dhody
S. Karunanithi
Huawei Technologies
A. Farrel
Juniper Networks, Inc
C. Zhou
Cisco Systems
June 18, 2018

PCEP Procedures and Protocol Extensions for Using PCE as a Central
Controller (PCECC) of LSPs
draft-zhao-pce-pcep-extension-for-pce-controller-08

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 December 20, 2018.

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	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	10
5.4.2.1. Label Download	10
5.4.2.2. Label Cleanup	11
5.4.3. PCE Initiated PCECC LSP	12
5.4.4. PCECC LSP Update	14
5.4.5. Re Delegation and Cleanup	16
5.4.6. Synchronization of Central Controllers Instructions .	16

5.4.7.	PCECC LSP State Report	16
6.	PCEP messages	16
6.1.	The PCInitiate message	17
6.2.	The PCRpt message	18
7.	PCEP Objects	19
7.1.	OPEN Object	19
7.1.1.	PCECC Capability sub-TLV	19
7.2.	PATH-SETUP-TYPE TLV	20
7.3.	CCI Object	20
7.3.1.	Address TLVs	21
8.	Security Considerations	23
8.1.	Malicious PCE	23
9.	Manageability Considerations	23
9.1.	Control of Function and Policy	23
9.2.	Information and Data Models	23
9.3.	Liveness Detection and Monitoring	23
9.4.	Verify Correct Operations	23
9.5.	Requirements On Other Protocols	23
9.6.	Impact On Network Operations	24
10.	IANA Considerations	24
10.1.	PCEP TLV Type Indicators	24
10.2.	New Path Setup Type Registry	24
10.3.	PCEP Object	24
10.4.	CCI Object Flag Field	24
10.5.	PCEP-Error Object	25
11.	Acknowledgments	25
12.	References	25
12.1.	Normative References	25
12.2.	Informative References	26
Appendix A.	Contributor Addresses	29
	Authors' Addresses	29

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

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 specifies 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) [I-D.ietf-pce-lsp-setup-type] 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 [I-D.ietf-pce-lsp-setup-type] 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.

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.

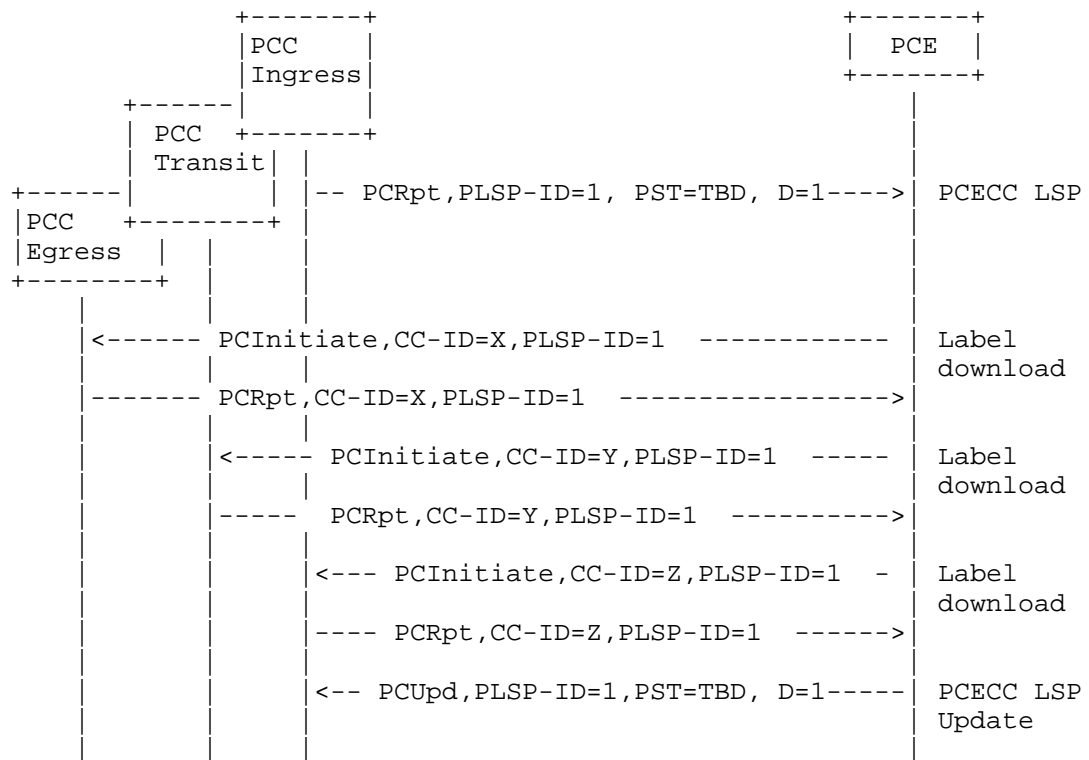


Figure 2: Basic PCECC LSP setup

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.

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.

