Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE)
draft-ietf-pce-hierarchy-extensions-11

Abstract

The Hierarchical Path Computation Element (H-PCE) architecture is defined in RFC 6805. It provides a mechanism to derive an optimum end-to-end path in a multi-domain environment by using a hierarchical relationship between domains to select the optimum sequence of domains and optimum paths across those domains.

This document defines extensions to the Path Computation Element Protocol (PCEP) to support Hierarchical PCE procedures.

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 2, 2019.

Copyright Notice

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

1. Introduction ........................................... 3
   1.1. Scope .............................................. 4
   1.2. Terminology ....................................... 5
   1.3. Requirements Language ............................... 5
2. Requirements for H-PCE .................................. 5
   2.1. Path Computation Request ............................ 6
      2.1.1. Qualification of PCEP Requests ................. 6
      2.1.2. Multi-domain Objective Functions ............... 6
      2.1.3. Multi-domain Metrics ............................ 7
   2.2. Parent PCE Capability Advertisement .................. 7
   2.3. PCE Domain Identification ........................... 7
   2.4. Domain Diversity ................................... 7
3. PCEP Extensions ......................................... 8
   3.1. Applicability to PCC-PCE Communications .............. 8
   3.2. OPEN Object ....................................... 8
      3.2.1. H-PCE Capability TLV ........................... 8
      3.2.1.1. Backwards Compatibility ..................... 9
      3.2.2. Domain-ID TLV ................................ 10
   3.3. RP Object ......................................... 11
      3.3.1. H-PCE-FLAG TLV ................................ 11
      3.3.2. Domain-ID TLV ................................ 12
   3.4. Objective Functions ................................ 12
      3.4.1. OF Codes ...................................... 12
      3.4.2. OF Object ..................................... 13
   3.5. Metric Object ...................................... 14
   3.6. SVEC Object ....................................... 15
   3.7. PCEP-ERROR Object ................................... 15
      3.7.1. Hierarchy PCE Error-Type ....................... 15
   3.8. NO-PATH Object ..................................... 16
4. H-PCE Procedures ....................................... 16
   4.1. OPEN Procedure between Child PCE and Parent PCE ...... 16
   4.2. Procedure to Obtain Domain Sequence .................. 17
5. Error Handling .......................................... 17
6. Manageability Considerations ............................ 18
   6.1. Control of Function and Policy ...................... 18
      6.1.1. Child PCE ...................................... 18
      6.1.2. Parent PCE .................................... 18
## 1. Introduction

The Path Computation Element communication Protocol (PCEP) provides a mechanism for Path Computation Elements (PCEs) and Path Computation Clients (PCCs) to exchange requests for path computation and responses that provide computed paths.

The capability to compute the routes of end-to-end inter-domain MPLS Traffic Engineering (MPLS-TE) and GMPLS Label Switched Paths (LSPs) is expressed as requirements in [RFC4105] and [RFC4216]. This capability may be realized by a PCE [RFC4655]. The methods for establishing and controlling inter-domain MPLS-TE and GMPLS LSPs are documented in [RFC4726].

[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 architecture, the parent PCE is used to compute a multi-domain path based on the domain connectivity information. A child PCE may be responsible for single or multiple domains and is used to compute the intra-domain path based on its own domain topology information.

The H-PCE end-to-end domain path computation procedure is described below:

- A path computation client (PCC) sends the inter-domain path computation requests to the child PCE responsible for its domain;
- The child PCE forwards the request to the parent PCE;
- The parent PCE computes the likely domain paths from the ingress domain to the egress domain;
- The parent PCE sends the intra-domain path computation requests (between the domain border nodes) to the child PCEs which are responsible for the domains along the domain path;
- The child PCEs return the intra-domain paths to the parent PCE;
- The parent PCE constructs the end-to-end inter-domain path based on the intra-domain paths;
- The parent PCE returns the inter-domain path to the child PCE;
- The child PCE forwards the inter-domain path to the PCC.

The parent PCE may be requested to provide only the sequence of domains to a child PCE so that alternative inter-domain path computation procedures, including Per Domain (PD) [RFC5152] and Backwards Recursive Path Computation (BRPC) [RFC5441], may be used.

This document defines the PCEP extensions for the purpose of implementing Hierarchical PCE procedures, which are described in [RFC6805].

1.1. Scope

The following functions are out of scope of this document:

- Determination of Destination Domain (section 4.5 of [RFC6805]):
  * via a collection of reachability information from child domain;
  * via requests to the child PCEs to discover if they contain the
destination node;
* or any other methods.

- Parent Traffic Engineering Database (TED) methods (section 4.4 of [RFC6805]), although suitable mechanisms include:
  * YANG-based management interfaces;
  * BGP-LS [RFC7752];
  * Future extension to PCEP (such as [I-D.dhodylee-pce-pcep-ls]).

