PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 16, 2019

F. Zhang
Q. Zhao
Huawei
O. Gonzalez de Dios
Telefonica I+D
R. Casellas
CTTC
D. King
Old Dog Consulting

July 15, 2018

Extensions to Path Computation Element Communication Protocol (PCEP) for Hierarchical Path Computation Elements (PCE)

draft-ietf-pce-hierarchy-extensions-05

Abstract

The Hierarchical Path Computation Element (H-PCE) architecture RFC 6805, provides a mechanism to allow the optimum sequence of domains to be selected, and the optimum end-to-end path to be derived through the use of a hierarchical relationship between domains.

This document defines the Path Computation Element Protocol (PCEP) extensions for the purpose of implementing necessary Hierarchical PCE procedures and protocol extensions.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

The Path Computation Element communication Protocol (PCEP) provides mechanisms for Path Computation Elements (PCEs) to perform path computations in response to Path Computation Clients' (PCCs) requests.

The 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 a single domain or multiple domains, it 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:

- o A path computation client (PCC) sends the inter-domain path computation requests to the child PCE responsible for its domain;

- o The child PCE forwards the request to the parent PCE;

- o The parent PCE computes the likely domain paths from the ingress domain to the egress domain;

- o 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;

- o The child PCEs return the intra-domain paths to the parent PCE;

- o The parent PCE constructs the end-to-end inter-domain path based on the intra-domain paths;

- o The parent PCE returns the inter-domain path to the child PCE;

- o The child PCE forwards the inter-domain path to the PCC.

In addition, 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.

- o Determination of Destination Domain (section 4.5 of [RFC6805])

* via 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.

o Parent Traffic Engineering Database (TED) methods (section 4.4 of [RFC6805])

o Learning of Domain connectivity and boundary nodes (BN) addresses.

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

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 of the PCEP protocol 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) messages are used by a PCC or 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:

o Qualification of PCE Requests;

o Multi-domain Objective Functions (OF);

o 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:

- It MUST be possible 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. It SHOULD be possible to specify the destination domain information in a PCEP request, if it is known.

- It MAY be possible to indicate that the inter domain path computed by parent PCE should disallow domain re-entry.

2.1.2. Multi-domain Objective Functions

For inter-domain path computation, there is one new objective Function which is defined in section 1.3.1 and 4.1 of [RFC6805]:

- Minimize the number of domains crossed. A domain can be either an Autonomous System (AS) or an Internal Gateway Protocol (IGP) area depending on the type of multi-domain network hierarchical PCE is applied to.

Another objective Function to minimize the number of border nodes is also defined in this document.

During the PCEP session establishment procedure, the parent PCE needs to be capable of indicating the Objective Functions (OF) [RFC5541] capability in the Open message. This capability information may then be announced by child PCEs, and used for selecting the PCE when a PCC wants a path that satisfies one or multiple inter-domain objective functions.

When a PCC requests a PCE to compute an inter-domain path, the PCC needs to be capable of indicating the new objective functions for inter-domain path. Note that a given child PCE may also act as a parent PCE (for some other child PCE).

For the reasons described previously, new OF codes need to be defined for the new inter-domain objective functions. Then the PCE can notify its new inter-domain objective functions to the PCC by carrying them in the OF-list TLV which is carried in the OPEN object.

The PCC can specify which objective function code to use, which is carried in the OF object when requesting a PCE to compute an inter-domain path.

A parent PCE MUST be capable of ensuring homogeneity, across domains, when applying OF codes for strict OF intra-domain requests.

2.1.3. Multi-domain Metrics

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

- o Domain count (number of domains crossed);

- o 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

Parent and child PCE relationships are likely to be configured. However, as mentioned in [RFC6805], it would assist network operators if the child and parent PCEs could indicate their H-PCE capabilities.

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. Also, during the PCEP session establishment procedure, the parent PCE needs to be capable of indicating whether its parent capability can be provided or not.

A PCEP Speaker (Parent PCE or Child PCE or PCC) includes the "H-PCE Capability" TLV, described in Section 3.1.1, in the OPEN Object to advertise its support for PCEP extensions for H-PCE Capability.

