Stream:	Internet Engineering Task Force (IETF)				
RFC:	8685				
Category:	Standards Track				
Published:	November 2019				
ISSN:	2070-1721				
Authors:	F. Zhang	Q. Zhao	O. Gonzalez de Dios	R. Casellas	D. King
	Huawei	*Huawei*	*Telefonica I+D*	*CTTC*	*Old Dog Consulting*

RFC 8685
Path Computation Element Communication Protocol (PCEP) Extensions for the Hierarchical Path Computation Element (H-PCE) Architecture

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 Communication Protocol (PCEP) to support H-PCE procedures.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc8685.

Copyright Notice

Table of Contents

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 Path Computation Element (H-PCE) architecture that can be used for computing end-to-end paths for inter-domain MPLS-TE and GMPLS LSPs.

In the H-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 PCC sends the inter-domain Path Computation Request (PCReq) messages [RFC5440] 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 PCReq messages (between the domain border nodes) to the child PCEs that 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) path computation [RFC5152] and Backward-Recursive PCE-Based Computation (BRPC) [RFC5441], may be used.

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

1.1. Scope

The following functions are out of scope for this document:

- Determination of the destination domain (Section 4.5 of [RFC6805]):
  - via a collection of reachability information from child domains,
  - via requests to the child PCEs to discover if they contain the destination node, or
  - via any other methods.
- Parent Traffic Engineering Database (TED) methods (Section 4.4 of [RFC6805]), although suitable mechanisms include:
  - YANG-based management interfaces.
  - BGP - Link State (BGP-LS) [RFC7752].
  - Future extensions to PCEP (for example, see [PCEP-LS]).
- Learning of domain connectivity and border node addresses. Methods to achieve this function include:
  - YANG-based management interfaces.
  - BGP-LS [RFC7752].
  - Future extensions to PCEP (for example, see [PCEP-LS]).
- Stateful PCE operations. (Refer to [STATEFUL-HPCE].)
- Applicability of the H-PCE model 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 Label Switching Routers (LSRs) is limited. As highlighted in [RFC7399], applying the H-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] and [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 the H-PCE Architecture

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

2.1. Path Computation Requests

The PCReq messages [RFC5440] 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 (OFs).
- 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 as follows:

- The ability for a child PCE to indicate that a PCReq message 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 PD path computation per [RFC5152] or a BRPC procedure [RFC5441].
- As stated in [RFC6805], Section 4.5, if a PCC knows the egress domain, it can supply this information as part of the PCReq message. 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 a parent PCE should be capable of disallowing re-entry into a specified domain.

2.1.2. Multi-domain Objective Functions

For H-PCE inter-domain path computation, there are three new OFs 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 OF code to use when requesting inter-domain path computation. It may also include intra-domain OFs, such as Minimum Cost Path (MCP) [RFC5541], which must be considered by participating child PCEs.

2.1.3. Multi-domain Metrics

For inter-domain path computation, there are two 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. See Section 3.4 for details.

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 this fact and negotiate its role with its PCEP peers. It does this using the "H-PCE Capability" TLV, as described in Section 3.2.1, in the OPEN object [RFC5440] to advertise its support for PCEP extensions for the 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 Autonomous System (AS).

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

2.4. Domain Diversity

"Domain diversity" in the context of a multi-domain environment is defined in [RFC6805] and described as follows:

> 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 because of some geo-political reasons and commercial relationships that would require domain diversity. For example, a pair of paths should choose different transit ASes because of certain 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 domain-diversity techniques (node, link, Shared Risk Link Group (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 Request Parameters (RP) object ([RFC5440] and Section 3.3), the OF codes (Section 3.4.1), and the OF object (Section 3.4.2). A PCC and a child PCE could also exchange the H-PCE 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

This document defines two new TLVs to be carried in 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 the H-PCE capability of PCEP speakers.