- Learning of Domain connectivity and boundary nodes (BN) addresses, methods to achieve this function include:
  * YANG-based management interfaces;
  * BGP-LS [RFC7752];
  * Future extension to PCEP (such as [I-D.dhodylee-pce-pcep-ls]).

- Stateful PCE Operations (Refer [I-D.ietf-pce-stateful-hpce])

- Applicability of hierarchical PCE to large multi-domain environments.

  * The hierarchical relationship model is described in [RFC6805]. It is applicable to environments with small groups of domains where visibility from the ingress LSRs is limited. As highlighted in [RFC7399] applying the hierarchical PCE model to very large groups of domains, such as the Internet, is not considered feasible or desirable.

1.2. Terminology

This document uses the terminology defined in [RFC4655], [RFC5440] and the additional terms defined in Section 1.4 of [RFC6805].

1.3. 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. Requirements for H-PCE
This section compiles the set of requirements to the PCEP extensions to support the H-PCE architecture and procedures. [RFC6805] identifies high-level requirements of PCEP extensions required to support the hierarchical PCE model.

2.1. Path Computation Request

The Path Computation Request (PCReq) [RFC5440] messages are used by a PCC or a PCE to make a path computation request to a PCE. In order to achieve the full functionality of the H-PCE procedures, the PCReq message needs to include:

- Qualification of PCE Requests (Section 4.8.1. of [RFC6805]);
- Multi-domain Objective Functions (OF);
- Multi-domain Metrics.

2.1.1. Qualification of PCEP Requests

As described in Section 4.8.1 of [RFC6805], the H-PCE architecture introduces new request qualifications, which are:

- The ability for a child PCE to indicate that a path computation request sent to a parent PCE should be satisfied by a domain sequence only, that is, not by a full end-to-end path. This allows the child PCE to initiate a per-domain (PD) [RFC5152] or a backward recursive path computation (BRPC) [RFC5441].
- As stated in [RFC6805], Section 4.5, if a PCC knows the egress domain, it can supply this information as the path computation request. The PCC may also want to specify the destination domain information in a PCEP request, if it is known.
- An inter domain path computed by parent PCE should be capable of disallowing specific domain re-entry.

2.1.2. Multi-domain Objective Functions

For H-PCE inter-domain path computation, there are three new Objective Functions defined in this document:

- Minimize the number of Transit Domains (MTD)
- Minimize the number of border nodes (MBN)
- Minimize the number of Common Transit Domains (MCTD)

The PCC may specify the multi-domain Objective Function code to use when requesting inter-domain path computation, it may also
include intra-domain OFs, such as Minimum Cost Path (MCP) [RFC5441], which must be considered by participating child PCEs.

2.1.3. Multi-domain Metrics

For inter-domain path computation, there are several path metrics of interest.

- Domain count (number of domains crossed);
- Border Node count.

A PCC may be able to limit the number of domains crossed by applying a limit on these metrics. Details in Section 3.4.

2.2. Parent PCE Capability Advertisement

A PCEP Speaker (Parent PCE or Child PCE) that supports and wishes to use the procedures described in this document must advertise the fact and negotiate its role with its PCEP peers. It does this using the "H-PCE Capability" TLV, described in Section 3.2.1, in the OPEN Object to advertise its support for PCEP extensions for H-PCE Capability.

During the PCEP session establishment procedure, the child PCE needs to be capable of indicating to the parent PCE whether it requests the parent PCE capability or not.

2.3. PCE Domain Identification

A PCE domain is a single domain with an associated PCE. Although it is possible for a PCE to manage multiple domains simultaneously. The PCE domain could be an IGP area or AS.

The PCE domain identifiers MAY be provided during the PCEP session establishment procedure.

2.4. Domain Diversity

In a multi-domain environment, Domain Diversity is defined in [RFC6805] and described as "A pair of paths are domain-diverse if they do not transit any of the same domains. A pair of paths that share a common ingress and egress are domain-diverse if they only share the same domains at the ingress and egress (the ingress and egress domains). Domain diversity may be maximized for a pair of paths by selecting paths that have the smallest number of shared domains."

The main motivation behind domain diversity is to avoid fate sharing, but it can also be because of some geo-political reasons and commercial relationships that would require domain diversity. For example, a pair of paths should choose different transit Autonomous System (AS) because of some policy considerations.

In the case when full domain diversity could not be achieved, it is helpful to minimize the commonly shared domains. Also, it is interesting to note that other scope of diversity (node, link, SRLG etc.) can still be applied inside the commonly shared domains.

3. PCEP Extensions

This section defines extensions to PCEP [RFC5440] to support the H-PCE procedures.

3.1 Applicability to PCC-PCE Communications

Although the extensions defined in this document are intended primarily for use between a child PCE and a parent PCE, they are also applicable for communications between a PCC and its PCE.

