

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2011

H. Chen
Huawei Technology, Inc.
October 18, 2010

Extensions to the Path Computation Element Communication Protocol (PCEP)
for Backup Egress of a Traffic Engineering Label Switched Path
draft-chen-pce-compute-backup-egress-00.txt

Abstract

This document presents extensions to the Path Computation Element Communication Protocol (PCEP) for a PCC to send a request for computing a backup egress for an MPLS TE P2MP LSP or an MPLS TE P2P LSP to a PCE and for a PCE to compute the backup egress and reply to the PCC with a computation result for the backup egress.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 21, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

described in the Simplified BSD License.

Table of Contents

| | |
|--|----|
| 1. Introduction | 3 |
| 2. Terminology | 3 |
| 3. Conventions Used in This Document | 3 |
| 4. Extensions to PCEP | 3 |
| 4.1. Backup Egress Capability Advertisement | 4 |
| 4.1.1. Capability TLV in Existing PCE Discovery Protocol | 4 |
| 4.1.2. Open Message Extension | 6 |
| 4.2. Request and Reply Message Extension | 7 |
| 4.2.1. RP Object Extension | 7 |
| 4.2.2. External Destination Nodes Object | 8 |
| 4.2.3. Constraints between Egress and Backup Egress | 10 |
| 4.2.4. Constraints for Backup Path | 11 |
| 4.2.5. Backup Egress Node | 11 |
| 4.2.6. Backup Egress PCEP Error Objects and Types | 11 |
| 4.2.7. Request Message Format | 12 |
| 4.2.8. Reply Message Format | 12 |
| 5. Security Considerations | 13 |
| 6. IANA Considerations | 13 |
| 6.1. Backup Egress Capability Flag | 13 |
| 6.2. Backup Egress Capability TLV | 14 |
| 6.3. Request Parameter Bit Flags | 14 |
| 6.4. PCEP Objects | 14 |
| 7. Acknowledgement | 15 |
| 8. References | 15 |
| 8.1. Normative References | 15 |
| 8.2. Informative References | 15 |
| Author's Address | 15 |

1. Introduction

"A Path Computation Element-(PCE) Based Architecture" RFC4655 describes a set of building blocks for constructing solutions to compute Point-to-Point (P2P) Traffic Engineering (TE) label switched paths across multiple areas or Autonomous System (AS) domains. A typical PCE-based system comprises one or more path computation servers, traffic engineering databases (TED), and a number of path computation clients (PCC). A routing protocol is used to exchange traffic engineering information from which the TED is constructed. A PCC sends a Point-to-Point traffic engineering Label Switched Path (LSP) computation request to the path computation server, which uses the TED to compute the path and responses to the PCC with the computed path. A path computation server is named as a PCE. The communications between a PCE and a PCC for Point-to-Point label switched path computations follow the PCE communication protocol (PCEP).

"Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths" RFC6006 describes extensions to the PCE communication Protocol (PCEP) to handle requests and responses for the computation of paths for P2MP TE LSPs.

This document defines extensions to the Path Computation Element Communication Protocol (PCEP) for a PCC to send a request for computing a backup egress node for an MPLS TE P2MP LSP or an MPLS TE P2P LSP to a PCE and for a PCE to compute the backup egress node and reply to the PCC with a computation result for the backup egress node.

2. Terminology

This document uses terminologies defined in RFC5440, and RFC4875.

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 RFC 2119.

4. Extensions to PCEP

This section describes the extensions to PCEP for computing a backup egress of an MPLS TE P2MP LSP and an MPLS TE P2P LSP.

4.1. Backup Egress Capability Advertisement

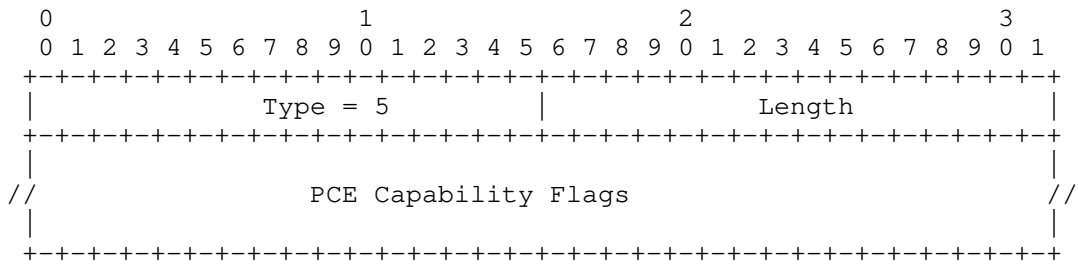
4.1.1. Capability TLV in Existing PCE Discovery Protocol

There are two options for advertising a PCE capability for computing a backup egress for an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

The first option is to define a new flag in the OSPF and IS-IS PCE Capability Flags to indicate the capability that a PCE is capable to compute both a backup egress for an MPLS TE P2MP LSP and a backup egress for an MPLS TE P2P LSP.

The second option is to define two new flags. One new flag in the OSPF and IS-IS PCE Capability Flags indicates the capability that a PCE is capable to compute a backup egress for an MPLS TE P2MP LSP; and another new flag in the OSPF and IS-IS PCE Capability Flags indicates the capability that a PCE is capable to compute a backup egress for an MPLS TE P2P LSP.

The format of the PCE-CAP-FLAGS sub-TLV is as follows:



Type: 5
 Length: Multiple of 4 octets
 Value: This contains an array of units of 32-bit flags
 numbered from the most significant as bit zero, where
 each bit represents one PCE capability.

The following capability bits have been assigned by IANA:

| Bit | Capabilities |
|-------|--|
| 0 | Path computation with GMPLS link constraints |
| 1 | Bidirectional path computation |
| 2 | Diverse path computation |
| 3 | Load-balanced path computation |
| 4 | Synchronized path computation |
| 5 | Support for multiple objective functions |
| 6 | Support for additive path constraints (max hop count, etc.) |
| 7 | Support for request prioritization |
| 8 | Support for multiple requests per message |
| 9 | Global Concurrent Optimization (GCO) |
| 10 | P2MP path computation |
| 11-31 | Reserved for future assignments by IANA. |

Reserved bits SHOULD be set to zero on transmission and MUST be ignored on receipt.

For the first option, one bit such as bit 13 may be assigned to indicate that a PCE is capable to compute both a backup egress for an MPLS TE P2MP LSP and a backup egress for an MPLS TE P2P LSP.

| Bit | Capabilities |
|-------|--|
| 13 | Backup egress computation for P2MP LSP and P2P LSP |
| 14-31 | Reserved for future assignments by IANA. |

For the second option, one bit such as bit 13 may be assigned to indicate that a PCE is capable to compute a backup egress for an MPLS TE P2MP LSP and another bit such as bit 14 may be assigned to indicate that a PCE is capable to compute a backup egress for an MPLS TE P2P LSP.

| Bit | Capabilities |
|-------|--|
| 13 | Backup egress computation for P2MP LSP |
| 14 | Backup egress computation for P2P LSP |
| 15-31 | Reserved for future assignments by IANA. |

4.1.2. Open Message Extension

If a PCE does not advertise its backup egress computation capability during discovery, PCEP should be used to allow a PCC to discover, during the Open Message Exchange, which PCEs are capable of supporting backup egress computation.

To achieve this, we extend the PCEP OPEN object by defining a new optional TLV to indicate the PCE's capability to perform backup egress computation for an MPLS TE P2MP LSP and an MPLS TE P2P LSP.

We request IANA to allocate a value such as 8 from the "PCEP TLV Type Indicators" subregistry, as documented in Section below ("Backup Egress Capability TLV"). The description is "backup egress capable", and the length value is 2 bytes. The value field is set to indicate the capability of a PCE for backup egress computation for an MPLS TE LSP in details.

There are two options to indicate a PCE's capability for computing a backup egress for an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

The first option is to use one bit such as bit 2 in the value field to indicate that a PCE is capable to compute both a backup egress for an MPLS TE P2MP LSP and a backup egress for an MPLS TE P2P LSP.

The second option is to use one bit such as bit 2 in the value field to indicate that a PCE is capable to compute a backup egress for an MPLS TE P2MP LSP; and another bit such as bit 3 in the value field to indicate that a PCE is capable to compute a backup egress for an MPLS TE P2P LSP.

The inclusion of this TLV in an OPEN object indicates that the sender can perform backup egress computation for an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

The capability TLV is meaningful only for a PCE, so it will typically appear only in one of the two Open messages during PCE session establishment. However, in case of PCE cooperation (e.g., inter-domain), when a PCE behaving as a PCC initiates a PCE session it SHOULD also indicate its path computation capabilities.

4.2. Request and Reply Message Extension

This section describes extensions to the existing RP (Request Parameters) object to allow a PCC to request a PCE for computing a backup egress of an MPLS TE P2MP LSP or an MPLS TE P2P LSP when the PCE receives the PCEP request.

4.2.1. RP Object Extension

The following flags are added into the RP Object:

The T bit is added in the flag bits field of the RP object to tell the receiver of the message that the request/reply is for computing a backup egress of an MPLS TE P2MP LSP and an MPLS TE P2P LSP.

- o T (Backup Egress bit - 1 bit):

- 0: This indicates that this is not PCReq/PCRep for backup egress.

- 1: This indicates that this is PCReq or PCRep message for backup egress.

The IANA request is referenced in Section below (Request Parameter Bit Flags) of this document.

This T bit with the N bit defined in RFC 6006 can indicate whether a request/reply is for a backup egress of an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

- o T = 1 and N = 1: This indicates that this is a PCReq/PCRep message for backup egress of an MPLS TE P2MP LSP.

- o T = 1 and N = 0: This indicates that this is a PCReq/PCRep message for backup egress of an MPLS TE P2P LSP.

4.2.2. External Destination Nodes Object

In addition to the information about the path that an MPLS TE P2MP LSP or an MPLS TE P2P LSP traverses, a request message may comprise other information that may be used for computing the backup egress for the P2MP LSP or P2P LSP. For example, the information about an external destination node, to which data traffic is delivered from an egress node of the P2MP LSP or P2P LSP, is useful for computing a backup egress node.

The PCC can specify an external destination nodes (EDN) Object. In order to represent the external destination nodes efficiently, we define two types of encodes for the external destination nodes in the object.

One encode indicates that the EDN object contains an external destination node for every egress node of an MPLS TE P2MP LSP or an MPLS TE P2P LSP. The order of the external destination nodes in the object is the same as the egress node(s) of the P2MP LSP or P2P LSP contained in the PCE messages.

Another encode indicates that the EDN object contains a list of egress node and external destination node pairs. For an egress node and external destination node pair, the data traffic is delivered to the external destination node from the egress node of the LSP.

The format of the external destination nodes (EDN) object body for IPv4 with the first type of encodes is illustrated as follows:

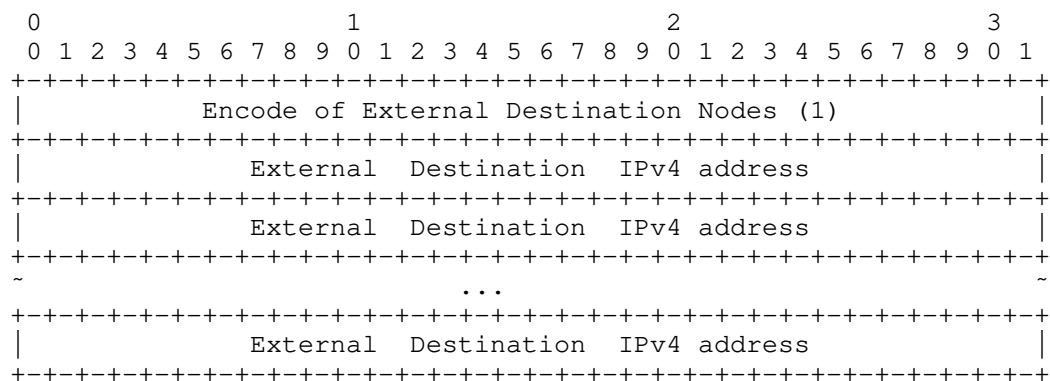


Figure 1: Format of EDN Object with one Encode for IPv4

The format of the external destination nodes (EDN) object body for IPv4 with the second type of encodes is illustrated below:

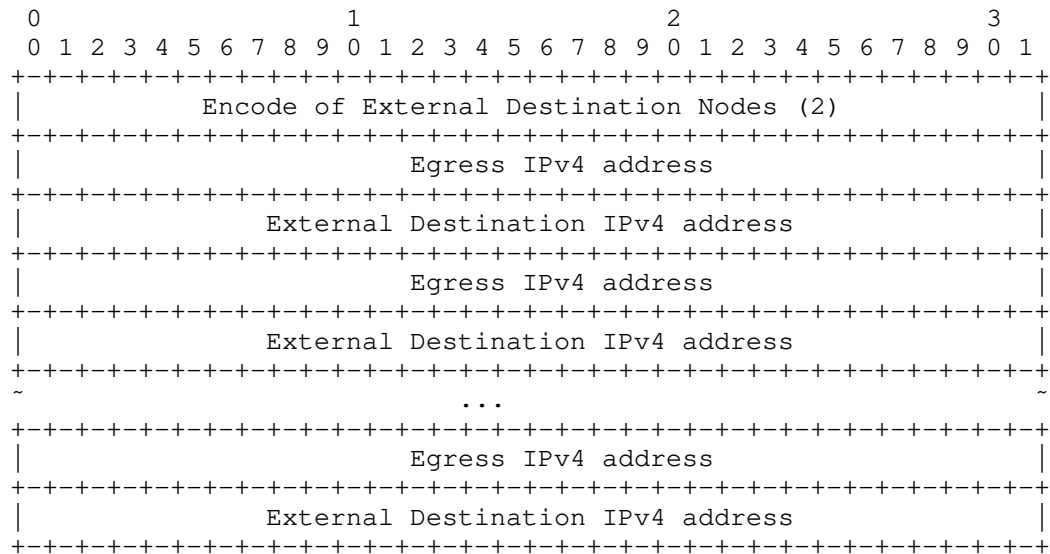


Figure 2: Format of EDN Object with another Encode for IPv4

The format of the external destination nodes (EDN) object body for IPv6 with the first type of encodes is illustrated as follows:

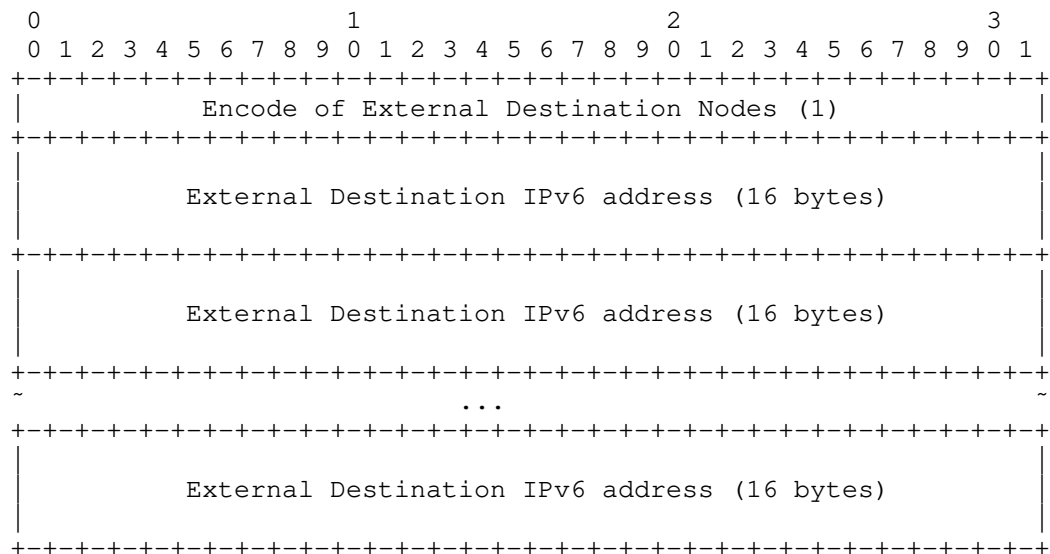


Figure 3: Format of EDN Object with one Encode for IPv6

The format of the external destination nodes (EDN) object body for IPv6 with the second type of encodes is illustrated below:

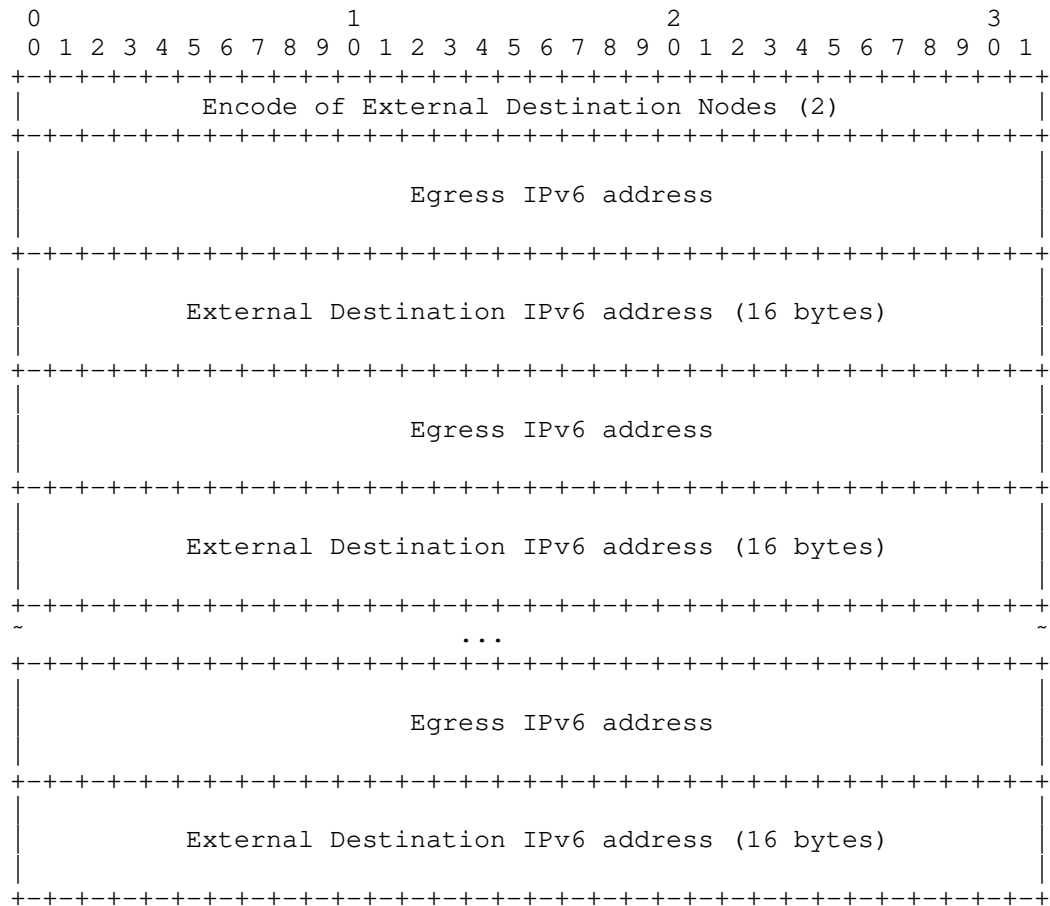


Figure 4: Format of EDN Object with another Encode for IPv6

The object can only be carried in a PCReq message. A Path Request may carry at most one external destination nodes Object.

The Object-Class and Object-types will need to be allocated by IANA. The IANA request is documented in Section below (PCEP Objects).

4.2.3. Constraints between Egress and Backup Egress

A request message sent to a PCE from a PCC for computing a backup egress of an MPLS TE P2MP LSP or an MPLS TE P2P LSP may comprise a

constraint indicating that there must be a path from the backup egress node to be computed to the egress node of the P2MP LSP or P2P LSP and that the length of the path is within a given hop limit such as one hop.

This constraint can be considered as default by a PCE or explicitly sent to the PCE by a PCC [TBD].

4.2.4. Constraints for Backup Path

A request message sent to a PCE from a PCC for computing a backup egress of a P2MP LSP or P2P LSP may comprise a constraint indicating that the backup egress node to be computed may not be a node on the P2MP LSP or P2P LSP. In addition, the request message may comprise a list of nodes, each of which is a candidate for the backup egress node.

A request message sent to a PCE from a PCC for computing a backup egress of a P2MP LSP or P2P LSP may comprise a constraint indicating that there must be a path from the previous hop node of the egress node of the P2MP LSP or P2P LSP to the backup egress node to be computed and that there is not an internal node of the path from the previous hop node of the egress node of the P2MP LSP or P2P LSP to the backup egress that is on the path of the P2MP LSP or P2P LSP.

Most of these constraints for the backup path can be considered as default by a PCE. The constraints for the backup path may be explicitly sent to the PCE by a PCC [TBD].

4.2.5. Backup Egress Node

The PCE may send a reply message to the PCC in return to the request message for computing a new backup egress node or a number of backup egress nodes. The reply message may comprise information about the computed backup egress node(s), which is contained in the path(s) from the previous-hop node of the egress node of the P2MP LSP or P2P LSP to the backup egress node(s) computed.

4.2.6. Backup Egress PCEP Error Objects and Types

In some cases, the PCE may not complete the backup egress computation as requested, for example based on a set of constraints. As such, the PCE may send a reply message to the PCC that indicates an unsuccessful backup egress computation attempt. The reply message may comprise a PCEP-error object, which may comprise an error-value, error-type and some detail information.

4.2.7. Request Message Format

The PCReq message is encoded as follows using RBNF as defined in [RFC5511].

Below is the message format for a request message:

```
<PCReq Message> ::= <Common Header>
                    [<svec-list>]
                    <request>
<request> ::= <RP>
              <end-point-rro-pair-list>
              [<OF>]
              [<LSPA>]
              [<BANDWIDTH>]
              [<metric-list>]
              [<EDNO>]
              [<IRO>]
              [<LOAD-BALANCING>]
```

where:

<EDNO> is an external destination nodes object.

Figure 5: The Format for a Request Message

The definitions for svec-list, RP, end-point-rro-pair-list, OF, LSPA, BANDWIDTH, metric-list, IRO, and LOAD-BALANCING are described in RFC5440 and RFC6006.

4.2.8. Reply Message Format

The PCRep message is encoded as follows using RBNF as defined in [RFC5511].

Below is the message format for a reply message:

```

<PCRep Message> ::= <Common Header>
                      <response>
<response> ::= <RP>
                <end-point-path-pair-list>
                [<NO-PATH>]
                [<attribute-list>]
where:

<end-point-path-pair-list> ::=
    [<END-POINTS>] <path> [<end-point-path-pair-list>]

<path> ::= (<ERO> | <SERO>) [<path>]

<attribute-list> ::= [<OF>]
                    [<LSPA>]
                    [<BANDWIDTH>]
                    [<metric-list>]
                    [<IRO>]

```

Figure 6: The Format for a Reply Message

The definitions for RP, NO-PATH, END-POINTS, OF, LSPA, BANDWIDTH, metric-list, IRO, and SERO are described in RFC5440, RFC6006 and RFC4875.

5. Security Considerations

The mechanism described in this document does not raise any new security issues for the PCEP, OSPF or IS-IS protocols.

6. IANA Considerations

This section specifies requests for IANA allocation.

6.1. Backup Egress Capability Flag

Two new OSPF Capability Flags are defined in this document to indicate the capabilities for computing a backup egress for an MPLS TE P2MP LSP and an MPLS TE P2P LSP. IANA is requested to make the assignment from the "OSPF Parameters Path Computation Element (PCE) Capability Flags" registry:

| Bit | Description | Reference |
|-----|----------------------------|-----------|
| 13 | Backup egress for P2MP LSP | This I-D |
| 14 | Backup egress for P2P LSP | This I-D |

6.2. Backup Egress Capability TLV

A new backup egress capability TLV is defined in this document to allow a PCE to advertize its backup egress computation capability. IANA is requested to make the following allocation from the "PCEP TLV Type Indicators" sub-registry.

| Value | Description | Reference |
|-------|-----------------------|-----------|
| 8 | Backup egress capable | This I-D |

6.3. Request Parameter Bit Flags

A new RP Object Flag has been defined in this document. IANA is requested to make the following allocation from the "PCEP RP Object Flag Field" Sub-Registry:

| Bit | Description | Reference |
|-----|-----------------------|-----------|
| 15 | Backup egress (T-bit) | This I-D |

6.4. PCEP Objects

An External Destination Nodes Object-Type is defined in this document. IANA is requested to make the following Object-Type allocation from the "PCEP Objects" sub-registry:

| | |
|--------------------|--|
| Object-Class Value | 34 |
| Name | External Destination Nodes |
| Object-Type | 1: IPv4 2: IPv6 3-15: Unassigned |
| Reference | This I-D |

7. Acknowledgement

The author would like to thank Quintin Zhao and others 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, March 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.
- [RFC6006] Zhao, Q., King, D., Verhaeghe, F., Takeda, T., Ali, Z., and J. Meuric, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 6006, September 2010.

8.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5862] Yasukawa, S. and A. Farrel, "Path Computation Clients (PCC) - Path Computation Element (PCE) Requirements for Point-to-Multipoint MPLS-TE", RFC 5862, June 2010.

Author's Address

Huaimo Chen
Huawei Technology, Inc.
Boston, MA
US

Email: Huaimochen@huawei.com

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2011

H. Chen
Huawei Technology, Inc.
October 18, 2010

Extensions to the Path Computation Element Communication Protocol (PCEP)
for Backup Ingress of a Traffic Engineering Label Switched Path
draft-chen-pce-compute-backup-ingress-00.txt

Abstract

This document presents extensions to the Path Computation Element Communication Protocol (PCEP) for a PCC to send a request for computing a backup ingress for an MPLS TE P2MP LSP or an MPLS TE P2P LSP to a PCE and for a PCE to compute the backup ingress and reply to the PCC with a computation result for the backup ingress.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 21, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

described in the Simplified BSD License.

Table of Contents

| | |
|--|----|
| 1. Introduction | 3 |
| 2. Terminology | 4 |
| 3. Conventions Used in This Document | 4 |
| 4. Extensions to PCEP | 4 |
| 4.1. Backup Ingress Capability Advertisement | 4 |
| 4.1.1. Capability TLV in Existing PCE Discovery Protocol | 4 |
| 4.1.2. Open Message Extension | 6 |
| 4.2. Request and Reply Message Extension | 7 |
| 4.2.1. RP Object Extension | 7 |
| 4.2.2. External Source Node Object | 8 |
| 4.2.3. Constraints between Ingress and Backup Ingress | 8 |
| 4.2.4. Constraints for Backup Path | 8 |
| 4.2.5. Backup Ingress Node | 9 |
| 4.2.6. Backup Ingress PCEP Error Objects and Types | 9 |
| 4.2.7. Request Message Format | 9 |
| 4.2.8. Reply Message Format | 10 |
| 5. Security Considerations | 10 |
| 6. IANA Considerations | 10 |
| 6.1. Backup Ingress Capability Flag | 11 |
| 6.2. Backup Ingress Capability TLV | 11 |
| 6.3. Request Parameter Bit Flags | 11 |
| 6.4. PCEP Objects | 11 |
| 7. Acknowledgement | 12 |
| 8. References | 12 |
| 8.1. Normative References | 12 |
| 8.2. Informative References | 13 |
| Author's Address | 13 |

1. Introduction

"Fast Reroute Extensions to RSVP-TE for LSP Tunnels" RFC4090 describes two methods to backup P2P LSP tunnels or paths at local repair points. The local repair points may comprise a number of intermediate nodes between an ingress node and an egress node along the path. The first method is a one-to-one backup method, where a detour backup P2P LSP for each protected P2P LSP is created at each potential point of local repair. The second method is a facility bypass backup protection method, where a bypass backup P2P LSP tunnel is created using MPLS label stacking to protect a potential failure point for a set of P2P LSP tunnels. The bypass backup tunnel can protect a set of P2P LSPs that have similar backup constraints.