Its format is shown in the following figure:

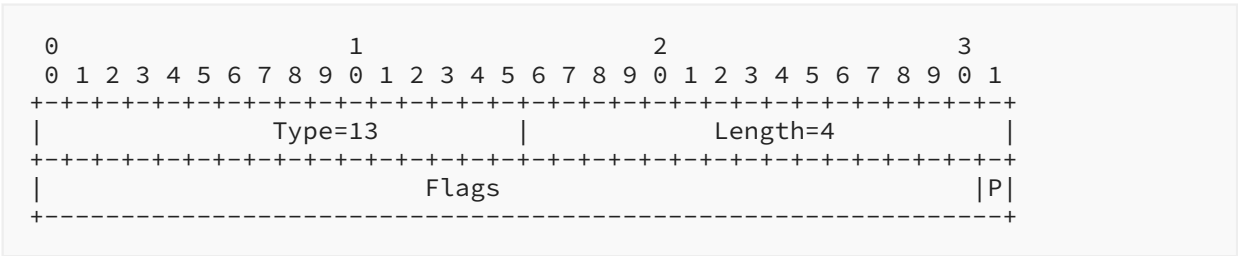

Figure 1: H-PCE-CAPABILITY TLV Format

The type of the TLV is 13, 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 a PCErr message using Error-Type 1 (PCEP session establishment failure) as per [RFC5440].

If the PCE understands the H-PCE PCReq message but did not advertise its H-PCE capability, it **MUST** send a PCErr message with Error-Type=28 (H-PCE Error) and Error-Value=1 (H-PCE Capability not advertised).

3.2.1.1. Backwards Compatibility

Section 7.1 of [RFC5440] specifies the following requirement: "Unrecognized TLVs **MUST** be ignored."

The OPEN object [RFC5440] contains the necessary PCEP information between the PCE entities, including session information and PCE capabilities via TLVs (including if H-PCE is supported). If the PCE does not support this document but receives an Open message containing an OPEN object that includes an H-PCE-CAPABILITY TLV, it will ignore that TLV and continue to attempt to establish a PCEP session. However, it will 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 an H-PCE relationship), the receiving PCE will ignore the H-PCE related parameters because they are all encoded in TLVs in 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 provide the domain information to the parent PCE.

The Domain-ID TLV is defined below:

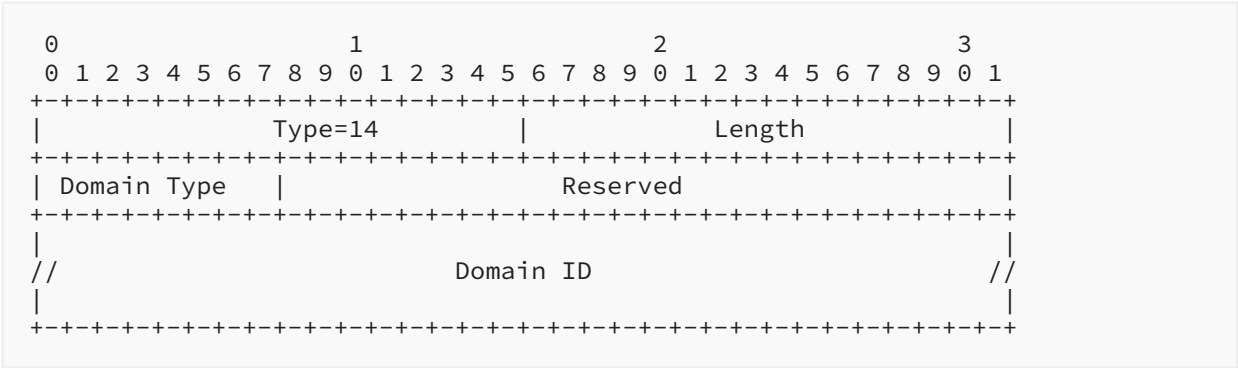

Figure 2: Domain-ID TLV Format

The type of the TLV is 14, and it has a variable Length of the value portion. The value part comprises the following:

Domain Type (8 bits): Indicates the domain type. Four types of domains 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 a 2-byte Area-Len and a variable-length IS-IS area ID. Padded with trailing zeros to a 4-byte boundary.