Thus, the information that may be encoded in a PCReq can be sent from a PCC towards the child PCE. This includes the RP object (Section 3.3) and the Objective Function (OF) codes and objects (Section 3.4). A PCC and a child PCE could also exchange the capability (Section 3.2.1) during its session.

This allows a PCC to request paths that transit multiple domains utilizing the capabilities defined in this document.

3.2. OPEN Object

Two new TLVs are defined in this document to be carried within an OPEN object. This way, during the PCEP session establishment, the H-PCE capability and Domain information can be advertised.

3.2.1. H-PCE Capability TLV

The H-PCE-CAPABILITY TLV is an optional TLV associated with the OPEN Object [RFC5440] to exchange H-PCE capability of PCEP speakers.

Its format is shown in the following figure:
The type of the TLV is TBD1 (to be assigned by IANA), and it has a
fixed length of 4 octets.

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

  P (Parent PCE Request bit): if set, will signal that the child PCE
  wishes to use the peer PCE as a parent PCE.

Unassigned bits MUST be set to 0 on transmission and MUST be ignored
on receipt.

The inclusion of this TLV in an OPEN object indicates that the H-PCE
extensions are supported by the PCEP speaker. The child PCE MUST
include this TLV and set the P flag. The parent PCE MUST include
this TLV and unset the P flag.

The setting of the P flag (parent PCE request bit) would mean that
the PCEP speaker wants the peer to be a parent PCE, so in the case
of a PCC to Child-PCE relationship, neither entity would set the P
flag.

If both peers attempt to set the P flag then the session
establishment MUST fail, and the PCEP speaker MUST respond with PCErr
message using Error-Type 1: "PCEP Session Establishment Failure" as
per [RFC5440].

If the PCE understands the H-PCE path computation request but did not
advertise its H-PCE capability, it MUST send a PCErr message with
Error-Type=TBD8 ("H-PCE error") and Error-Value=1 ("H-PCE Capability
not advertised").

3.2.1.1 Backwards Compatibility

Section 7.1 of [RFC5440] requires that "Unrecognized TLVs MUST be
ignored.

That means that a PCE that does not support this document but that
receives an Open Message containing an Open Object that includes
an H-PCE-CAPABILITIES TLV will ignore that TLV and will continue to attempt to establish a PCEP session. It will, however, not include the TLV in the Open message that it sends, so the H-PCE relationship will not be created.

If a PCE does not support the extensions defined in this document but receives them in a PCEP message (notwithstanding the fact that the session was not established as supporting a H-PCE relationship), the receiving PCE will ignore the H-PCE related parameters because they are all encoded in TLVs within standard PCEP objects.

3.2.2. Domain-ID TLV

The Domain-ID TLV, when used in the OPEN object, identifies the domains served by the PCE. The child PCE uses this mechanism to inform the domain information to the parent PCE.

The Domain-ID TLV is defined below:

```
0                   1                   2                   3
+-------------------+-------------------+-------------------+-------------------+
|       Type= TBD2  |        Length      |
+-------------------+-------------------+-------------------+-------------------+
|   Domain Type    |       Reserved    |
|                  +-------------------+-------------------+-------------------+
|                  |                      Domain ID                      |
|                  +-------------------+-------------------+-------------------+
```

Figure 2: Domain-ID TLV format

The type of the TLV is TBD2 (to be assigned by IANA), and it has a variable Length of the value portion. The value part comprises:

- Domain Type (8 bits): Indicates the domain type. Four types of domain are currently defined:
  - Type=1: the Domain ID field carries a 2-byte AS number. Padded with trailing zeros to a 4-byte boundary.
  - Type=2: the Domain ID field carries a 4-byte AS number.
  - Type=3: the Domain ID field carries a 4-byte OSPF area ID.
  - Type=4: the Domain ID field carries (2-byte Area-Len, variable length IS-IS area ID). Padded with trailing zeros to a 4-byte
3.3.1. H-PCE-FLAG TLV