"Extensions to RSVP-TE for P2MP TE LSPs" RFC4875 describes how to use the one-to-one backup method and facility bypass backup method to protect a link or intermediate node failure on the path of a P2MP LSP.

However, there is no mention of locally protecting an ingress node failure in a protected P2MP LSP or P2P LSP.

The methods for protecting an ingress node of a P2MP LSP or P2P LSP may be classified into two categories.

A first category uses a backup P2MP LSP that is from a backup ingress node to the number of destination nodes for the P2MP LSP, and a backup P2P LSP that is from a backup ingress node to the destination node for the P2P LSP. The disadvantages of this class of methods include more network resource such as computer power and link bandwidth consumption since the backup P2MP LSP or P2P LSP is from the backup ingress node to the number of destination nodes or the destination respectively.

A second category uses a local P2MP LSP or P2P LSP for protecting the ingress of a P2MP LSP or P2P LSP locally. The local P2MP LSP is from a backup ingress node to the next hop nodes of the ingress of the P2MP LSP. The local P2P LSP is from a backup ingress node to the next hop node of the ingress of the P2P LSP. It is desirable to let PCE compute these backup ingress nodes.

This document defines extensions to the Path Computation Element Communication Protocol (PCEP) for a PCC to send a request for computing a backup ingress node for an MPLS TE P2MP LSP or an MPLS TE P2P LSP to a PCE and for a PCE to compute the backup ingress node and reply to the PCC with a computation result for the backup ingress node.

2. Terminology

This document uses terminologies defined in RFC5440, RFC4090, and RFC4875.

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 computing a backup ingress of an MPLS TE P2MP LSP and an MPLS TE P2P LSP.

4.1. Backup Ingress Capability Advertisement

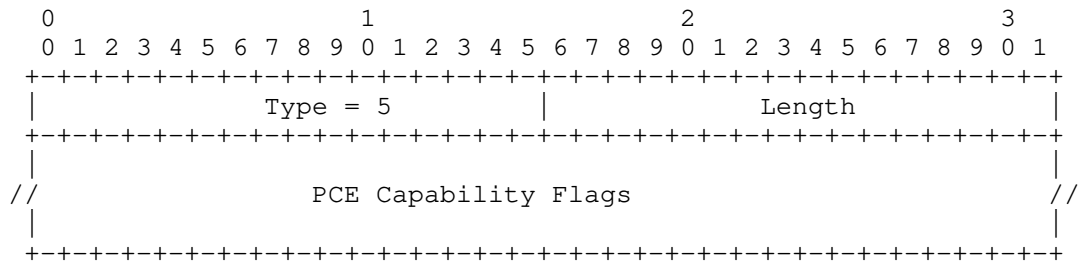
4.1.1. Capability TLV in Existing PCE Discovery Protocol

There are two options for advertising a PCE capability for computing a backup ingress for an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

The first option is to define a new flag in the OSPF and ISIS PCE Capability Flags to indicate the capability that a PCE is capable to compute both a backup ingress for an MPLS TE P2MP LSP and a backup ingress for an MPLS TE P2P LSP.

The second option is to define two new flags. One new flag in the OSPF and ISIS PCE Capability Flags indicates the capability that a PCE is capable to compute a backup ingress for an MPLS TE P2MP LSP; and another new flag in the OSPF and ISIS PCE Capability Flags indicates the capability that a PCE is capable to compute a backup ingress for an MPLS TE P2P LSP.

The format of the PCE-CAP-FLAGS sub-TLV is as follows:



Type: 5
 Length: Multiple of 4 octets
 Value: This contains an array of units of 32-bit flags
 numbered from the most significant as bit zero, where
 each bit represents one PCE capability.

The following capability bits have been assigned by IANA:

| Bit | Capabilities |
|-------|--|
| 0 | Path computation with GMPLS link constraints |
| 1 | Bidirectional path computation |
| 2 | Diverse path computation |
| 3 | Load-balanced path computation |
| 4 | Synchronized path computation |
| 5 | Support for multiple objective functions |
| 6 | Support for additive path constraints (max hop count, etc.) |
| 7 | Support for request prioritization |
| 8 | Support for multiple requests per message |
| 9 | Global Concurrent Optimization (GCO) |
| 10 | P2MP path computation |
| 11-31 | Reserved for future assignments by IANA. |

Reserved bits SHOULD be set to zero on transmission and MUST be ignored on receipt.

For the first option, one bit such as bit 11 may be assigned to indicate that a PCE is capable to compute both a backup ingress for an MPLS TE P2MP LSP and a backup ingress for an MPLS TE P2P LSP.

| Bit | Capabilities |
|-------|---|
| 11 | Backup ingress computation for P2MP LSP and P2P LSP |
| 12-31 | Reserved for future assignments by IANA. |

For the second option, one bit such as bit 11 may be assigned to indicate that a PCE is capable to compute a backup ingress for an MPLS TE P2MP LSP and another bit such as bit 12 may be assigned to indicate that a PCE is capable to compute a backup ingress for an MPLS TE P2P LSP.

| Bit | Capabilities |
|-------|--|
| 11 | Backup ingress computation for P2MP LSP |
| 12 | Backup ingress computation for P2P LSP |
| 13-31 | Reserved for future assignments by IANA. |

4.1.2. Open Message Extension

If a PCE does not advertise its backup ingress computation capability during discovery, PCEP should be used to allow a PCC to discover, during the Open Message Exchange, which PCEs are capable of supporting backup ingress computation.

To achieve this, we extend the PCEP OPEN object by defining a new optional TLV to indicate the PCE's capability to perform backup ingress computation for an MPLS TE P2MP LSP and an MPLS TE P2P LSP.

We request IANA to allocate a value such as 8 from the "PCEP TLV Type Indicators" subregistry, as documented in Section below ("Backup Ingress Capability TLV"). The description is "backup ingress capable", and the length value is 2 bytes. The value field is set to indicate the capability of a PCE for backup ingress computation for an MPLS TE LSP in details.

There are two options to indicate a PCE's capability for computing a backup ingress for an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

The first option is to use one bit such as bit 0 in the value field to indicate that a PCE is capable to compute both a backup ingress for an MPLS TE P2MP LSP and a backup ingress for an MPLS TE P2P LSP.

The second option is to use one bit such as bit 0 in the value field to indicate that a PCE is capable to compute a backup ingress for an MPLS TE P2MP LSP; and another one bit such as bit 1 in the value field to indicate that a PCE is capable to compute a backup ingress for an MPLS TE P2P LSP.

The inclusion of this TLV in an OPEN object indicates that the sender can perform backup ingress computation for an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

The capability TLV is meaningful only for a PCE, so it will typically appear only in one of the two Open messages during PCE session establishment. However, in case of PCE cooperation (e.g., inter-domain), when a PCE behaving as a PCC initiates a PCE session it SHOULD also indicate its path computation capabilities.

4.2. Request and Reply Message Extension

This section describes extensions to the existing RP (Request Parameters) object to allow a PCC to request a PCE for computing a backup ingress of an MPLS TE P2MP LSP or an MPLS TE P2P LSP when the PCE receives the PCEP request.

4.2.1. RP Object Extension

The following flags are added into the RP Object:

The I bit is added in the flag bits field of the RP object to tell the receiver of the message that the request/reply is for computing a backup ingress of an MPLS TE P2MP LSP and an MPLS TE P2P LSP.

- o I (Backup Ingress bit - 1 bit):

- 0: This indicates that this is not PCReq/PCRep for backup ingress.

- 1: This indicates that this is PCReq or PCRep message for backup ingress.

The IANA request is referenced in Section below (Request Parameter Bit Flags) of this document.

This I bit with the N bit defined in RFC6006 can indicate whether the request/reply is for a backup ingress of an MPLS TE P2MP LSP or an MPLS TE P2P LSP.

- o I = 1 and N = 1: This indicates that this is a PCReq/PCRep message for backup ingress of an MPLS TE P2MP LSP.

- o I = 1 and N = 0: This indicates that this is a PCReq/PCRep message for backup ingress of an MPLS TE P2P LSP.

4.2.2. External Source Node Object

In addition to the information about the path that an MPLS TE P2MP LSP or an MPLS TE P2P LSP traverses, a request message may comprise other information that may be used for computing the backup ingress for the P2MP LSP or P2P LSP. For example, the information about an external source node, from which data traffic is delivered to the ingress node of the P2MP LSP or P2P LSP and transported to the egress node(s) via the P2MP LSP or P2P LSP, is useful for computing a backup ingress node.

The PCC can specify an external source node (ESN) Object. The ESN Object has the same format as the IRO object defined in [RFC5440] except that it only supports IPv4 and IPv6 prefix sub-objects.

The object can only be carried in a PCReq message. A Path Request may carry at most one external source node Object.

The Object-Class and Object-types will need to be allocated by IANA. The IANA request is documented in Section below. (PCEP Objects).

4.2.3. Constraints between Ingress and Backup Ingress

A request message sent to a PCE from a PCC for computing a backup ingress of an MPLS TE P2MP LSP or an MPLS TE P2P LSP may comprise a constraint indicating that there must be a path from the backup ingress node to be computed to the ingress node of the P2MP LSP or P2P LSP and that the length of the path is within a given hop limit such as one hop.

This constraint can be considered as default by a PCE or explicitly sent to the PCE by a PCC [TBD].

4.2.4. Constraints for Backup Path

A request message sent to a PCE from a PCC for computing a backup ingress of a P2MP LSP or P2P LSP may comprise a constraint indicating that the backup ingress node to be computed may not be a node on the P2MP LSP or P2P LSP. In addition, the request message may comprise a list of nodes, each of which is a candidate for the backup ingress node.

A request message sent to a PCE from a PCC for computing a backup ingress of a P2MP LSP or P2P LSP may comprise a constraint indicating that there must be a path from the backup ingress node to be computed to the next-hop nodes of the ingress node of the P2MP LSP or P2P LSP and that there is not an internal node of the path from the backup ingress to the next-hop nodes on the P2MP LSP or P2P LSP .

Most of these constraints for the backup path can be considered as default by a PCE. The constraints for the backup path may be explicitly sent to the PCE by a PCC [TBD].

4.2.5. Backup Ingress Node

The PCE may send a reply message to the PCC in return to the request message for computing a new backup ingress node. The reply message may comprise information about the computed backup ingress node, which is contained in the path from the backup ingress node to the next-hop node(s) of the ingress node of the P2MP LSP or P2P LSP.

The backup ingress node is the root or head node of the backup path computed.

4.2.6. Backup Ingress PCEP Error Objects and Types

In some cases, the PCE may not complete the backup ingress computation as requested, for example based on a set of constraints. As such, the PCE may send a reply message to the PCC that indicates an unsuccessful backup ingress computation attempt. The reply message may comprise a PCEP-error object, which may comprise an error-value, error-type and some detail information.

4.2.7. Request Message Format

The PCReq message is encoded as follows using RBNF as defined in [RFC5511].

Below is the message format for a request message:

```

<PCReq Message> ::= <Common Header>
                        [<svec-list>]
                        <request>
<request> ::= <RP>
                <end-point-rro-pair-list>
                [<OF>]
                [<LSPA>]
                [<BANDWIDTH>]
                [<metric-list>]
                [<ESNO>]
                [<IRO>]
                [<LOAD-BALANCING>]
where:
    <ESNO> is an external source node object.

```

Figure 1: The Format for a Request Message

The definitions for svec-list, RP, end-point-rro-pair-list, OF, LSPA, BANDWIDTH, metric-list, IRO, and LOAD-BALANCING are described in RFC5440 and RFC6006.

4.2.8. Reply Message Format

The PCRep message is encoded as follows using RBNF as defined in [RFC5511].

Below is the message format for a reply message:

```

    <PCRep Message> ::= <Common Header>
                        <response>
    <response> ::= <RP>
                  <end-point-path-pair-list>
                  [<NO-PATH>]
                  [<attribute-list>]
  where:

    <end-point-path-pair-list> ::=
      [<END-POINTS>] <path> [<end-point-path-pair-list>]

    <path> ::= (<ERO> | <SERO>) [<path>]

    <attribute-list> ::= [<OF>]
                        [<LSPA>]
                        [<BANDWIDTH>]
                        [<metric-list>]
                        [<IRO>]

```

Figure 2: The Format for a Reply Message

The definitions for RP, NO-PATH, END-POINTS, OF, LSPA, BANDWIDTH, metric-list, IRO, and SERO are described in RFC5440, RFC6006 and RFC4875.

5. Security Considerations

The mechanism described in this document does not raise any new security issues for the PCEP, OSPF and IS-IS protocols.

6. IANA Considerations

This section specifies requests for IANA allocation.

6.1. Backup Ingress Capability Flag

Two new OSPF Capability Flags are defined in this document to indicate the capabilities for computing a backup ingress for an MPLS TE P2MP LSP and an MPLS TE P2P LSP. IANA is requested to make the assignment from the "OSPF Parameters Path Computation Element (PCE) Capability Flags" registry:

| Bit | Description | Reference |
|-----|-----------------------------|-----------|
| 11 | Backup ingress for P2MP LSP | This I-D |
| 12 | Backup ingress for P2P LSP | This I-D |

6.2. Backup Ingress Capability TLV

A new backup ingress capability TLV is defined in this document to allow a PCE to advertize its backup ingress computation capability. IANA is requested to make the following allocation from the "PCEP TLV Type Indicators" sub-registry.

| Value | Description | Reference |
|-------|------------------------|-----------|
| 8 | Backup ingress capable | This I-D |

6.3. Request Parameter Bit Flags

A new RP Object Flag has been defined in this document. IANA is requested to make the following allocation from the "PCEP RP Object Flag Field" Sub-Registry:

| Bit | Description | Reference |
|-----|------------------------|-----------|
| 16 | Backup ingress (I-bit) | This I-D |

6.4. PCEP Objects

An External Source Node Object-Type is defined in this document. IANA is requested to make the following Object-Type allocation from the "PCEP Objects" sub-registry:

| | |
|--------------------|----------------------|
| Object-Class Value | 33 |
| Name | External Source Node |
| Object-Type | 1: IPv4 |
| | 2: IPv6 |
| | 3-15: Unassigned |
| Reference | This I-D |

7. Acknowledgement

The author would like to thank Quintin Zhao and others 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, March 1997.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.
- [RFC6006] Zhao, Q., King, D., Verhaeghe, F., Takeda, T., Ali, Z., and J. Meuric, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 6006, September 2010.

8.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5862] Yasukawa, S. and A. Farrel, "Path Computation Clients (PCC) - Path Computation Element (PCE) Requirements for Point-to-Multipoint MPLS-TE", RFC 5862, June 2010.

Author's Address

Huaimo Chen
Huawei Technology, Inc.
Boston, MA
US

Email: Huaimochen@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 9, 2011

D. Dhody
U. Pallé
Q. Zhao
Huawei Technology
D. King
Old Dog Consulting
October 6, 2010

Management Information Base for the PCE Communications Protocol (PCEP)
for Path-Key-Based Inter-Domain Path Computation
draft-dhody-pce-pcep-pathkey-mib-00

Abstract

This memo defines an experimental portion of the Management Information Base for use with network management protocols in the Internet community. In particular, it describes managed objects for modeling of the Path Computation Element communication Protocol (PCEP) for communications between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between two PCEs when path-key-based inter-domain path computation is requested.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 9, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

This Internet-Draft will expire on April 9, 2011.

Table of Contents

| | |
|---|----|
| 1. Introduction | 3 |
| 2. Terminology | 3 |
| 3. The Internet-Standard Management Framework | 4 |
| 4. PCEP Pathkey MIB Module Architecture | 4 |
| 5. Example of the PCEP PathKey MIB module usage | 4 |
| 6. Object definitions | 5 |
| 6.1. PCE-PCEP-PATHKEY-DRAFT-MIB | 5 |
| 6.2. Objects for inclusion in module PCE-PCEP-DRAFT-MIB | 15 |
| 7. IANA Considerations | 15 |
| 8. Security Considerations | 15 |
| 9. References | 16 |
| 9.1. Normative References | 16 |
| 9.2. Informative References | 17 |

1. Introduction

The Path Computation Element (PCE) defined in [RFC4655] is an entity that is capable of computing a network path or route based on a network graph, and applying computational constraints. A Path Computation Client (PCC) may make requests to a PCE for paths to be computed.

The PCE communication protocol (PCEP) is designed as a communication protocol between PCCs and PCEs for point-to-point (P2P) path computations and is defined in [RFC5440].

If confidentiality is required between domains, Path-Key-Based mechanism is described in [RFC 5520]. For preserving the confidentiality of the "Confidential Path Segment (CPS)"; the PCE returns a path containing a loose hop in place of the segment that must be kept confidential.

[PCE-PCEP-DRAFT-MIB] defines a portion of the Management Information Base (MIB) for use with network management protocols in the Internet community for P2P path computations.

This memo defines an experimental portion of the Management Information Base for use with network management protocols in the Internet community. In particular, it describes managed objects for modeling of Path Computation Element communication Protocol (PCEP) [RFC5440] for communications between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between two PCEs in path-key-based inter-domain path computations.

Some objects maybe moved to [PCE-PCEP-DRAFT-MIB] after consensus with the authors and working group, these are defined in section 6.2.

2. Terminology

The following terminology is used in this document.

CPS: Confidential Path Segment. A segment of a path that contains nodes and links that the AS policy requires to not be disclosed outside the AS.

Domain: Any collection of network elements within a common sphere of address management or path computational responsibility. Examples of domains include Interior Gateway Protocol (IGP) areas and Autonomous Systems (ASs).

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

PCC: Path Computation Client: any client application requesting a path computation to be performed by a Path Computation Element.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

P2P: Point-to-Point

3. The Internet-Standard Management Framework

For a detailed overview of the documents that describe the current Internet-Standard Management Framework, please refer to section 7 of RFC 3410 [RFC3410].

Managed objects are accessed via a virtual information store, termed the Management Information Base or MIB. MIB objects are generally accessed through the Simple Network Management Protocol (SNMP). Objects in the MIB are defined using the mechanisms defined in the Structure of Management Information (SMI). This memo specifies a MIB module that is compliant to the SMIv2, which is described in STD 58, RFC 2578 [RFC2578] and STD 58, RFC 2580 [RFC2580].

4. PCEP Pathkey MIB Module Architecture

The PCEP Pathkey MIB will contain the following information:

- o PCEP Pathkey counters, timers and configurations
- o PCEP Pathkey table of CPS related information.

5. Example of the PCEP PathKey MIB module usage

In this section we provide an example (pcePcepPathKeyTable 1) of using the MIB objects described in Section 6 (Object definitions) to monitor. While this example is not meant to illustrate every permutation of the MIB, it is intended as an aid to understanding some of the key concepts. It is meant to be read after going through the MIB itself.

pcePcepPathKeyTable 1 of the PCE-PCEP-PATHKEY-DRAFT-MIB module :

```
{
    pcePcepPathKey                (4512),
    pcePcepPathKeyPath            (10.1.1.1 S
                                   10.1.1.2 S),
    pcePcepPathKeyRequestSource   (x.x.x.x),
    pcePcepPathKeyRequestId       (10),
    pcePcepPathKeyRetrieved       (1),
    pcePcepPathKeyRetrieveSource   (y.y.y.y),
    pcePcepPathKeyDiscardTime     (10),
    pcePcepPathKeyReuseTime       (30)
}
```

6. Object definitions

6.1. PCE-PCEP-PATHKEY-DRAFT-MIB

This MIB module makes references to the following documents.

[RFC2578], [RFC2580], [RFC3411], [RFC2863], [RFC3813].

PCE-PCEP-PATHKEY-DRAFT-MIB DEFINITIONS ::= BEGIN

IMPORTS

```
MODULE-IDENTITY, OBJECT-TYPE, NOTIFICATION-TYPE,
Unsigned32,
Counter32,
OCTET STRING,
experimental
    FROM SNMPv2-SMI                -- [RFC2578]
```

```
PcePcepIdentifier,
    FROM PCE-TC-STD-MIB
```

```
MODULE-COMPLIANCE,
OBJECT-GROUP,
NOTIFICATION-GROUP
    FROM SNMPv2-CONF;            -- [RFC2580]
```

pcePcepPathkeyDraftMIB MODULE-IDENTITY

LAST-UPDATED "201009171200Z" --Sep 17, 2010

ORGANIZATION "Path Computation Element (PCE) Working Group"

CONTACT-INFO "

Dhruv Dhody
Udayasree Palle
Quintin Zhao
Huawei Technology
Daniel King
OldDog Consulting

EMail: dhruvd@huawei.com

EMail: udayasreepalle@huawei.com

EMail: qzhao@huawei.com

EMail: daniel@oldog.co.uk

EMail comments directly to the PCE WG Mailing List at pce@ietf.org

WG-URL: <http://www.ietf.org/html.charters/pce-charter.html>

"

DESCRIPTION

"This MIB module defines a collection of objects for managing PCE communication protocol(PCEP) for Path-Key-Based Inter-Domain Path Computation"

-- Revision history

REVISION

"201009171200Z" -- 17 Sep 2010 12:00:00 EST

DESCRIPTION

"draft-00 version"

::= { experimental 9999 } --

-- Notifications --

pcePcepPathKeyNotifications OBJECT IDENTIFIER ::=

{ pcePcepPathKeyDraftMIB 0 }

pcePcepPathKeyMIBObjects OBJECT IDENTIFIER ::=

{ pcePcepPathKeyDraftMIB 1 }

pcePcepPathKeyConformance OBJECT IDENTIFIER ::=

{ pcePcepPathKeyDraftMIB 2 }

pcePcepPathKeyObjects OBJECT IDENTIFIER ::=

{ pcePcepPathKeyMIBObjects 1 }

```
--  
-- PCE Pathkey Objects  
--  
pcePcepPathKeyDiscardTimer OBJECT-TYPE  
    SYNTAX  Unsigned32  
    UNITS   "minutes"  
    MAX-ACCESS read-create  
    STATUS  mandatory  
    DESCRIPTION  
        "The value which indicates a period of time after the  
        expiration of which a PCE discard unwanted path-keys."  
    ::= { pcePcepPathKeyObjects 1 }  
  
pcePcepPathKeyReUseTimer OBJECT-TYPE  
    SYNTAX  Unsigned32  
    UNITS   "minutes"  
    MAX-ACCESS read-create  
    STATUS  mandatory  
    DESCRIPTION  
        "The value which indicates a period of time which  
        should expire before an old path-key could be  
        reused for a new CPS."  
    ::= { pcePcepPathKeyObjects 2 }  
  
pcePcepPathKeyRetainStatus OBJECT-TYPE  
    SYNTAX  INTEGER {  
        enabled(1),  
        disabled(2)  
    }  
    MAX-ACCESS read-create  
    STATUS  optional  
    DESCRIPTION  
        "The path-key retain status of this PCE to retain the  
        path-key and CPS for debugging purposes."  
    ::= { pcePcepPathKeyObjects 3 }  
  
pcePcepPathKeysGenerated OBJECT-TYPE  
    SYNTAX  Counter32  
    MAX-ACCESS read-only  
    STATUS  mandatory  
    DESCRIPTION  
        "The number of path-keys generated by this PCE."  
    ::= { pcePcepPathKeyObjects 4 }
```

```
pcePcepPathKeyExpandUnknown OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of attempts to expand an unknown
        path-key."
    ::= { pcePcepPathKeyObjects 5 }

pcePcepPathKeyExpandExpired OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of attempts to expand an expired
        path-key."
    ::= { pcePcepPathKeyObjects 6 }

pcePcepPathKeyExpandSame OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of attempts to expand the same
        path-key."
    ::= { pcePcepPathKeyObjects 7 }

pcePcepPathKeyExpiredNoExpansion OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of path-keys expired without any attempt
        to expand it."
    ::= { pcePcepPathKeyObjects 8 }

pcePcepPathKeyExpansionSuccess OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of path-key expansion requests (PCReq)
        which had successful retrieval."
    ::= { pcePcepPathKeyObjects 9 }
```



```
pcePcepPathKeyExpansionFailures OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of path-key expansion requests (PCReq)
         which had failed retrieval."
    ::= { pcePcepPathKeyObjects 10 }

pcePcepPathKeyConfig OBJECT-TYPE
    SYNTAX INTEGER {
        enabled(1),
        disabled(2)
    }
    MAX-ACCESS read-create
    STATUS mandatory
    DESCRIPTION
        "The path-key based inter domain computation
         configuration."
    ::= { pcePcepPathKeyObjects 11 }

pcePcepPathKeyTable OBJECT-TYPE
    SYNTAX SEQUENCE OF pcePcepPathKeyEntry
    MAX-ACCESS not-accessible
    STATUS current
    DESCRIPTION
        "This table contains information about the
         Pathkey CPS of PCE."
    ::= { pcePcepPathKeyObjects 12 }

pcePcepPathKeyEntry OBJECT-TYPE
    SYNTAX pcePcepPathKeyEntry
    MAX-ACCESS not-accessible
    STATUS current
    DESCRIPTION
        "An entry in this table represents a path-key and CPS.
         An entry is only created when a path-key generated by
         PCE during inter-domain computation."

    INDEX { pcePcepPathKey }

    ::= { pcePcepPathKeyTable 1 }
```

```
pcePcepPathKeyEntry ::= SEQUENCE {
    pcePcepPathKey                Unsigned32,
    pcePcepPathKeyPath            OCTET STRING,
    pcePcepPathKeyRequestSource   PcePcepIdentifier,
    pcePcepPathKeyRequestId      Unsigned32,
    pcePcepPathKeyRetrieved       INTEGER,
    pcePcepPathKeyRetrieveSource  PcePcepIdentifier,
    pcePcepPathKeyDiscardTime     Unsigned32,
    pcePcepPathKeyReuseTime       Unsigned32,
}

pcePcepPathKey OBJECT-TYPE
    SYNTAX      Unsigned32
    MAX-ACCESS  read-only
    STATUS      mandatory
    DESCRIPTION
        "The path-key value to identify a CPS."
    ::= { pcePcepPathKeyEntry 1 }

pcePcepPathKeyPath OBJECT-TYPE
    SYNTAX      OCTET STRING (SIZE (0..1024))
    MAX-ACCESS  read-only
    STATUS      mandatory
    DESCRIPTION
        "The CPS associated with the pathkey .
        This field is a displayable string in the
        format of XXX.XXX.XXX.XXX <space> S/L <newline>
        repeated for each hop address. The S/L character
        stands for Strict/Loose route.
        This field is meaningless unless pcePcepPathKey
        is not empty."
    ::= { pcePcepPathKeyEntry 2 }

pcePcepPathKeyRequestSource OBJECT-TYPE
    SYNTAX      PcePcepIdentifier
    MAX-ACCESS  read-only
    STATUS      mandatory
    DESCRIPTION
        "Source that issued the original request that led
        to the creation of the path-key."
    ::= { pcePcepPathKeyEntry 3 }
```

```
pcePcepPathKeyRequestId OBJECT-TYPE
    SYNTAX  Unsigned32
    MAX-ACCESS read-only
    STATUS  mandatory
    DESCRIPTION
        "The request ID of the original PCReq that led
         to the creation of the path-key."
    ::= { pcePcepPathKeyEntry 4 }

pcePcepPathKeyRetrieved OBJECT-TYPE
    SYNTAX  INTEGER {
        TRUE(1),
        FALSE(2)
    }
    MAX-ACCESS read-only
    STATUS  mandatory
    DESCRIPTION
        "It specifies whether the path-key is retrieved
         or not."

pcePcepPathKeyRetrieveSource OBJECT-TYPE
    SYNTAX  PcePcepIdentifier
    MAX-ACCESS read-only
    STATUS  mandatory
    DESCRIPTION
        "If the path-key is retrieved then by which
         PCC."
    ::= { pcePcepPathKeyEntry 6 }

pcePcepPathKeyDiscardTime OBJECT-TYPE
    SYNTAX  Unsigned32
    MAX-ACCESS read-only
    STATUS  mandatory
    DESCRIPTION
        "The time after which the path segment associated
         with the path-key will be discarded."
    ::= { pcePcepPathKeyEntry 7 }

pcePcepPathKeyReuseTime OBJECT-TYPE
    SYNTAX  Unsigned32
    MAX-ACCESS read-only
    STATUS  mandatory
    DESCRIPTION
        "The time after which the path-key will be available
         for re-use."
    ::= { pcePcepPathKeyEntry 8 }
```

--- Notifications

pcePcepPathKeyExpandUnknownNtf NOTIFICATION-TYPE

OBJECTS {
pcePcepPathKeyExpandUnknown
}

STATUS mandatory

DESCRIPTION

"This notification is sent when an attempt to expand an unknown path-key is made. The value of the counter pcePcepPathKeyExpandUnknown is also increased at this time."