Reserved: Zero at transmission; ignored on receipt.

Domain ID (variable): Indicates an IGP area ID or AS number as per the Domain Type field. 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 the case of IS-IS, it includes the Area-Len as well.

In the case where a PCE serves more than one domain, multiple Domain-ID TLVs are included for each domain it serves.

3.3. RP Object

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 PCReq message and options.

Its format is shown in the following figure:

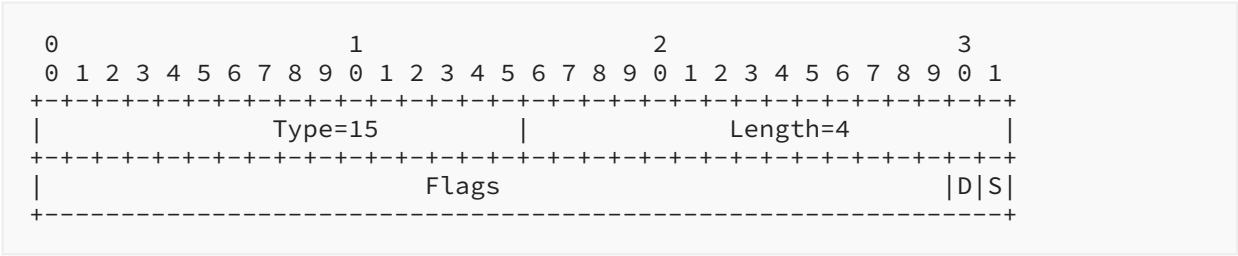

Figure 3: H-PCE-FLAG TLV Format

The type of the TLV is 15, and it has a fixed length of 4 octets.

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

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

S (Domain Sequence bit):
: If set, will signal that the child PCE wishes to get only the domain sequence in the Path Computation Reply (PCRep) message [RFC5440]. Refer to Section 3.7 of [RFC7897] for details.

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 managed domain (or a list of managed domains) and is described in Section 3.2.2. 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. Section 3.2.2 also defines the format for this TLV and the procedure for using it.

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 the NO-PATH-VECTOR TLV should be used. A new bit number is assigned to indicate "Destination is not found in the indicated domain" (see Section 3.8).

3.4. Objective Functions

3.4.1. OF Codes

[RFC5541] defines a mechanism to specify an OF that is used by a PCE when it computes a path. Three new OFs are defined for the H-PCE model; these are:

- MTD

  Name: Minimize the number of Transit Domains (MTD)

OF code: 12

Description: Find a path P such that it passes through the least number of transit domains.

- OFs are formulated using the following terminology:
  - A network comprises a set of N domains {Di, (i=1...N)}.
  - A path P passes through K unique domains {Dpi, (i=1...K)}.
  - Find a path P such that the value of K is minimized.
- MBN

  Name: Minimize the number of Border Nodes.

  OF code: 13

  Description: Find a path P such that it passes through the least number of border nodes.
- OFs are formulated using the following terminology:
  - A network comprises a set of N links {Li, (i=1...N)}.
  - A path P is a list of K links {Lpi, (i=1...K)}.
  - D(Lpi) is a function that determines if the links Lpi and Lpi+1 belong to different domains. D(Li) = 1 if link Li and Li+1 belong to different domains; D(Lk) = 0 if link Lk and Lk+1 belong to the same domain.
  - The number of border nodes in a path P is denoted by B(P), where B(P) = sum{D(Lpi), (i=1...K-1)}.
  - Find a path P such that B(P) is minimized.

There is one OF that applies to a set of synchronized PCReq messages to increase the domain diversity:

- MCTD

  Name: Minimize the number of Common Transit Domains.

  OF code: 14

  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 {Di, (i=1...N)}.
  - A path P passes through K unique domains {Dpi, (i=1...K)}.
  - A set of paths {P1...Pm} has L transit domains that are common to more than one path {Dpi, (i=1...L)}.
  - Find a set of paths such that the value of L is minimized.