The H-PCE-FLAG TLV is an optional TLV associated with the RP Object [RFC5440] to indicate the H-PCE path computation request and options. Its format is shown in the following figure:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Type= TBD3      |             Length=4          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Flags                             |D|S|
|----------------------------------------------------------------|
```

Figure 3: H-PCE-FLAG TLV format

The type of the TLV is TBD3 (to be assigned by IANA), and it has a fixed length of 4 octets.

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

- S (Domain Sequence bit): if set, will signal that the child PCE wishes to get only the domain sequence in the path computation reply. Refer to Section 3.7 of [RFC7897] for details.

- D (Disallow Domain Re-entry bit): if set, will signal that the computed path does not enter a domain more than once.

Unassigned bits MUST be set to 0 on transmission and MUST be ignored on receipt.

The presence of the TLV indicates that the H-PCE based path
computation is requested as per this document.

3.3.2. Domain-ID TLV

The Domain-ID TLV, carried in an OPEN object, is used to indicate a (list of) managed domains and is described in Section 3.3.1. This TLV, when carried in an RP object, indicates the destination domain ID. If a PCC knows the egress domain, it can supply this information in the PCReq message. The format and procedure of this TLV are defined in Section 3.2.2.

If a Domain-id TLV is used in the RP object, and the destination is not actually in the indicated domain, then the parent PCE should respond with a NO-PATH object and NO-PATH VECTOR TLV should be used. A new bit number is assigned to indicate "Destination not found in the indicated domain" (see Section 3.7).

3.4. Objective Functions

3.4.1. OF Codes

[RFC5541] defines a mechanism to specify an Objective Function that is used by a PCE when it computes a path. Three new Objective Functions are defined for H-PCE, these are:

- **MTD**
  - Name: Minimize the number of Transit Domains (MTD)
  - Objective Function Code - TBD4 (to be assigned by IANA)
  - Description: Find a path P such that it passes through the least number of transit domains.
  - Objective functions are formulated using the following terminology:
    + A network comprises a set of N domains \( \{D_i, (i=1...N)\} \).
    + A path P passes through K unique domains \( \{D_{pi}, (i=1...K)\} \).
    + Find a path P such that the value of K is minimized.

- **MBN**
  - Name: Minimize the number of border nodes.
  - Objective Function Code - TBD5 (to be assigned by IANA)
* Description: Find a path $P$ such that it passes through the least number of border nodes.

* Objective functions are formulated using the following terminology:

  + A network comprises a set of $N$ links \{L$_i$, $(i=1...N)$\}.
  
  + A path $P$ is a list of $K$ links \{L$_{pi}$, $(i=1...K)$\}.
  
  + D(L$_{pi}$) is a function that determines if the links L$_{pi}$ and L$_{pi+1}$ belong to different domains, D(L$_i$) = 1 if link L$_i$ and L$_{i+1}$ belong to different domains, D(L$_k$) = 0 if link L$_k$ and L$_{k+1}$ belong to the same domain.
  
  + The number of border node in a path $P$ is denoted by $B(P)$, where $B(P) = \sum{D(L_{pi})}$, $(i=1...K-1)$.
  
  + Find a path $P$ such that $B(P)$ is minimized.

There is one objective function that applies to a set of synchronized path computation requests to increase the domain diversity:

  o MCTD

    * Name: Minimize the number of Common Transit Domains
    
    * Objective Function Code - TBD13 (to be assigned by IANA)
    
    * Description: Find a set of paths such that it passes through the least number of common transit domains.
    
      + A network comprises a set of $N$ domains \{D$_i$, $(i=1...N)$\}.
      
      + A path $P$ passes through $K$ unique domains \{D$_{pi}$, $(i=1...K)$\}.
      
      + A set of paths \{P$_1$...P$_m$\} have $L$ transit domains that are common to more than one path \{D$_{pi}$, $(i=1...L)$\}.
      
      + Find a set of paths such that the value of $L$ is minimized.

3.4.2. OF Object

The OF (Objective Function) object [RFC5541] is carried within a PCReq message so as to indicate the desired/required objective function to be applied by the PCE during path computation. As per Section 3.2 of [RFC5541] a single OF object may be included in a path.
computation request.

The new OF codes described in Section 3.4.1 are applicable at the inter-domain path computation performed by the parent PCE, it is also necessary to specify the OF code that may be applied for the intra-domain path computation performed by the child PCE. To accommodate this, the OF-List TLV (described in Section 2.1. of [RFC5541]) is included in the OF object as an optional TLV.

The OF-List TLV allows encoding of multiple OF codes. When this TLV is included inside the OF object, only the first OF-code in the OF-LIST TLV is considered. The parent PCE MUST use this OF code in the OF object when sending the intra domain path computation request to the child PCE. If the OF list TLV is included in the OF Object, the OF Code inside the OF Object MUST include one of the H-PCE Objective Functions defined in this document, the OF Code inside the OF List TLV MUST NOT include an H-PCE Objective Function. If this condition is not met, the PCEP speaker MUST respond with a PCErr message with Error-Type=10 (Reception of an invalid object) and Error-Value=TBD15 (Incompatible OF codes in H-PCE).