::= { pcePcepPathKeyNotifications 1 }

pcePcepPathKeyExpandExpiredNtf NOTIFICATION-TYPE

OBJECTS {
pcePcepPathKeyExpandExpired
}

STATUS mandatory

DESCRIPTION

"This notification is sent when an attempt to expand an expired path-key is made. The value of the counter pcePcepPathKeyExpandExpired is also increased at this time."

::= { pcePcepPathKeyNotifications 2 }

pcePcepPathKeyExpandSameNtf NOTIFICATION-TYPE

OBJECTS {
pcePcepPathKeyExpandSame
}

STATUS optional

DESCRIPTION

"This notification is sent when a duplicate attempt to expand the same path-key is made. The value of the counter pcePcepPathKeyExpandSame is also increased at this time."

::= { pcePcepPathKeyNotifications 3 }

```

pcePcepPathKeyExpandSameNtf NOTIFICATION-TYPE
    OBJECTS      {
        pcePcepPathKeyExpiredNoExpansion
    }
    STATUS      optional
    DESCRIPTION
        "This notification is sent when path-key expires without
        any attempt to expand it. The value of the counter
        pcePcepPathKeyExpiredNoExpansion is also increased at
        this time."
    ::= { pcePcepPathKeyNotifications 4 }

--*****
-- Module Conformance Statement
--*****

pcePcepPathKeyGroups
    OBJECT IDENTIFIER ::= { pcePcepPathKeyConformance 1 }

pcePcepPathKeyCompliances
    OBJECT IDENTIFIER ::= { pcePcepPathKeyConformance 2 }

--
-- Full Compliance
--

pcePcepPathKeyModuleFullCompliance MODULE-COMPLIANCE
    STATUS current
    DESCRIPTION
        "The Module is implemented with support
        for read-create and read-write. In other
        words, both monitoring and configuration
        are available when using this MODULE-COMPLIANCE."

    MODULE -- this module
        MANDATORY-GROUPS { pcePcepPathKeyGeneralGroup,
                            pcePcepPathKeyNotificationsGroup
        }

    ::= { pcePcepPathKeyCompliances 1 }

```

```
--
-- Read-Only Compliance
--

pcePcepPathKeyModuleReadOnlyCompliance MODULE-COMPLIANCE
    STATUS current
    DESCRIPTION
        "The Module is implemented with support
        for read-only. In other words, only monitoring
        is available by implementing this MODULE-COMPLIANCE."

    MODULE -- this module
        MANDATORY-GROUPS { pcePcepPathKeyGeneralGroup,
                           }
    ::= { pcePcepPathKeyCompliances 2 }

-- units of conformance

pcePcepPathKeyGeneralGroup OBJECT-GROUP
    OBJECTS {
        pcePcepPathKeyDiscardTimer,
        pcePcepPathKeyReUseTimer,
        pcePcepPathKeysGenerated,
        pcePcepPathKeyExpandUnknown,
        pcePcepPathKeyExpandExpired,
        pcePcepPathKeyConfig,
        pcePcepPathKey,
        pcePcepPathKeyPath,
        pcePcepPathKeyRequestSource,
        pcePcepPathKeyRequestId,
        pcePcepPathKeyRetrieved,
        pcePcepPathKeyRetrieveSource,
        pcePcepPathKeyDiscardTime,
        pcePcepPathKeyReuseTime
    }
    STATUS current
    DESCRIPTION
        "Objects that apply to all PCEP Pathkey MIB
        implementations."

    ::= { pcePcepPathKeyGroups 1 }
```

```
pcePcepPathKeyNotificationsGroup NOTIFICATION-GROUP
    NOTIFICATIONS { pcePcepPathKeyExpandUnknownNtf,
                    pcePcepPathKeyExpandExpiredNtf
                    }
    STATUS    current

    DESCRIPTION
        "The notifications for a PCEP Pathkey MIB implementation."
        ::= { pcePcepPathKeyGroups 2 }

    END
```

6.2. Objects for inclusion in module PCE-PCEP-DRAFT-MIB

Following object maybe moved to [PCE-PCEP-DRAFT-MIB] after consensus with the authors and working group.

pcePcepPathKeyConfig

7. IANA Considerations

TBD

8. Security Considerations

This MIB module can be used for configuration of certain objects, and anything that can be configured can be incorrectly configured, with potentially disastrous results.

There are a number of management objects defined in this MIB module with a MAX-ACCESS clause of read-create. Such objects may be considered sensitive or vulnerable in some network environments. The support for SET operations in a non-secure environment without proper protection can have a negative effect on network operations. These are the tables and objects and their sensitivity/vulnerability:

- o pcePcepPathKeyDiscardTimer: Setting this value incorrectly may cause the expiration of Pathkey before attempt to retrieve the CPS.
- o pcePcepPathKeyReUseTimer: Setting this value incorrectly may cause the re-use of pathkey which may not guarantee the uniqueness of path-key values.

The user of the PCE-PCEP-PATHKEY-DRAFT-MIB module must therefore be aware that support for SET operations in a non-secure environment without proper protection can have a negative effect on network

operations.

The readable objects in the PCE-PCEP-PATHKEY-DRAFT-MIB module (i.e., those with MAX-ACCESS other than not-accessible) may be considered sensitive in some environments since, collectively, they provide information about the amount and frequency of path computation requests and responses within the network and can reveal some aspects of their configuration.

In such environments it is important to control also GET and NOTIFY access to these objects and possibly even to encrypt their values when sending them over the network via SNMP.

SNMP versions prior to SNMPv3 did not include adequate security. Even if the network itself is secure (for example by using IPsec), even then, there is no control as to who on the secure network is allowed to access and GET/SET (read/change/create/delete) the objects in this MIB module.

It is RECOMMENDED that implementers consider the security features as provided by the SNMPv3 framework (see [RFC3410], section 8), including full support for the SNMPv3 cryptographic mechanisms (for authentication and privacy).

Further, deployment of SNMP versions prior to SNMPv3 is NOT RECOMMENDED. Instead, it is RECOMMENDED to deploy SNMPv3 and to enable cryptographic security. It is then a customer/operator responsibility to ensure that the SNMP entity giving access to an instance of this MIB module is properly configured to give access to the objects only to those principals (users) that have legitimate rights to indeed GET or SET (change/create/delete) them.

9. References

9.1. Normative References

- [RFC2578] McCloghrie, k., Perkins, D., Schoenwaelder, J., Case, J., Rose, M., and S. Waldbusser, "Structure of Management Information Version 2 (SMIv2)", April 1999.
- [RFC2580] McCloghrie, k., Perkins, D., Schoenwaelder, J., Case, J., Rose, M., and S. Waldbusser, "Conformance Statements for SMIv2", April 1999.
- [RFC2863] McCloghrie, k. and F. Kastenholz, "The Interfaces Group MIB", June 2000.

- [RFC3411] Harrington, D., Presuhn, R., and B. Wijnen, "An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks", December 2002.
- [RFC3813] Srinivasan, C., Viswanathan, A., and T. Nadeau, "MPLS Multiprotocol Label Switching (MPLS) Label Switch Router Management Information Base", June 2004.
- [RFC5440] Ayyangar, A ., Farrel, A ., Oki, E., Atlas, A., Dolganow, A., Ikejiri, Y., Kumaki, K., Vasseur, J., and J. Roux, "Path Computation Element (PCE) communication Protocol (PCEP)", March 2009.

9.2. Informative References

- [PCE-PCEP-DRAFT-MIB] Kiran Koushik, A S., Stephan, E., Zhao, Q., and D. King, "PCE communication protocol (PCEP) Management Information Base", July 2010.
- [RFC3410] Case, J ., Mundy, R., Partain, D., and B. Stewart, "Introduction and Applicability Statements for Internet-Standard Management Framework", December 2002.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", August 2006.
- [RFC5520] Bradford, R., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", April 2009.

Authors' Addresses

Dhruv Dhody
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruvd@huawei.com

Udayasree Palle
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: Udayasreepalle@huawei.com

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

EMail: qzhao@huawei.com

Daniel King
Old Dog Consulting
UK

EMail: daniel@olddog.co.uk

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 18, 2011

O. Gonzalez de Dios, Ed.
Telefonica Investigacion y
Desarrollo
R. Casellas
CTTC - Centre Tecnologic de
Telecomunicacions de Catalunya
F. Jimenez Chico
Telefonica Investigacion y
Desarrollo
October 15, 2010

PCEP Extensions for Temporary Reservation of Computed Path Resources and
Support for Limited Context State in PCE
draft-gonzalezdedios-pce-reservation-state-00

Abstract

The Path Computation Element (PCE) provides path computation functions in support of traffic engineering in Multi-Protocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

A limited form of statefulness is useful to improve PCE functionality in situations in which the local TED might not be up to date, or in the case of concurrent requests where most of the LSPs are computed before the end of the set-up of the LSPs when the TED is updated. The PCE can retain some context from the resources assigned to Path Requests during a certain period of time, so that it avoids suggesting the use of the same resources for subsequent TE LSPs.

This document proposes an extension to the PCEP protocol to allow the PCC to request the PCE to block or reserve the resources computed in a path request of a TE LSP for subsequent requests for a certain time.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference

material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 18, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

| | |
|--|----|
| 1. Introduction | 4 |
| 2. PCEP Requirements | 6 |
| 3. PCEP Extensions (Encoding) | 7 |
| 3.1. Requesting a Reservation of Resources | 7 |
| 3.2. Replying a reservation status | 9 |
| 3.3. Cancelling a Reservation | 9 |
| 3.4. RESERVATION object format | 10 |
| 3.5. RESERVATION_CONF object format | 11 |
| 3.6. RESERVATION_ID TLV | 12 |
| 4. Protocol procedures | 12 |
| 5. Use cases | 13 |
| 5.1. Multiple LSP restoration | 13 |
| 5.2. Domain path selection | 14 |
| 5.3. Multidomain path computation | 14 |
| 6. Manageability Considerations | 14 |
| 6.1. Control of Function and Policy | 14 |
| 6.2. Information and Data Models | 15 |
| 6.3. Liveness Detection and Monitoring | 15 |
| 6.4. Verifying Correct Operation | 15 |
| 6.5. Requirements for Other Protocols and Functional Components | 15 |
| 6.6. Impact on Network Operation | 15 |
| 7. Security Considerations | 16 |
| 8. IANA Considerations | 16 |
| 8.1. RESERVATION object | 16 |
| 8.2. RESERVATION_CONF object | 16 |
| 8.3. RESERVATION_ID TLV | 16 |
| 8.4. PCEP Errors | 16 |
| 9. Acknowledgements | 17 |
| 10. Normative References | 17 |
| Authors' Addresses | 17 |

1. Introduction

According to [RFC4655], a PCE can be either stateful or stateless. In the former case, there is a strict synchronization between the PCE and not only the network states (in term of topology and resource information), but also the set of computed paths and reserved resources in use in the network.

In other words, the PCE utilizes information from the TED as well as information about existing paths (for example, TE LSPs) in the network when processing new requests. However, the maintenance and synchronization of a stateful database can be non-trivial, not only because it should verify the actual establishment of the computed paths, but also because it might not be the unique element to compute paths. Moreover, maintaining such a stateful database is not a function of the PCE, but rather of an NMS.

On the other hand, a stateless PCE does not keep track of any computed path, and each set of request(s) is processed independently of each other, typically using a local copy of the TED. Since a stateless PCE typically operates on a graph with computation constraints without tracking the state or history of path computations, independent requests will be processed on the same TED graph, until the graph is updated.

With a stateless PCE, there is a 'potential window of TED inaccuracy', where a stateless PCEs may compute paths based on current TED information, which could be out of sync with actual or potential network state changes given other recent PCE-computed paths.

For example, some sources for this potential TED inaccuracy are:

- o Control Plane link latencies, increasing: a) the time required for a PCC to obtain the paths after a successful computation, requiring several Round-Trip-Times (RTT) as per TCP; b) the setup delay and c) the time it takes for the PCE to update the local TED given IGP update times.
- o IGP (i.e. OSPF-TE) may operate with timers for LSA updates, to avoid excessive control plane overhead.
- o Concurrent requests that arrive during the time window, between a response is sent and the LSP is setup and the topology changes flooded. Even for very fast networks with low latency, there may be 'batched' requests: several path computation requests within a PCReq message or, in dynamic restoration without pre-planning, several LSPs need to be rerouted avoiding a failed link.

- o Local PCE contention, where the PCE needs to concurrently serve path computation requests and update the LSA (e.g. parsing OSPF-TE LSA updates). A PCE implementation may need to find a trade-off, when synchronizing access to the local TED: favor OSPF-TE parsing which means that some path computations are slightly delayed to allow an 'update' to be processed, or give strict priority to computation requests.

In consequence, a stateless PCE may assign the same (or a subset of the same) resources to several requests, which may result in contention and degraded network performance. The effects are detected late, typically during path signaling, causing path blocking and excessive crank-backs and retries.

Note that a PCC may include a set of previously computed paths in its request, in order to take them into account, for instance, to avoid double bandwidth accounting or to try to minimize changes (minimum perturbation problem).

Section 6.8 of RFC 4655 [RFC4655] suggests that a limited form of statefulness might be applied within an otherwise stateless PCE. The PCE may retain some context from paths it has recently computed so that it avoids suggesting the use of the same resources for other TE LSPs, using heuristics / forecasting for improved resource (i.e. wavelength) allocation.

This document proposes a set of extensions to the PCEP protocol to allow the PCC to request the PCE to block or reserve the resources associated to a path computation for a given path request. By reservation, it is implied that a set of resources which have been associated to such computation are excluded for subsequent path computations for a given time period.

Associated resources include (but not limit to): the bandwidth computed for the path in PSC or L2SC layers, a specific time slot (SDH) or tributary slot (OTN ODU-k) in TDM networks or a given wavelength or regenerator (WSON or OTN OCh).

This document also presents some illustrative use cases where the PCC would want the PCE to retain some context or state, like multiple LSP restoration, and counterexamples where the PCC does not have the intention to immediately set up the LSP, i.e., multidomain cases where the PCE is probing different paths to decide the sequence of domains.

2. PCEP Requirements

This section provides the set of requirements, both for PCCs and PCEs, to support context awareness.

When requesting a path computation (PCReq) to a PCE, the PCC should be able to indicate:

- o Whether the resources computed in the request should be blocked for further requests.
- o The amount of time the resources should be blocked, i.e. not used for subsequent requests.
- o The type and granularity of the resources to be blocked in the request. The type refers to the actual resources blocked such as path bandwidth or timeslot, wavelength, fiber... The granularity refers to the possibility of not only reserving the resource computed for the path but whether the associated links/nodes/SRLGs may need to be reserved too.

The PCE should be able to:

- o Apply policies whether a reservation request can be applied or not.
- o Compute one or more paths according to the request parameters and, based on the PCC indications, prevent that (part of) the resources involved in the computed route be used in subsequent computations for a given period.
- o If the request is allowed, the given reservation period SHOULD be no less than the requested period by the PCC (e.g. for the cases where the PCE is only able to reserve for multiples of a given value). This does not preclude the fact that, if configured by policy, a PCE MAY limit the period to a lower period. Alternatively, a PCE MAY be configured to reply with a PCEP_ERROR stating the cause of the failed computation/reservation.
- o The PCE MAY decide to apply a different granularity for the reservation request (e.g. block a given Time Slot or wavelength but not the TE links). In this case, the PCE MUST reply with the actual reservation.

Note that, the means by which a PCE can perform this are out of the scope of the present document but could include, for example, marking the resources as 'reserved', applying internal exclude objects etc.

The PCE should be able to respond (PCRep) to the PCC the following:

- o If the resources have been effectively blocked, and the final allocated reservation period, which may be different from the requested one.
- o The granularity of the reservation, which may be different from the requested one.
- o Provide a means to allow a PCC to request the cancellation of an active reservation (for example an identification of the reservation to allow its cancellation).

The PCC should be able to request the cancellation of an active resource reservation.

3. PCEP Extensions (Encoding)

3.1. Requesting a Reservation of Resources

A PCC that wants to request a PCE to temporarily reserve or block resources does so by including a RESERVATION object along with a client PCC_ID_REQ in the PCReq message.

Analogously to [RFC5886] the PCC-ID-REQ object is used to specify the IP address of the requesting PCC. The PCC-ID-REQ MUST be inserted within a PCReq message to specify the IP address of the requesting PCC. In [RFC5886] two PCC-ID-REQ objects (for IPv4 and IPv6) are defined.

A PCE that receives a PCReq message with a RESERVATION object MUST act according to the P-bit in the object header: if the P-bit is set, the object MUST be treated as mandatory and the request must either be processed using the contents of the object or rejected as defined in [RFC5440]. If the P-bit is clear, the object MAY be used by the PCE or MAY be ignored.

The RESERVATION object is optional in a PCReq message. Multiple instances of the object MUST NOT be used on a single PCReq message and if a PCE finds multiple instances of the object it MUST use the first one and discard the rest (Editors note: alternatively, it could send a PCErr). The RESERVATION object may appear either at an individual request level or within a SVEC. The latter means that the RESERVATION object applies to all requests involved in the SVEC object.

The PCReq ([RFC5440][RFC5541][RFC5557]) message is

```

<PCReq Message> ::= <Common Header>
                        [svec_list]
                        <request-list>

```

where

```

<svec-list> ::= <SVEC>
                [<OF>]
                [<GC>]
                [<XRO>]
                [<metric-list>]
                [<PCC-REQ-ID> <RESERVATION>]
                [<svec-list>]
<metric-list> ::= <METRIC>
                [<metric-list>]
<request-list> ::= <request>
                [<request-list>]
<request> ::= <RP>
                <END-POINTS>
                [<LSPA>]
                [<BANDWIDTH>]
                [<metric-list>]
                [<OF>]
                [<PCC_REQ_ID> <RESERVATION>]
                [<RRO> [<BANDWIDTH>]]
                [<IRO>]

```

[<LOAD-BALANCING>]

3.2. Replying a reservation status

If the PCE that receives the request applies the reservation, it indicates so using a RESERVATION_CONF object in the PCRep message.

The PCRep message is extended with regard to the one defined in [RFC5440] as follows:

<attribute-list>::=[<LSPA>]

[<BANDWIDTH>]

[<metric-list>]

[<IRO>]

[<RESERVATION_CONF>]

Note that the reservation applies at PATH level, and a RESERVATION_CONF object is included for all paths in a given response. This means distinct reservations for each path, which can be cancelled independently (Editor's Note: TDB, the PCC could indicate whether to have a single reservation or multiple reservation).

It is RECOMMENDED that the RESERVATION_CONF object appears the last attribute for a Path (or as an optional object in the attribute-list associated to a NO_PATH object).

3.3. Cancelling a Reservation

A PCC that wishes to cancel a reservation may send an unsolicited notification to the PCE, including the identifier of the reservation.

The PCNtf message used for one or more cancellations has no RP object. As with [RFC5440], the PCNtf message MUST carry at least one NOTIFICATION object and MAY contain several NOTIFICATION objects should the PCE or the PCC intend to notify of multiple events:

```

<PCNtf Message>::=<Common Header>

    <notify-list>

    <notify-list>::=<notify> [<notify-list>]

    <notify>::= <notification-list>

    <notification-list>::=<NOTIFICATION>[<notification-list>]

```

NOTIFICATION objects used for the purposes of cancelling an active reservation MUST include the RESERVATION_ID TLV. It is RECOMMENDED to use dedicated PCNtf messages for the purposes of cancelling reservations.

Both the Notification-type and Notification-value are TBD by IANA

The following Notification-type and Notification-value values are currently defined:

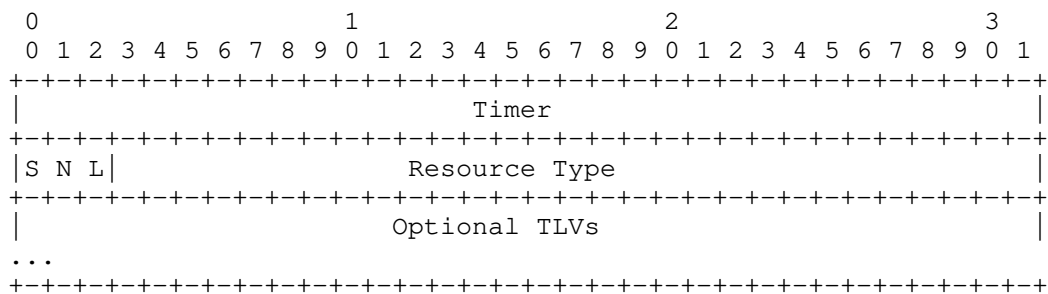
- o Notification-type=TBD: Pending Reservation cancelled
- o Notification-value=TBD (sug 1): PCC cancels a set of reservation requests.

3.4. RESERVATION object format

RESERVATION Object-Class is to be assigned by IANA.

RESERVATION Object-Type is to be assigned by IANA (recommended value=1)

The RESERVATION object indicates the intention of the PCC to set up the requested path and request the PCE to reserve the resources of the computed path to avoid being used by other requests.



- o Timer is the value in ms of the time that the resources should be blocked, encoded as a 32 bit unsigned integer.
- o Resource Type indicates the type of resource to be reserved. A value of 0 means the default resource type:
 - * Bandwidth (PSC, L2SC, ...)
 - * Time Slot (Sonet/SDH TDM)
 - * Tributary Slot (G709 OTN ODU-k TDM)
 - * Wavelength (G709 OTN OCh or WSON LSC)
- o Bit L: if set, TE Links should be part of the reservation, and excluded from subsequent request.
- o Bit N: if set, Nodes should be part of the reservation.
- o Bit S: if set, the set of SRLG (Shared Risk Link Group) deduced from the associated resources (i.e., the union of SRLGs of the links) should be part of the reservation.

Currently no TLVs are defined.

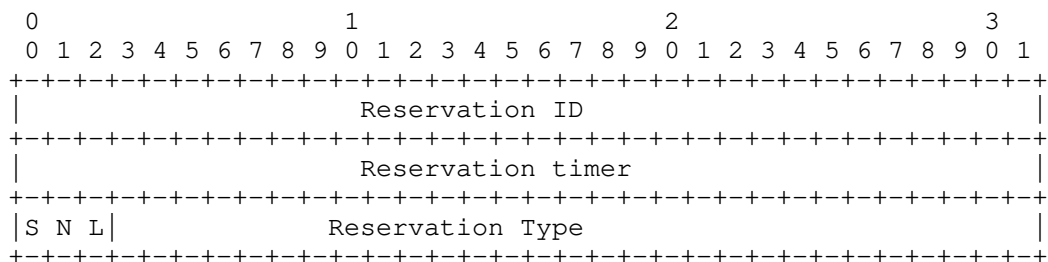
3.5. RESERVATION_CONF object format

The RESERVATION_CONF object is optional. The RESERVATION_CONF object indicates that the PCE has reserved the resources of computed path to avoid being used by other requests. The RESERVATION_CONF object is sent in the PCRep.

The RESERVATION_CONF Object-Class is to be assigned by IANA.

The RESERVATION_CONF Object-Type is to be assigned by IANA (recommended value=1)

The format of the RESERVATION_CONF object body is:



Timer is the value in ms of the time that the resources are blocked. The PCE May decide to apply a different value that the one requested by the PCC.

A PCC MUST NOT send a RESERVE_RESPONSE object if the client has not requested a RESERVATION in the PCReq message. A PCE MAY apply reservations as a means of internal policy and/or operation.

3.6. RESERVATION_ID TLV

The TLV indicates the reservation ID (Type TBA by IANA).

```

      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
      +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
      |                               Reservation ID                               |
      +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
  
```

4. Protocol procedures

A client that wishes to request a path computation with reservation shall:

- o Include a PCC_REQ_ID and RESERVATION objects in the involved Request within the PCReq message.
- o Specify what level of reservation to apply after the request.

Upon receiving a PCReq with a Resource Reservation object, the PCE will:

- o Perform the Path Computation using the local Traffic Engineering Database which has been extended to account for resources that have been marked reserved or blocked and which SHOULD not be used while blocked. This includes both synchronized / dependent path computations via SVEC or individual Path Computations requested in the request_list.
- o For the successful path computations, and for all paths corresponding to a given Request, determine the type of resources to be blocked (marked as reserved) with the granularity requested by the client once mapped to PCE policies.
- o It will start a local timer associated with this blocking action.

- o Send the Responses (successful or not) using PCRep message(s) and, where appropriate, indicate the level of reservation and associated period.
- o For subsequent requests, perform path computation as detailed above, updating the local TED with potential new reservations.

Whenever a timer expires, the PCE will:

- o Remove the reservation status / blocking that affected the reservation (e.g. add the previously subtracted unreserved bandwidth, mark the label, wavelength or time slot as available, etc).
- o Delete any data related with this blocking action.

5. Use cases

This section aims to show the use cases of the proposed possibility to activate the limited context awareness.

5.1. Multiple LSP restoration

One of the most challenging scenarios for a PCE-based architecture is the one of PCE-based dynamic multiple LSP restoration without pre-planning. In the event of a network failure affecting a high number of LSPs (e.g. a fibre cut), a PCE could potentially receive a significant amount of restoration requests in a short period of time from different PCCs.

One of the various challenges in this scenario is the fact that the PCE needs to sequentially perform multiple independent path computations. In this scenario, a stateless PCE would rely on TED information, which could potentially be up-to-date before the first incoming request (e.g. in case the routing algorithm has disseminated the failure event), but will definitely be outdated for subsequent requests.

It might be expected that the paths calculated for different connections would rely on the same nodes, TE links or even labels. It might occur at the signaling phase that multiple connection requests are contending for the same resources. After the eventual failure in the establishment of some of the connections, subsequent requests to the PCE would be triggered. After a number of loops, the PCE-based restoration would be eventually solved, but the potential number of retries could be significantly high.

The main issue is that the stateless PCE relied on an outdated TED to perform path computation. As the subsequent connection request is expected to be computed immediately, there is either no time for the routing algorithm to update the TED after a successful signaling or for the signaling process to successfully finish.

In this context, the availability of a limited context aware PCE could potentially solve the issue in a graceful fashion. Each of the restoration path requests will have an associated Resource Reservation object, which will state the kind of resources and the amount of time they should be blocked.

The PCE will then temporarily 'mark' the resources as blocked, so as not to consider them in subsequent connection requests, and thus avoiding the contention at the signaling phase. The timer should be in line with the LSP set up time and TED time to update.