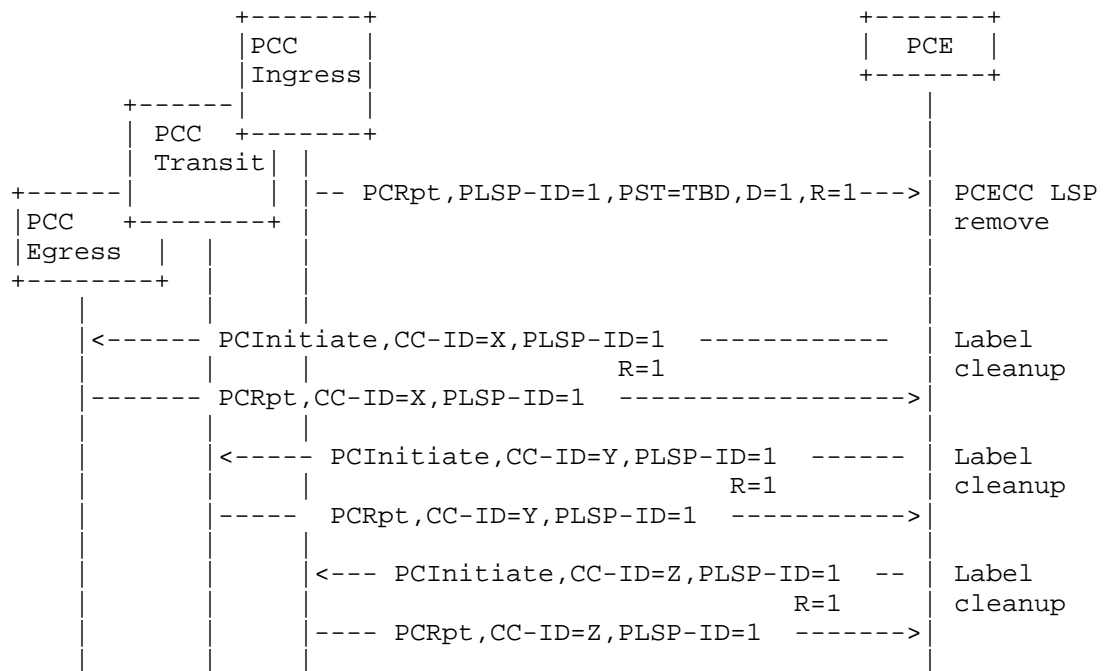
New PCEP object for central control instructions (CCI) is defined in Section 7.3.

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.

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.

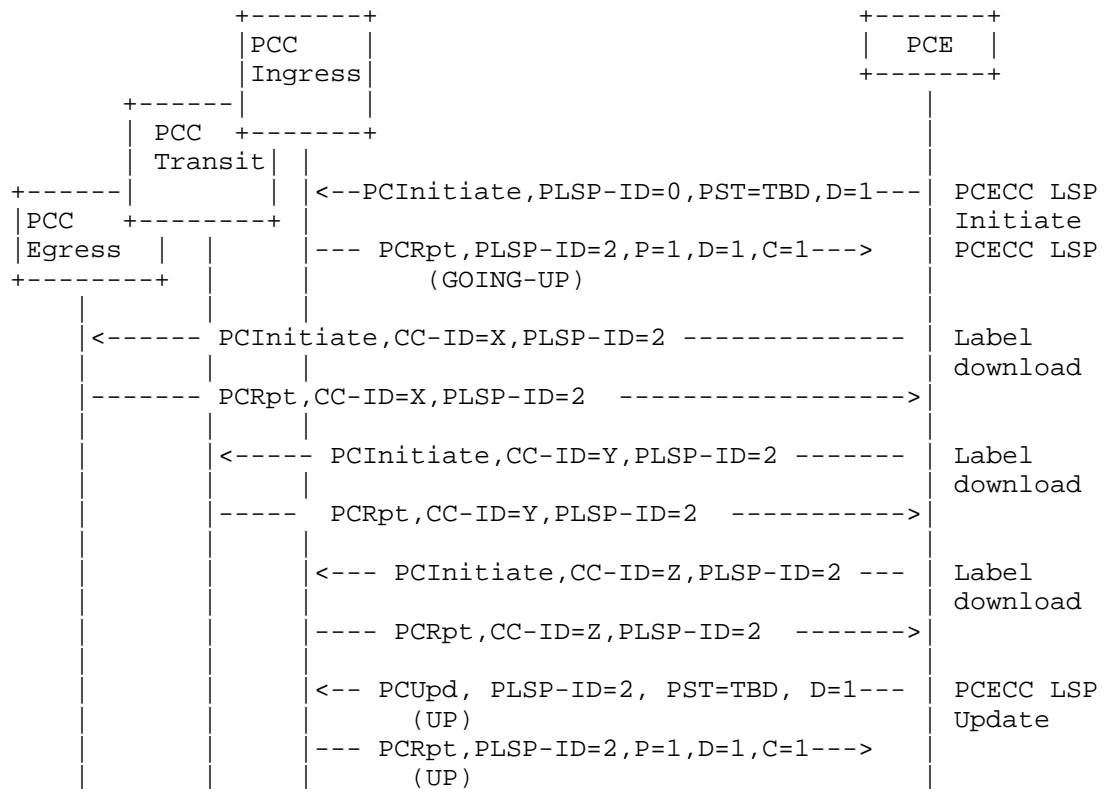
The Ingress PCC MUST also set D (Delegate) flag (see [RFC8231]) and C (Create) flag (see [RFC8281]) in LSP object of PCRpt message. The PCC responds with first PCRpt message with the status as "GOING-UP" and assigned PLSP-ID.