If the Objective Functions defined in this document are unknown or unsupported by a PCE, then the procedure as defined in [RFC5541] is followed.

3.5. Metric Object

The METRIC object is defined in Section 7.8 of [RFC5440], comprising of metric-value, metric-type (T field) and flags. This document defines the following types for the METRIC object for H-PCE:

- T=TBD6: Domain count metric (number of domains crossed);
- T=TBD7: Border Node count metric (number of border nodes crossed).

The domain count metric type of the METRIC object encodes the number of domains crossed in the path. The border node count metric type of the METRIC object encodes the number of border nodes in the path. If a domain is re-entered, then domain should be double counted.

A PCC or child PCE MAY use the metric in a PCReq message for an inter-domain path computation, meeting the number of domain or border nodes crossing requirement. As per [RFC5440], in this case, the B bit is set to suggest a bound (a maximum) for the metric that must not be exceeded for the PCC to consider the computed path as acceptable.

A PCC or child PCE MAY also use this metric to ask the PCE to optimize the metric during inter-domain path computation. In this
case, the B flag is cleared, and the C flag is set.

The Parent PCE MAY use the 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 MAY also use this metric to send the computed end to end metric value in a reply message.

3.6. SVEC Object

[RFC5440] defines SVEC object which includes flags for the potential dependency between the set of path computation requests (Link, Node and SRLG diverse). This document defines a new flag O for domain diversity.

The following new bit is added to the Flags field:

- Domain Diverse O-bit - TBD14 : when set, this indicates that the computed paths corresponding to the requests specified by the following RP objects MUST NOT have any transit domains in common.

The Domain Diverse O-bit can be used in Hierarchical PCE path computation to compute synchronized domain diverse end to end path or diverse domain sequences.

When domain diverse O bit is set, it is applied to the transit domains. The other bit in SVEC object (N, L, S etc.) MAY be set and MUST still be applied in the ingress and egress shared domain.

3.7. PCEP-ERROR Object

3.7.1. Hierarchy PCE Error-Type

A new PCEP Error-Type [RFC5440] is used for the H-PCE extension as defined below:

+-----------------+-----------------------------------------+
| Error-Type      | Meaning                                 |
|                 |                                         |
| TBD8            | H-PCE error                             |
| Error-value=1   | H-PCE capability was not advertised      |
| Error-value=2   | parent PCE capability cannot be provided |
|                 |                                         |
+-----------------+-----------------------------------------+

Figure 4: H-PCE error
3.8. NO-PATH Object

To communicate the reason(s) for not being able to find a multidomain path or domain sequence, the NO-PATH object can be used in the PCRep message. [RFC5440] defines the format of the NO-PATH object. The object may contain a NO-PATH-VECTOR TLV to provide additional information about why a path computation has failed.

Three new bit flags are defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object.

- Bit number TBD9: When set, the parent PCE indicates that destination domain unknown;
- Bit number TBD10: When set, the parent PCE indicates unresponsive child PCE(s);
- Bit number TBD11: When set, the parent PCE indicates no available resource available in one or more domains.
- Bit number TBD12: When set, the parent PCE indicates that the destination is not found in the indicated domain.

4. H-PCE Procedures

The H-PCE path computation procedure is described in [RFC6805].

4.1. OPEN Procedure between Child PCE and Parent PCE

If a child PCE wants to use the peer PCE as a parent, it MUST set the P (parent PCE request flag) in the H-PCE-CAPABILITY TLV inside the OPEN object carried in the Open message during the PCEP session initialization procedure.

The child PCE MAY also report its list of domain IDs, to the parent PCE, by specifying them in the Domain-ID TLVs in the OPEN object. This object is carried in the OPEN message during the PCEP session initialization procedure.

The OF codes defined in this document can be carried in the OF-list TLV of the OPEN object. If the OF-list TLV carries the OF codes, it means that the PCE is capable of implementing the corresponding objective functions. This information can be used for selecting a proper parent PCE when a child PCE wants to get a path that satisfies a certain Objective Function.

When a child PCE sends a PCReq to a peer PCE, which requires parental
activity and H-PCE capability flags TLV but which were not included in the session establishment procedure described above, the peer PCE SHOULD send a PCErr message to the child PCE and MUST specify the error-type=TBD8 (H-PCE error) and error-value=1 (H-PCE capability was not advertised) in the PCEP-ERROR object.

When a specific child PCE sends a PCReq to a peer PCE, that requires parental activity and the peer PCE does not want to act as the parent for it, the peer PCE SHOULD send a PCErr message to the child PCE and MUST specify the error-type=TBD8 (H-PCE error) and error-value=2 (Parent PCE capability cannot be provided) in the PCEP-ERROR object.

4.2. Procedure to Obtain Domain Sequence