5.2. Domain path selection

When selecting the set of domains of a multidomain path, a PCE may request paths to several PCEs of different domains. Thus, the intention of the request is not to establish a LSP, but to obtain a hint on the domain path. Thus, in this case, no Reservation Object would be sent.

5.3. Multidomain path computation

Once the domain path is known, when computing the actual path, the reservation object can be used. Note that multidomain paths may take a long time to be established, as it involves several AS or domains with different behavior and policies. Thus, it is a way to guarantee the availability of resources.

6. Manageability Considerations

Standard PCEP [RFC5440] describes various manageability considerations in PCEP, and most of the manageability requirements are already covered there. Specific aspects are detailed in this section.

6.1. Control of Function and Policy

In addition to PCE configuration parameters listed in [RFC5440], the following additional parameters might be required:

- o The ability to enable or disable reservations on the PCE.

- o The ability to retrieve a list of reservations currently active in the PCE.
- o The ability to configure which PCCs are allowed to perform reservations and the ability to configure limits on the timer periods requested. This includes, for example, the configuration of IP based access lists for PCCs.
- o The ability to configure which PCCs are allowed to perform reservations for single-domain and multi-domain scenarios, typically according to pre-defined agreements.
- o The ability to configure which reservation granularity a given PCC group is able to request, and the associated action (error or downgrade).
- o TDB: Advertisements of capabilities via IGP and configurability

6.2. Information and Data Models

A number of MIB objects have been defined for general PCEP control and monitoring of P2P computations in [PCEP-MIB]. For the time being, no extra models are considered although it could be possible that current means to retrieve information from the PCE be extended to include eventual resource reservations.

6.3. Liveness Detection and Monitoring

Other than the considerations expressed in [RFC5440], a PCE could provide extensions to [MONITORING] to verify reservation status, and to obtain statistics on the system.

6.4. Verifying Correct Operation

There are no additional requirements for verifying the correct operation of the PCEP sessions. Future MIB objects could facilitate verification of correct operation and reporting of reservations and errors.

6.5. Requirements for Other Protocols and Functional Components

The method for the PCC to obtain information about a PCE capable of reservation may include extensions to IGP protocols.

6.6. Impact on Network Operation

It is expected that the use of PCEP extensions specified in this document will not significantly increase the level of operational

traffic. However, mis-configured, excessive reservation requests, excessive reservation periods, or excessive granularity may increase the number of failed requests or cause the PCE to provide sub-optimal routes due to existing reservations. Coarse reservations may also limit the resources that are available for a PCE in order to serve requests.

An excessive number of reservation requests and reservation cancellations may degrade server performance. A PCE SHOULD provide a means to control the rate of messages with reservation, extending the proposed mechanism of [RFC5440].

7. Security Considerations

In the event of an unauthorized path computation request with mandatory resource reservation, or in case of a (distributed) denial of service attack, the subsequent state/context managed within the PCE may be disruptive to the network, resulting in performance degradation or sub-optimal computed routes. Implementations should conform to the relevant security requirements of [RFC5440] that specifically help to control unauthorized requests. These mechanisms include securing the PCEP session requests and responses using TCP security techniques, authenticating the PCEP requests and responses to ensure the message is intact and sent from an authorized node, providing policy control by explicitly defining which PCCs are allowed to perform resource reservations to the PCE and disallowing reservation requests that may block an excessive amount of resources.

8. IANA Considerations

IANA maintains a registry of PCEP parameters. A number of IANA considerations have been highlighted in previous sections of this document.

8.1. RESERVATION object

8.2. RESERVATION_CONF object

8.3. RESERVATION_ID TLV

8.4. PCEP Errors

For the RESERVATION object, the default error procedures regarding supported unknown objects defined in [RFC5440] apply

- o Unsupported Reservation Option
- o Reservation Forbidden by Policy
- o Unknown Reservation Request

9. Acknowledgements

The authors thank Cyril Margaria for the discussions and suggestions in the topic.

10. Normative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, June 2009.
- [RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, July 2009.
- [RFC5886] Vasseur, JP., Le Roux, JL., and Y. Ikejiri, "A Set of Monitoring Tools for Path Computation Element (PCE)-Based Architecture", RFC 5886, June 2010.

Authors' Addresses

Oscar Gonzalez de Dios (editor)
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Ramon Casellas
CTTC - Centre Tecnologic de Telecomunicacions de Catalunya
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Phone:
Email: ramon.casellas@cttc.es

Francisco Javier Jimenez Chico
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28043
Spain

Phone: +34 91 3379037
Email: fjjc@tid.es

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 27, 2011

C. Margaria, Ed.
Nokia Siemens Networks
O. Gonzalez de Dios, Ed.
Telefonica Investigacion y
Desarrollo
F. Zhang, Ed.
Huawei Technologies
October 24, 2010

PCEP extensions for GMPLS
draft-ietf-pce-gmpls-pcep-extensions-01

Abstract

This memo provides extensions for the Path Computation Element communication Protocol (PCEP) for the support of GMPLS control plane.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 27, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

described in the Simplified BSD License.

Table of Contents

| | |
|--|----|
| 1. Introduction | 3 |
| 1.1. Contributing Authors | 3 |
| 1.2. PCEP requirements for GMPLS | 3 |
| 1.3. PCEP existing objects related to GMPLS | 4 |
| 1.4. Requirements Language | 6 |
| 2. PCEP objects and extensions | 7 |
| 2.1. RP object extension | 8 |
| 2.2. Traffic parameters encoding, GENERALIZED-BANDWIDTH | 9 |
| 2.3. Traffic parameters encoding, GENERALIZED-LOAD-BALANCING | 11 |
| 2.4. END-POINTS Object extensions | 14 |
| 2.4.1. Generalized endpoint Object Type | 14 |
| 2.4.2. END-POINTS TLVs extensions | 17 |
| 2.5. LABEL-SET object | 20 |
| 2.6. SUGGESTED-LABEL-SET object | 20 |
| 2.7. LSPA extensions | 21 |
| 2.8. NO-PATH Object Extension | 21 |
| 2.8.1. Extensions to NO-PATH-VECTOR TLV | 21 |
| 3. Additional Error Type and Error Values Defined | 23 |
| 4. Manageability Considerations | 25 |
| 5. IANA Considerations | 26 |
| 5.1. PCEP Objects | 26 |
| 5.2. New PCEP TLVs | 27 |
| 5.3. New PCEP Error Codes | 27 |
| 6. Security Considerations | 29 |
| 7. Contributing Authors | 30 |
| 8. Acknowledgments | 32 |
| 9. References | 33 |
| 9.1. Normative References | 33 |
| 9.2. Informative References | 34 |
| Authors' Addresses | 36 |

1. Introduction

PCEP RFCs [RFC5440], [RFC5521], [RFC5541], [RFC5520] are focused on path computation requests in MPLS networks. [RFC4655] defines the PCE framework also for GMPLS networks. This document complements these RFCs by providing some consideration of GMPLS applications and routing requests, for example for OTN and WSON networks.

The requirements on PCE extensions to support those characteristics are described in [I-D.ietf-pce-gmpls-aps-req] and [I-D.ietf-pce-wson-routing-wavelength].

1.1. Contributing Authors

Elie Sfeir, Franz Rambach (Nokia Siemens Networks) Francisco Javier Jimenez Chico (Telefonica Investigacion y Desarrollo) Suresh BR, Young Lee, SenthilKumar S, Jun Sun (Huawei Technologies), Ramon Casellas (CTTC)

1.2. PCEP requirements for GMPLS

This section provides a set of PCEP requirements to support GMPLS LSPs and assure signal compatibility in the path. When requesting a path computation (PCReq) to PCE, the PCC should be able to indicate, according to [I-D.ietf-pce-gmpls-aps-req] and to RSVP procedures like explicit label control (ELC), the following additional attributes:

(1) Switching capability: for instance PSC1-4, L2SC, TDM, LSC, FSC

(2) Encoding type: as defined in [RFC4202], [RFC4203], e.g., Ethernet, SONET/SDH, Lambda, etc.

(3) Signal Type: Indicates the type of elementary signal that constitutes the requested LSP. A lot of signal types with different granularity have been defined in SONET/SDH and G.709 ODUk, such as VC11, VC12, VC2, VC3 and VC4 in SDH, and ODU1, ODU2 and ODU3 in G.709 ODUk [RFC4606], [RFC4328] and other signal types like the one defined in [I-D.ceccarelli-ccamp-gmpls-ospf-g709] or [I-D.zhang-ccamp-gmpls-evolving-g709] .

(4) Concatenation Type: In SDH/SONET and G.709 OTN networks, two kinds of concatenation modes are defined: contiguous concatenation which requires co-route for each member signal and requires all the interfaces along the path to support this capability, and virtual concatenation which allows diverse routes for the member signals and only requires the ingress and egress interfaces to support this capability. Note that for the virtual concatenation, it also may specify co-routed or separated-routed. See [RFC4606]

and [RFC4328] about concatenation information.

(5) Concatenation Number: Indicates the number of signals that are requested to be contiguously or virtually concatenated. See also [RFC4606] and [RFC4328].

(6) Technology specific label(s) such as wavelength label as defined in [I-D.ietf-ccamp-gmpls-g-694-lambda-labels]

(7) e2e Path protection type: as defined in [RFC4872], e.g., 1+1 protection, 1:1 protection, (pre-planned) rerouting, etc.

(8) Link Protection type: as defined in [RFC4203]

(9) Support for unnumbered interfaces: as defined in [RFC3477]

(10) Support for asymmetric bandwidth requests.

(11) Indicate the requested granularity for the path ERO: node, link, label to allow the use of the explicit/suggested label control of RSVP.

We describe in this document a proposal to fulfill those requirements.

1.3. PCEP existing objects related to GMPLS

PCEP as of [RFC5440], [RFC5521] and [I-D.ietf-pce-inter-layer-ext], supports the following information (in the PCReq and PCRep) related to the described requirements.

From [RFC5440]:

- o numbered endpoints
- o bandwidth (encoded as IEEE float)
- o ERO
- o LSP attributes (setup and holding priorities)
- o Request attribute (include some LSP attributes)

From [RFC5521]:

- o Extensions to PCEP for Route Exclusions, definition of a XRO object and a new semantic (F bit or Fail bit) indicating that the existing route is failed and resources present in the RRO can be

reused. This object also allows to exclude (strict or not) resources; XRO include the diversity level (node, link, SRLG). The requested diversity is expressed in the XRO.

From [I-D.ietf-pce-inter-layer-ext]:

- o INTER-LAYER : indicates if inter-layer computation is allowed
- o SWITCH-LAYER : indicates which layer(s) should be considered, can be used to represent the RSVP-TE generalized label request
- o REQ-ADAP-CAP : indicates the adaptation capabilities requested, can also be used for the endpoints in case of mono-layer computation

The shortcomings of the existing PCEP information are:

The BANDWIDTH and LOAD-BALANCING objects do not describe the details of the traffic request (for example NVC, multiplier) in the context of GMPLS networks, for instance TDM or OTN networks.

The END-POINTS object does not allow specifying an unnumbered interface, nor the labels on the interface. Those parameters are of interest in case of switching constraints.

Current attributes do not allow to express the requested link level protection and end-to-end protection attributes.

The covered PCEP extensions are:

New objects are introduced (GENERALIZED-BANDWIDTH and GENERALIZED-LOAD-BALANCING) for flexible bandwidth encoding,

A new object type is introduced for the END-POINTS object (generalized-endpoint),

A new TLV is added to the LSPA object.

In order to allow to restrict the range of labels returned, an additional object is added: LABEL-SET

In order to indicate the mandatory routing granularity in the response, a new flag in the RP object is added.

1.4. Requirements Language

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 RFC 2119.

2. PCEP objects and extensions

This section describes the required PCEP objects and extensions. The PCReq and PCRep messages are defined in [RFC5440]. The format of the request and response messages with the proposed extensions (GENERALIZED-BANDWIDTH, SUGGESTED-LABEL-SET and LABEL-SET) is as follows:

```

<request> ::= <RP>
               <segment-computation> | <path-key-expansion>
<segment-computation> ::=
  <ENDPOINTS>
  [<LSPA>]
  [<BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>]
  [<metric-list>]
  [<OF>]
  <RRO> [<BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>]
  [<IRO>]
  [<SUGGESTED-LABEL-SET>]
  [<LABEL-SET>]
  [<LOAD-BALANCING>]
  [<GENERALIZED-LOAD-BALANCING>]
  [<GENERALIZED-LOAD-BALANCING>]
  [<XRO>]

<path-key-expansion> ::= <PATH-KEY>

<response> ::= <RP>
  [<NO-PATH>]
  [<attribute-list>]
  [<path-list>]

<path-list> ::= <path> [<path-list>]
<path> ::= <ERO> <attribute-list>
<metric-list> ::= <METRIC> [<metric-list>]

```

For point-to-multipoint (P2MP) computations, the proposed grammar is:

```

<segment-computation> ::=
  <end-point-rro-pair-list>
  [<LSPA>]
  [<BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>] [<GENERALIZED-BANDWIDTH>]
  [<metric-list>]
  [<IRO>]
  [<SUGGESTED-LABEL-SET>]
  [<LABEL-SET>]
  [<LOAD-BALANCING>]
  [<GENERALIZED-LOAD-BALANCING>]
  [<GENERALIZED-LOAD-BALANCING>]
  [<XRO>]

<end-point-rro-pair-list> ::=
  <END-POINTS> [<RRO-List>] [<BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>]
  [<end-point-rro-pair-list>]

<RRO-List> ::= <RRO> [<BANDWIDTH>]
  [< GENERALIZED-BANDWIDTH>] [<RRO-List>]

```

Where:

```

<attribute-list> ::= [<LSPA>]
  [<BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>]
  [<GENERALIZED-BANDWIDTH>]
  [<metric-list>]
  [<IRO>]

```

2.1. RP object extension

Explicit label control (ELC) is a procedure supported by RSVP-TE, where the outgoing label(s) is(are) encoded in the ERO. In consequence, the PCE may be able to provide such label(s) directly in the path ERO. The PCC, depending on policies or switching layer, may be required to use explicit label control or expect explicit link, thus it need to indicate in the PCEReq which granularity it is expecting in the ERO. The possible granularities can be node, link, label. Those granularities are dependent, i.e link granularity imply that the nodes are provided, label granularity that the links and nodes are provided in the ERO

A new 2-bit routing granularity (RG) flag is defined in the RP object (IANA suggestion : bit 17 and 16). The values are defined as follows

00 : node
01 : link
02 : label
03 : reserved

When the RP object appears in a request within a PCReq message the flag indicates the requested route granularity. The PCE SHOULD try to follow this granularity and MAY return a NO-PATH if the requested granularity cannot be provided. The PCE MAY return more details on the route based on its policy. The PCC can decide if the ERO is acceptable based on its content.

When the RP object appears in a response within a PCRep message the flag indicates the granularity provided in the response. The PCE MAY indicates the granularity of the returned ERO. The RG flag is backward-compatible with previous RFCs: the value sent by implementation not supporting it will indicate a node granularity. this flag is optional for responses. A new capability flag in the PCE-CAP-FLAGS from RFC [RFC5088] and [RFC5089] may be added.

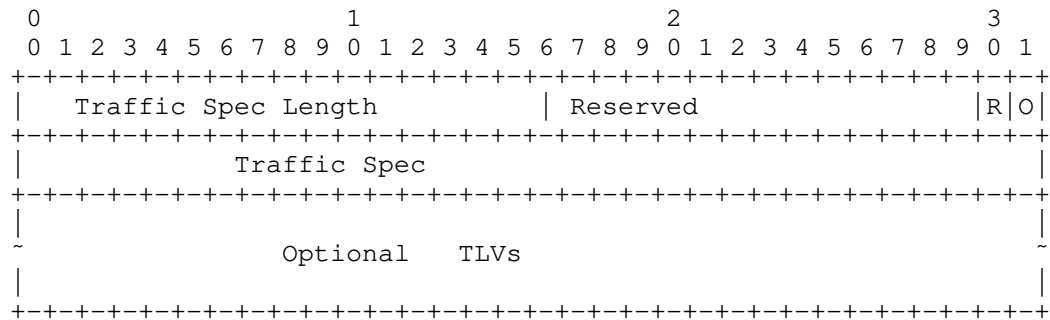
2.2. Traffic parameters encoding, GENERALIZED-BANDWIDTH

The PCEP BANDWIDTH does not describe the details of the signal (for example NVC, multiplier), hence the bandwidth information should be extended to use the RSVP Tspec object encoding. The PCEP BANDWIDTH object defines two types: 1 and 2. C-Type 2 is representing the existing bandwidth in case of re-optimization.

The following possibilities cannot be represented in the BANDWIDTH object:

- o Asymmetric bandwidth (different bandwidth in forward and reverse direction), as described in [RFC5467]
- o GMPLS (SDH/SONET, G.709, ATM, MEF etc) parameters are not supported.

According to [RFC5440] the BANDWIDTH object has no TLV and has a fixed size of 4 bytes. This definition does not allows extending it with the required information. To express this information, a new Object named GENERALIZED-BANDWIDTH having the following format is defined:



The GENERALIZED-BANDWIDTH has a variable length. The Traffic spec length field indicate the length of the Traffic spec field. The bits R and O have the following meaning:

O bit : set when the value refer to the previous bandwidth in case of re-optimization

R bit : set when the value refer to the bandwidth of the reverse direction

The Object type determines which type of bandwidth is represented by the object. The following object types are defined:

1. Intserv
2. SONET/SDH
3. G.709
4. Ethernet

The encoding of the field Traffic Spec is the same as in RSVP-TE, it can be found in the following references.

| Object Type | Name | Reference |
|-------------|-----------|-----------|
| 0 | Reserved | |
| 1 | Reserved | |
| 2 | Intserv | [RFC2210] |
| 3 | Reserved | |
| 4 | SONET/SDH | [RFC4606] |
| 5 | G.709 | [RFC4328] |
| 6 | Ethernet | [RFC6003] |

Traffic Spec field encoding

The GENERALIZED-BANDWIDTH MAY appear more than once in a PCReq message. If more than one GENERALIZED-BANDWIDTH have the same Object Type, Reserved, R and O values, only the first one is processed, the others are ignored. On the response the object that were considered in the processing SHOULD be included.

When a PCC needs to get a bi-directional path with asymmetric bandwidth, it SHOULD specify the different bandwidth in forward and reverse directions through two separate GENERALIZED-BANDWIDTH objects. The PCE MUST compute a path that satisfies the asymmetric bandwidth constraint and return the path to PCC if the path computation is successful.

Optional TLVs may be included within the object body to specify more specific BW requirements. The specification of such TLVs is outside the scope of this document.

2.3. Traffic parameters encoding, GENERALIZED-LOAD-BALANCING

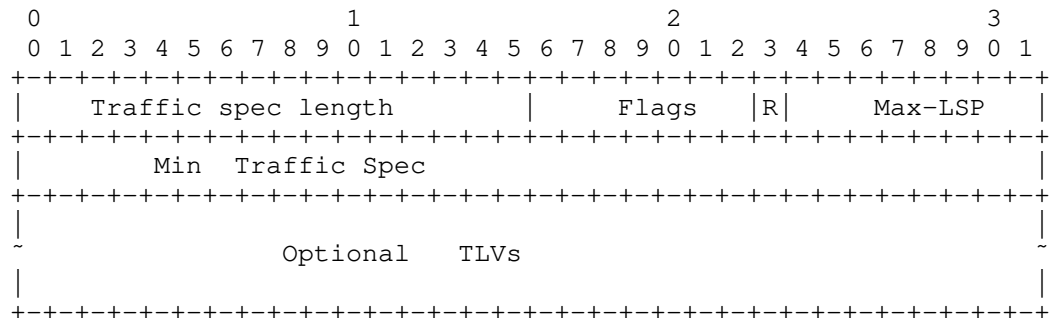
The PCEP LOAD-BALANCING follows the bandwidth encoding of the BANDWIDTH object, it does not describe enough details for the traffic specification expected by GMPLS, hence this bandwidth information should be extended to use the RSVP Tspec object encoding.

According to [RFC5440] the LOAD-BALANCING object has no TLV and has a fixed size of 8 bytes. This definition does not allows extending it with the required information. To express this information, a new Object named GENERALIZED-LOAD-BALANCING is defined

The GENERALIZED-LOAD-BALANCING object is optional.

GENERALIZED-LOAD-BALANCING Object-Class is To be assigned by IANA.

The format of the GENERALIZED-LOAD-BALANCING object body is as follows:



Traffic spec length (16 bits): the length of the min traffic spec length, also including the eventual TLV present in RSVP-TE traffic specification.

Flags (8 bits): The undefined Flags field MUST be set to zero on transmission and MUST be ignored on receipt. The following flag is defined:

R Flag : (1 bit) set when the value refer to the bandwidth of the reverse direction

Max-LSP (8 bits): maximum number of TE LSPs in the set.

Min-Traffic spec (variable): Specifies the minimum traffic spec of each element of the set of TE LSPs.

The GENERALIZED-LOAD-BALANCING has a variable length. The Object type determines which type of minimum bandwidth is represented by the object. The following object types are defined:

1. Intserv
2. SONET/SDH
3. G.709
4. Ethernet

The encoding of the field Traffic Spec is the same as in RSVP-TE, it can be found in the following references.

| Object Type | Name | Reference |
|-------------|-----------|-----------|
| 2 | Intserv | [RFC2210] |
| 4 | SONET/SDH | [RFC4606] |
| 5 | G.709 | [RFC4328] |
| 6 | Ethernet | [RFC6003] |

Traffic Spec field encoding

The GENERALIZED-LOAD-BALANCING MAY appear more than once in a PCReq message. If more than one GENERALIZED-LOAD-BALANCING have the same Object Type, and R Flag, only the first one is processed, the others are ignored. On the response the object that were considered in the processing SHOULD be included.

When a PCC needs to get a bi-directional path with asymmetric bandwidth, it SHOULD specify the different bandwidth in forward and reverse directions through two separate GENERALIZED-LOAD-BALANCING objects with different R Flag. The PCE MUST compute a path that satisfies the asymmetric bandwidth constraint and return the path to PCC if the path computation is successful.

Optional TLVs may be included within the object body to specify more specific bandwidth requirements. The specification of such TLVs is outside the scope of this document.

The GENERALIZED-LOAD-BALANCING object has the same semantic as the LOAD-BALANCING object, If a PCC requests the computation of a set of TE LSPs so that the total of their generalized bandwidth is X, the maximum number of TE LSPs is N, and each TE LSP must at least have a bandwidth of B, it inserts a GENERALIZED-BANDWIDTH object specifying X as the required bandwidth and a GENERALIZED-LOAD-BALANCING object with the Max-LSP and Min-traffic spec fields set to N and B, respectively.

For example a request for one co-signaled VCAT members will not use the GENERALIZED-LOAD-BALANCING. In case the VCAT member can be diversely routed, the GENERALIZED-BANDWIDTH will contain a traffic specification indicating the complete VCAT group and the GENERALIZED-LOAD-BALANCING the minimum co-signaled members. For a SDH network, a request to have a VC4 VCAT group with 10 VC4 container, diversely routed with 2VC4 container on each path minimum, can be represented with a GENERALIZED-BANDWIDTH object with OT=4, the content of the Traffic specification is ST=6,RCC=0,NCC=0,NVC=10,MT=1. The GENERALIZED-LOAD-BALANCING, OT=4,R=0,Max-LSP=5, min Traffic spec is

(ST=6,RCC=0,NCC=0,NVC=2,MT=1). The PCE can respond with a response with maximum 5 path, each of then having a GENERALIZED-BANDWIDTH OT=4,R=0, and traffic spec matching the minimum traffic spec from the GENERALIZED-LOAD-BALANCING object of the corresponding request

2.4. END-POINTS Object extensions

The END-POINTS object is used in a PCReq message to specify the source and destination of the path for which a path computation is requested. From [RFC3471] the source IP address and the destination IP address are used to identify those. A new Object Type is defined to address the following possibilities:

- o Different endpoint types.
- o Label restrictions on the endpoint.
- o Specification of unnumbered endpoints type as seen in GMPLS networks.

The Object encoding is described in the following sections.

2.4.1. Generalized endpoint Object Type

In GMPLS context the endpoints can:

- o Be unnumbered
- o Have label(s) associated to them
- o May have different switching capabilities

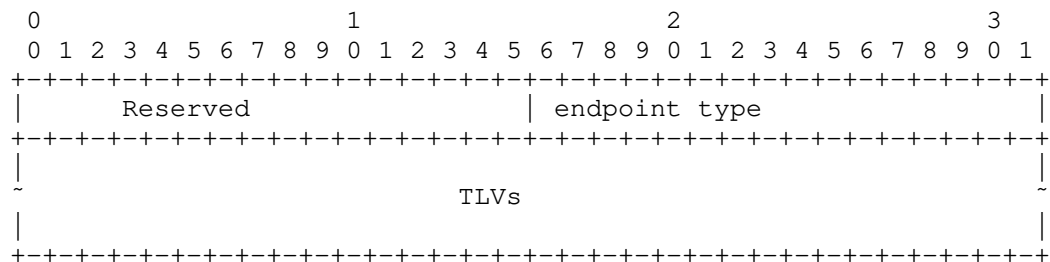
The IPV4 and IPV6 endpoints are used to represent the source and destination IP addresses. The scope of the IP address (Node or Link) is not explicitly stated. It should also be possible to request a Path between a numbered link and an unnumbered link, or a P2MP path between different type of endpoints.

Since the PCEP ENDPOINTS object only support endpoints of the same type new C-Types are proposed that support different endpoint types, including unnumbered. This new C-Type also supports the specification of constraints on the endpoint label to be use. The PCE might know the interface restrictions but this is not a requirement. On the path calculation request only the TSPEC and SWITCH layer need to be coherent, the endpoint labels could be different (supporting a different TSPEC). Hence the label restrictions include a Generalized label request in order to interpret the labels.

The proposed object format consists of a body and a list of TLVs with the following defined TLVs (described in Section 2.4.2).

1. IPV4 address.
2. IPV6 address.
3. Unnumbered endpoint.
4. Label request.
5. Label.
6. Label set.
7. Suggested label set.

The Object is encoded as follow:



Reserved bits should be set to 0 when a message is sent and ignored when the message is received

the endpoint type is defined as follow:

| Value | Type | Meaning |
|-------------|---------------------|---|
| 0 | Point-to-Point | |
| 1 | Point-to-Multipoint | New leaves to add |
| 2 | | Old leaves to remove |
| 3 | | Old leaves whose path can be modified/reoptimized |
| 4 | | Old leaves whose path must be left unchanged |
| 5-32767 | Reserved | |
| 32768-65535 | Experimental range | |

Endpoint type 0 MUST be accepted by the PCE, other endpoint type MAY be supported if the PCE implementation supports P2MP path calculation. The TLVs present in the object body should follow the following grammar:

```

<generalized-endpoint-tlvs> ::=
  <p2p-endpoints> | <p2mp-endpoints>

<p2p-endpoints> ::=
  <endpoint>
  [<endpoint-restrictions>]
  <endpoint>
  [<endpoint-restrictions>]

<p2mp-endpoints> ::=
  <endpoint> [<endpoint-restrictions>]
  [<endpoint> [<endpoint-restrictions>] ...]

```

Private TLV MAY be inserted at any place and SHOULD be ignored if not supported by the PCE

For endpoint type Point-to-Point the first endpoint and optional endpoint-restriction is the ingress endpoint. The second endpoint and optional endpoint-restriction is the egress endpoint. The further endpoint and endpoint-restriction are ignored

For endpoint type Point-to-Multipoint several endpoint objects may be present in the message and represent a leave, exact meaning depend on the endpoint type defined of the object.