2.3. PCE Domain Discovery

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]. A pair of paths are domain-diverse if they do not traverse any of the same transit domains. Domain diversity may be maximized for a pair of paths by selecting paths that have the smallest number of shared domains. Path computation should facilitate the selection of domain diverse paths as a way to reduce the risk of shared failure and automatically helps to ensure path diversity for most of the route of a pair of LSPs.

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

3. PCEP Extensions

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

3.1. OPEN object

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

3.1.1. 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:

```
     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= TBD1      |            Length=4           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                             Flags                         |I|R|
    +---------------------------------------------------------------+
```

Figure 1: H-PCE-CAPABILITY TLV format

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):

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

I (Parent PCE Indication bit): if set, will signal that the PCE can be used as a parent PCE by the peer PCE.

The inclusion of this TLV in an OPEN object indicate that the H-PCE extensions are supported by the PCEP speaker. The PCC MAY include this TLV to indicate that it understands the H-PCE extensions. The Child PCE MUST include this TLV and set the R flag (and unset the I flag) on the PCEP session towards the Parent PCE. The Parent PCE MUST include this TLV and set the I flag and unset the R flag on the PCEP session towards the child PCE. The parent-child PCEP session is set to be established only when this capability is advertised.

If such capability is not exchanged and the parent PCE receive a "H-PCE path computation request", it MUST send a PCErr message with Error-Type=TBD8 (H-PCE error) and Error-Value=1 (Parent PCE Capability not advertised).

3.1.2. Domain-ID TLV

The Domain-ID TLV when used in OPEN object identify the domain(s) 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
     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= 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 of -

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 an 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 boundary.

Reserved: Zero at transmission; ignored at receipt.

Domain ID (variable): Indicates an IGP Area ID or AS number. It can be 2 bytes, 4 bytes or variable length depending on the domain identifier used. It is padded with trailing zeros to a 4-byte boundary.

In case a PCE serves more than one domain, multiple Domain-ID TLV is included for each domain it serves.

3.2. RP object

3.2.1. 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 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 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.

3.2.2. Domain-ID TLV

The usage of Domain-ID TLV carried in an OPEN object is used to indicate a (list of) managed domains and is described in Section 3.1.2. This TLV when carried in a 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 is defined in Section 3.1.2.

3.3. Objective Functions

3.3.1. OF Codes

[RFC5541] defines a mechanism to specify an objective function that is used by a PCE when it computes a path. Two new objective functions are defined for the H-PCE experiment.

o 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 {Di, (i=1...N)}.

         + A path P passes through K domains {Dpi,(i=1...K)}.

         + Find a path P such that the value of K is minimized.

   o 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 nodes {Ni, (i=1...N)}.

         + A path P is a list of K nodes {Npi,(i=1...K)}.

         + B(N) if a function that determine if the node is a border node. B(Ni) = 1 if Ni is border node; B(Nk) = 0 if Nk is not a border node.

         + The number of border node in a path P is denoted by B(P), where B(P) = sum{B(Npi),(i=1...K)}.

         + Find a path P such that B(P) is minimized.

   MCTD

   o Name: Minimize the number of Common Transit Domains.

   o Objective Function Code: TBD13

   o Description: Find a set of paths such that it passes through the least number of common transit domains.

3.3.2. 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 code described in Section 3.3.1 are applicable at the inter-domain level (parent), it is also necessary to specify the OF code that may be applied at the intra-domain (child) path computation level. To accommodate this, the OF-List TLV (described in section 2.1. of [RFC5541]) is included in the OF object as an optional TLV.

OF-List TLV allow 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 objective functions defined in this document are unknown/ unsupported by a PCE, then the procedure as defined in [RFC5541] is followed.

3.4. Metric Object

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

o T=TBD6: Domain count metric (number of domains crossed);

o T=TBD7: Border Node count metric (number of border nodes crossed).

The domain count metric type of the METRIC object encodes the number of domain crossed in the path. The border node count metric type of the METRIC object encodes the number of border nodes in the path.

A PCC or child PCE MAY use these metric in PCReq message an inter-domain path meeting the number of domain or border nodes 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.

The Parent PCE MAY use these 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 in a reply message.