If a child PCE only wants to get the domain sequence for a multi-domain path computation from a parent PCE, it can set the Domain Path Request bit in the H-PCE-FLAG TLV in the RP object carried in a PCReq message. The parent PCE which receives the PCReq message tries to compute a domain sequence for it (instead of the E2E path). If the domain path computation succeeds the parent PCE sends a PCRep message which carries the domain sequence in the Explicit Route Object (ERO) to the child PCE. Refer to [RFC7897] for more details about domain sub-objects in the ERO. Otherwise, it sends a PCReq message which carries the NO-PATH object to the child PCE.

5. Error Handling

A PCE that is capable of acting as a parent PCE might not be configured or willing to act as the parent for a specific child PCE. This fact could be determined when the child sends a PCReq that requires parental activity, and could result in a negative response in a PCEP Error (PCErr) message and indicate the hierarchy PCE error-type=TBD8 (H-PCE error) and suitable error-value. (Section 3.7)

Additionally, the parent PCE may fail to find the multi-domain path or domain sequence due to one or more of the following reasons:

- A child PCE cannot find a suitable path to the egress;
- The parent PCE does not hear from a child PCE for a specified time;
- The Objective Functions specified in the path request cannot be met.

In this case, the parent PCE MAY need to send a negative path computation reply specifying the reason. This can be achieved by
including NO-PATH object in the PCRep message. Extension to NO-PATH object is needed to include the aforementioned reasons described in Section 3.7.

6. Manageability Considerations

General PCE and PCEP management considerations are discussed in [RFC4655] and [RFC5440]. There are additional management considerations for H-PCE which are described in [RFC6805], and repeated in this section.

The administrative entity responsible for the management of the parent PCEs must be determined for the following cases:

- multi-domains (e.g., IGP areas or multiple ASes) within a single service provider network, the management responsibility for the parent PCE would most likely be handled by the service provider,

- multiple ASes within different service provider networks, it may be necessary for a third party to manage the parent PCEs according to commercial and policy agreements from each of the participating service providers.

6.1. Control of Function and Policy

Control and function will need to be carefully managed in an H-PCE network. A child PCE will need to be configured with the address of its parent PCE. It is expected that there will only be one or two parents of any child.

The parent PCE also needs to be aware of the child PCEs for all child domains that it can see. This information is most likely to be configured (as part of the administrative definition of each domain).

Discovery of the relationships between parent PCEs and child PCEs do not form part of the hierarchical PCE architecture. Mechanisms that rely on advertising or querying PCE locations across domain or provider boundaries are undesirable for security, scaling, commercial, and confidentiality reasons. The specific behaviour of the child and parent PCE are described in the following sub-sections.

6.1.1. Child PCE

Support of the hierarchical procedure will be controlled by the management organization responsible for each child PCE. A child PCE must be configured with the address of its parent PCE in order for it to interact with its parent PCE. The child PCE must also be
authorized to peer with the parent PCE.

6.1.2. Parent PCE

The parent PCE MUST only accept path computation requests from authorized child PCEs. If a parent PCE receives requests from an unauthorized child PCE, the request SHOULD be dropped. This means that a parent PCE MUST be able to cryptographically authenticate requests from child PCEs.

Multi-party shared key authentication schemes are not recommended for inter-domain relationships because of the potential for impersonation and repudiation and for the operational difficulties should revocation be required.

The choice of authentication schemes to employ may be left to implementers of H-PCE and are not discussed further in this document.

6.1.3. Policy Control

It may be necessary to maintain a policy module on the parent PCE [RFC5394]. This would allow the parent PCE to apply commercially relevant constraints such as SLAs, security, peering preferences, and monetary costs.

It may also be necessary for the parent PCE to limit the end-to-end path selection by including or excluding specific domains based on commercial relationships, security implications, and reliability.

6.2. Information and Data Models

A MIB module for PCEP was published as RFC 7420 [RFC7420] that describes managed objects for modelling of PCEP communication. A YANG module for PCEP has also been proposed [I-D.ietf-pce-pcep-yang].

Additionally, H-PCE MIB module, or additional data model, will be required to report parent PCE and child PCE information, including:

- parent PCE configuration and status,
- child PCE configuration and information,
- notifications to indicate session changes between parent PCEs and child PCEs, and
- notification of parent PCE TED updates and changes.
6.3. Liveness Detection and Monitoring

The hierarchical procedure requires interaction with multiple PCEs. Once a child PCE requests an end-to-end path, a sequence of events occurs that requires interaction between the parent PCE and each child PCE. If a child PCE is not operational, and an alternate transit domain is not available, then the failure must be reported.