An endpoint is defined as follow:

```

<endpoint>::=<IPv4-ADDRESS>|<IPv6-ADDRESS>|<UNNUMBERED-ENDPOINT>
<endpoint-restrictions> ::= <LABEL-REQUEST><label-restriction>
                               [<endpoint-restrictions>]
<label-restriction> ::= ((<LABEL><UPSTREAM-LABEL>)|
                          <LABEL-SET>|
                          <SUGGESTED-LABEL-SET>)
                          [<label-restriction>]

```

2.4.2. END-POINTS TLVs extensions

2.4.2.1. IPV4-ADDRESS

The format of the END-POINTS TLV object for IPv4 (TLV-Type=To be assigned) is as follows:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------|---|---|---|---|---|---|---|---|---|--------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Type | | | | | | | | | | Length | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IPv4 address | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

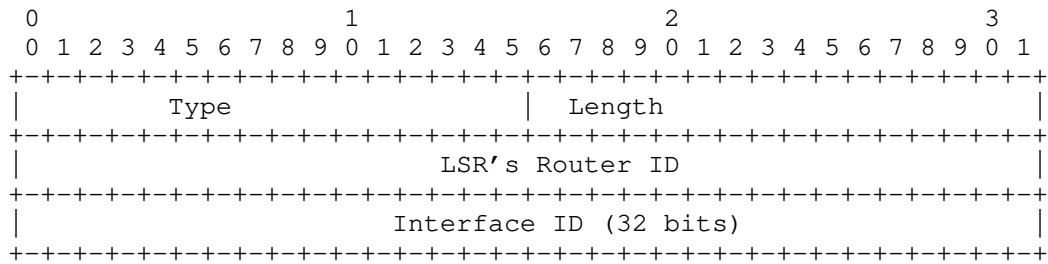
2.4.2.2. IPV6-ADDRESS TLV

The format of the END-POINTS TLV object for IPv6 (TLV-Type=To be assigned) is as follows:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------------------------|---|---|---|---|---|---|---|---|---|--------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Type | | | | | | | | | | Length | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| IPv6 address (16 bytes) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

2.4.2.3. UNNUMBERED-ENDPOINT TLV

This TLV represent an unnumbered interface. This TLV has the same semantic as in [RFC3477]



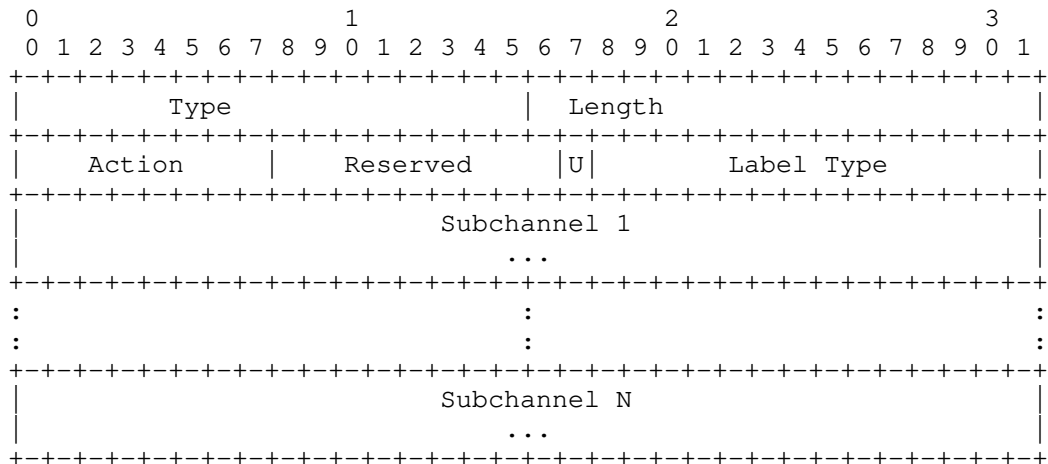
2.4.2.4. LABEL-REQUEST TLV

The LABEL-REQUEST indicate the and encoding of the LABEL restriction present in the ENDPOINTS its format is the same as described in [RFC3471] Section 3.1 Generalized label request

2.4.2.5. LABELS TLV

Label or label range may be specified for the TE-LSP endpoints. Those are encoded in the TLVs. The label value cannot be interpreted without a description on the Encoding and switching type. The REQ-ADAP-CAP object from [I-D.ietf-pce-inter-layer-ext] can be used in case of mono-layer request, however in case of multilayer it is possible to have in the future more than one object, so it is better to have a dedicated TLV for the label (the scope is then more clear). TLVs are encoded as follow (following [RFC5440]) :

- o LABEL TLV, Type = TBA by IANA, Length is variable, Encoding is as [RFC3471] Section 3.2 Generalized label. This represent the downstream label
- o UPSTEAM-LABEL TLV , Type = TBA by IANA, Length is variable, Encoding is as [RFC3471] Section 3.2 Generalized label. This represent the upstream label
- o LABEL-SET TLV, Type = TBA by IANA , Length is variable, Encoding follow :[RFC3471] Section 3.5 Label set with the addition of a U bit, the U bit is set for upstream direction in case of bidirectional LSP.



- o SUGGESTED-LABEL-SET TLV Set, Type = TBA by IANA, Length is variable, Encoding is as Label Set.

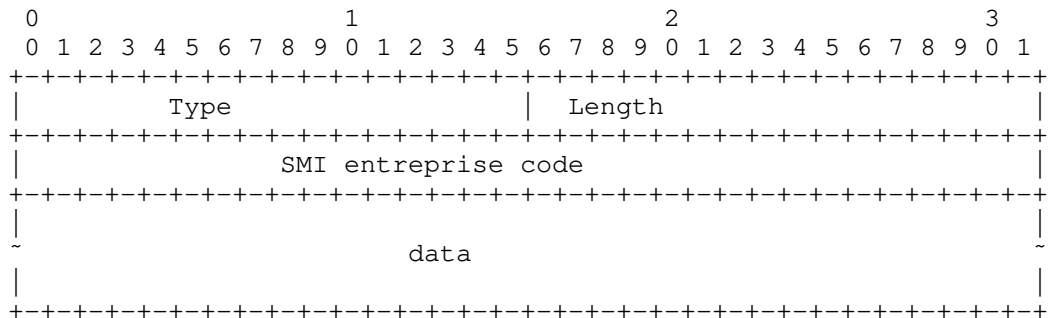
A LABEL TLV represent the label used on the unnumbered interface, bits I and U are used to indicate which exact unnumbered interface/direction is considered. the fields are encoded as in the RSVP-TE. The Encoding Type indicates the encoding type, e.g., SONET/SDH/GigE etc., that will be used with the data associated with the LSP. The Switching type indicates the type of switching that is being requested on the link. G-PID identifies the payload of the TE-LSP. The label type indicates which type of label (2) for generalized label is carried. A LABEL-SET TLV represents a set of possible labels that can be used on the unnumbered interface. The action parameter in the Label set indicates the type of list provided. Those parameters are described by [RFC3471] A SUGGESTED-LABEL-SET TLV has the same encoding as the LABEL-SET TLV, it includes the preferred (ordered) set of label to be used.

The U bit has the following meaning:

U: Upstream direction: set when the label or label set is in the reverse direction

2.4.2.6. Private TLVs

The format of the private TLV object is described as follow:



The length is at minimum 4 bytes.

2.5. LABEL-SET object

The LABEL-SET object is carried in a request within a PCReq message to restrict the set of labels to be assigned during the path computation. Any label included in the ERO object on the response must comply with the restrictions stated in the LABEL-SET, whose encoding is defined as following

`<LABEL-SET-OBJECT> ::= <LABEL-REQUEST><LABEL-SET>[<LABEL-SET>]`

The LABEL-REQUEST and LABEL-SET TLVs are as defined in Section 2.4.2.5, See also [RFC3471] and [RFC3473] for the definitions of the fields.

It is allowed to have more than one LABEL-SET object per request within a PCReq message (for example in case of multiple SWITCH-LAYER present).

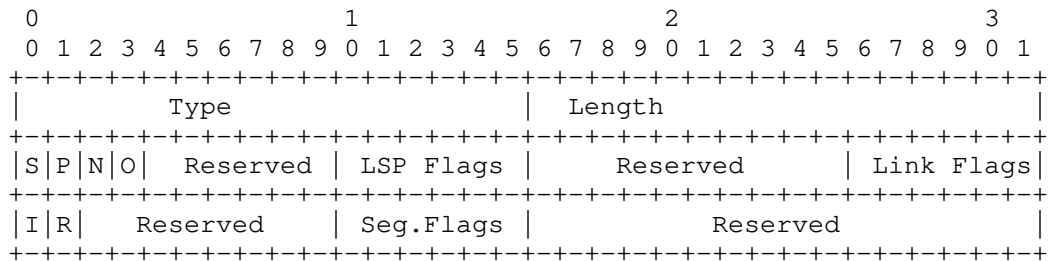
In the case of unsuccessful path computation the LABEL-SET object MAY be used to indicate the set of constraint that could not be satisfied.

2.6. SUGGESTED-LABEL-SET object

The SUGGESTED-LABEL-SET object is carried within a PCReq or PCRep message to indicate the preferred set of labels to be assigned during the path computation. The encoding is the same as the LABEL-SET object. It is allowed to have more than one SUGGESTED LABEL-SET object per PCReq (for example in case of multiple SWITCH-LAYER present).

2.7. LSPA extensions

The LSPA carries the LSP attributes. In the end-to-end protection context this also includes the protection state information. The LSPA object can be extended by a protection TLV type: Type TBA by IANA: protection attribute



The content is as defined in [RFC4872], [RFC4873].

LSP Flags can be considered for routing policy based on the protection type. The other attributes are only meaningful for a stateful PCE.

2.8. NO-PATH Object Extension

The NO-PATH object is used in PCRep messages in response to an unsuccessful path computation request (the PCE could not find a path by satisfying the set of constraints). In this scenario, PCE MUST include a NO-PATH object in the PCRep message. The NO-PATH object carries the NO-PATH-VECTOR TLV that specifies more information on the reasons that led to a negative reply. In case of GMPLS networks there could be some more additional constraints that led to the failure like protection mismatch, lack of resources, and so on. Few new flags have been introduced in the 32-bit flag field of the NO-PATH-VECTOR TLV and no modifications have been made in the NO-PATH object.

2.8.1. Extensions to NO-PATH-VECTOR TLV

The current NO-PATH-VECTOR TLV carry the following information:

Bit number 31 - PCE currently unavailable [RFC5440]

Bit number 30 - Unknown destination [RFC5440]

Bit number 29 - Unknown source [RFC5440]

Bit number 28 - BRPC Path computation chain unavailable [RFC5440]

Bit number 27 - PKS expansion failure [RFC5520]

Bit number 26 - No GCO migration path found [RFC5557]

Bit number 25 - No GCO solution found [RFC5557]

Bit number 24 - P2MP Reachability Problem [RFC5440]

The modified NO-PATH-VECTOR TLV carrying the additional information is as follows: New fields PM and NR are defined in the 23th and 22th bit of the Flags field respectively.

Bit number 23 (TBA by IANA) - Protection Mismatch (1-bit).
Specifies the mismatch of the protection type in the request.

Bit number 22 (TBA by IANA) - No Resource (1-bit). Specifies that the resources are not currently sufficient to provide the path.

Bit number 21 (TBA by IANA) - Granularity not supported (1-bit).
Specifies that the PCE is not able to provide a route with the requested granularity.

3. Additional Error Type and Error Values Defined

A PCEP-ERROR object is used to report a PCEP error and is characterized by an Error-Type that specifies the type of error and an Error-value that provides additional information about the error type. An additional error type and few error values are defined to represent some of the errors related to the newly identified objects related to SDH networks. For each PCEP error, an Error-Type and an Error-value are defined. Error-Type 1 to 10 are already defined in [RFC5440]. Additional Error- values are defined for Error-Type 10 and A new Error-Type 14 is introduced.

Error-Type Error-value

| | | |
|----|--------------------------------|---|
| 10 | Reception of an invalid object | |
| | Error-value=1: | Bad Generalized Bandwidth Object value. |
| | Error-value=2: | Unsupported LSP Protection Type in protection attribute TLV. |
| | Error-value=3: | Unsupported LSP Protection Flags in protection attribute TLV. |
| | Error-value=4: | Unsupported Secondary LSP Protection Flags in protection attribute TLV. |
| | Error-value=5: | Unsupported Link Protection Type in protection attribute TLV. |
| | Error-value=6: | Unsupported Link Protection Type in protection attribute TLV. |
| 14 | Path computation failure | |
| | Error-value=1: | Unacceptable request message. |
| | Error-value=2: | Generalized bandwidth object not supported. |
| | Error-value=3: | Label Set constraint could not be met. |
| | Error-value=4: | Label constraint could not be met. |
| | Error-value=5: | Unsupported endpoint type in END-POINTS GENERALIZED-ENDPOINTS object type |

Error-value=6: Unsupported TLV present in END-POINTS
 GENERALIZED-ENDPOINTS object type

Error-value=7: Unsupported granularity in the RP object
 flags

4. Manageability Considerations

Liveness Detection and Monitoring This document makes no change to the basic operation of PCEP and so there are no changes to the requirements for liveness detection and monitoring set out in [RFC4657] and [RFC5440].

5. IANA Considerations

IANA assigns values to the PCEP protocol objects and TLVs. IANA is requested to make some allocations for the newly defined objects and TLVs introduced in this document. Also, IANA is requested to manage the space of flags that are newly added in the TLVs.

5.1. PCEP Objects

As described in Section 2.2 and Section 2.3 new Objects are defined IANA is requested to make the following Object-Type allocations from the "PCEP Objects" sub-registry:

Object Class to be assigned

Name GENERALIZED-BANDWIDTH

Object-Type 0 to 6

Reference This document (section Section 2.2)

Object Class to be assigned

Name GENERALIZED-LOAD-BALANCING

Object-Type 0 to 6

Reference This document (section Section 2.3)

As described in Section 2.4.1 a new Object type is defined IANA is requested to make the following Object-Type allocations from the "PCEP Objects" sub-registry:

Object Class 4

Name END-POINTS

Object-Type 5 : Generalized Endpoint

6-15 : unassigned

Reference This document (section Section 2.2)

5.2. New PCEP TLVs

IANA is requested to create a registry for the following TLVs:

| Value | Meaning | Reference |
|-------|--------------------------------|---|
| x | IPV4 endpoint | This document (section Section 2.4.2.1) |
| x | IPV6 endpoint | This document (section Section 2.4.2.2) |
| x | Unnumbered endpoint | This document (section Section 2.4.2.3) |
| x | Label request | This document (section Section 2.4.2.4) |
| x | Requested GMPLS Label | This document (section Section 2.4.2.5) |
| x | Requested GMPLS Upstream Label | This document (section Section 2.4.2.5) |
| x | Requested GMPLS Label Set | This document (section Section 2.4.2.5) |
| x | Suggested GMPLS Label Set | This document (section Section 2.4.2.5) |
| x | LSP Protection Information | This document (section Section 2.7) |

5.3. New PCEP Error Codes

As described in Section Section 3, new PCEP Error-Type and Error Values are defined. IANA is requested to manage the code space of the Error object.

| Error-Type | Error-value |
|----------------|---|
| 10 | Reception of an invalid object |
| Error-value=1: | Bad Generalized Bandwidth Object value. |
| Error-value=2: | Unsupported LSP Protection Type in protection attribute TLV. |
| Error-value=3: | Unsupported LSP Protection Flags in protection attribute TLV. |
| Error-value=4: | Unsupported Secondary LSP Protection Flags in protection attribute TLV. |
| Error-value=5: | Unsupported Link Protection Type in protection attribute TLV. |
| Error-value=6: | Unsupported Link Protection Type in protection attribute TLV. |
| 14 | Path computation failure |
| Error-value=1: | Unacceptable request message. |
| Error-value=2: | Generalized bandwidth object not supported. |
| Error-value=3: | Label Set constraint could not be met. |
| Error-value=4: | Label constraint could not be met. |
| Error-value=5: | Unsupported endpoint type in END-POINTS GENERALIZED-ENDPOINTS object type |
| Error-value=6: | Unsupported TLV present in END-POINTS GENERALIZED-ENDPOINTS object type |
| Error-value=7: | Unsupported granularity in the RP object flags |

6. Security Considerations

None.

7. Contributing Authors

Nokia Siemens Networks:

Elie Sfeir
St Martin Strasse 76
Munich, 81541
Germany

Phone: +49 89 5159 16159
Email: elie.sfeir@nsn.com

Franz Rambach
St Martin Strasse 76
Munich, 81541
Germany

Phone: +49 89 5159 31188
Email: franz.rambach@nsn.com

Francisco Javier Jimenez Chico
Telefonica Investigacion y Desarrollo
C/ Emilio Vargas 6
Madrid, 28043
Spain

Phone: +34 91 3379037
Email: fjjc@tid.es

Huawei Technologies

Suresh BR
Shenzhen
China
Email: sureshbr@huawei.com

Young Lee
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

SenthilKumar S
Shenzhen
China
Email: senthilkumars@huawei.com

Jun Sun
Shenzhen
China
Email: johnsun@huawei.com

CTTC - Centre Tecnologic de Telecomunicacions de Catalunya

Ramon Casellas
PMT Ed B4 Av. Carl Friedrich Gauss 7
08860 Castelldefels (Barcelona)
Spain
Phone: (34) 936452916
Email: ramon.casellas@cttc.es

8. Acknowledgments

The research of Ramon Casellas, Francisco Javier Jimenez Chico, Oscar Gonzalez de Dios, Cyril Margaria, and Franz Rambach leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n. 247674.

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated Services", RFC 2210, September 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol -Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.
- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4328] Papadimitriou, D., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC4606] Mannie, E. and D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", RFC 4606, August 2006.
- [RFC4872] Lang, J., Rekhter, Y., and D. Papadimitriou, "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.
- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, May 2007.
- [RFC5088] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang,

- "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5520] Bradford, R., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, April 2009.
- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, April 2009.
- [RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, June 2009.
- [RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, July 2009.
- [RFC6003] Papadimitriou, D., "Ethernet Traffic Parameters", RFC 6003, October 2010.

9.2. Informative References

- [I-D.ceccarelli-ccamp-gmpls-ospf-g709]
Ceccarelli, D., Caviglia, D., Zhang, F., Li, D., Xu, Y., Belotti, S., Grandi, P., and J. Drake, "Traffic Engineering Extensions to OSPF for Generalized MPLS (GMPLS) Control of Evolving G.709 OTN Networks", draft-ceccarelli-ccamp-gmpls-ospf-g709-04 (work in progress), October 2010.
- [I-D.ietf-ccamp-gmpls-g-694-lambda-labels]
Otani, T., Rabbat, R., Shiba, S., Guo, H., Miyazaki, K., Caviglia, D., Li, D., and T. Tsuritani, "Generalized Labels for Lambda-Switching Capable Label Switching Routers", draft-ietf-ccamp-gmpls-g-694-lambda-labels-07 (work in progress), April 2010.

- [I-D.ietf-pce-gmpls-aps-req]
Otani, T., Ogaki, K., Caviglia, D., and F. Zhang,
"Document: draft-ietf-pce-gmpls-aps-req-03.txt",
draft-ietf-pce-gmpls-aps-req-03 (work in progress),
October 2010.
- [I-D.ietf-pce-inter-layer-ext]
Oki, E., Takeda, T., Roux, J., and A. Farrel, "Extensions
to the Path Computation Element communication Protocol
(PCEP) for Inter-Layer MPLS and GMPLS Traffic
Engineering", draft-ietf-pce-inter-layer-ext-04 (work in
progress), July 2010.
- [I-D.ietf-pce-wson-routing-wavelength]
Lee, Y., Bernstein, G., Martensson, J., Takeda, T., and T.
Tsuritani, "PCEP Requirements for WSON Routing and
Wavelength Assignment",
draft-ietf-pce-wson-routing-wavelength-02 (work in
progress), August 2010.
- [I-D.zhang-ccamp-gmpls-evolving-g709]
Zhang, F., Zhang, G., Belotti, S., Ceccarelli, D., Lin,
Y., Xu, Y., Grandi, P., and D. Caviglia, "Generalized
Multi-Protocol Label Switching (GMPLS) Signaling
Extensions for the evolving G.709 Optical Transport
Networks Control",
draft-zhang-ccamp-gmpls-evolving-g709-06 (work in
progress), October 2010.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE)
Communication Protocol Generic Requirements", RFC 4657,
September 2006.
- [RFC5467] Berger, L., Takacs, A., Caviglia, D., Fedyk, D., and J.
Meuric, "GMPLS Asymmetric Bandwidth Bidirectional Label
Switched Paths (LSPs)", RFC 5467, March 2009.

Authors' Addresses

Cyril Margaria (editor)
Nokia Siemens Networks
St Martin Strasse 76
Munich, 81541
Germany

Phone: +49 89 5159 16934
Email: cyril.margaria@nsn.com

Oscar Gonzalez de Dios (editor)
Telefonica Investigacion y Desarrollo
C/ Emilio Vargas 6
Madrid, 28043
Spain

Phone: +34 91 3374013
Email: ogondio@tid.es

Fatai Zhang (editor)
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen, 518129
P.R.China

Email: zhangfatai@huawei.com

Network Working Group
Internet Draft
Intended status: Standard Track
Expires: February 2011

Y. Lee
Huawei

G. Bernstein
Grotto Networking

Jonas Martensson
Acreo

T. Takeda
NTT

T. Tsuritani
KDDI

August 23, 2010

PCEP Requirements for WSON Routing and Wavelength Assignment

draft-ietf-pce-wson-routing-wavelength-02.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on December 23, 2010.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Abstract

This memo provides application-specific requirements for the Path Computation Element communication Protocol (PCEP) for the support of Wavelength Switched Optical Networks (WSON). Lightpath provisioning in WSONs requires a routing and wavelength assignment (RWA) process. From a path computation perspective, wavelength assignment is the process of determining which wavelength can be used on each hop of a path and forms an additional routing constraint to optical light path computation. Requirements for Optical impairments will be addressed in a separate document.

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 RFC-2119 0.

Table of Contents

| | |
|---|---|
| 1. Introduction..... | 3 |
| 1.1. WSON RWA Processes..... | 4 |
| 2. WSON PCE Architectures and Requirements..... | 5 |
| 2.1. RWA PCC to PCE Interface..... | 6 |
| 2.1.1. A new RWA path request/reply..... | 6 |
| 2.1.2. Bulk RWA path request/reply..... | 6 |
| 2.1.3. An RWA path re-optimization request/reply..... | 7 |
| 2.1.4. Wavelength Range Constraint..... | 7 |
| 2.1.5. Wavelength Policy Constraint..... | 7 |

| | |
|---|----|
| 3. Manageability Considerations..... | 8 |
| 3.1. Control of Function and Policy..... | 8 |
| 3.2. Information and Data Models, e.g. MIB module..... | 8 |
| 3.3. Liveness Detection and Monitoring..... | 8 |
| 3.4. Verifying Correct Operation..... | 9 |
| 3.5. Requirements on Other Protocols and Functional Components..... | 9 |
| 3.6. Impact on Network Operation..... | 9 |
| 4. Security Considerations..... | 9 |
| 5. IANA Considerations..... | 9 |
| 6. Acknowledgments..... | 9 |
| 7. References..... | 10 |
| 7.1. Normative References..... | 10 |
| 7.2. Informative References..... | 10 |
| Authors' Addresses..... | 11 |
| Intellectual Property Statement..... | 11 |
| Disclaimer of Validity..... | 12 |

1. Introduction

[RFC4655] defines the PCE based Architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is shown to be any network component that makes such a request and may be for instance an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communications Protocol (PCEP) is the communication protocol used between PCC and PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

This document provides a set of application-specific PCEP requirements for support of path computation in Wavelength Switched Optical Networks (WSON). WSON refers to WDM based optical networks in which switching is performed selectively based on the wavelength of an optical signal.

The path in WSON is referred to as a lightpath. A lightpath may span multiple fiber links and the path should be assigned a wavelength for

each link. A transparent optical network is made up of optical devices that can switch but not convert from one wavelength to another. In a transparent optical network, a lightpath operates on the same wavelength across all fiber links that it traverses. In such case, the lightpath is said to satisfy the wavelength-continuity constraint. Two lightpaths that share a common fiber link can not be assigned the same wavelength. To do otherwise would result in both signals interfering with each other. Note that advanced additional multiplexing techniques such as polarization based multiplexing are not addressed in this document since the physical layer aspects are not currently standardized. Therefore, assigning the proper wavelength on a lightpath is an essential requirement in the optical path computation process.

When a switching node has the ability to perform wavelength conversion the wavelength-continuity constraint can be relaxed, and a lightpath may use different wavelengths on different links along its route from origin to destination. It is, however, to be noted that wavelength converters may be limited due to their relatively high cost, while the number of WDM channels that can be supported in a fiber is also limited. As a WSON can be composed of network nodes that cannot perform wavelength conversion, nodes with limited wavelength conversion, and nodes with full wavelength conversion abilities, wavelength assignment is an additional routing constraint to be considered in all lightpath computation.

In this document we first review the processes for routing and wavelength assignment (RWA) used when wavelength continuity constraints are present and then specify requirements for PCEP to support RWA.

The remainder of this document uses terminology from [RFC4655].

1.1. WSON RWA Processes

In [WSON-Frame] three alternative process architectures were given for performing routing and wavelength assignment. These are shown schematically in Figure 1.

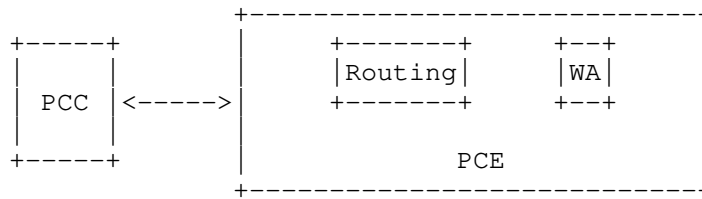


Figure 2 Combined Process (R&WA) architecture

2.1. RWA PCC to PCE Interface

The requirements for the PCC to PCE interface of Figure 2 are specified in this section.

2.1.1. A new RWA path request/reply

1. The PCReq Message MUST include the path computation type. This can be:
 - (i) Both Routing and Wavelength Assignment (RWA), or
 - (ii) Routing only.

This requirement is needed to differentiate between the currently supported routing with distributed wavelength assignment option and combined RWA. In case of distributed wavelength assignment option, wavelength assignment will be performed at each node of the route.

2. The PCRep Message MUST include the route, wavelengths assigned to the route (i.e., each hop of the route must be assigned a wavelength).
3. In the case where a valid path is not found, the PCRep Message MUST include why the path is not found (e.g., no route, wavelength not found, optical quality check failed, etc.)

2.1.2. Bulk RWA path request/reply

1. The PCReq Message MUST be able to specify an option for bulk RWA path request. Bulk path request is an ability to request a number of simultaneous RWA path requests.

2. The PCRep Message MUST include the route, wavelength assigned to the route for each RWA path request specified in the original bulk PCReq Message.

2.1.3. An RWA path re-optimization request/reply

1. For a re-optimization request, the PCReq Message MUST provide the path to be re-optimized and include the following options:
 - a. Re-optimize the path keeping the same wavelength(s)
 - b. Re-optimize wavelength(s) keeping the same path
 - c. Re-optimize allowing both wavelength and the path to change
2. The corresponding PCRep Message for the re-optimized request MUST provide the Re-optimized path and wavelengths.
3. In case that the path is not found, the PCRep Message MUST include why the path is not found (e.g., no route, wavelength not found, both route and wavelength not found, etc.)

