Path Computation Element Communication Protocol (PCEP) Extensions for Using the PCE as a Central Controller (PCECC) of point-to-multipoint (P2MP) LSPs
draft-dhody-pce-pcep-extension-pce-controller-p2mp-09

Abstract

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

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

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

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

Status of This Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.
Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 11 January 2023.

Copyright Notice

Copyright (c) 2022 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

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

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

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

In early PCE implementations, where the PCE was used to derive paths for MPLS Label Switched Paths (LSPs), paths were requested by network elements (known as Path Computation Clients (PCCs)), and the results of the path computations were supplied to network elements using the Path Computation Element Communication Protocol (PCEP) [RFC5440]. This protocol was later extended to allow a PCE to send unsolicited requests to the network for LSP establishment [RFC8281]. [RFC8283] introduces the architecture for PCE as a central controller as an extension of the architecture described in [RFC4655] and assumes the continued use of PCEP as the protocol used between PCE and PCC. [RFC8283] further examines the motivations and applicability for PCEP as a Southbound Interface (SBI), and introduces the implications for the protocol.

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

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

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

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

2. Terminology

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

2.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] when, and only when, they appear in all capitals, as shown here.
3. Basic PCECC Mode

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

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

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

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

4.1. Stateful PCE Model

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

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

4.2. PCECC Capability Advertisement

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

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

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

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

* terminate the session.

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

4.3. LSP Operations

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

4.3.1. PCE-Initiated PCECC LSP

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

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

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

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

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

### 4.3.2. PCC-Initiated PCECC LSP

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

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

### 4.3.3. Central Control Instructions

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

#### 4.3.3.1. Label Download CCI

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

PCInitiate message. As per [RFC9050], there are 2 instances of CCI object in the PCInitiate message in a transit node for the P2P LSP. For PCECC-P2MP operations, multiple instances of CCI object for out-labels is allowed at the branch node. Similarly to acknowledge the central controller instructions, the PCRpt message allows multiple instances of CCI object for PCECC-P2MP operations.

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

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

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

```
+----------+
|    R1    | Root node of the P2MP TE LSP
+----------+

Transit Node   |    R2    |
branch         +----------+

  *    |   *    *    *
  9001* |   *    +9002
        |  *    *

  +-----------------+  *  +-----------------+
  |    R4    |  *  |    R5    | Transit Nodes
  +-----------------+  *  +-----------------+

  *  |  *  |  *    *
  9003* |  *    +9004
        |  *    *

  +-----------------+  *  +-----------------+
  |    R3    |  *  |    R6    | Leaf Node
  +-----------------+  *  +-----------------+

  9005|

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

4.3.3.2. Label Cleanup CCI

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

4.3.4. PCECC LSP Update

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

4.3.5. Re-delegation and Cleanup

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

4.3.6. Synchronization of Central Controllers Instructions

The procedure and instructions are as per [RFC9050].

4.3.7. PCECC LSP State Report

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

4.3.8. PCC-Based Allocations

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

5.  PCEP Messages

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

6.  PCEP Objects

6.1.  OPEN Object

6.1.1.  PCECC Capability sub-TLV

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

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

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

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

6.2.  PATH-SETUP-TYPE TLV

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

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

If a node (PCC) receives a PCInitiate message with more than one CCI with 0-bit set for the outgoing label and the node does not support the P2MP branch/replication capability, it MUST respond with PCErr message with Error-Type=2 (Capability not supported) (defined in [RFC5440]).

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

7. Security Considerations

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

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

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

8.1. Control of Function and Policy

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

8.2. Information and Data Models

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

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

8.3. Liveness Detection and Monitoring

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

8.4. Verify Correct Operations

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

8.5. Requirements On Other Protocols

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

8.6. Impact On Network Operations

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

9. IANA Considerations

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

<table>
<thead>
<tr>
<th>Bit</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD1</td>
<td>P2MP</td>
<td>This document</td>
</tr>
</tbody>
</table>

Table 1

9.2. PCEP-Error Object

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

<table>
<thead>
<tr>
<th>Error-Type</th>
<th>Meaning</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>Invalid operation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Error-value = TBD2: P2MP</td>
<td>This</td>
</tr>
<tr>
<td></td>
<td>capability was not advertised</td>
<td>document</td>
</tr>
</tbody>
</table>

Table 2

10. References

10.1. Normative References


10.2. Informative References


Appendix A. Contributor Addresses

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

EMail: dhruv.ietf@gmail.com

Udayasree Palle

EMail: udayasreereddy@gmail.com

Authors' Addresses

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

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

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