6.4. Verify Correct Operations

Verifying the correct operation of a parent PCE can be performed by monitoring a set of parameters. The parent PCE implementation should provide the following parameters monitored at the parent PCE:

- number of child PCE requests,
- number of successful hierarchical PCE procedures completions on a per-PCE-peer basis,
- number of hierarchical PCE procedure completion failures on a per-PCE-peer basis, and
- number of hierarchical PCE procedure requests from unauthorized child PCEs.

6.5. Requirements On Other Protocols

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

6.6. Impact On Network Operations

The hierarchical PCE procedure is a multiple-PCE path computation scheme. Subsequent requests to and from the child and parent PCEs do not differ from other path computation requests and should not have any significant impact on network operations.

7. IANA Considerations

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" registry. This document requests IANA actions to allocate code points for the protocol elements defined in this document.

7.1. PCEP TLV Type Indicators

IANA Manages the PCEP TLV code point registry (see [RFC5440]). This is maintained as the "PCEP TLV Type Indicators" sub-registry of the
This document defines three new PCEP TLVs. IANA is requested to make the following allocation:

<table>
<thead>
<tr>
<th>Type</th>
<th>TLV name</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD1</td>
<td>H-PCE-CAPABILITY TLV</td>
<td>This I-D</td>
</tr>
<tr>
<td>TBD2</td>
<td>Domain-ID TLV</td>
<td>This I-D</td>
</tr>
<tr>
<td>TBD3</td>
<td>H-PCE-FLAG TLV</td>
<td>This I-D</td>
</tr>
</tbody>
</table>

### 7.2. H-PCE-CAPABILITY TLV Flags

This document requests that a new sub-registry, named "H-PCE-CAPABILITY TLV Flag Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field in the H-PCE-CAPABILITY TLV of the PCEP OPEN 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:

<table>
<thead>
<tr>
<th>Bit</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>P (Parent PCE Request bit)</td>
<td>This I.D.</td>
</tr>
</tbody>
</table>

### 7.3. Domain-ID TLV Domain type

This document requests that a new sub-registry, named "Domain-ID TLV Domain type", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Domain-Type field of the Domain-ID TLV. The allocation policy for this new sub-registry is IETF Review [RFC8126].

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2-byte AS number</td>
</tr>
<tr>
<td>2</td>
<td>4-byte AS number</td>
</tr>
<tr>
<td>3</td>
<td>4-byte OSPF area ID</td>
</tr>
<tr>
<td>4</td>
<td>Variable length IS-IS area ID</td>
</tr>
</tbody>
</table>
7.4. H-PCE-FLAG TLV Flags

This document requests that a new sub-registry, named "H-PCE-FLAG TLV Flag Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field in the H-PCE-FLAGS TLV of the PCEP RP 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:

<table>
<thead>
<tr>
<th>Bit</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>31</td>
<td>S (Domain Sequence bit)</td>
<td>This I.D.</td>
</tr>
<tr>
<td>30</td>
<td>D (Disallow Domain Re-entry bit)</td>
<td>This I.D.</td>
</tr>
</tbody>
</table>

7.5. OF Codes

IANA maintains a registry of Objective Function (described in [RFC5541]) at the sub-registry "Objective Function". Three new Objective Functions have been defined in this document.

IANA is requested to make the following allocations:

<table>
<thead>
<tr>
<th>Code Point</th>
<th>Name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD4</td>
<td>Minimum number of Transit Domains (MTD)</td>
<td>This I.D.</td>
</tr>
<tr>
<td>TBD5</td>
<td>Minimize number of Border Nodes (MBN)</td>
<td>This I.D.</td>
</tr>
<tr>
<td>TBD13</td>
<td>Minimize the number of Common Transit Domains (MCTD)</td>
<td>This I.D.</td>
</tr>
</tbody>
</table>

7.6. METRIC Types

IANA maintains one sub-registry for "METRIC object T field". Two new metric types are defined in this document for the METRIC object
IANA is requested to make the following allocations:

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD6</td>
<td>Domain Count metric</td>
<td>This I.D.</td>
</tr>
<tr>
<td>TBD7</td>
<td>Border Node Count metric</td>
<td>This I.D.</td>
</tr>
</tbody>
</table>

### 7.7. New PCEP Error-Types and 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:

<table>
<thead>
<tr>
<th>Error-Type</th>
<th>Meaning and error values</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD8</td>
<td>H-PCE Error</td>
<td>This I.D.</td>
</tr>
</tbody>
</table>

  - Error-value=1 H-PCE Capability not advertised
  - Error-value=2 Parent PCE Capability cannot be provided

- 10 Reception of an invalid object [RFC5440]
  - Error-value=TBD15: Incompatible OF codes in H-PCE

### 7.8. New NO-PATH-VECTOR TLV Bit Flag

IANA maintains a sub-registry "NO-PATH-VECTOR TLV Flag Field" of bit flags carried in the PCEP NO-PATH-VECTOR TLV in the PCEP NO-PATH object as defined in [RFC5440]. IANA is requested to assign three new bit flag as follows:

<table>
<thead>
<tr>
<th>Bit Number</th>
<th>Name Flag</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD9</td>
<td>Destination Domain unknown</td>
<td>This I.D.</td>
</tr>
<tr>
<td>TBD10</td>
<td>Unresponsive child PCE(s)</td>
<td>This I.D.</td>
</tr>
<tr>
<td>TBD11</td>
<td>No available resource in one or more domain</td>
<td>This I.D.</td>
</tr>
<tr>
<td>TBD12</td>
<td>Destination is not found in the indicated domain.</td>