2.1.4. Wavelength Range Constraint

For any PCReq Message that is associated with a request for wavelength assignment the requester (PCC) MUST be able to specify a restriction on the wavelengths to be used.

Note that the requestor (PCC) is NOT required to furnish any range restrictions. This restriction is to be interpreted by the PCE as a constraint on the tuning ability of the origination laser transmitter.

2.1.5. Wavelength Policy Constraint

The PCReq Message May include specific operator's policy information for WA (E.g., random assignment, descending order, ascending order, etc.)

3. Manageability Considerations

Manageability of WSON Routing and Wavelength Assignment (RWA) with PCE must address the following considerations:

3.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [PCEP], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- o The ability to send a WSON RWA request.

In addition to the parameters already listed in Section 8.1 of [PCEP], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- o The support for WSON RWA.
- o The maximum number of synchronized path requests associated with WSON RWA per request message.
- o A set of WSON RWA specific policies (authorized sender, request rate limiter, etc).

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

3.2. Information and Data Models, e.g. MIB module

Extensions to the PCEP MIB module defined in [PCEP-MIB] should be defined, so as to cover the WSON RWA information introduced in this document. A future revision of this document will list the information that should be added to the MIB module.

3.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in section 8.3 of [PCEP].

3.4. Verifying Correct Operation

Mechanisms defined in this document do not imply any new verification requirements in addition to those already listed in section 8.4 of [PCEP]

3.5. Requirements on Other Protocols and Functional Components

The PCE Discovery mechanisms ([RFC5089] and [RFC5088]) may be used to advertise WSON RWA path computation capabilities to PCCs.

3.6. Impact on Network Operation

Mechanisms defined in this document do not imply any new network operation requirements in addition to those already listed in section 8.6 of [PCEP].

4. Security Considerations

This document has no requirement for a change to the security models within PCEP [PCEP]. However the additional information distributed in order to address the RWA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

5. IANA Considerations

A future revision of this document will present requests to IANA for codepoint allocation.

6. Acknowledgments

The authors would like to thank Adrian Farrel for many helpful comments that greatly improved the contents of this draft.

This document was prepared using 2-Word-v2.0.template.dot.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol", RFC 5440, March 2009.

7.2. Informative References

- [WSON-IMP] Lee, Y. and Bernstein, G. (Editors), D. Li and G. Martinelli "A Framework for the Control and Measurement of Wavelength Switched Optical Networks (WSON) with Impairments, draft-ietf-ccamp-wson-impairments, work in progress.
- [WSON-Frame] Lee, Y. and Bernstein, G. (Editors), and W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-framework, work in progress.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.

[RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.

Authors' Addresses

Young Lee (Ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075, USA
Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Greg Bernstein (Ed.)
Grotto Networking
Fremont, CA, USA
Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Jonas Martensson
Acreo
Email: Jonas.Martensson@acreo.se

Tomonori Takeda
NTT Corporation
3-9-11, Midori-Cho
Musashino-Shi, Tokyo 180-8585, Japan
Email: takeda.tomonori@lab.ntt.co.jp

Takehiro Tsuritani
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Kamifukuoka Saitama, 356-8502. Japan
Phone: +81-49-278-7357
Email: tsuri@kddilabs.jp

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license

under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet Draft
Intended status: Standard Track
Expires: April 2011

Y. Lee
Huawei

G. Bernstein
Grotto Networking

Jonas Martensson
Acreo

T. Tsuritani
KDDI

Oscar Gonzalez de Dios
Telefonica

October 22, 2010

PCEP Extensions in support of WSON Signal Compatibility Constraints

draft-lee-pce-wson-signal-compatibility-03.txt

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

This Internet-Draft will expire on April 22, 2009.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Abstract

This memo provides the Path Computation Element communication Protocol (PCEP) extensions for the support of signal compatibility constraints in Wavelength Switched Optical Networks (WSON).

Signal compatibility is an essential path computation constraint in path computation of WSON networks where network elements can be limited to processing WSON signals with specific characteristics and attributes.

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 RFC-2119 0.

Table of Contents

| | |
|---|----|
| 1. Introduction..... | 3 |
| 2. PCEP Requirements..... | 4 |
| 3. PCEP Extensions (Encoding)..... | 5 |
| 3.1. The PCC to PCE Interface..... | 5 |
| 3.1.1. Signal Compatibility Indicator in the RP Object in the PC Request Message..... | 5 |
| 3.1.2. Modulation Type List sub-TLV..... | 5 |
| 3.1.3. Channel Spacing sub-TLV..... | 7 |
| 3.1.4. FEC Type List sub-TLV..... | 8 |
| 3.1.5. The GPID Type Sub-TLV..... | 11 |

| | |
|--|----|
| 3.2. The PCE to PCC Interface..... | 11 |
| 3.2.1. Modulation Type sub-TLV..... | 12 |
| 3.2.2. FEC Type sub-TLV..... | 12 |
| 3.2.3. Regeneration Point sub-TLV..... | 12 |
| 4. Manageability Considerations..... | 13 |
| 5. Security Considerations..... | 13 |
| 6. IANA Considerations..... | 13 |
| 7. Acknowledgments..... | 13 |
| 8. References..... | 14 |
| 8.1. Normative References..... | 14 |
| 8.2. Informative References..... | 14 |
| Authors' Addresses..... | 15 |
| Intellectual Property Statement..... | 16 |
| Disclaimer of Validity..... | 16 |

1. Introduction

[RFC4655] defines the PCE based Architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is shown to be any network component that makes such a request and may be for instance an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communications Protocol (PCEP) is the communication protocol used between PCC and PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

This document provides the Path Computation Element communication Protocol (PCEP) extensions for the support of signal compatibility constraints in Wavelength Switched Optical Networks (WSON). Signal compatibility is an essential path computation constraint in path computation of WSON networks where network elements can be limited to processing WSON signals with specific characteristics and attributes.

Signals used in a WSON are not always compatible with common network elements including regenerators, OEO switches, wavelength converters, etc. [WSON-Frame] defines the GMPLS control plane framework that

allows both multiple WSON signal types and common hybrid electro optical systems. Reference [WSON-Frame] characterizes WSON signals in line with ITU-T standards, and adds attributes describing signal compatibility constraints to WSON network elements.

[CompatOSPF] provides GMPLS OSPF routing enhancements to support signal compatibility constraints associated with WSON network elements. On a high-level the following network element information would be required for the path computation element (PCE) to be able to compute a constrained path that satisfies signal compatibility and processing constraints:

- . Input Compatibility: the type of signals it can receive (modulation types, Channel Spacing, FEC types)
- . Regeneration Capability: the types of processing/enhancement it can perform (1R, 2R, 3R)
- . The types of conversions it can perform (modulation types, FEC types)
- . Output Format: the type of signals it can transmit (modulation types, Channel Spacing, FEC types)

2. PCEP Requirements

This section provides a set of PCEP requirements to support signal compatibility constraints.

When requesting a path computation (PCReq) to PCE, the PCC should be able to indicate the following:

- o The acceptable signal attributes at the transmitter (at the source): (i) Modulation types; (ii) Channel Spacing; (iii) FEC types
- o The acceptable signal attributes at the receiver (at the sink): (i) Modulation types; (ii) Channel Spacing; (iii) FEC types
- o The ability to specify if regeneration is allowed in the computed path; if allowed, the maximum number of regenerators allowed in the computed path should be indicated.

The PCE should be able to respond (PC Rep) to the PCC with the following:

- o The conformity of the requested optical characteristics associated with the resulting LSP with the source, sink and Network Element (NE) along the LSP.
- o Additional LSP attributes modified along the path (e.g., modulation format change, etc.)
- o Special node processing with the resulting LSP (e.g., regeneration point)

3. PCEP Extensions (Encoding)

This section provides PCEP encoding to support the identified requirements in Section 2.

3.1. The PCC to PCE Interface

This section provides the enhancements to the Request Parameter (RP) Object and its associated sub-TLV in the PC Request message from the PCC to the PCE interface to support signal compatibility constraints.

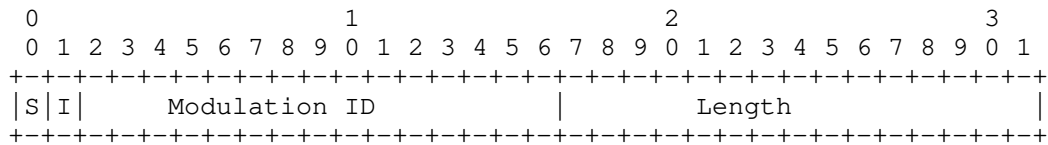
3.1.1. Signal Compatibility Indicator in the RP Object in the PC Request Message

The RP object should have a bit indicating that signal compatibility check (say SC bit) should be performed for a path computation.

3.1.2. Modulation Type List sub-TLV

When the SC bit in the RP Object is set to 1, then the RP object should include the Modulation Type List TLV associated with the request. Modulation types listed in the Modulation Type List TLV indicate allowable modulation types in both the source (transmitter) and the sink (receiver).

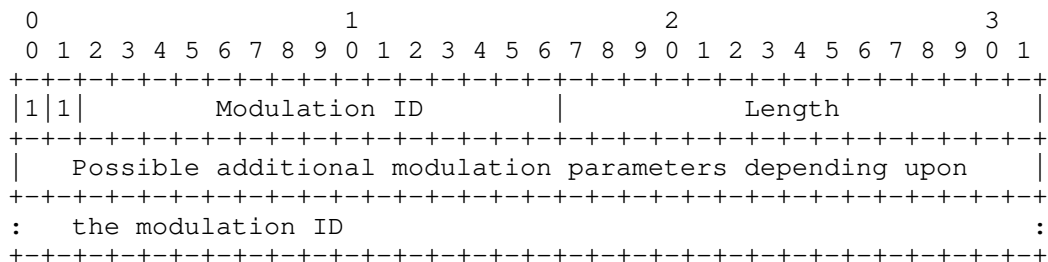
The modulation type list sub-TLV may consist of two different types of fields: a standard modulation field or a vendor specific modulation field. Both start with the same 32 bit header shown below.



Where S bit set to 1 indicates a standardized modulation format and S bit set to 0 indicates a vendor specific modulation format. The length is the length in bytes of the entire modulation type field.

Where I bit set to 1 indicates an input modulation format and where I bit set to 0 indicates an output modulation format. Note that the source modulation type is implied when I bit is set to 0 and that the sink modulation type is implied when I bit is set to 1.

The format for the standardized type for the input modulation at the sink is given by:



Modulation ID

Takes on the following currently defined values:

- 0 Reserved
- 1 optical tributary signal class NRZ 1.25G
- 2 optical tributary signal class NRZ 2.5G
- 3 optical tributary signal class NRZ 10G
- 4 optical tributary signal class NRZ 40G
- 5 optical tributary signal class RZ 40G

Note that future modulation types may require additional parameters in their characterization.

The format for vendor specific input modulation field at the sink is given by:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
| 0 | 1 |   Vendor Modulation ID   |           Length           |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Enterprise Number          |
+-----+-----+-----+-----+-----+-----+-----+-----+
:   Any vendor specific additional modulation parameters       :
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Vendor Modulation ID

This is a vendor assigned identifier for the modulation type.

Enterprise Number

A unique identifier of an organization encoded as a 32-bit integer. Enterprise Numbers are assigned by IANA and managed through an IANA registry [RFC2578].

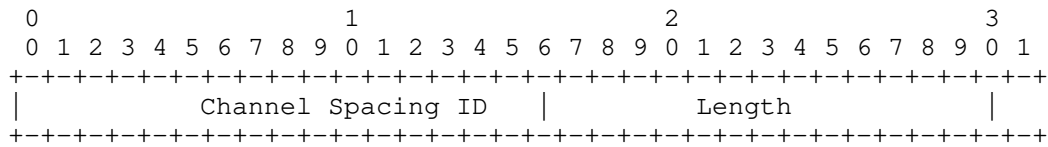
Vendor Specific Additional parameters

There can be potentially additional parameters characterizing the vendor specific modulation.

3.1.3. Channel Spacing sub-TLV

When the SC bit in the RP Object is set to 1, then the RP object should include the Channel Spacing Type List TLV associated with the request. Spacing types listed in the Channel Spacing Type List TLV indicate allowable Channel Spacing types in both source (transmitter) and the sink (receiver).

The format for the channel spacing field is given by:



Channel Spacing ID

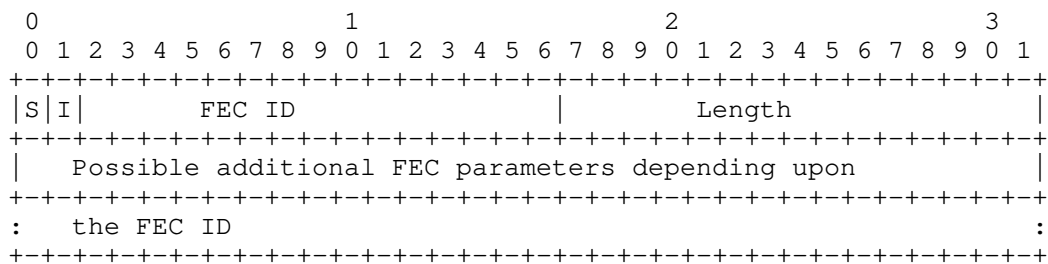
Takes on the following currently defined values:

| | |
|------|------------|
| 0 | Reserved |
| 1 | 100 GHz |
| 2 | 50 GHz |
| 3 | 25 GHz |
| 4 | 12.5 GHz |
| 5-15 | Future Use |

3.1.4. FEC Type List sub-TLV

When the SC bit in the RP Object is set to 1, then the RP object should include the FEC Type List TLV associated with the request. FEC types listed in the FEC Type List TLV indicate allowable FEC types in both the source (transmitter) and the sink (receiver).

The FEC type list sub-TLV may consist of two different types of fields: a standard FEC field or a vendor specific FEC field. Both start with the same 32 bit header shown below.

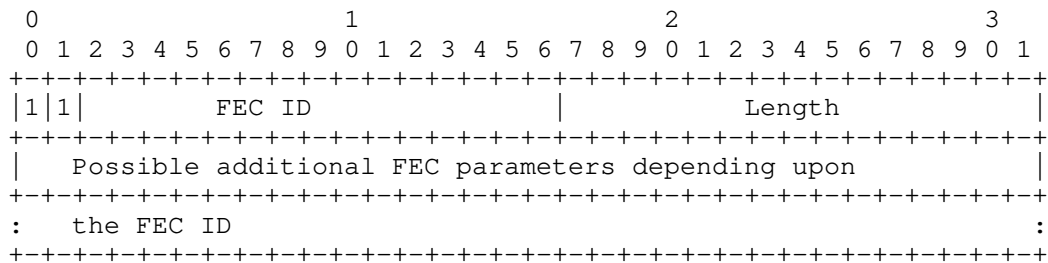


Where S bit set to 1 indicates a standardized FEC format and S bit set to 0 indicates a vendor specific FEC format.

Where the length is the length in bytes of the entire FEC type field.

Where I bit set to 1 indicates an input FEC format and where I bit set to 0 indicates an output FEC format. Note that the source FEC type is implied when I bit is set to 0 and that the sink FEC type is implied when I bit is set to 1.

The format for input standard FEC field at the sink is given by:



Takes on the following currently defined values for the standard FEC ID:

| | |
|---|---|
| 0 | Reserved |
| 1 | G.709 RS FEC |
| 2 | G.709V compliant Ultra FEC |
| 3 | G.975.1 Concatenated FEC (RS(255,239)/CSOC(n0/k0=7/6,J=8)) |
| 4 | G.975.1 Concatenated FEC (BCH(3860,3824)/BCH(2040,1930)) |
| 5 | G.975.1 Concatenated FEC (RS(1023,1007)/BCH(2407,1952)) |
| 6 | G.975.1 Concatenated FEC (RS(1901,1855)/Extended Hamming Product Code (512,502)X(510,500)) |
| 7 | G.975.1 LDPC Code |

- 8 G.975.1 Concatenated FEC (Two orthogonally concatenated BCH codes)
- 9 G.975.1 RS(2720,2550)
- 10 G.975.1 Concatenated FEC (Two interleaved extended BCH (1020,988) codes)

Where RS stands for Reed-Solomon and BCH for Bose-Chaudhuri-Hocquengham.

The format for input vendor-specific FEC field at the sink is given by:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|0|1|           Vendor FEC ID           |           Length           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Enterprise Number           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
:   Any vendor specific additional FEC parameters   :
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Vendor FEC ID

This is a vendor assigned identifier for the FEC type.

Enterprise Number

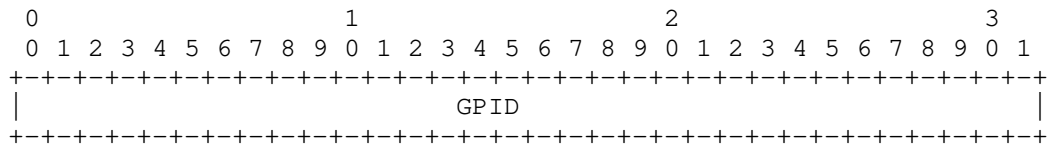
A unique identifier of an organization encoded as a 32-bit integer. Enterprise Numbers are assigned by IANA and managed through an IANA registry [RFC2578].

Vendor Specific Additional FEC parameters

There can be potentially additional parameters characterizing the vendor specific FEC.

3.1.5. The GPID Type Sub-TLV

When the SC bit in the RP Object is set to 1, then the RP object should include the GPID Type sub-TLV with the path request. The GPID Type should be one of Generalized Protocol Identifiers (GPIDs). GPIDs are assigned by IANA and many are defined in [RFC3471] and [RFC4328].



Where GPID is an identifier encoded as a 32-bit integer.

3.2. The PCE to PCC Interface

This section provides the enhancements to the RP Object and its associated sub-TLV in the PC Reply message from the PCE to the PCC interface to support signal compatibility constraints.

The PCE MUST specify the detail signal compatibility information in response to the "SC" (Signal Compatibility) request made by the PCC. The ERO object in the PC Rep message SHOULD include the following sub-TLV if the SC-bit is set in the RP object in the PC Req message:

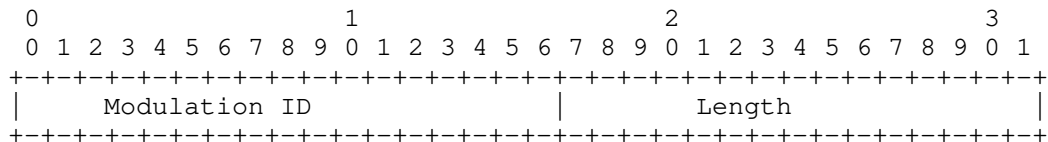
- o Modulation Type sub-TLV
- o FEC Type sub-TLV
- o Regeneration Point sub-TLV

Note that each of the TLV defined above would be in an ERO as subobjects placed after the node identifier (IP address).

In addition, the PC Rep message SHOULD specify a list of Modulation/FEC types (per node identifier) in the ERO object in the case when more than one is compatible with all constraints.

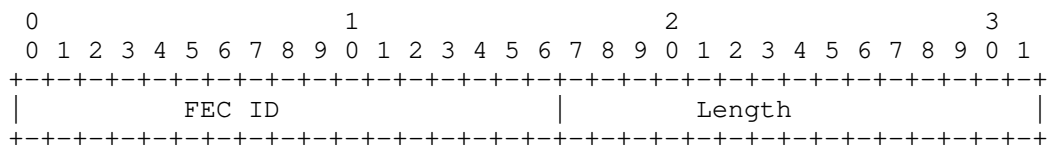
3.2.1. Modulation Type sub-TLV

The modulation type sub-TLV indicates the output modulation type associated with the node identifier. It starts with the same 32 bit header shown below.



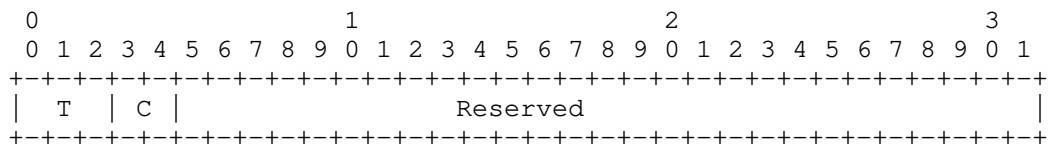
3.2.2. FEC Type sub-TLV

The FEC type sub-TLV indicates the output FEC type associated with the node identifier. It starts with the same 32 bit header shown below.



3.2.3. Regeneration Point sub-TLV

The Regeneration Point sub-TLV indicates this particular node is a regeneration point.



Where T bit indicates the type of regenerator:

T=0: Reserved

T=1: 1R Regenerator

T=2: 2R Regenerator

T=3: 3R Regenerator

Where C bit indicates the capability of regenerator:

C=0: Reserved

C=1: Fixed Regeneration Point

C=2: Selective Regeneration Pools

Note that when the capability of regenerator is indicated to be Selective Regeneration Pools, regeneration pool properties such as ingress and egress restrictions and availability need to be specified. This encoding is to be determined in the later revision.

4. Manageability Considerations

This document does not add additional manageability considerations.

5. Security Considerations

This document has no requirement for a change to the security models within PCEP [PCEP]. However the additional information distributed in order to address the RWA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

6. IANA Considerations

A future revision of this document will present requests to IANA for codepoint allocation.

7. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol (PCEP) - Version 1", RFC 5440, March 2009.
- [CompatOSPF] Lee, Y. and Bernstein, G., "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", draft-lee-ccamp-wson-signal-compatibility-ospf, work in progress.

8.2. Informative References

- [WSON-IMP] Lee, Y. and Bernstein, G. (Editors), D. Li, G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSON) with Impairments", draft-ietf-ccamp-wson-impairments, work in progress.
- [WSON-Frame] Lee, Y. and Bernstein, G. (Editors), and W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-framework, work in progress.

- [CompatOSPF] Lee, Y. and Bernstein, G. (Editors), J. Martensson, and Takehiro Tsuritani, "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signal-compatibility-ospf, work in progress.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [G.709] ITU-T Recommendation G.709, Interfaces for the Optical Transport Network(OTN), March 2003.
- [G.975.1] ITU-T Recommendation G.975.1, Forward error correction for high bit-rate DWDM submarine systems, February, 2004.

Authors' Addresses

Young Lee (Ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075, USA
Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Greg Bernstein (Ed.)
Grotto Networking
Fremont, CA, USA
Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Jonas Martensson
Acreo
Email:Jonas.Martensson@acreo.se

Takehiro Tsuritani
2-1-15 Ohara, Fujimino, Saitama, 356-8502, JAPAN
KDDI R&D Laboratories Inc.
Phone: +81-49-278-7806
Email: tsuri@kddilabs.jp

Oscar Gonzalez de Dios
Telefonica
Email : ogondio@tid.es

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

PCE
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2011

W. Lu
S. Kini
S. Narayanan
Ericsson
October 18, 2010

Relayed CSPF for Multi-Area Multi-AS PCE
draft-lu-relayed-cspf-00

Abstract

For LSPs that span across multiple areas or multiple autonomous systems (AS), the path computation element (PCE) in each area or each AS can cooperate and conclude the optimal results if so exist. An upstream PCE, though incapable to carry out the path computation for the tailend outside of its domain, can provide the history of the computation to its downstream PCEs which will assume the computation job till it is accomplished or relay the baton to the next runner.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 21, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must

include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

| | |
|--|----|
| 1. Introduction | 3 |
| 1.1. Requirements Language | 4 |
| 1.2. Acronyms | 4 |
| 2. CSPF Seed | 4 |
| 2.1. Multiple Seeds | 4 |
| 2.2. Heap Equivalence | 4 |
| 2.2.1. SPT Equivalence | 5 |
| 2.2.2. Seed Deposit Timing | 6 |
| 2.2.3. Seed Set Reduction | 6 |
| 3. Multi-Area Path Computation | 6 |
| 3.1. Inter-area PCE | 6 |
| 3.2. Initial Seed Set | 7 |
| 3.3. PCE Relay | 7 |
| 3.4. Relay Content | 8 |
| 3.5. PCE Elect | 9 |
| 4. Multi-Home Tail-End | 9 |
| 4.1. Relay Timer | 10 |
| 5. Multi-AS Path Computation | 11 |
| 5.1. Information Hiding | 11 |
| 5.2. Transit Link | 12 |
| 5.3. AS Number | 12 |
| 6. Other Cases | 12 |
| 6.1. Backups | 12 |
| 6.2. SRLG | 12 |
| 6.3. Loose ERO | 13 |
| 6.3.1. Pre-computed EROs | 13 |
| 6.3.2. Re-Query | 13 |
| 7. Acknowledgements | 13 |
| 8. IANA Considerations | 13 |
| 9. Security Considerations | 13 |
| 10. References | 14 |
| 10.1. Normative References | 14 |
| 10.2. Informative References | 14 |
| Authors' Addresses | 14 |

1. Introduction

The demand for multi-area multi-AS path computation ability has evolved from the academic discussion to carrier network's feature request.

A few solutions have been proposed and there are three major variations in this field.

A global TE database is ideal and the simplest for serving multi-area or multi-as path request. This idea is apparently prohibitive for two reasons.

1. Such database may be too big and negate the purpose of having multiple areas or ASes;
2. This violates the information hiding and confidentiality requirement and is unacceptable by ISPs.

A crankback method is more practical as it is an exhaustive search based mechanism and will find an LSP if it exists. The drawback nevertheless is obvious:

1. It does not scale, for one that it often requires and wastes more than one tryout to find a qualified LSP, and for two it is RSVP signaling based which is by its nature poor in scaling;
2. The extra signaling messages add burdens on the existing network;
3. The path, if found, is not guaranteed optimal;
4. It requires lots of manual configurations to specify border routers and hence is labor intensive;
5. It requires substantial RSVP changes, in both protocol and operation.

BRPC [RFC5441] is another idea in coping with the aforementioned problems. It assumes that the destination is known in a particular domain and area. This assumption is not always true. Besides the destination may be multi-homed, meaning reachable through different areas and domains. The RFC method cannot handle this. Further more, the RFC's procedure mandates a PCEP [RFC5440] extension which understands the Virtual Shortest Path Tree (VSPT). This further complicates the method. Further more, VSPT approach only address one destination at a time.

This document describes a relayed CSPF based PCE scheme which offers

a solution with optimality, scalability, and simplicity.

1.1. Requirements Language

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 RFC 2119 [RFC2119].

1.2. Acronyms

CSPF - Constrained Shortest Path First
PCE - Path Computation Element
PCReq - Path Computation Request
PCRep - Path Computation Reply
AS - Autonomous System
LSP - Label Switched Path
RSVP - Resource ReserVation Protocol
PCEP - PCE to PCE Communication Protocol
SPT - Shortest Path Tree
VSPT - Virtual Shortest Path Tree

2. CSPF Seed

Call the node initially deposited into the heap seed. CSPF, or more generically SPF, is a seed based algorithm. The entire SPT is built upon this seed.

2.1. Multiple Seeds

It is not necessary, however, that the heap can have only one seed. In fact, most of the time during the SPF expansion, the heap contains many nodes which can be perceived as seeds for further expansion.

2.2. Heap Equivalence

An SPF heap possesses following properties:

1. A heap with one initial seed is equivalent to that with multiple intermediate seeds in any SPF stages for the destinations that have not yet been reached.
2. The deposit time of seeds is insensitive to destinations that have not yet been reached, provided that the seeds carry correct attributes values such as cost and nexthop.
3. The multiple seeds in property 1 can further be reduced to those that constitute a set of nodes besides which the destinations are not viable.