3.5. 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 proposes a new flag O for domain diversity.

Following new bit is added in the Flags field:

- O (Domain diverse) bit - TBD12 : when set, this indicates that the computed paths corresponding to the requests specified by the following RP objects MUST NOT have any transit domain(s) 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.

3.6. PCEP-ERROR object

3.6.1. 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: parent PCE capability was not advertised Error-value=2: parent PCE capability cannot be provided

Figure 4: H-PCE error

3.7. NO-PATH Object

To communicate the reason(s) for not being able to find a multi-domain 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.

- o Bit number TBD9: When set, the parent PCE indicates that destination domain unknown;
- o Bit number TBD10: When set, the parent PCE indicates unresponsive child PCE(s);
- o Bit number TBD11: When set, the parent PCE indicates no available resource available in one or more domain(s).

4. H-PCE Procedures

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 R (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.

If the parent PCE can provide the parent function to the peer PCE, it MUST set the I (parent PCE indication flag) in the H-PCE-CAPABILITY TLV inside the OPEN object carried in the Open message during the PCEP session creation 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 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 specific child PCE sends a PCReq to a peer PCE that requires parental activity and H-PCE capability flags were not set in the

session establishment procedure as described above, the peer PCE should send a PCErr message to the child PCE and specify the error-type=TBD (H-PCE error) and error-value=1 (parent 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 specify the error-type=TBD (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 for E2E path). If the domain path computation succeeds the parent PCE sends a PCRep message which carries the domain sequence in the ERO to the child PCE. Refer [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 3.6)

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

- o A child PCE cannot find a suitable path to the egress;
- o The parent PCE do not hear from a child PCE for a specified time;
- o 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:

- o 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,

- o 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 a 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 does 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. Specific behavior 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 configured with the identities and security credentials of all of its child PCEs, or there must be some form of shared secret that allows an unknown child PCE to be authorized by the parent PCE.

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 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 modeling of PCEP communication. A YANG module for PCEP has also been proposed [I-D.ietf-pce-pcep-yang].

A 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 a 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 by the parent PCE:

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

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 "Path Computation Element Protocol (PCEP) Numbers" registry.

This document defines three new PCEP TLVs. IANA is requested to make the following allocation:

Type	TLV name	References
TBD1	H-PCE-CAPABILITY TLV	This I-D
TBD2	Domain-ID TLV	This I-D
TBD3	H-PCE-FLAG TLV	This I-D

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 [RFC5226]. 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

The following values are defined in this document:

Bit	Description	Reference
31	R (Parent PCE Request bit)	This I.D.
30	I (Parent PCE Indication bit)	This I.D.

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.

Value	Meaning
1	2-byte AS number
2	4-byte AS number
3	4-byte OSPF area ID
4	Variable length IS-IS area ID

7.4. H-PCE-FLAG TLV Flags

This document requests that a new sub-registry, named "H-PCE-FLAGS 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 [RFC5226]. 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

The following values are defined in this document:

Bit	Description	Reference
31	S (Domain Sequence bit)	This I.D.
30	D (Disallow Domain Re-entry bit)	This I.D.

7.5. OF Codes

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

IANA is requested to make the following allocations:

Code Point	Name	Reference
TBD4	Minimum number of Transit Domains (MTD)	This I.D.
TBD5	Minimize number of Border Nodes (MBN)	This I.D.
TBD13	Minimize the number of Common Transit Domains. (MCTD)	This I.D.

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 (specified in [RFC5440]).

IANA is requested to make the following allocations:

Value	Description	Reference
TBD6	Domain Count metric	This I.D.
TBD7	Border Node Count metric	This I.D.

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:

Error-Type	Meaning and error values	Reference
TBD8	H-PCE Error	This I.D.
	Error-value=1 Parent PCE Capability not advertised	
	Error-value=2 Parent PCE Capability not supported	

7.8. New NO-PATH-VECTOR TLV Bit Flag

IANA maintains a registry 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:

Bit Number	Name Flag	Reference
TBD9	Destination Domain unknown	This I.D.
TBD10	Unresponsive child PCE(s)	This I.D.
TBD11	No available resource in one or more domain	This I.D.