3.4.2. OF Object

The OF object [RFC5541] is carried in a PCReq message so as to indicate the desired/required OF 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 PCReq message.

The new OF codes described in Section 3.4.1 are applicable to 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 the 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 PCReq message 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 OFs defined in this document. The OF code inside the OF-List TLV **MUST NOT** include an H-PCE OF. 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=23 (Incompatible OF codes in H-PCE).

If the OFs defined in this document are unknown or unsupported by a PCE, then the procedure as defined in [RFC5440] is followed.

3.5. METRIC Object

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

T=20: Domain Count metric (number of domains crossed).

T=21: 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 the 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 requirement for the number of domains or border nodes being crossed. 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 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 that meets 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 the Synchronization Vector (SVEC) object, which includes flags for the potential dependency between the set of PCReq messages (Link, Node, and SRLG diverse). This document defines a new flag (the O-bit) for domain diversity.

The following new bit is added to the Flags field:

Domain Diverse O-bit - 18:
: When set, this indicates that the computed paths corresponding to the requests specified by any RP objects that might be provided **MUST NOT** have any transit domains in common.

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

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

3.7. PCEP-ERROR Object

3.7.1. Hierarchical PCE Error-Type

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

Error-Type	Meaning
28	H-PCE Error Error-Value=1: H-PCE Capability not advertised Error-Value=2: Parent PCE Capability cannot be provided

Table 1: H-PCE Error

3.8. 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.

This document defines four new bit flags in the "NO-PATH-VECTOR TLV Flag Field" subregistry. These flags are to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH object.

Bit number 22: When set, the parent PCE indicates that the destination domain is unknown.

Bit number 21: When set, the parent PCE indicates that one or more child PCEs are unresponsive.

Bit number 20: When set, the parent PCE indicates that no resources are available in one or more domains.

Bit number 19: 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-flag (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 OFs. This information can be used for selecting a proper parent PCE when a child PCE wants to get a path that satisfies a certain OF.

When a child PCE sends a PCReq to a peer PCE that requires parental activity and the H-PCE-CAPABILITY TLV but these items were not taken into account in the session establishment procedure described above, the peer PCE **SHOULD** send a PCErr message to the child PCE and **MUST** specify Error-Type=28 (H-PCE Error) and Error-Value=1 (H-PCE Capability 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 Error-Type=28 (H-PCE Error) and Error-Value=2 (Parent PCE Capability cannot be provided) in the PCEP-ERROR object.

4.2. Procedure for Obtaining the 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 that receives the PCReq message tries to compute a domain sequence for it (instead of the end-to-end path). If the domain path computation succeeds, the parent PCE sends a PCRep message that carries the domain sequence in the Explicit Route Object (ERO) to the child PCE. Refer to [RFC7897] for more details about domain subobjects in the ERO. Otherwise, it sends a PCReq message that 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. When the child PCE sends a PCReq that requires parental activity, a negative response in the form of a PCEP Error (PCErr) message that includes H-PCE Error-Type=28 (H-PCE Error) and an applicable Error-Value (Section 3.7) might result.

Additionally, the parent PCE may fail to find the multi-domain path or domain sequence for 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 OFs specified in the path request cannot be met.

In this case, the parent PCE **MAY** need to send a negative PCRep message specifying the reason for the failure. This can be achieved by including the NO-PATH object in the PCRep message. An extension to the NO-PATH object is needed in order to include the reasons defined in Section 3.8.

6. Manageability Considerations

General PCE and PCEP management/manageability considerations are discussed in [RFC4655] and [RFC5440]. There are additional management considerations for the H-PCE model; these 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:

- Multiple domains (e.g., IGP areas or multiple ASes) in a single service provider network. The management responsibility for the parent PCE would most likely be handled by the service provider.
- Multiple ASes in 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 of H-PCE function will need to be carefully managed via configuration and interaction policies, which may be controlled via a policy module on the H-PCE. 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 H-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 behavior of the child and parent PCEs is described in the following subsections.

