Workgroup: PCE Working Group
Internet-Draft: draft-chen-pce-h-discovery-10
Published: 10 January 2022
Intended Status: Standards Track
Expires: 14 July 2022
Authors: H. Chen, Futurewei; M. Toy, Verizon; X. Liu, Volta Networks; L. Liu, Fujitsu; Z. Li, China Mobile

Hierarchical PCE Determination

Abstract

This document presents extensions to the Path Computation Element Communication Protocol (PCEP) for determining parent child relations and exchanging the information between a parent and a child PCE in a hierarchical PCE system.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

A hierarchical PCE architecture is described in RFC 6805, in which a parent PCE has a number of child PCEs. A child PCE may also be a parent PCE, which has multiple child PCEs.

For a parent PCE, it needs to obtain the information about each of its child PCEs. The information about a child PCE comprises the address or ID of the PCE and the domain for which the PCE is responsible. It may also include the position of the PCE, which indicates whether the PCE is a leaf (i.e., only a child) or branch (i.e., a child and also a parent). In addition, the information may indicate whether the child PCE and its responsible domain is in a same organization as the parent PCE.

For a child PCE, it needs to obtain the information about its parent PCE, which includes the address or ID of the parent PCE. The information may also indicate whether the parent PCE is in a same organization as the child PCE.

After a user configures a parent PCE and a child PCE over a session, this parent child PCE relation needs to be determined in the protocol level. This is similar to OSPF and BGP. After an adjacency between two OSPF routers is configured by a user, the OSPF protocol (refer to RFC 2328, Section 7) will determine whether the adjacency is allowed based on the parameters configured, and forms the OSPF adjacency after the determination. After a peer relation between two BGP routers is configured by a user, the BGP protocol (refer to RFC 4271, Section 8) will determine whether the peer is allowed based on the parameters configured, and forms the BGP peer relation after the determination.

For a parent child PCE relation determination, the PCE protocol needs to check or confirm whether the parent child PCE relation is allowed based on the parameters configured. If so, the child PCE has to send its parent PCE the information about it and vice versa.

This document presents extensions to the Path Computation Element Communication Protocol (PCEP) for determining parent child relations and exchanging the information between a parent and a child PCE in a hierarchical PCE system.

2. Terminology

The following terminology is used in this document.

Parent Domain: A domain higher up in a domain hierarchy such that it contains other domains (child domains) and potentially other links and nodes.

Child Domain: A domain lower in a domain hierarchy such that it has a parent domain.

Parent PCE: A PCE responsible for selecting a path across a parent domain and any number of child domains by coordinating with child PCEs and examining a topology map that shows domain inter-connectivity.

Child PCE: A PCE responsible for computing the path across one or more specific (child) domains. A child PCE maintains a relationship with at least one parent PCE.

TED: Traffic Engineering Database.

This document uses terminology defined in [RFC5440].

3. Conventions Used in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

4. Extensions to PCEP

This section describes the extensions to PCEP for determining the relation between a parent PCE and a child PCE and exchanging the information between a parent and a child PCE in a hierarchical PCE system. A child PCE is simply called a child and a parent PCE is called a parent in the following sections.

4.1. Determination of Parent Child Relation

During a PCEP session establishment between two PCEP speakers, each of them advertises its capabilities for Hierarchical PCE (H-PCE for short) through the Open Message with the Open Object containing a new TLV to indicate its capabilities for H-PCE. This new TLV is called H-PCE capability TLV. It has the following format.

```
  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             |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |P|C|S|B|                Capability Flags                       |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |                       Optional  Sub-TLVs                      |
 ~                                                               ~
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The type of the TLV is TBD1. It has a length of 4 octets plus the size of optional Sub-TLVs. The value of the TLV comprises a capability flags field of 32 bits, which are numbered from the most significant as bit zero. Some of them are defined as follows. The others are not defined and MUST be set to zero.

- P (Parent - 1 bit): Bit 0 is used as P flag. It is set to 1 indicating a parent.

- C (Child - 1 bit): Bit 1 is used as C flag. It is set to 1 indicating a child.

- S (Same Org - 1 bit): Bit 2 is used as S flag. It is set to 1 indicating a PCE in a same organization as its remote peer.

- B (Both - 1 bit): Bit 3 is used as B flag. It is set to 1 indicating a PCE as both a child and a parent.

The following Sub-TLVs are defined:

- A Domain Sub-TLV containing an AS number and optional area, and

- PCE-ID Sub-TLV containing the ID of a PCE.

4.2. Sub-TLVs

When a child sends its parent a Open message, it places the information about it in the message through using some optional Sub-TLVs. When a parent sends each of its child PCEs a Open message, it puts the information about it in the message.

4.2.1. Domain Sub-TLV

A domain is an AS or an area in an AS. An AS is identified by an AS number. An area in an AS is identified by the combination of the AS and the area. The former is indicated by an AS number and the latter by an area number. A domain is represented by a domain Sub-TLV containing an AS number and a optional area number.

The format of the domain Sub-TLV is shown 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 (tTBD1)          |             Length            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       AS Number (4 bytes)                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                      Optional  Area Number                    ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where Length is four plus size of area number.

An AS is represented by a domain Sub-TLV containing only the AS number of the AS. In this case, the Length is four. An area in an AS is represented by a domain Sub-TLV containing the AS number of the AS and the area number of the area. In this case, the Length is eight.

4.2.2. PCE ID Sub-TLV

An Identifier (ID) of a PCE (PCE ID for short) is a 32-bit number that uniquely identifies the PCE among all PCEs. This 32-bit number for PCE ID SHOULD NOT be zero.

The format of the PCE ID Sub-TLV is shown 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 (tTBD3)          |           Length (4)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         PCE ID (4 bytes)                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The PCE ID Sub-TLV specifies an non zero number as the identifier of the

Alternatively, an IP address attached to a PCE can also be used as an identifier of the PCE. The format of an IPv4 address Sub-TLV is shown 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 (tTBD4)          |           Length (4)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      IPv4 Address (4 bytes)                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The IPv4 address Sub-TLV specifies an IPv4 address associated with the PCE, which is used as the identifier of the PCE.

The format of an IPv6 address Sub-TLV is shown 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 (tTBD5)          |          Length (16)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      IPv6 Address (16 bytes)                  |
~                                                               ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The IPv6 Sub-TLV specifies an IPv6 address associated with the PCE, which is used as the identifier of the PCE.

4.3. Procedures

For two PCEs A and B configured as parent and child, they determine parent child relation through Open messages in the initialization phase. The following is a sequence of events related.