7.9. SVEC Flag

IANA maintains a registry of bit flags carried in the PCEP SVEC object as defined in [RFC5440]. IANA Is requested to assign one new bit flag as follows:

Bit Number	Name Flag	Reference
TBD13	Domain Diverse	This I.D.

8. Security Considerations

The hierarchical PCE procedure relies on PCEP and inherits the security requirements defined in [RFC5440]. As PCEP operates over TCP, it may also make use of TCP security mechanisms, such as TCP-AO or [RFC8253].

H-PCE operation also relies on information used to build the TED. Attacks on a parent or child PCE may be achieved by falsifying or impeding this flow of information. If the child PCE listens to the IGP or BGP-LS for populating the TED, then normal IGP or BGP-LS security measures may be applied, and it should be noted that an IGP routing system is generally assumed to be a trusted domain such that router subversion is not a risk. The parent PCE TED is constructed as described in this document and may involve:

- o multiple parent-child relationships using PCEP
- o the parent PCE listening to child domain IGPs (with the same security features as a child PCE listening to its IGP)
- o an external mechanism (such as [RFC7752]), which will need to be authorized and secured.

Any multi-domain operation necessarily involves the exchange of information across domain boundaries. This is bound to represent a significant security and confidentiality risk especially when the child domains are controlled by different commercial concerns. PCEP allows individual PCEs to maintain 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. Implementation Status

The H-PCE architecture and protocol procedures describe in this I-D were implemented and tested for a variety of optical research applications.

9.1. Inter-layer traffic engineering with H-PCE

This work was led by:

- o Ramon Casellas [ramon.casellas@cttc.es]

o 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. Martinez, R. Munoz, L. Liu, T. Tsuritani, I. Morita, "Inter-layer traffic engineering with hierarchical-PCE in MPLS-TP over wavelength switched optical networks" , Optics Express, Vol. 20, No. 28, December 2012.

R. Casellas, R. Martinez, R. Munoz, L. Liu, T. Tsuritani, I. Morita, M. Msurusawa, "Dynamic virtual link mesh topology aggregation in multi-domain translucent WSON with hierarchical-PCE", Optics Express Journal, Vol. 19, No. 26, December 2011.

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 OpenFlow/GMPLS Flexi-grid Networks with a Stateful Hierarchical PCE", in Proceedings of Optical Fiber Communication Conference and Exposition (OFC), 9-13 March, 2014, San Francisco (EEUU). Extended Version to appear in Journal Of Optical Communications and Networking January 2015

F. Paolucci, O. Gonzalez de Dios, R. Casellas, S. Duhovnikov, P. Castoldi, R. Munoz, R. Martinez, "Experimenting Hierarchical PCE Architecture in a Distributed Multi-Platform Control Plane Testbed" , in Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic

Engineers Conference (NFOEC), 4-8 March, 2012, Los Angeles, California (USA).

R. Casellas, R. Martinez, R. Munoz, L. Liu, T. Tsuritani, I. Morita, M. Tsurusawa, "Dynamic Virtual Link Mesh Topology Aggregation in Multi-Domain Translucent WSON with Hierarchical-PCE", in Proceedings of 37th European Conference and Exhibition on Optical Communication (ECOC 2011), 18-22 September 2011, Geneve (Switzerland).

R. Casellas, R. Munoz, R. Martinez, "Lab Trial of Multi-Domain Path Computation in GMPLS Controlled WSON Using a Hierarchical PCE", in Proceedings of OFC/NFOEC Conference (OFC2011), 10 March 2011, Los Angeles (USA).

9.2. 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:

o Oscar Gonzalez de Dios [oscar.gonzalezdedios@telefonica.com]

o Victor Lopez Alvarez [victor.lopezalvarez@telefonica.com]

o Telefonica I+D, Madrid, Spain

The PCE code is publicly available in a GitHub repository:

o https://github.com/telefonicaid/netphony-pce

The PCEP protocol encodings are located in the following repository:

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

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

o 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:

- O. Gonzalez de Dios, R. Casellas, F. Paolucci, A. Napoli, L. Gifre, A. Dupas, E, Hugues-Salas, R. Morro, S. Belotti, G. Meloni, T. Rahman, V.P Lopez, R. Martinez, F. Fresi, M. Bohn, S. Yan, L. Velasco, . Layec and J. P. Fernandez-Palacios: Experimental Demonstration of Multivendor and Multidomain EON With Data and Control Interoperability Over a Pan-European Test Bed, in Journal of Lightwave Technology, Dec. 2016, Vol. 34, Issue 7, pp. 1610-1617.