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 -

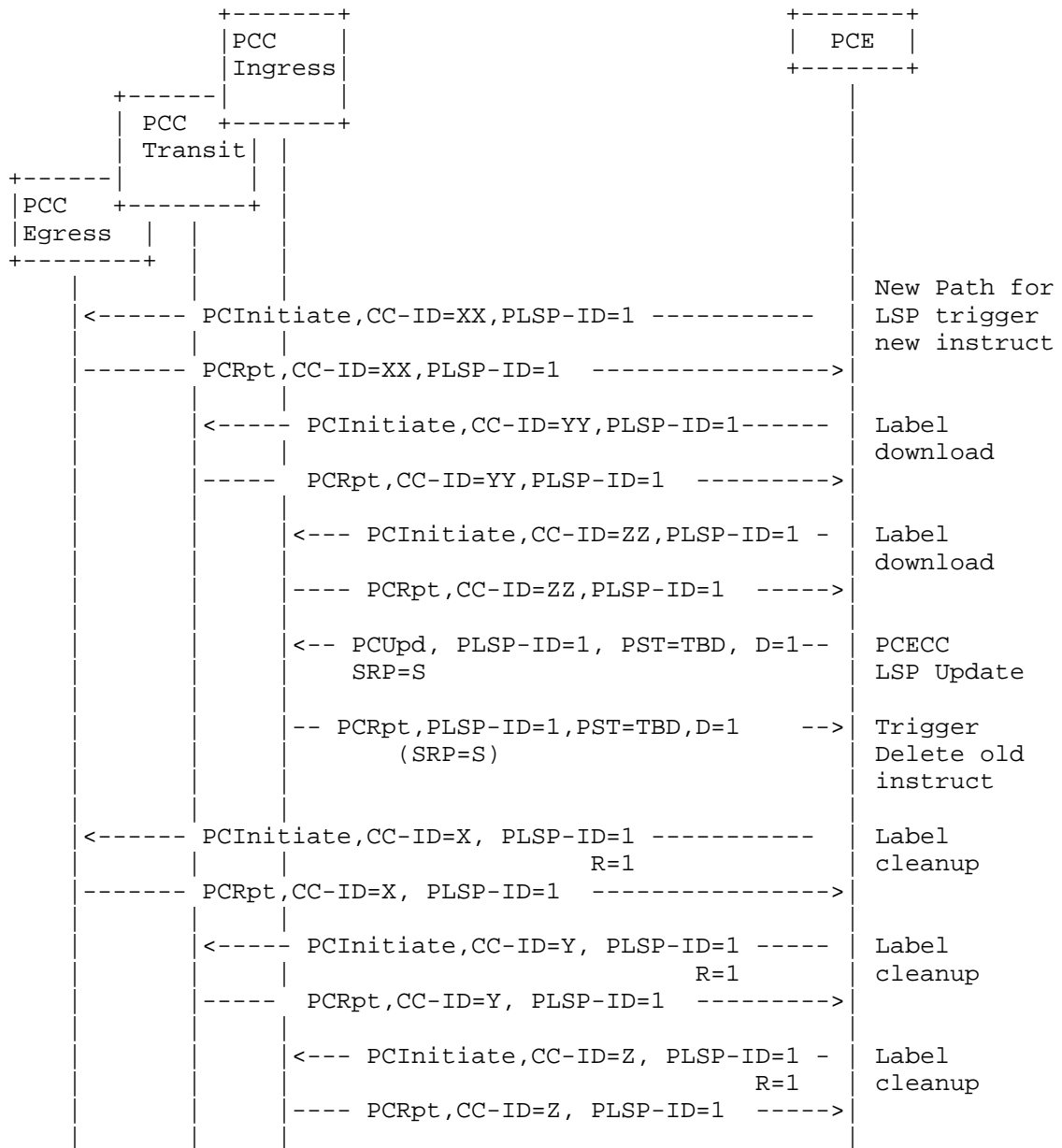


Once the label operations are completed, the PCE SHOULD send the PCUpd message to the Ingress PCC. The PCUpd message is as per [RFC8231].

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.

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.

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].

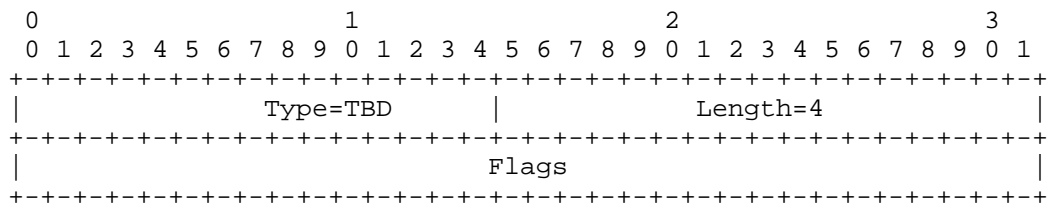
7.1. OPEN Object

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 [I-D.ietf-pce-lsp-setup-type]; this document defines a new PST value:

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

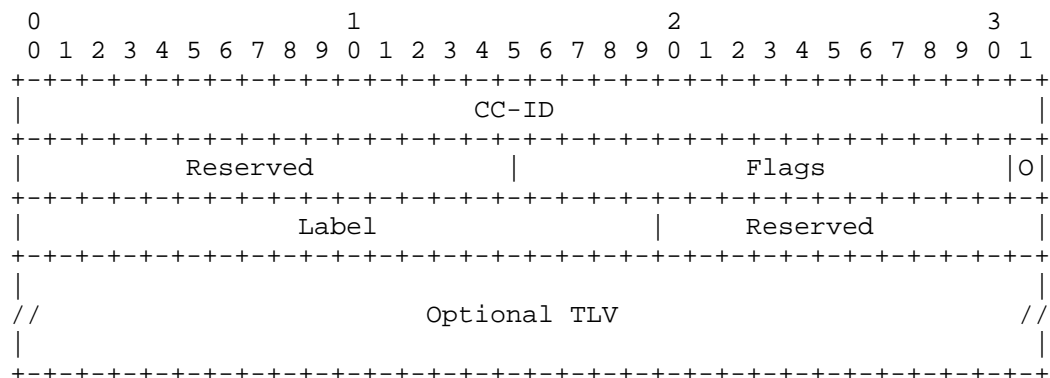
On a PCRpt/PCUpd/PCInitiate message, the PST=TBD in PATH-SETUP-TYPE TLV in SRP object indicates that this LSP was setup via a PCECC based mechanism.

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).

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.

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>>.

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., Ke, Z., Fang, L., Zhou, C., Communications, T., and A. Rachitskiy, "The Use Cases for Using PCE as the Central Controller(PCECC) of LSPs", draft-ietf-teas-pcecc-use-cases-01 (work in progress), May 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.
- [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-07 (work in progress), March 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-02 (work
in progress), March 2018.

Appendix A. Contributor Addresses

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

EMail: udayasreereddy@gmail.com

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

EMail: mahendrasingh@huawei.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

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

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

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.

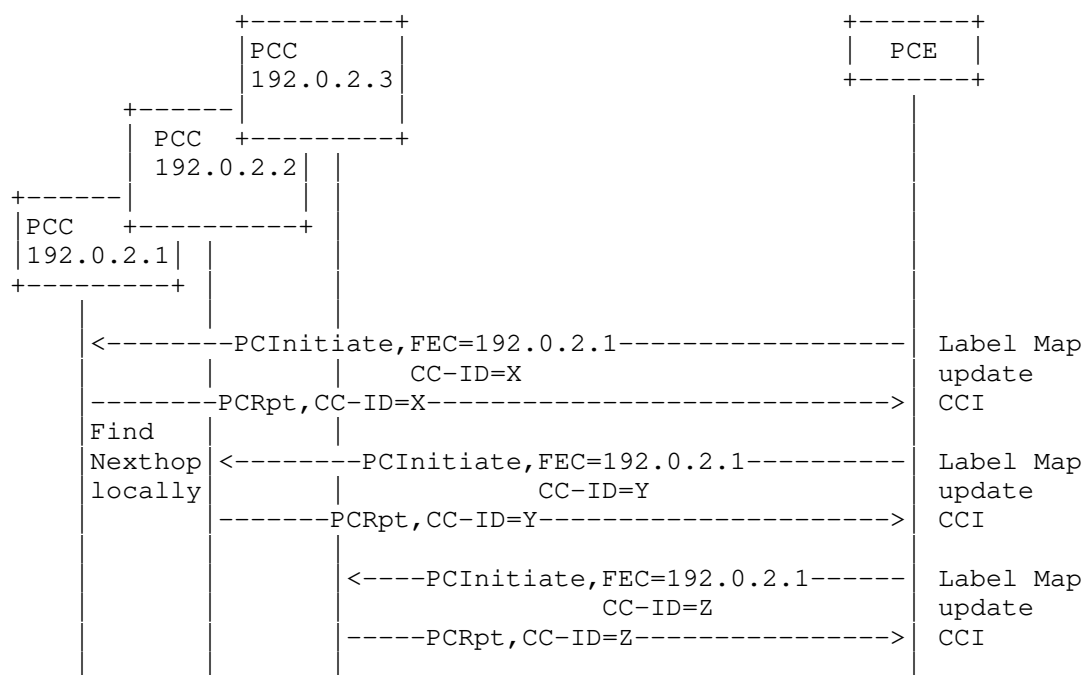
5.5.1.1. PCECC SR Node/Prefix SID allocation

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.