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 PCReq messages 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 (1) the potential for impersonation and repudiation and (2) operational difficulties should revocation be required.

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

6.1.3. Policy Control

It may be necessary to maintain H-PCE policy [RFC5394] via a policy control module on the parent PCE. 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

[RFC7420] provides a MIB module for PCEP and describes managed objects for the modeling of PCEP communication. A YANG module for PCEP has also been proposed [PCEP-YANG].

An H-PCE MIB module or an additional data model will also be required for reporting 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. Verifying 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 H-PCE procedure completions on a per-PCE-peer basis,
- number of H-PCE procedure-completion failures on a per-PCE-peer basis, and
- number of H-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 H-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. IANA has allocated code points for the protocol elements defined in this document.

7.1. PCEP TLV Type Indicators

IANA maintains the "PCEP TLV Type Indicators" subregistry (see [RFC5440]) within the "Path Computation Element Protocol (PCEP) Numbers" registry.

IANA has allocated the following three new PCEP TLVs:

Type	TLV Name	Reference
13	H-PCE-CAPABILITY	RFC 8685
14	Domain-ID	RFC 8685
15	H-PCE-FLAG	RFC 8685

Table 2: New PCEP TLVs

7.2. H-PCE-CAPABILITY TLV Flags

IANA has created the "H-PCE-CAPABILITY TLV Flag Field" subregistry 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 assigned by Standards Action [RFC8126]. Each registered bit should include the following information:

- Bit number (counting from bit 0 as the most significant bit)
- Capability description
- Defining RFC

The following value is defined in this document:

Bit	Description	Reference
31	P (Parent PCE Request bit)	RFC 8685

Table 3: Parent PCE Request Bit

7.3. Domain-ID TLV Domain Type

IANA has created the "Domain-ID TLV DomainType" subregistry 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 subregistry is IETF Review [RFC8126].

The following value is defined in this document:

Value	Meaning
0	Reserved
1	2-byte AS number
2	4-byte AS number
3	4-byte OSPF area ID

Value	Meaning
4	Variable-length IS-IS area ID
5-255	Unassigned

Table 4: Parameters for Domain-ID TLV Domain Type

7.4. H-PCE-FLAG TLV Flags

IANA has created the "H-PCE-FLAG TLV Flag Field" subregistry within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field in the H-PCE-FLAG TLV of the PCEP RP object. New values are to be assigned by Standards Action [RFC8126]. Each registered bit should include the following information:

- Bit number (counting from bit 0 as the most significant bit)
- Capability description
- Defining RFC

The following values are defined in this document:

Bit	Description	Reference
30	D (Disallow Domain Re-entry bit)	RFC 8685
31	S (Domain Sequence bit)	RFC 8685

Table 5: New H-PCE-FLAG TLV Flag Field Entries

7.5. OF Codes

IANA maintains a list of OFs (described in [RFC5541]) in the "Objective Function" subregistry within the "Path Computation Element Protocol (PCEP) Numbers" registry.

IANA has allocated the following OFs:

Code Point	Name	Reference
12	Minimize the number of Transit Domains (MTD)	RFC 8685
13	Minimize the number of Border Nodes (MBN)	RFC 8685
14	Minimize the number of Common Transit Domains (MCTD)	RFC 8685

Table 6: New OF Codes

7.6. METRIC Object Types

IANA maintains the "METRIC Object T Field" subregistry [RFC5440] within the "Path Computation Element Protocol (PCEP) Numbers" registry.

The following two new metric types for the METRIC object are defined in this document:

Value	Description	Reference
20	Domain Count metric	RFC 8685
21	Border Node Count metric	RFC 8685

Table 7: New METRIC Object Types

7.7. New PCEP Error-Types and Values

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

IANA has allocated the following:

Error-Type	Meaning and Error Values	Reference
28	H-PCE Error Error-Value=1: H-PCE Capability not advertised Error-Value=2: Parent PCE Capability cannot be provided	RFC 8685
10	Reception of an invalid object Error-Value=23: Incompatible OF codes in H-PCE	RFC 5440 RFC 8685

Table 8: New PCEP Error-Types and Values

7.8. New NO-PATH-VECTOR TLV Bit Flag

IANA maintains the "NO-PATH-VECTOR TLV Flag Field" subregistry, which contains a list of bit flags carried in the PCEP NO-PATH-VECTOR TLV in the PCEP NO-PATH object as defined in [RFC5440].