- O. Gonzalez de Dios, R. Casellas, R. Morro, F. Paolucci, V. Lopez, R. Martinez, R. Munoz, R. Villalta, P. Castoldi: "Multi-partner Demonstration of BGP-LS enabled multi-domain EON, in Journal of Optical Communications and Networking, Dec. 2015, Vol. 7, Issue 12, pp. B153-B162.

- F. Paolucci, O. Gonzalez de Dios, R. Casellas, S. Duhovnikov, P. Castoldi, R. Munoz, R. Martinez, "Experimenting Hierarchical PCE Architecture in a Distributed Multi-Platform Control Plane Testbed" , in Proceedings of Optical Fiber Communication Conference and Exposition (OFC) and The National Fiber Optic Engineers Conference (NFOEC), 4-8 March, 2012, Los Angeles, California (USA).

9.3. Implementation 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.

10. 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

11. References

11.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC5152] Vasseur, JP., Ed., Ayyangar, A., Ed., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, DOI 10.17487/RFC5152, February 2008, <https://www.rfc-editor.org/info/rfc5152>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/info/rfc5440>.

[RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, DOI 10.17487/RFC5541, June 2009, <https://www.rfc-editor.org/info/rfc5541>.

[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>.

11.2. Informative References

[RFC4105] Le Roux, J., Ed., Vasseur, J., Ed., and J. Boyle, Ed., "Requirements for Inter-Area MPLS Traffic Engineering", RFC 4105, DOI 10.17487/RFC4105, June 2005, <https://www.rfc-editor.org/info/rfc4105>.

[RFC4216] Zhang, R., Ed. and J. Vasseur, Ed., "MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements", RFC 4216, DOI 10.17487/RFC4216, November 2005, <https://www.rfc-editor.org/info/rfc4216>.

[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>.

[RFC4726] Farrel, A., Vasseur, J., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, DOI 10.17487/RFC4726, November 2006, <https://www.rfc-editor.org/info/rfc4726>.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 5226, DOI 10.17487/RFC5226, May 2008, <https://www.rfc-editor.org/info/rfc5226>.

[RFC5376] Bitar, N., Zhang, R., and K. Kumaki, "Inter-AS Requirements for the Path Computation Element Communication Protocol (PCECP)", RFC 5376, DOI 10.17487/RFC5376, November 2008, <https://www.rfc-editor.org/info/rfc5376>.

[RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, DOI 10.17487/RFC5394, December 2008, <https://www.rfc-editor.org/info/rfc5394>.

[RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, DOI 10.17487/RFC5520, April 2009, <https://www.rfc-editor.org/info/rfc5520>.

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

[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>.

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

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[RFC7897] Dhody, D., Palle, U., and R. Casellas, "Domain Subobjects for the Path Computation Element Communication Protocol (PCEP)", RFC 7897, DOI 10.17487/RFC7897, June 2016, <https://www.rfc-editor.org/info/rfc7897>.

[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-08 (work in progress), June 2018.

[I-D.ietf-pce-stateful-hpce]
Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., King, D., and O. Dios, "Hierarchical Stateful Path Computation Element (PCE).", draft-ietf-pce-stateful-hpce-05 (work in progress), June 2018.

Authors' Addresses

Fatai Zhang
Huawei
Huawei Base, Bantian, Longgang District
Shenzhen 518129
China

EMail: zhangfatai@huawei.com

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

EMail: quintin.zhao@huawei.com

Oscar Gonzalez de Dios
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid 28045
Spain

EMail: ogondio@tid.es

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n.7
Barcelona, Castelldefels
Spain

EMail: ramon.casellas@cttc.es

Daniel King
Old Dog Consulting
UK

EMail: daniel@olddog.co.uk