2.2.1. SPT Equivalence

The property 1 is not difficult to comprehend. During normal SPF cycles, the heap will change and the path tree will grow. At any SPF cycle, the path tree records reachability to certain destinations. This is important to generic SPF applications such as IGP routing protocols. With IGP, either OSPF [RFC2328] or IS-IS [RFC1195][ISO.10589.1992], all destinations must be included, whether they come out of the heap early or late. None of those can be neglected.

Nevertheless in CSPF, since only the targeted destinations matter, non-relevant records can be thrown away. The early path tree records are insignificant and can be disregarded. Therefore for the selected destinations, the expanded heap with multiple seeds is equivalent with the heap at the initial stage.

Figure 1 gives a simple illustration of this concept. With normal SPF computation, the Headend node "H" is used as an initial seed to the SPF heap. The final shortest path tree (SPT) will provide reachability to nodes "A", "B" and the Tailend node "T". If one uses nodes "A" and "B" as deposit seeds, he will conclude a SPT with the same reachability for "T". The difference between the two SPTs is the reachabilities from "H" to nodes "A" and "B".

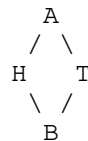


Figure 1: SPT Equivalence

2.2.2. Seed Deposit Timing

The second property indicates that the seed deposit time does not change its SPT contribution for destinations that have not yet been reached. Figure 2 shows this concept in detail.

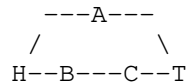


Figure 2: Seed Deposit Timing

For destination "T", using seed set "A" and "B", or "A" and "C" will produce identical results.

2.2.3. Seed Set Reduction

The third property allows us to remove the seeds that will not contribute to the path to the destinations. The statement is the key point that this proposal is based upon.

As shown in Figure 3, for the reachability to "T", the initial seed "H" can be replaced with "A and B", per property 1. The two seeds can further be replaced with "A, C and D", per property 2. Now since "D" will not contribute to the path to "T", the seed set "A, C, D" can further be reduced to "A and C", which is the property 3.

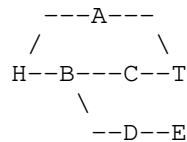


Figure 3: Seed Set Reduction

These properties are the base of multi-area multi-AS path computation method described in this document.

3. Multi-Area Path Computation

3.1. Inter-area PCE

Per IGP nature, the Tailend in a different area are not visible to the computing PCE if the PCE only knows the TE database of the Headend area. There exist options for multi-area TE database or even

a global multi-AS TE-database. But that will have scalability and confidentiality consequences. And this option is beyond of the scope of this document. Nevertheless, using the method of this document, the cross-area path can be achieved with no need of global TE database, nor much manual intervene.

3.2. Initial Seed Set

Area border routers, or BN (Border Nodes) for short, lie in the necessary paths to the destination in next area, or areas beyond. They are natural choices of the initial seed set.

3.3. PCE Relay

The path computation proceeds as a relayed CSPF job, one per each area. Take Figure 4 for example, assuming routers "A" and "C" are BNs. The column line divides the topology into two areas, area "west" on the left, and "east" on the right.

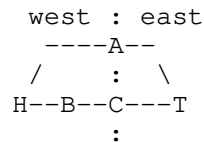


Figure 4: PCE Relay

Following steps describe the procedure:

- a. PCE in area "west", or PCE-west for short, computes paths to "A" and "C";
- b. PCE-west sends PCReq to PCE of "east" for relayed CSPF. The path information to "A" and "C" are encoded in the same PCReq message. The path information contains path attributes such as cost, bandwidth, admin-group, hop-count, etc.
- c. PCE-east uses the path information to BNs "A" and "C" and deposits them to its SPF heap. It then starts its SPF, and concludes the path computation to "T". The path to "T" if found, will be the segment in area "east".
- d. PCE-east then sends PCRep back to PCE-west with its path segment information.
- e. PCE-west maps the path segment to one of BNs. It then stitches the segment to the one in area "west" and forms a complete path.

If the destination is not in east, but further in an area eastward, PCE-east will relay the PCReq to a PCE further downstream, which will either find the target or continue to relay the job. The stitching will proceed likewise, but in a reversed order.

3.4. Relay Content

Besides standard PCReq format, an extension to PCEP protocol is needed to allow encoding of seed information. Following table describes the minimum data fields:

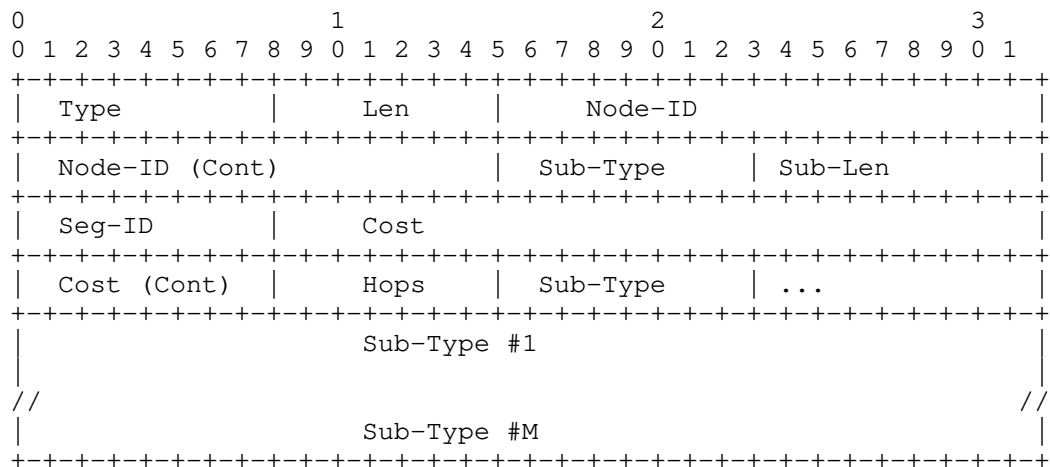


Figure 5: Relay Content

TLV:

- Type - 1 byte, value TBD (RELAYED-SEED)
- Len - 1 byte, actual length in value
- Node-ID - 4 bytes, BN's node ID

This TLV MAY have zero or more occurrences

Sub-TLV:

- Sub-Type - 1 byte, value TBD (SEGMENT)

- Sub-Len - 1 byte, value 6
- Seg-ID - Segment number, 1 byte, a BN can have multiple
 segment paths to it. This is to accommodate additive
 constraints;
- Cost - 4 bytes, starting cost
- Hops - 1 byte, starting hop count

This TLV MAY have one or more occurrences

There is no need to carry admin-group or bandwidth information. These can be learned from the standard path request information fields.

3.5. PCE Elect

An area PCE has to know all BNs which connect to another area. This can be achieved with the help from the IGP protocols and their TE extensions. The method itself however is beyond the scope of this document.

Once the list of BNs is known, the area PCE can send PCReq to PCEs in other areas. For a particular downstream area, only one PCE is necessary. Among eligible PCEs, an election mechanism may be necessary to avoid the waste and contention of computation resources. The elect PCE can be any router or a dedicated PCE. It does not have to be the transit router.

Any tie-breaker algorithm can be used in such election. One simple method is to choose the one with the highest (or lowest) router ID.

The election is a local decision of the requesting PCE. It can be proprietary and does not require standardization.

4. Multi-Home Tail-End

A destination may be viable through multiple areas. This happens typically in dual-homed or multi-homed destination cases. To maximize the path availability and optimality, the PCReq SHOULD be sent to each neighboring area and to each area's PCE elect.

As shown in Figure 6, the Headend in area A may reach out through its borders with area B and area C. The PCReq message it sends is different per recipient PCE's service area, either B or C. For PCE of area B, the message encodes the seeds of BNs between area A and B.

Likewise the PCReq for PCE of area C lists seeds of BNs between A and C.

Both PCEs may find paths to the tailend, and send PCRep back to the Headend in A. To accommodate the race condition to two possible paths, the requester needs to implement a timer to allow reasonable wait time to collect all possible PCRep, details in the next section.

The sending of multiple PCReq is not limited to the headend. It can happen to any intermediate PCEs as far as they have multiple exit borders to other areas.

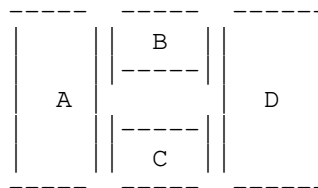


Figure 6: Multi-Homed Tailend

4.1. Relay Timer

A PCE should always send PCRep back to its requester whether it finds the path successfully or not. However the PCRep may take time and may be lost at all. For this reason it is advisable for the requester to instantiate a relay timer so that it will not wait indefinitely if the PCRep never comes. The timer also facilitates it to collect and compare all returning PCRep. Figure 7 is a sample pseudo code of the timer logic.

```

Sending PCReq:
For (PCE_Elect of each border area) {
    send_PCReq(PCE_Elect, path_Req_Info, seeds_of_BNs);
    add_to_PCReq_Pending_List(PCE_Elect);
    if (relay_timer==NULL) {
        create_relay_timer();
    }
}

Receiving PCRep:
PCE_Elect = lookup(PCRep's source address);
path = retrieve_from(PCRep, PCE_Elect);
best_path = better_path(best_path, path);
remove_from_PCReq_Pending_List(PCE_Elect);
if (Pending_List==NULL) {
    Cancel_timer(relay_timer);
    if (self==Headend) {
        terminate;
    } else {
        send_PCRep(best_path, upstream_PCE);
    }
}

Relay Timer Expires:
cleanup_Pending_List();
if (Pending_List==NULL) {
    Cancel_timer(relay_timer);
    if (self==Headend) {
        terminate;
    } else {
        send_PCRep(best_path, upstream_PCE);
    }
}

```

Figure 7: Relay Timer Pseudo Code

5. Multi-AS Path Computation

For path computation across different autonomous systems, the methods for multi-area in section 3 mostly apply as well except a few special handlings described below.

5.1. Information Hiding

When a PCE sends a PCRep to the requester which is in a different AS, it does not send explicit hop-by-hop EROs. Instead it sends a loose ERO with path level characters such as cost metric. This ensures the

information hiding from different ASes while the End-to-End path can still be established.

5.2. Transit Link

Unlike the multi-area setup where an area border router sits over both areas, two AS border routes need a transit link to connect them together.

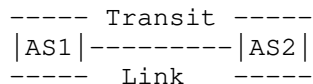


Figure 8: Transit Link

As shown in Figure 8, the transit link needs to be considered for an ASBR when it is to send PCReq to its peer ASBR. The link's characters such as metric, hop count, bandwidth, MUST be taken into account of the seed value. The PCReq relay logic remains the same.

5.3. AS Number

The BN election concept does not apply in cross AS BNs. There is no need to carry the AS number into the TE database. The PCReq and the corresponding seeds SHOULD be sent to each viable AS peer node.

6. Other Cases

6.1. Backups

Backup LSPs, or bypass LSPs, or pass re-optimization, or any path computation that requires the knowledge of an existing LSP MUST have the information of that LSP passed along with the PCReq and seeds. In case of multi-AS path computation, the upstream LSR MUST hide the path segment detail from the downstream LSR.

6.2. SRLG

The SRLG handling should be no different than that of single AS single area path computation. As long as the SRLG information is available, through for example GMPLS, each relayed PCE should compute correct PCE segment, and the end-to-end path should meet SRLG requirement.

6.3. Loose ERO

In multi-AS case, since the explicit EROs are not passed back, the AS border router has to recover the path segment that it promised upstream LSR when the LSP request reaches it. Following two approaches can be used for this purpose.

6.3.1. Pre-computed EROs

The LSR remembers and makes a record of the path segment. When RSVP request comes into the ASBR LSR, it already has resolved EROs stored locally. This is quick but requires RSVP implementation change.

6.3.2. Re-Query

The LSR does not need to remember the explicit path segment. It is kept stateless. When requested, the ASBR LSR will query its own PCE, because it does not remember the previous query and forget its response. The result nevertheless should be the same, provided the topology has no changes since.

If the answer is different or no longer available, the network has dramatically changed. The whole path computation process needs redo. And this has to be done anyway regardless of the method.

7. Acknowledgements

TBD

8. IANA Considerations

This document defines the following additional TLVs to the PCEP protocol:

| Type | Name | Source |
|------|-------------------|---------------|
| TBD | RELAYED-SEED | This document |
| TBD | SEGMENT (Sub-TLV) | This document |

9. Security Considerations

There are no specific security considerations within the scope of this document.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

10.2. Informative References

- [ISO.10589.1992]
International Organization for Standardization,
"Intermediate system to intermediate system intra-domain-
routing routine information exchange protocol for use in
conjunction with the protocol for providing the
connectionless-mode Network Service (ISO 8473)",
ISO Standard 10589, 1992.
- [RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and
dual environments", RFC 1195, December 1990.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element
(PCE) Communication Protocol (PCEP)", RFC 5440,
March 2009.
- [RFC5441] Vasseur, JP., 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, April 2009.

Authors' Addresses

Wenhu Lu
Ericsson
300 Holger Way
San Jose, California 95134
USA

Phone: 408 750-5436
Email: wenhu.lu@ericsson.com

Sriganesh Kini
Ericsson
300 Holger Way
San Jose, California 95134
USA

Phone: 408 750-5210
Email: sriganesh.kini@ericsson.com

Srikanth Narayanan
Ericsson
300 Holger Way
San Jose, California 95134
USA

Phone: 408 750-8567
Email: srikanth.narayanan@ericsson.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 28, 2011

Kexin Tang
Zhihong Wang
Yuanlin Bao
Xuerong Wang
Gang Lu
ZTE Corporation
Oct 25, 2010

Stateful PCE
draft-tang-pce-stateful-pce-01.txt

Abstract

A PCE can be either stateful or stateless. The information carried in stateful PCE are more detailed than that of stateless PCE. With the state capability of PCEs, the PCCs may make advanced and informed choices about the PCEs to use. This draft focus on stateful PCE, describes the applicability of stateful PCE and gives the IGP and PCEP extensions to support stateful PCE.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 28, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

| | |
|---|---|
| 1. Introduction | 3 |
| 1.1. Conventions used in this document | 3 |
| 2. Terminology | 3 |
| 3. Applicability of stateful PCE | 4 |
| 3.1. stateful PCE in support of GCO | 4 |
| 3.2. stateful PCE in support of resources restoration | 4 |
| 4. Requirements | 5 |
| 5. PCE Discovery and PCEP Extensions | 6 |
| 5.1. PCED Extensions | 6 |
| 5.2. PCEP Extensions | 6 |
| 6. Security Considerations | 7 |
| 7. IANA Consideration | 7 |
| 8. Normative References | 7 |
| Authors' Addresses | 8 |

1. Introduction

As defined in section 6.8 of RFC4655 [RFC4655], a PCE can be either stateful or stateless. For stateful PCE, there is a strict synchronization between the PCE and not only the network states (in term of topology and resource information), but also the set of computed paths and reserved resources in use in the network. So stateful PCE has more network information, and it can be used to do some complicated work, such as supporting GCO as well as resources restoration.

Since the information carried in stateful PCEs are more detailed than that of stateless PCEs, having knowledge of the state capability of PCEs, the PCC may make advanced and informed choices about which PCE to use. However, the existing PCE discovery ([RFC5088], [RFC5089]) and PCEP don't support stateful PCE, and the PCC have no knowledge of the state of PCE. So, this document focus on stateful PCE, describes the applicability of stateful PCE and gives the IGP and PCEP extensions to support stateful PCE.

1.1. 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].

2. Terminology

- o PCC: Path Computation Client. A client application requesting a path computation to be performed by the Path Computation Element.
- o PCE: Path Computation Element. An entity that is capable of computing a network path or route based on a network graph, and of applying computational constraints during the computation.
- o PCED: PCE Discovery.
- o PCEP: Path Computation Element communication Protocol.
- o TED: Traffic Engineering Database, which contains the topology and resource information of the domain. The TED may be fed by Interior Gateway Protocol (IGP) extensions or potentially by other means.
- o GCO: Global Concurrent Optimization. A concurrent path computation application where a set of TE paths are computed concurrently in order to optimize network resources. A GCO path

computation is able to simultaneously consider the entire topology of the network and the complete set of existing TE LSPs, and their respective constraints, and look to optimize or reoptimize the entire network to satisfy all constraints for all TE LSPs. A GCO path computation can also provide an optimal way to migrate from an existing set of TE LSPs to a reoptimized set (Morphing Problem).

3. Applicability of stateful PCE

As mentioned in the preceding part of this document, stateful PCE utilizes information from the TED as well as information about existing paths (for example, TE LSPs) in the network. Since stateful PCE has more network information, it can be used to solve the problem of resources conflict. Typical use cases of stateful PCE are listed in this section.

3.1. stateful PCE in support of GCO

As mentioned in RFC5557 [RFC5557], when computing or reoptimizing the routes of a set of Traffic Engineering Label Switched Paths (TE LSPs) through a network, it may be advantageous to perform bulk path computations in order to avoid blocking problems and to achieve more optimal network-wide solutions. Such bulk optimization is termed Global Concurrent Optimization (GCO). A GCO is able to simultaneously consider the entire topology of the network and the complete set of existing TE LSPs, and their respective constraints, and look to optimize or reoptimize the entire network to satisfy all constraints for all TE LSPs.

Since stateful PCE realized not only network states (in term of topology and resource information), but also the set of computed paths and reserved resources in use in the network, it can help a GCO realize the entire topology of the network and complete set of existing TE LSPs, so as to make a GCO to achieve the optimal network, particularly when there are several LSP needed to build, if a stateful PCE have computed a end-to-end path successfully, and hold the resources needed by this path, as a stateful PCE, therefore it could realize the newly path and reserved resources, so it can inform other PCEs involved in the GCO not to consider the same resources it just hold for the path, so stateful PCE can avoid unnecessary retries in GCO, so as to make a GCO sufficiently.

3.2. stateful PCE in support of resources restoration

Another important scenario for using the state of PCEs is that in resources restoration. A serious situation of network failure as

fiber cutting may rise to a huge number of resources restoration requests in a short time from the PCC.

In the restoration, if a stateful PCE have computed a LSP in its own domain successfully, and hold the resources needed by this LSP, as a stateful PCE, therefore it could realize the newly path and reserved resources, so it can inform other PCEs involved in the end-to-end path computation not to consider the same resources it just hold for the LSP, so stateful PCE can avoid unnecessary retries in resources restoration.

And if a stateful PCE failed to compute an end-to-end LSP, it will informed the newly path states to other stateful PCE, and release the resources it just hold, so the other stateful PCE can use the resources.

4. Requirements

As mentioned before, stateful PCE synchronized with not only the network states (in term of topology and resource information), but also the set of computed paths and reserved resources in use in the network. So the PCC need to tell stateful PCE the path state (created or deleted). However, having no knowledge of the state of PCEs, the PCC have no idea of which PCE to send the path state. In this situation, there are two possibilities for the PCC: send the path state to all the PCEs whatever the state of them, or not send to any of the PCE, and every stateful PCE query the path state information when needed. In the former case, there would be lots of unnecessary operation; and in the second case, it would increasing the complexity of the realization of the control plane and PCE. Therefore there are requirement of having knowledge of the state of PCEs for PCC. Knowing the state of PCEs, the PCC only send the path state information to stateful PCEs.

[RFC5088] defines extensions to OSPFv2 [RFC2328] and OSPFv3 [RFC2740] to allow a PCE in an OSPF routing domain to advertise some information useful to a PCC for PCE selection. It defines a new TLV (named the PCE Discovery TLV (PCED TLV)) to be carried within the OSPF Router Information LSA ([RFC4970]). The type 5 sub-TLV of PCED TLV, which named PCE-CAP-FLAGS sub-TLV, used to indicate PCE capabilities. It contains eight capabilities, but not includes the state capability of a PCE. So the PCE in an OSPF routing domain cannot advertise its state capability information to a PCC for PCE selection.

5. PCE Discovery and PCEP Extensions

This section provides protocol extensions for support of stateful PCE. Protocol extensions discussed in this section including PCED and PCEP.

5.1. PCED Extensions

To support stateful PCE, PCC SHOULD know a PCE is stateful or not. Therefore, the PCE discovery message SHOULD indicate whether the PCE advertises this message is a stateful PCE. Since PCE-CAP-FLAGS Sub-TLV ([RFC5088] for OSPF, [RFC5089] for IS-IS) contains PCE Capability Flags, this document defines a new flag, Stateful PCE Capability Flag, as follows (need to be assigned by IANA):

| Bit | Capabilities |
|-----|--------------|
| 9 | Stateful PCE |

5.2. PCEP Extensions

A PCC that wishes to inform a successful end-to-end path computation and end-to-end path connection may send an unsolicited notification to the PCE involved in the end-to-end path computation. New Notification-type and Notification-value are currently defined as follows (need to be assigned by IANA):

- o Notification-type=3: end-to-end path computation result
 - * Notification-value=1: end-to-end path computation successful.
When an end-to-end path is successfully computed, the PCC SHOULD send a notification message with Notification-type=3 and Notification-value=1 to all the PCE which involved in the end-to-end path computation.
 - * Notification-value=2: end-to-end path computation failure.
When an end-to-end path is unsuccessfully computed, the PCC SHOULD send a notification message with Notification-type=3 and Notification-value=2 to all the PCE which involved in the end-to-end path computation.
- o Notification-type=4: end-to-end connection result
 - * Notification-value=1: end-to-end path connection is success.
When an end-to-end path is successfully connected, the PCC SHOULD send a NOTIFICATION message with Notification-type=4 and Notification-value=1 to all the PCE which involved in the end-to-end path computation.

- * Notification-value=2: end-to-end path connection failure. When an end-to-end path is unsuccessfully connected, the PCC SHOULD send a notification message with Notification-type=4 and Notification-value=2 to all the PCE which involved in the end-to-end path connection.

6. Security Considerations

The extensions of this draft is baed on PCEP and OSPF, only some optional protocol elements are added which will not change the security of existing network.

7. IANA Consideration

TBD.

8. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.
- [RFC2740] Coltun, R., Ferguson, D., and J. Moy, "OSPF for IPv6", RFC 2740, December 1999.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4970] Lindem, A., Shen, N., Vasseur, JP., Aggarwal, R., and S. Shaffer, "Extensions to OSPF for Advertising Optional Router Capabilities", RFC 4970, July 2007.
- [RFC5088] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, July 2009.

Authors' Addresses

Kexin Tang
ZTE Corporation
No.68 ZiJingHua Road,Yuhuatai District
Nanjing, Jiangsu 210012
P.R.China

Phone: +86-025-52871745
Email: tang.kexin@zte.com.cn
URI: <http://www.zte.com.cn/>

Zhihong Wang
ZTE Corporation
12F, ZTE Plaza, No.19 East Huayuan Road,Haidian District
Beijing 100191
P.R.China

Phone: +86-010-59932453
Email: wang.zhihong@zte.com.cn
URI: <http://www.zte.com.cn/>

Yuanlin Bao
ZTE Corporation
5F, R&D Building 3, ZTE Industrial Park, XiLi LiuXian Road,
Nanshan District, Shenzhen 518055
P.R.China

Phone: +86-755-26773731
Email: bao.yuanlin@zte.com.cn
URI: <http://www.zte.com.cn/>

Xuerong Wang
ZTE Corporation
R&D Building 3, ZTE Industrial Zone, Liuxian Road,Nanshan District
Shenzhen 518055
P.R.China

Phone: +86-755-26773926
Email: wang.xuerong@zte.com.cn
URI: <http://www.zte.com.cn/>

Gang Lu
ZTE Corporation
2/F, ZTE Plaza, North Huashiyuan Road, East Lake Zone
Wuhan 430223
P.R.China

Phone: +86-027-51811033
Email: lu.gang2@zte.com.cn
URI: <http://www.zte.com.cn/>

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 18, 2011

F. Zhang
Q. Zhao
Huawei Technologies
O. Gonzalez de Dios
Telefonica I+D
R. Casellas
CTTC
D. King
Old Dog Consulting
October 18, 2010

Extensions to
Path Computation Element Communication Protocol (PCEP)
for Hierarchical Path Computation Elements (PCE)
draft-zhang-pcep-hierarchy-extensions-00

Abstract

The hierarchical Path Computation Element (PCE) architecture defined in [PCE-HIERARCHY-FWK] allows 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 hierarchical PCE procedures which are described in [PCE-HIERARCHY-FWK].

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 18, 2011.

Copyright Notice

Zhang, et al.

[Page 1]

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

| | |
|--|----|
| 1. Introduction..... | 3 |
| 2. PCEP Extension Requirements..... | 4 |
| 2.1. New Objective Functions..... | 4 |
| 2.2. PCEP Request Qualifiers..... | 4 |
| 2.3. Discovery Between Parent and Child PCEs..... | 4 |
| 2.3.1. Parent PCE Capability Discovery..... | 5 |
| 2.3.2. PCE Domain and PCE ID Discovery..... | 5 |
| 2.4. Domain Connectivity Information Collection..... | 5 |
| 2.5. Error Case Handling..... | 6 |
| 3. PCEP Extensions..... | 6 |
| 3.1. Extensions to OPEN Object..... | 7 |
| 3.1.1. OF Codes..... | 7 |
| 3.1.2. OPEN Object Flags..... | 7 |
| 3.1.3. Domain-ID TLV..... | 7 |
| 3.1.4. PCE-ID TLV..... | 8 |
| 3.1.5. Procedures..... | 8 |
| 3.2. Extensions to RP Object..... | 9 |
| 3.2.1. RP Object Flags..... | 9 |
| 3.2.2. Domain-ID TLV..... | 9 |
| 3.2.3. Procedures..... | 9 |
| 3.3. Extensions to NOTIFICATION Object..... | 9 |
| 3.3.1. Notification Types..... | 10 |
| 3.3.2. Inter-domain Link TLV..... | 10 |
| 3.3.3. Inter-domain Node TLV..... | 11 |
| 3.3.4. Domain-ID TLV..... | 11 |
| 3.3.5. PCE-ID TLV..... | 11 |
| 3.3.6. Procedures..... | 12 |
| 3.4. Extensions to PCEP-ERROR Object..... | 12 |
| 3.4.1. Hierarchy PCE Error-Type..... | 12 |
| 3.4.2. Procedures..... | 12 |
| 4. Manageability Considerations..... | 13 |
| 5. IANA Considerations..... | 13 |
| 6. Security Considerations..... | 13 |
| 7. References..... | 13 |
| 7.1. Normative References..... | 13 |
| 7.2. Informative References..... | 13 |

[PCE-HIERARCHY-FWK] describes a hierarchy PCE architecture which can be used for computing the end-to-end paths of inter-domain MPLS Traffic Engineering (TE) and GMPLS Label Switched Paths (LSPs). In the hierarchy PCE architecture, the parent PCE can compute a domain path based on the domain connectivity information and the child PCE can compute the intra-domain path based on the domain topology information. The end-to-end domain path computing procedures can be abstracted as follows:

- o The PCC requests a child PCE to return an inter-domain path.
- o The child PCE forwards the request to the parent PCE.
- o The parent PCE computes one or multiple 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(s).
- 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 and returns the inter-domain path to the child PCE.
- o The child PCE forwards the inter-domain path to the PCC.

This document defines the PCEP extensions for the purpose of implementing hierarchy PCE procedures which are described in [PCE-HIERARCHY-FWK].

The document also uses a number of editor notes to describe options and alternative solutions. These options and notes will be removed before publication.

1.1. Terminology

This document uses the terminology defined in [RFC4655] and [RFC5440] and [PCE-HIERARCHY-FWK].

1.2. Requirements Language

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 RFC 2119 [RFC2119].