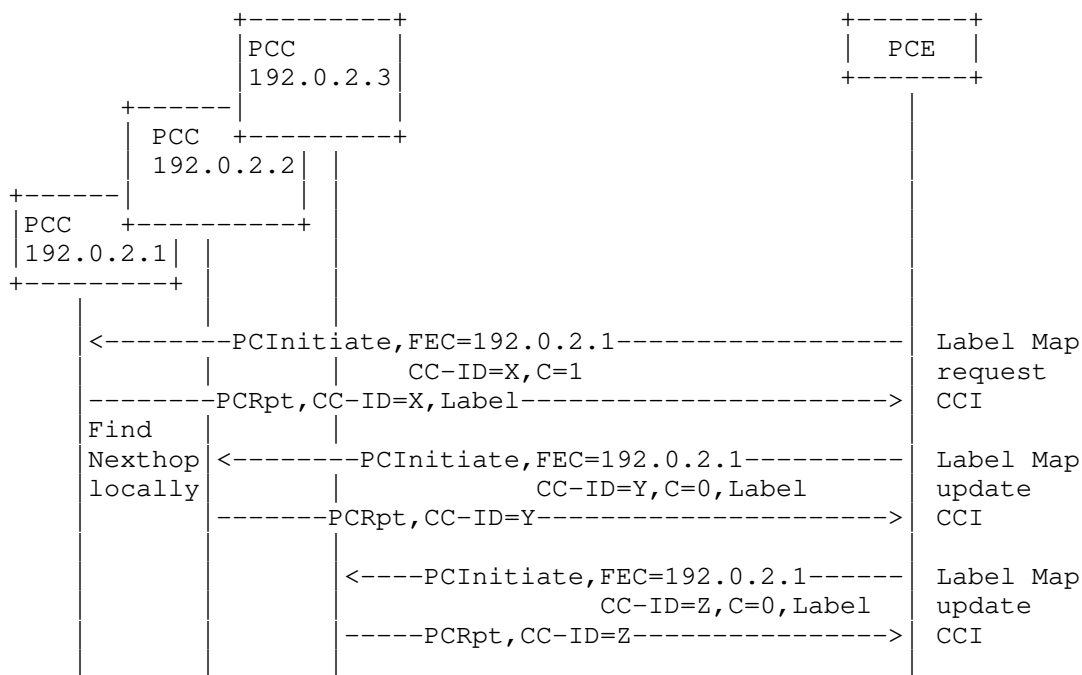


The forwarding behavior and the end result is similar to IGP based "Node-SID" in SR. Thus, from anywhere in the domain, it enforces the ECMP-aware shortest-path forwarding of the packet towards the related node as per [RFC8402].

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.

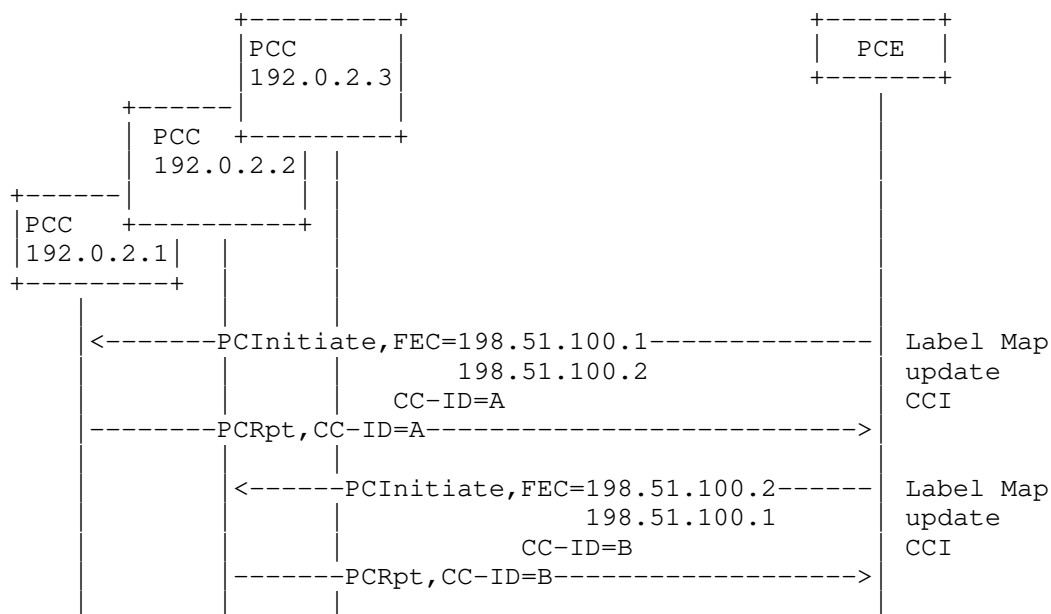
In the case where the label/SID allocation is made by the PCC itself (see Section 5.5.1.6), the PCE could request an allocation to be made by the PCC, and where the PCC would send a PCRpt with the allocated label/SID encoded in the CC-ID object as shown in Figure 2.



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.



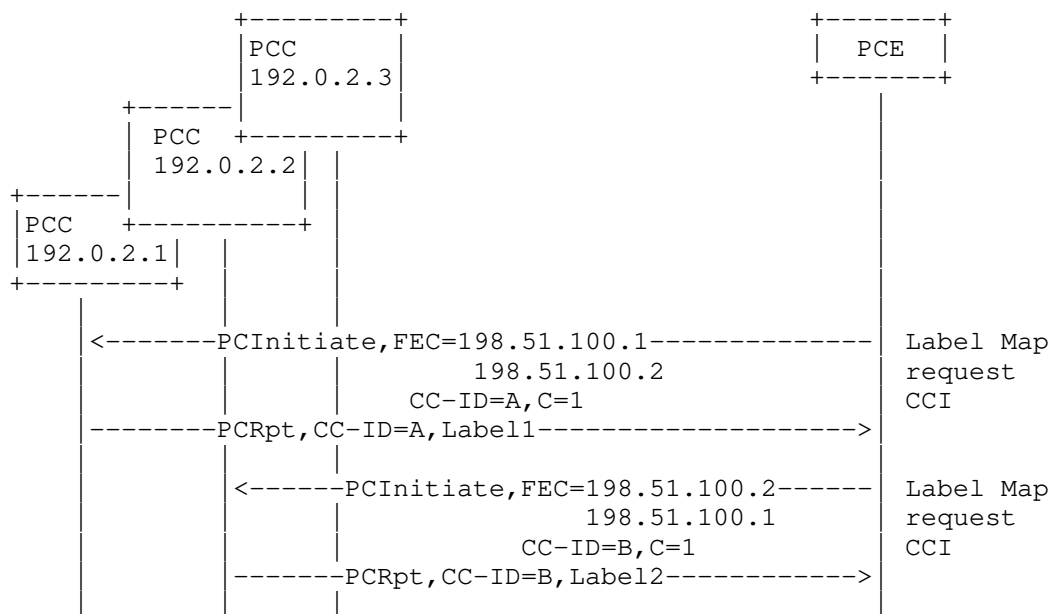
The forwarding behavior and the end result is similar to IGP based "Adj-SID" in SR.