IANA has allocated four new bit flags as follows:

Bit Number	Description	Reference
22	Destination domain unknown	RFC 8685
21	Unresponsive child PCE(s)	RFC 8685
20	No available resource in one or more domains	RFC 8685
19	Destination is not found in the indicated domain	RFC 8685

Table 9: PCEP NO-PATH Object Flags

7.9. SVEC Flag

IANA maintains the "SVEC Object Flag Field" subregistry, which contains a list of bit flags carried in the PCEP SVEC object as defined in [RFC5440].

IANA has allocated the following new bit flag:

Bit Number	Description	Reference
18	Domain Diverse O-bit	RFC 8685

Table 10: Domain Diverse O-Bit

8. Security Considerations

The H-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 the TCP Authentication Option (TCP-AO) [RFC5925] or Transport Layer Security (TLS) [RFC8253] [RFC8446].

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 information related to domain topology and capabilities as possible.

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

9. References

9.1. Normative References

[RFC2119]

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

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

[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>.

9.2. Informative References

[RFC4105] Le Roux, J.-L., Ed., Vasseur, J.-P., 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.-P. 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.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <https://www.rfc-editor.org/info/rfc4655>.

[RFC4726] Farrel, A., Vasseur, J.-P., 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>.

[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>.

[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>.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <https://www.rfc-editor.org/info/rfc5925>.

[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>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <https://www.rfc-editor.org/info/rfc7399>.

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

[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>.

[RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <https://www.rfc-editor.org/info/rfc8126>.

[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>.

[RFC8446] Rescorla, E., "The Transport Layer Security (TLS) Protocol Version 1.3", RFC 8446, DOI 10.17487/RFC8446, August 2018, <https://www.rfc-editor.org/info/rfc8446>.

[PCEP-YANG] Dhody, D., Hardwick, J., Beeram, V., and J. Tantsura, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", Work in Progress, Internet-Draft, draft-ietf-pce-pcep-yang-13, 31 October 2019, <https://tools.ietf.org/html/draft-ietf-pce-pcep-yang-13>.

[STATEFUL-HPCE] Dhody, D., Lee, Y., Ceccarelli, D., Shin, J., and D. King, "Hierarchical Stateful Path Computation Element (PCE)", Work in Progress, Internet-Draft, draft-ietf-pce-stateful-hpce-15, 20 October 2019, <https://tools.ietf.org/html/draft-ietf-pce-stateful-hpce-15>.

[PCEP-LS] Dhody, D., Lee, Y., and D. Ceccarelli, "PCEP Extension for Distribution of Link-State and TE Information.", Work in Progress, Internet-Draft, draft-dhodylee-pce-pcep-ls-14, 21 October 2019, <https://tools.ietf.org/html/draft-dhodylee-pce-pcep-ls-14>.

Acknowledgements

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

Contributors

The following people contributed substantially to the content of this document and should be considered coauthors:

```
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
```

Authors' Addresses

Fatai Zhang
China
518129
Shenzhen
Huawei Base, Bantian, Longgang District
Huawei
Email: zhangfatai@huawei.com

Quintin Zhao
Huawei
125 Nagog Technology Park
Acton, MA 01719
United States of America
Email: quintinzhao@gmail.com

Oscar Gonzalez de Dios
Telefonica I+D
Don Ramon de la Cruz 82-84
28045 Madrid
Spain
Email: oscar.gonzalezdedios@telefonica.com

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n.7
Castelldefels Barcelona
Spain
Email: ramon.casellas@cttc.es

Daniel King
Old Dog Consulting
United Kingdom
Email: daniel@olddog.co.uk