<td>This I.D.</td>
</tr>
</tbody>
</table>
7.9. SVEC Flag

IANA maintains a sub-registry "SVEC Object Flag Field" of bit flags carried in the PCEP SVEC object as defined in [RFC5440]. IANA is requested to assign one new bit flag as follows:

<table>
<thead>
<tr>
<th>Bit Number</th>
<th>Name Flag</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD14</td>
<td>Domain Diverse O-bit</td>
<td>This I.D.</td>
</tr>
</tbody>
</table>

8. Security Considerations

The hierarchical PCE procedure relies on PCEP and inherits the security considerations defined in [RFC5440]. As PCEP operates over TCP, it may also make use of TCP security mechanisms, such as TCP Authentication Option (TCP-AO) [RFC5925] or Transport Layer Security (TLS) [RFC8253].

Any multi-domain operation necessarily involves the exchange of information across domain boundaries. This may represent a significant security and confidentiality risk especially when the child domains are controlled by different commercial concerns. PCEP allows individual PCEs to maintain the confidentiality of their domain path information using path-keys [RFC5520], and the H-PCE architecture is specifically designed to enable as much isolation of domain topology and capabilities information as is possible.

For further considerations of the security issues related to inter-AS path computation, see [RFC5376].

9. Contributing Authors

Xian Zhang
Huawei
EMail: zhang.xian@huawei.com

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

EMail: dhruv.ietf@gmail.com
10. Acknowledgements

The authors would like to thank Mike McBride, Kyle Rose, Roni Even for their detailed review, comments and suggestions which helped improve this document.

11. References

11.1. Normative References


11.2. Informative References


Appendix

A1. Implementation Status

Zhang, et al.          Expires December, 2019
The H-PCE architecture and protocol procedures describe in this I-D were implemented and tested for a variety of optical research applications.

The Appendix should be removed before publication.

A1.1. Inter-layer traffic engineering with H-PCE

This work was led by:

- Ramon Casellas [ramon.casellas@cttc.es]
- Centre Tecnologic de Telecomunicacions de Catalunya (CTTC)

The H-PCE instances (parent and child) were multi-threaded asynchronous processes. Implemented in C++11, using C++ Boost Libraries. The targeted system used to deploy and run H-PCE applications was a POSIX system (Debian GNU/Linux operating system).

Some parts of the software may require a Linux Kernel, the availability of a Routing Controller running collocated in the same host and the usage of libnetfilter / libipq and GNU/Linux firewalling capabilities. Most of the functionality, including algorithms is done by means of plugins (e.g., as shared libraries or .so files in Unix systems).

The CTTC PCE supports the H-PCE architecture, but also supports stateful PCE with active capabilities, as an OpenFlow controller, and has dedicated plugins to support monitoring, BRPC, P2MP, path keys, back end PCEs. Management of the H-PCE entities was supported via HTTP and CLI via Telnet.

Further details of the H-PCE prototyping and experimentation can be found in the following scientific papers:


R. Casellas, R. Munoz, R. Martinez, R. Vilalta, L. Liu, T. Tsuritani, I. Morita, V. Lopez, O. Gonzalez de Dios, J. P. Fernandez-Palacios, "SDN based Provisioning Orchestration of


A1.2. Telefonica Netphony (Open Source PCE)

The Telefonica Netphony PCE is an open source Java-based implementation of a Path Computation Element, with several flavours, and a Path Computation Client. The PCE follows a modular architecture and allows to add customized algorithms. The PCE has also stateful and remote initiation capabilities. In current version, three components can be built, a domain PCE (aka child PCE), a parent PCE (ready for the H-PCE architecture) and a PCC (path computation client).

This work was led by:

- Oscar Gonzalez de Dios [oscar.gonzalezdedios@telefonica.com]
- Victor Lopez Alvarez [victor.lopezalvarez@telefonica.com]
- Telefonica I+D, Madrid, Spain

The PCE code is publicly available in a GitHub repository:

- https://github.com/telefonicaid/netphony-pce
The PCEP protocol encodings are located in the following repository:

- https://github.com/telefonicaid/netphony-network protocols

The traffic engineering database and a BGP-LS speaker to fill the database is located in:

- https://github.com/telefonicaid/netphony-topology

The parent and child PCE are multi-threaded java applications. The path computation uses the jgrapht free Java class library (0.9.1) that provides mathematical graph-theory objects and algorithms. Current version of netphony PCE runs on java 1.7 and 1.8, and has been tested in GNU/Linux, Mac OS-X and Windows environments. The management of the parent and domain PCEs is supported though CLI via Telnet, and configured via XML files.

Further details of the netphony H-PCE prototyping and experimentation can be found in the following research papers:


A1.3. H-PCE Proof of Concept developed by Huawei

Huawei developed this H-PCE on the Huawei Versatile Routing Platform (VRP) to experiment with the hierarchy of PCE. Both end to end path computation as well as computation for domain-sequence are supported.
This work was led by:

- Udayasree Pallee [udayasreereddy@gmail.com]
- Dhruv Dhody [dhruv.ietf@gmail.com]
- Huawei Technologies, Bangalore, India

Further work on stateful H-PCE [I-D.ietf-pce-stateful-hpce] is being carried out on ONOS.