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).

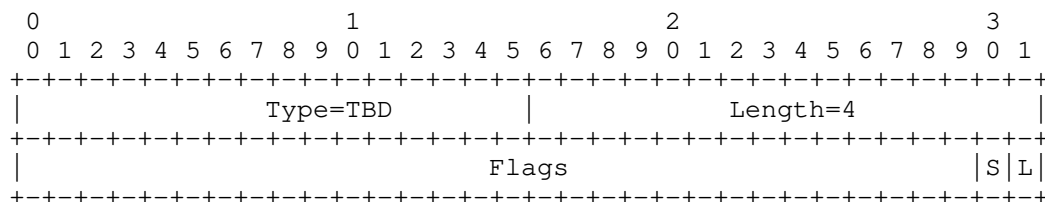
7. PCEP Objects

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.

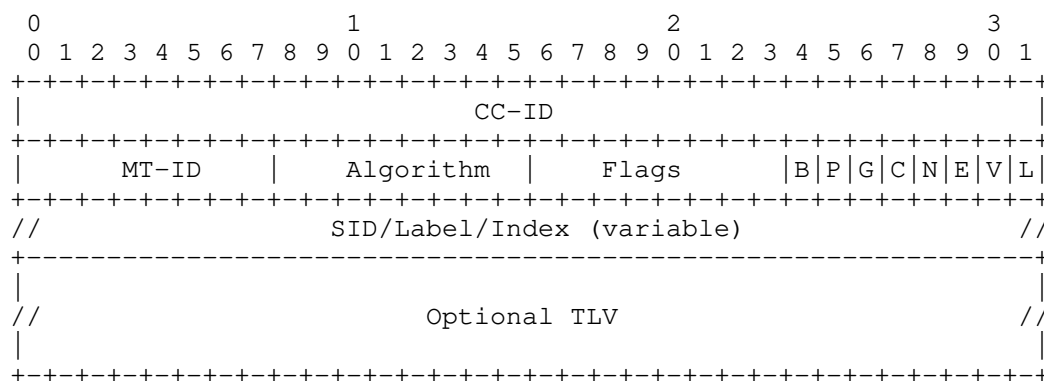
7.2. SR-TE Path Setup

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.

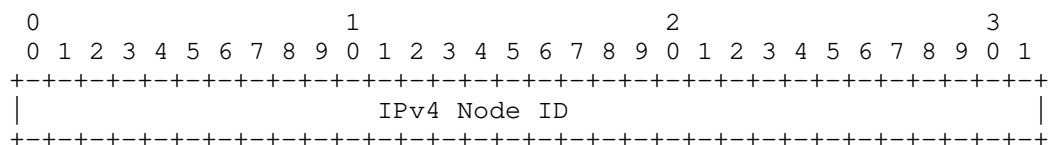
A 24-bit label where the 20 rightmost bits are used for encoding the label value.

7.4. FEC Object

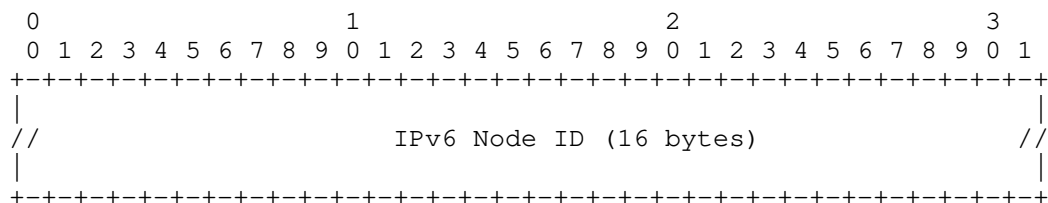
The FEC Object is used to specify the FEC information and MAY be carried within PCInitiate or PCRpt message.

FEC Object-Class is TBD3.

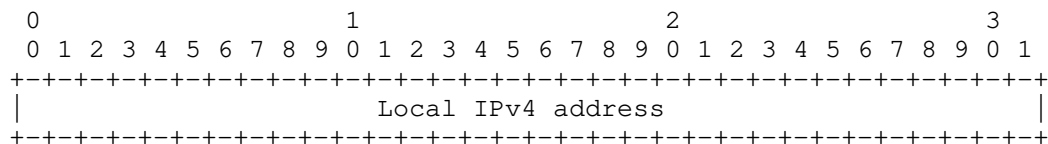
FEC Object-Type is 1 'IPv4 Node ID'.



FEC Object-Type is 2 'IPv6 Node ID'.