```
       A                                      B
     Configure B                          Configure A
     as its child                         as its parent
                     Open (P=1, A's ID)
                    --------------------> Same as configured
                                          A is B's parent
                     Open (C=1, B's ID)
 Same as configured <--------------------
   B is A's child
```

A sends B a Open message with P=1 and A's ID after B is configured as its child on it. B sends A a Open message with C=1 and B's ID after A is configured as its parent on it.

When A receives the Open message from B and determines C=1 and the PCE ID of B in the message is the same as the PCE ID of the child locally configured, B is A's child.

When B receives the Open message from A and determines P=1 and the PCE ID of A in the message is the same as the PCE ID of the parent locally configured, A is B's parent.

The Open message from child B to its parent A contains B's domain, which is represented by a domain Sub-TLV in the H-PCE capability TLV. If child B is also a parent, the B flag in the TLV is set to 1.

The PCE ID in a Open message may be represented in one of the following ways:

- The source IP address of the message (i.e., PCE session).

- A PCE ID Sub-TLV in the H-PCE capability TLV.

- An IP address Sub-TLV in the H-PCE capability TLV.

When the IP address Sub-TLV is used, the address in the Sub-TLV MUST be the same as the source IP address of the PCE session.

For a child that is a leaf, it is normally responsible for one domain, which is contained in the message to its parent.

For a child that is a branch (i.e., also a parent of multiple child PCEs), it may be directly responsible for one domain, which is contained in the message to its parent. In addition, it is responsible for the domains of its child PCEs. In other words, it is responsible for computing paths crossing the domains through working together with its child PCEs. If these domains are all areas of an AS, the AS is included in the message to its parent.

A parent stores the information about each of its child PCEs received. When the session to one of them is down, it removes the information about the child on that session.

A child stores the information about its parent received. When the session to the parent is down, it removes the information about the parent.

If there already exists a session between A and B and the configurations on parent and child are issued on them, the procedures above may be executed through bringing down the existing session and establishing a new session between them. Alternatively, they may determine parent child relation through using extended Notification messages in the same procedures as using Open messages described above without bringing down the existing session.

The following new Notification-type and Notification-value are defined for H-PCE:

o Notification-type=5 (TBD): Determination of H-PCE

  * Notification-value=1: The information about a parent PCE or a child PCE. A Notification-type=5, Notification-value=1 indicates that the PCE sends its peer the information about it and a TLV containing the information is in the Notification object. The format and contents of the TLV is the same as the H-PCE capability TLV described above. The only difference may be the type of the TLV.

5. Security Considerations

The mechanism described in this document does not raise any new security issues for the PCEP protocols.

6. IANA Considerations

This section specifies requests for IANA allocation.

7. Acknowledgement

The authors would like to Jescia Chen, Adrian Farrel for their valuable comments on this draft.

8. References

8.1. 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>.

[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>.

[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>.

8.2. Informative References

[RFC2328]

Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <https://www.rfc-editor.org/info/rfc2328>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

Authors' Addresses

Huaimo Chen
Futurewei
Boston, MA,
United States of America

Email: Huaimo.chen@futurewei.com

Mehmet Toy
Verizon
United States of America

Email: mehmet.toy@verizon.com

Xufeng Liu
Volta Networks
McLean, VA
United States of America

Email: xufeng.liu.ietf@gmail.com

Lei Liu
Fujitsu
United States of America

Email: liulei.kddi@gmail.com

Zhenqiang Li
China Mobile
No.32 Xuanwumenxi Ave., Xicheng District
Beijing
100032
P.R. China

Email: li_zhenqiang@hotmail.com