FEC Object-Type is 3 'IPv4 Adjacency'.



```

|----- Remote IPv4 address -----|
+-----+

```

FEC Object-Type is 4 'IPv6 Adjacency'.

```

      0               1               2               3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+
|
//               Local IPv6 address (16 bytes)               //
|
+-----+
|
//               Remote IPv6 address (16 bytes)               //
|
+-----+

```

FEC Object-Type is 5 'Unnumbered Adjacency with IPv4 NodeIDs'.

```

      0               1               2               3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+
|               Local Node-ID               |
+-----+
|               Local Interface ID           |
+-----+
|               Remote Node-ID              |
+-----+
|               Remote Interface ID          |
+-----+

```

FEC Object-Type is 6 'Linklocal IPv6 Adjacency'.

```

      0               1               2               3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+
//               Local IPv6 address (16 octets)               //
+-----+
|               Local Interface ID           |
+-----+
//               Remote IPv6 address (16 octets)               //
+-----+
|               Remote Interface ID          |
+-----+

```

The FEC objects are as follows:

IPv4 Node ID: where IPv4 Node ID is specified as an IPv4 address of the Node. FEC Object-type is 1, and the Object-Length is 4 in this case.

IPv6 Node ID: where IPv6 Node ID is specified as an IPv6 address of the Node. FEC Object-type is 2, and the Object-Length is 16 in this case.

IPv4 Adjacency: where Local and Remote IPv4 address is specified as pair of IPv4 addresses of the adjacency. FEC Object-type is 3, and the Object-Length is 8 in this case.

IPv6 Adjacency: where Local and Remote IPv6 address is specified as pair of IPv6 addresses of the adjacency. FEC Object-type is 4, and the Object-Length is 32 in this case.

Unnumbered Adjacency with IPv4 NodeID: where a pair of Node ID / Interface ID tuple is used. FEC Object-type is 5, and the Object-Length is 16 in this case.

Linklocal IPv6 Adjacency: where a pair of (global IPv6 address, interface ID) tuple is used. FEC object-type is 6, and the Object-Length is 40 in this case.

8. Implementation Status

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

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

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

It is up to the individual working groups to use this information as they see fit".

8.1. Huawei's Proof of Concept based on ONOS

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

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

9. Security Considerations

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

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

10. Manageability Considerations

10.1. Control of Function and Policy

A PCE or PCC implementation SHOULD allow to configure to enable/disable PCECC SR capability as a global configuration. The implementation SHOULD also allow setting the local IP address used by the PCEP session.

10.2. Information and Data Models

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

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

10.3. Liveness Detection and Monitoring

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

10.4. Verify Correct Operations

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

10.5. Requirements On Other Protocols

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

10.6. Impact On Network Operations

PCEP implementation SHOULD allow a limit to be placed on the rate of PCLabelUpd messages sent by PCE and processed by PCC. It SHOULD also allow sending a notification when a rate threshold is reached.

11. IANA Considerations

11.1. PCECC-CAPABILITY sub-TLV

[I-D.ietf-pce-pcep-extension-for-pce-controller] defines the PCECC-CAPABILITY sub-TLV and requests that IANA to create a new sub-registry to manage the value of the PCECC-CAPABILITY sub-TLV's Flag field.

IANA is requested to allocate a new bit in the PCECC-CAPABILITY sub-TLV Flag Field sub-registry, as follows:

Bit	Description	Reference
TBD1	SR	This document

11.2. PCEP Object

IANA is requested to allocate new code-points for the new FEC object and a new Object-Type for CCI object in "PCEP Objects" sub-registry as follows:

Object-Class Value	Name	Object-Type	Reference
TBD3	FEC	1: IPv4 Node ID	This document
		2: IPv6 Node ID	This document
		3: IPv4 Adjacency	This document
		4: IPv6 Adjacency	This document
		5: Unnumbered Adjacency with IPv4 NodeID	This document
		6: Linklocal IPv6 Adjacency	This document
TBD	CCI	TBD6: SR-MPLS	This document

11.3. PCEP-Error Object

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

Error-Type -----	Meaning -----	
6	Mandatory Object missing.	
19	Error-value = TBD5 :	FEC object missing
	Invalid operation.	
	Error-value = TBD4 :	SR capability was not advertised

11.4. CCI Object Flag Field for SR

IANA is requested to create a new sub-registry to manage the Flag field of the CCI Object-Type=TBD6 for SR called "CCI Object Flag Field for SR". New values are to be assigned by Standards Action [RFC8126]. Each bit should be tracked with the following qualities:

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

Following bits are defined for the CCI Object flag field for SR in this document as follows:

Bit	Description	Reference
0-7	Unassigned	This document
8	B-Bit - Backup	This document
9	P-Bit - Persistent	This document
10	G-Bit - Group	This document
11	C-Bit - PCC Allocation	This document
12	N-Bit - No-PHP	This document
13	E-Bit - Explicit-Null	This document
14	V-Bit - Value/Index	This document
15	L-Bit - Local/Global	This document

12. Acknowledgments

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

13. References

13.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <<https://www.rfc-editor.org/info/rfc3630>>.
- [RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <<https://www.rfc-editor.org/info/rfc4203>>.
- [RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, DOI 10.17487/RFC4915, June 2007, <<https://www.rfc-editor.org/info/rfc4915>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <<https://www.rfc-editor.org/info/rfc5305>>.
- [RFC5307] Kompella, K., Ed. and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, DOI 10.17487/RFC5307, October 2008, <<https://www.rfc-editor.org/info/rfc5307>>.

- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <<https://www.rfc-editor.org/info/rfc7752>>.
- [RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <<https://www.rfc-editor.org/info/rfc7942>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, DOI 10.17487/RFC8231, September 2017, <<https://www.rfc-editor.org/info/rfc8231>>.
- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <<https://www.rfc-editor.org/info/rfc8281>>.
- [RFC8408] Sivabalan, S., Tantsura, J., Minei, I., Varga, R., and J. Hardwick, "Conveying Path Setup Type in PCE Communication Protocol (PCEP) Messages", RFC 8408, DOI 10.17487/RFC8408, July 2018, <<https://www.rfc-editor.org/info/rfc8408>>.
- [RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <<https://www.rfc-editor.org/info/rfc8664>>.

[I-D.ietf-pce-pcep-extension-for-pce-controller]
Li, Z., Peng, S., Negi, M., Zhao, Q., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-ietf-pce-pcep-extension-for-pce-controller-08 (work in progress), November 2020.

13.2. Informative References

- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <<https://www.rfc-editor.org/info/rfc3031>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <<https://www.rfc-editor.org/info/rfc7399>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<https://www.rfc-editor.org/info/rfc7420>>.
- [RFC7491] King, D. and A. Farrel, "A PCE-Based Architecture for Application-Based Network Operations", RFC 7491, DOI 10.17487/RFC7491, March 2015, <<https://www.rfc-editor.org/info/rfc7491>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<https://www.rfc-editor.org/info/rfc7525>>.

- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [RFC8232] Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", RFC 8232, DOI 10.17487/RFC8232, September 2017, <<https://www.rfc-editor.org/info/rfc8232>>.
- [RFC8283] Farrel, A., Ed., Zhao, Q., Ed., Li, Z., and C. Zhou, "An Architecture for Use of PCE and the PCE Communication Protocol (PCEP) in a Network with Central Control", RFC 8283, DOI 10.17487/RFC8283, December 2017, <<https://www.rfc-editor.org/info/rfc8283>>.
- [RFC8402] Filtsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <<https://www.rfc-editor.org/info/rfc8402>>.
- [RFC8660] Bashandy, A., Ed., Filtsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/info/rfc8660>>.
- [RFC8665] Psenak, P., Ed., Previdi, S., Ed., Filtsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <<https://www.rfc-editor.org/info/rfc8665>>.
- [RFC8667] Previdi, S., Ed., Ginsberg, L., Ed., Filtsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <<https://www.rfc-editor.org/info/rfc8667>>.
- [I-D.ietf-teas-pcecc-use-cases]
Li, Z., Khasanov, B., Dhody, D., Zhao, Q., Ke, Z., Fang, L., Zhou, C., Communications, T., Rachitskiy, A., and A. Gulida, "The Use Cases for Path Computation Element (PCE) as a Central Controller (PCECC).", draft-ietf-teas-pcecc-use-cases-06 (work in progress), September 2020.

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

Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-15 (work in progress), October 2020.

[I-D.ietf-pce-binding-label-sid]

Sivabalan, S., Filsfils, C., Tantsura, J., Hardwick, J., Previdi, S., and C. Li, "Carrying Binding Label/Segment-ID in PCE-based Networks.", draft-ietf-pce-binding-label-sid-05 (work in progress), October 2020.

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

Litkowski, S., Sivabalan, S., Li, C., and H. Zheng, "Inter Stateful Path Computation Element (PCE) Communication Procedures.", draft-litkowski-pce-state-sync-09 (work in progress), November 2020.

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

Dhody, D., Peng, S., Lee, Y., Ceccarelli, D., Wang, A., and G. Mishra, "PCEP extensions for Distribution of Link-State and TE Information", draft-dhodylee-pce-pcep-ls-19 (work in progress), November 2020.

[I-D.dhody-pce-pcep-extension-pce-controller-srv6]

Li, Z., Peng, S., Geng, X., and M. Negi, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) for SRv6", draft-dhody-pce-pcep-extension-pce-controller-srv6-05 (work in progress), November 2020.

Appendix A. Contributor Addresses

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

EMail: dhruv.ietf@gmail.com

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

EMail: satishk@huawei.com

Adrian Farrel
Juniper Networks, Inc
UK

EMail: adrian@olddog.co.uk

Xuesong Geng
Huawei Technologies
China

Email: gengxuesong@huawei.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

Katherine Zhao
Huawei Technologies
2330 Central Expressway
Santa Clara, CA 95050
USA

EMail: katherine.zhao@huawei.com

Boris Zhang
Telus Ltd.
Toronto
Canada

EMail: boris.zhang@telus.com

Alex Tokar
Cisco Systems
Slovak Republic

EMail: atokar@cisco.com

Authors' Addresses

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

EMail: lizhenbin@huawei.com

Shuping Peng
Huawei Technologies
Huawei Bld., No.156 Beiqing Rd.
Beijing 100095
China

EMail: pengshuping@huawei.com

Mahendra Singh Negi
RtBrick Inc
N-17L, 18th Cross Rd, HSR Layout
Bangalore, Karnataka 560102
India

EMail: mahend.ietf@gmail.com

Quintin Zhao
Etheric Networks
1009 S CLAREMONT ST
SAN MATEO, CA 94402
USA

EMail: qzhao@ethericnetworks.com

Chao Zhou
HPE

EMail: chaozhou_us@yahoo.com