2. PCEP Extension Requirements

2.1. New Objective Functions

For inter-domain path computation, there are three new objective functions which are defined in section 1.3.1 of [PCE-HIERARCHY-FWK].

- o Minimize the number of boundary nodes used.
- o Limit the number of domains crossed.
- o Disallow domain re-entry.

During the PCEP session establishment procedure, the PCE needs to be capable of indicating the objective functions (OF) capability in the Open message. This information can be, in turn, announced by child PCEs and used for selecting the PCE when a PCC want a path that satisfies a certain inter-domain objective function.

When a PCC requests a PCE to compute an inter-domain path, the PCC needs also to be capable of indicating the new objective functions for inter-domain path.

For the reasons described above, 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.

2.2. PCEP Request Qualifiers

As described in section 5.8.1 of [PCE-HIERARCHY-FWK], support of the H-PCE architecture will introduce two new qualifications as follows:

- o It must be possible for a child PCE to indicate that the request it sends 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 per-domain or backward recursive path computation.
- o A parent PCE needs to be able to ask a child PCE whether a particular node address (the destination of an end-to-end path) is present in the domain that the child PCE serves.

To meet the above requirements, the PCEP PCReq message should be extended.

2.3. Discovery Between Parent and Child PCEs

In the H-PCE architecture, the parent PCE does not need to be aware

of each child domain topology. Therefore, it is possible that the parent PCE does not join the IGP instance of the child PCE domain, i.e. there is no IGP discovery mechanism between the parent PCE and child PCE.

Therefore there must be a discovery mechanism for basic PCE information between the parent and child PCEs. In this case, PCEP needs to provide discovery mechanisms that do not rely on IGP announcement/discovery procedures.

Editors note. A child PCE could forward the topology within PCNtf messages or any other mechanisms, without an IGP adjacency. Further discussion of the discovery mechanism and scope will be discussed in later versions of this document.

2.3.1. Parent PCE Capability Discovery

As described in [PCE-HIERARCHY-FWK], 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. The parent PCE needs also to be capable of indicating whether its parent capability can be provided to the child PCE or not.

2.3.2. PCE Domain and PCE ID Discovery

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

The PCE ID is an IPv4 and/or IPv6 address that is used to reach the parent/child PCE. It is RECOMMENDED to use an address that is always reachable if there is any connectivity to the PCE.

The PCE ID information and PCE domain identifiers may be provided during the PCEP session establishment procedure or the domain connectivity information collection procedure.

2.4. Domain Connectivity Information Collection

As described in [PCE-HIERARCHY-FWK], the parent PCE builds the domain topology map either from configuration or from information received from each child PCE. A child PCE may report its neighbor domain connectivity to its parent PCE. It is reasonable to use PCEP PCNtf message to do this procedure. If an IGP adjacency is established between parent and children, it could be used for this purpose.

There are two types of domain border for providing the domain connectivity information:

- o Domain border is a TE link, e.g. the inter-AS TE link which connects two ASs.
- o Domain border is a node, e.g. the IGP ABR which connects two IGP areas.

For the inter-AS TE links, the following information needs to be notified to the parent PCE:

- o Identifier of advertising child PCE.
- o Identifier of PCE's domain.
- o Identifier of the link.
- o TE properties of the link (metrics, bandwidth)
- o Other properties of the link (technology-specific).
- o Identifier of link end-points.
- o Identifier of adjacent domain.

For the ABR, the following information needs to be notified to the parent PCE:

- o Identifier of the ABR.
- o Identifier of the IGP Area IDs.

2.5. Error Case 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 (such as querying other child PCEs), and could result in a negative response in a PCEP Error (PCErr) message and indicate the hierarchy PCE error types.

3. PCEP Extensions

3.1. Extensions to OPEN Object

3.1.1. OF Codes

There are three new OF codes defined here for H-PCE:

- o Name: Minimize the number of Boundary Nodes used (MBN)
Objective Function Code: (to be assigned by IANA, recommended 9)

Description: Find a path P such that passes through the least boundary nodes.

- o Name: Minimize the number of Transit Domains (MTD) 10)
Objective Function Code: (to be assigned by IANA, recommended
Description: Find a path P such that passes through the least transit domains.
- o Name: Disallow Domain Re-entry (DDR)
Objective Function Code: (to be assigned by IANA, recommended 11)
Description: Find a path P such that does not entry a domain more than once.

3.1.2. OPEN Object Flags

There are two OPEN object flags defined here for H-PCE:

- o Parent PCE request bit (to be assigned by IANA, recommended bit 0):
if set it means the child PCE wishes to use the peer PCE as a parent PCE.
- o Parent PCE indication bit (to be assigned by IANA, recommended bit 1):
if set it means the PCE can be used as a parent PCE by the peer PCE.

Editors Note. It is possible that a parent PCE will also act as a child PCE.

3.1.3. Domain-ID TLV

The type of Domain-ID TLV is to be assigned by IANA (recommended 7). The length is 8 octets. The format of this TLV is defined below:

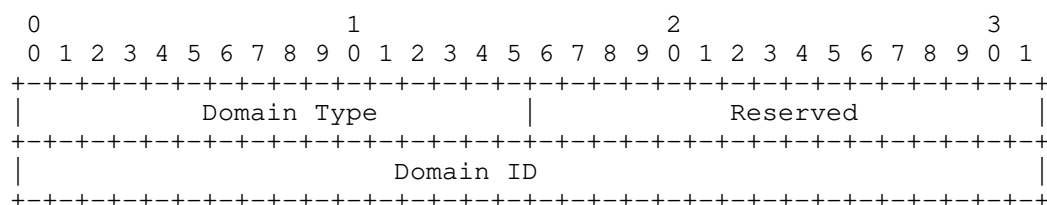


Figure 1: Domain-ID TLV

Domain Type (8 bits): Indicates the domain type. There are two types of domain defined currently:

- o Type=1: the Domain ID field carries an IGP Area ID.
- o Type=2: the Domain ID field carries an AS number.

Domain ID (32 bits): Indicates an IGP Area ID or AS number.

Editors note. It maybe necessary to support 64 bit domain IDs.

3.1.4. PCE-ID TLV

The type of PCE-ID TLV is to be assigned by IANA (recommended 8). The length is 4. The format of this TLV is defined below:

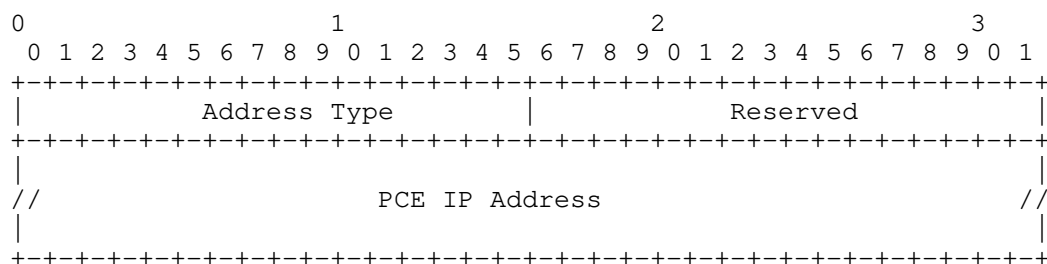


Figure 2: PCE ID TLV

Address Type (16 bits): Indicates the address type of PCE IP Address. 1 means IPv4 address type, 2 means IPv6 address type.

PCE IP Address: Indicates the reachable address of a PCE.

3.1.5. Procedures

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.

If a child PCE wants to use the peer PCE as a parent, it can set the parent PCE request bit in the OPEN object carried in the Open message during the PCEP session creation procedure. If the peer PCE does not want to provide the parent function to the child PCE, it must send a PCERR message to the child PCE and clear the parent PCE indication bit in the OPEN object.

If the parent PCE can provide the parent function to the peer PCE, it may set the parent PCE indication bit in the OPEN object carried in the Open message during the PCEP session creation procedure.

The PCE may also report its PCE ID and list of domain ID to the peer PCE by specifying them in the PCE-ID TLV and List of Domain-ID TLVs in the OPEN object carried in the Open message during the PCEP session creation procedure.

3.2. Extensions to RP Object

3.2.1. RP Object Flags

- o Domain Path Request bit (to be assigned by IANA, recommended bit 17): if set it means the child PCE wishes to get the domain sequence.
- o Destination Domain Query bit (to be assigned by IANA, recommended bit 16): if set it means the parent PCE wishes to get the destination domain ID.

3.2.2. Domain-ID TLV

The format of this TLV is defined in section 2.1.3. This TLV can be carried in an OPEN object to indicate a (list of) managed domains, or carried in a RP object to indicate the destination domain ID when a child PCE responds to the parent PCE's destination domain query by a PCRep message.

Editors note. In some cases, the Parent PCE may need to allocate a node which is not necessarily the destination node.

3.2.3. Procedures

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 RP object carried in a PCReq message. The parent PCE which receives the PCReq message tries to compute a domain sequence for it. 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. The domain sequence is specified as AS or AREA ERO sub-objects (type 32 for AS [RFC3209] or type. Otherwise it sends a PCReq message which carries the NO-PATH object to the child PCE.

The parent PCE can set the Destination Domain Query bit in a PCReq message to query the destination (which is specified in the END-POINTS objects) domain ID from a child PCE. If the child PCE knows the destination(s) domain ID, it sends a PCRep message to the parent PCE and specifies the domain ID in the Domain-ID TLV which is carried in the RP object. Otherwise it sends a PCRep message with a NO-PATH object to the parent PCE.

3.3. Extensions to NOTIFICATION Object

Because there will not be too many PCEP sessions between the child PCE(s) and parent PCE, it is recommended that the PCEP sessions between them keeping alive all the time. Then the child PCE can report all of the domain connectivity information to the parent PCE

when the PCEP session is established successfully. It can also notify the parent PCE to update or delete the domain connectivity information when it detects the changes.

3.3.1. Notification Types

There is a new notification type defined in this document :

- o Domain Connectivity Information notification-type (to be assigned by IANA, recommended 3).

Notification-value=0: sent from the parent to the child to query all of the domain connectivity information maintained by the child PCE.

Notification-value=1: sent from the child to the parent to update the domain connectivity information maintained by the child PCE.

Notification-value=2: sent from the child to the parent to delete the domain connectivity information maintained by the child PCE.

3.3.2. Inter-domain Link TLV

IGP in each neighbor domain can advertise its inter-domain TE link capabilities [RFC5316], [RFC5392]. This information can be collected by the child PCEs and forwarded to the parent PCE. PCEP Inter-domain Link TLV is used for carrying the inter-domain TE link attributes for this purpose. Each Inter-domain Link TLV can carry the attributes of one inter-domain link at the most.

The type of Inter-domain Link TLV is to be assigned by IANA (recommended 9). The length is variable. The format of this TLV is defined below:

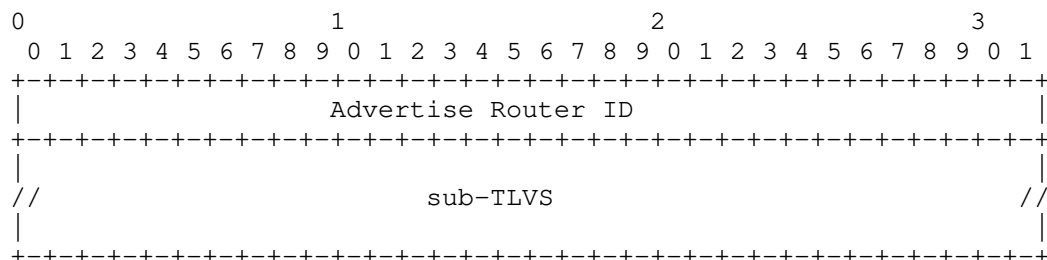


Figure 3: Inter-domain Link TLV

Advertise Router ID (32 bits): indicates the router ID which advertises the TE LSA or LSP.

Each inter-domain link is identified by the combination of advertise router ID and the link local IP address or link local unnumbered identifier. The PCNtf message which is used for notifying the parent PCE to update or delete a inter-domain link must contain the information identifies a TE link exclusively.

carried in a NOTIFICATION object to indicate the PCE ID of the PCE who sends the PCNtf message.

3.3.6. Procedures

When a parent PCE establishes a PCEP session with a child PCE successfully, the parent PCE may request the child PCE to report the domain connectivity information. This procedure can be done by sending a PCNtf message from the parent to the child, setting the notification-type to 3 and notification-value to 0 in the NOTIFICATION object.

When a child PCE receives the PCNtf message, it may send all of the domain connectivity information to the parent PCE by the PCNtf message(s). The notification-type is 3 and notification-value is 1 in the NOTIFICATION object. The NOTIFICATION object may carry the inter-domain link TLV and inter-domain node TLV to describe the inter-domain connectivity information. It is noted that if the child PCE does not support this function, it will ignore the received PCNtf message and the parent PCE will not receive the response.

The child PCE can also update the domain connectivity information by re-sending the PCNtf message(s) with the newly information.

When the child PCE detects a deletion of domain connectivity (e.g., the inter-domain link TLV is aged out), it must notify the parent PCE to delete the inter-domain link by sending the PCNtf message. The notification-type is 3 and notification-value is 2 in the NOTIFICATION object.

3.4. Extensions to PCEP-ERROR Object

3.4.1. Hierarchy PCE Error-Type

A new PCEP Error-Type is allocated for hierarchy PCE (to be assigned by IANA, recommended 11):

| Error-Type | Meaning |
|------------|---|
| 11 | Hierarchy PCE error Error-value=1: parent PCE capability can not be provided |

3.4.2. Procedures

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 (11) and error-value (1) in the PCEP-ERROR object.

4. Manageability Considerations

TBD.

5. IANA Considerations

TBD.

6. Security Considerations

TBD.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5440] Vasseur, JP., Ed., and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

7.2. Informative References

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC4655] Farrel, A., Vasseur, J.-P., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5316] M. Chen, R. Zhang, X. Duan, " ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering ", RFC 5316, December 2008.
- [RFC5392] M. Chen, R. Zhang, X. Duan, " OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering ", RFC 5316, January 2009.
- [PCE-HIERARCHY-FWK] D. King, A. Farrel, " The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS ", draft-king-pce-hierarchy-fwk-05, September 2010.

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Quintin Zhao
Huawei Technology
125 Nagog Technology Park
Acton, MA 01719
US
Email: qzhao@huawei.com

Oscar Gonzalez de Dios
Telefonica I+D
Emilio Vargas 6, Madrid
Spain
Email: ogondio@tid.es

Ramon Casellas
CTTC - Centre Tecnologic de Telecomunicacions de Catalunya
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain
Email: ramon.casellas@cttc.es

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

Internet-Draft

October 2010

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2011

Q. Zhao
Huawei Technology
Z. Ali
T. Saad
S. Sivabalan
Cisco Systems
D. King
Old Dog Consulting
R. Casellas
CTTC - Centre Tecnologic de
Telecomunicacions de Catalunya
October 25, 2010

PCE-based Computation Procedure To Compute Shortest Constrained P2MP
Inter-domain Traffic Engineering Label Switched Paths
draft-zhao-pce-pcep-inter-domain-p2mp-procedures-06

Abstract

The ability to compute paths for constrained point-to-multipoint (P2MP) Traffic Engineering Label Switched Paths (TE LSPs) across multiple domains has been identified as a key requirement for the deployment of P2MP services in MPLS and GMPLS networks. The Path Computation Element (PCE) has been recognized as an appropriate technology for the determination of inter-domain paths of P2MP TE LSPs.

This document describes the procedures and extensions to the PCE communication Protocol (PCEP) to handle requests and responses for the computation of inter-domain paths for P2MP TE LSPs.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 25, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

| | |
|---|----|
| 1. Introduction | 3 |
| 2. Terminology | 3 |
| 3. Problem Statement | 5 |
| 4. Assumptions | 7 |
| 5. Requirements | 8 |
| 6. Objective Functions | 9 |
| 7. P2MP Path Computation Procedures | 9 |
| 7.1. Core Trees | 9 |
| 7.2. Core Tree Computation Procedures | 11 |
| 7.3. Sub Tree Computation Procedures | 12 |
| 7.4. PCEP Protocol Extensions | 12 |
| 7.4.1. The Extension of RP Object | 13 |
| 7.4.2. The PCE Sequence Object | 13 |
| 7.5. Relationship with Hierarchical PCE | 15 |
| 7.6. Parallelism | 15 |
| 8. Manageability Considerations | 15 |
| 9. Control of Function and Policy | 16 |
| 10. Information and Data Models | 16 |
| 11. Liveness Detection and Monitoring | 16 |
| 12. Verifying Correct Operation | 16 |
| 13. Requirements on Other Protocols and Functional Components | 16 |
| 14. Impact on Network Operation | 16 |
| 15. Security Considerations | 17 |
| 16. IANA Considerations | 17 |
| 17. Acknowledgements | 17 |
| 18. References | 17 |
| 18.1. Normative References | 17 |
| 18.2. Informative References | 18 |
| Authors' Addresses | 18 |

1. Introduction

Multicast services are increasingly in demand for high-capacity applications such as multicast Virtual Private Networks (VPNs), IP-television (IPTV) which may be on-demand or streamed, and content-rich media distribution (for example, software distribution, financial streaming, or data-sharing). The ability to compute constrained Traffic Engineering Label Switched Paths (TE LSPs) for point-to-multipoint (P2MP) LSPs in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks across multiple domains. A domain can be defined as a collection of network elements within a common sphere of address management or path computational responsibility such as an IGP area or an Autonomous Systems.

The applicability of the Path Computation Element (PCE) [RFC4655] for the computation of such paths is discussed in [RFC5671], and the requirements placed on the PCE communications Protocol (PCEP) for this are given in [RFC5862].

This document describes how multiple PCE techniques can be combined to address the requirements. These mechanisms include the use of the per-domain path computation technique specified in [RFC5152], extensions to the backward recursive path computation (BRPC) technique specified in [RFC5441] for P2MP LSP path computation in an inter-domain environment, and a new procedure for core-tree based path computation defined in this document. These three mechanisms are suitable for different environments (topologies, administrative domains, policies, service requirements, etc.) and can also be effectively combined.

1.1. Requirements Language

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 RFC 2119 [RFC2119].

2. Terminology

Terminology used in this document is consistent with the related MPLS/GMPLS and PCE documents [RFC4461], [RFC4655], [RFC4875], [RFC5376], [RFC5440], [RFC5441]. [RFC5671], and [RFC5862].

ABR: Area Border Router. Router used to connect two IGP domains (areas in OSPF or levels in IS-IS).

ASBR: Autonomous System Border Router. Router used to connect together ASes of the same or different Service Providers via one or more Inter-AS links.

Boundary Node (BN): a boundary node is either an ABR in the context of inter-area Traffic Engineering or an ASBR in the context of inter-AS Traffic Engineering.

Core Tree: the core tree is a P2MP tree where the root is the ingress LSR, the transit nodes and branch nodes are the BNs of the transit domains and the leaf nodes are the leaf BNs of the leaf domains.

Destination: The lead Nodes can be in Root Domain, Transit Domain and Leaf Domain.

Entry BN of domain(n): a BN connecting domain(n-1) to domain(n) along a determined sequence of domains.

Exit BN of domain(n): a BN connecting domain(n) to domain(n+1) along a determined sequence of domains.

Inter-AS TE LSP: a TE LSP that crosses an AS boundary.

Inter-area TE LSP: a TE LSP that crosses an IGP area boundary.

Leaf Domain: a domain that does not have a downstream neighbor domain. Note that, with this definition, a domain with one or more leaf nodes is not necessarily a leaf domain.

Leaf Boundary Nodes: the entry boundary node in the leaf domain.

Leaf Nodes: the LSR which is the P2MP LSP's final.

LSR: Label Switching Router.

LSP: Label Switched Path.

OF: Objective Function. A set of one or more optimization criterion (criteria) used for the computation of paths either for single or for synchronized requests (e.g. path cost minimization), or the synchronized computation of a set of paths (e.g. aggregate bandwidth consumption minimization, etc.). See [RFC4655] and [RFC5441].

P2MP LSP Path Tree: A set of LSRs and TE links that comprise the path of a P2MP TE LSP from its ingress LSR to all of its egress LSRs.

Path Domain Sequence: The known sequence of domains for a path between root and leaf.

Path Domain Tree: The tree formed by the domains that the P2MP path crosses, where the source (ingress) domain is the root domain.

PCC: Path Computation Client. Any client application requesting a path computation to be performed by the Path Computation Element.

PCE (Path Computation Element): an entity (component, application or

network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

P2MP LSP Path Tree: A set of LSRs and TE links that comprise the path of a P2MP TE LSP from its ingress LSR to all of its egress LSRs.

Path Domain Sequence: the known sequence of domains for a path between the root node and a leaf node.

PCE Sequence: the known sequence of PCEs for calculating a path between the root node and a leaf node.

PCE Topology Tree: a list of PCE Sequences which has all the PCE Sequence for each path of the P2MP LSP path tree.

PCE(i): a PCE that performs path computations for domain(i).

Root Boundary Node: the egress LSR from the root domain on the path of the P2MP LSP.

Root Domain: the domain that includes the ingress (root) LSR.

TED: Traffic Engineering Database.

Transit/branch Domain: a domain that has an upstream and one or more downstream neighbour domain.

VSPT: Virtual Shortest Path Tree [RFC5441].

3. Problem Statement

The Path Computation Element (PCE) defined in [RFC4655] is an entity that is capable of computing a network path or route based on a network graph, and applying computational constraints. A Path Computation Client (PCC) may make requests to a PCE for paths to be computed.

[RFC4875] describes how to set up P2MP TE LSPs for use in MPLS and GMPLS networks. The PCE is identified as a suitable application for the computation of paths for P2MP TE LSPs [RFC5671].

[RFC5441] specifies a procedure relying on the use of multiple PCEs to compute (P2P) inter-domain constrained shortest paths across a predetermined sequence of domains, using a backward recursive path computation technique. The technique can be combined with the use of path keys [RFC5520] to preserve confidentiality across domains, which is sometimes required when domains are managed by different Service

Providers.

The PCE communication Protocol (PCEP) [RFC5440] is extended for point-to-multipoint (P2MP) path computation requests and in [RFC6006]. However, that specification does not provide all the necessary mechanisms to request the computation of inter-domain P2MP TE LSPs.

As discussed in [RFC4461], a P2MP tree is a graphical representation of all TE links that are committed for a particular P2MP LSP. In other words, a P2MP tree is a representation of the corresponding P2MP tunnel on the TE network topology. A sub-tree is a part of the P2MP tree describing how the root or an intermediate P2MP LSPs minimizes packet duplication when P2P TE sub-LSPs traverse common links. As described in [RFC5671] the computation of a P2MP tree requires three major pieces of information. The first is the path from the ingress LSR of a P2MP LSP to each of the egress LSRs, the second is the traffic engineering related parameters, and the third is the branch capability information.

Generally, an inter-domain P2MP tree (i.e., a P2MP tree with source and at least one destination residing in different domains) is particularly difficult to compute even for a distributed PCE architecture. For instance, while the BRPC recursive path computation may be well-suited for P2P paths, P2MP path computation involves multiple branching path segments from the source to the multiple destinations. As such, inter-domain P2MP path computation may result in a plurality of per-domain path options that may be difficult to coordinate efficiently and effectively between domains. That is, when one or more domains have multiple ingress and/or egress border nodes, there is currently no known technique for one domain to determine which border routers another domain will utilize for the inter-domain P2MP tree, and no way to limit the computation of the P2MP tree to those utilized border nodes.

A trivial solution to the computation of inter-domain P2MP tree would be to compute shortest inter-domain P2P paths from source to each destination and then combine them to generate an inter-domain, shortest-path-to-destination P2MP tree. This solution, however, cannot be used to trade cost to destination for overall tree cost (i.e., it cannot produce a MCT tree) and in the context of inter-domain P2MP LSPs it cannot be used to reduce the number of domain border nodes that are transited.

Computing P2P LSPs individually is not an acceptable solution for computing a P2MP tree. Even per domain path computation [RFC5152] can be used to compute P2P multi-domain paths, but it does not guarantee to find the optimal path which crosses multiple domains. Furthermore, constructing a P2MP tree from individual source to leaf

P2P LSPs does not guarantee to produce a least-cost tree. This approach may also be considered to have scaling issues during LSP setup. That is, the LSP to each leaf is signaled separately, and each border node must perform path computation for each leaf.

P2MP Minimum Cost Tree (MCT), i.e. one which guarantees the least cost resulting tree, is an NP-complete problem. Moreover, adding and/or removing a single destination to/from the tree may result in an entirely different tree. In this case, frequent MCT path computation requests may prove computationally intensive, and the resulting frequent tunnel reconfiguration may even cause network destabilization. There are several heuristic algorithms presented in the literature that approximate the result within polynomial time that are applicable within the context of a single-domain.

This document presents a solution, and procedures and extensions to PCEP to support P2MP inter-domain path computation.

4. Assumptions

It is assumed that, due to deployment and commercial limitations (e.g., inter-AS peering agreements), the sequence of domains for a path (the path domain tree) will be known in advance.

In the figure below, the P2MP tree spans 6 domains, with D1 being the root domain. The corresponding domain sequences which are assumed known would be: D1-D3-D6, D1-D3-D5 and D1-D2-D4.

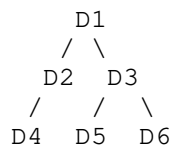


Figure 1: Domain Sequence Tree

The examples and scenarios used in this document are also based on the following assumptions:

- o The PCE that serves each domain in the path domain tree is known, and the set of PCEs and their relationships is propagated to each PCE during the first exchange of path computation requests; [Editors note - this assumption needs to be more explicit.
- o Each PCE knows about any leaf LSRs in the domain it serves;

- o The boundary nodes to use on the LSP are pre-determined and are part of the path domain tree. [Editors Note - In this version of the document we do not consider multi-homed domains.]

Additional assumptions are documented in [RFC5441] and will not be repeated here.

5. Requirements

This section summarizes the requirements specific to computing inter-domain P2MP paths. In these requirements we note that the actual computation times by any PCE implementation are outside the scope of this document, but we observe that reducing the complexity of the required computations has a beneficial effect on the computation time regardless of implementation. Additionally, reducing the number of message exchanges and the amount of information exchanged will reduce the overall computation time for the entire P2MP tree. We refer to the "Complexity of the computation" as the impact on these aspects of path computation time as various parameters of the topology and the P2MP LSP are changed.

Its also important that the solution preserves confidentiality across domains, which is required when domains are managed by different Service Providers.

Other than the requirements specified in [RFC5376], a number of requirements specific to P2MP are detailed below:

1. The computed P2MP LSP should be optimal when only considering the paths among the BNs.
2. Grafting and pruning of multicast destinations in a domain should have no impact on other domains and on the paths among BNs.
3. The complexity of the computation for each sub-tree within each domain should be dependent only on the topology of the domain and it should be independent of the domain sequence.
4. The number of PCEP request and reply messages should be independent of the number of multicast destinations in each domain.
5. Specifying the domain entry and exit nodes.
6. Specifying which nodes should be used as branch nodes.

7. Reoptimization of existing sub-trees.
8. Computation of P2MP paths that need to be diverse from existing P2MP paths.

6. Objective Functions

For the computation of a single or a set of P2MP TE LSPs, a request to meet specific optimization criteria, called an Objective Function (OF) may be indicated.

The computation of one or more P2MP TE-LSPs may be subject to an OF in order to select the "best" candidate paths. A variety of objective functions have been identified as being important during the computation of inter-domain P2MP LSPs. These include:

1. The sub-tree within each domain should be optimized, which can be either the Minimum cost tree [RFC5862] or Shortest path tree [RFC5862].
2. The P2MP LSP path, formed by considering only the entry and exit nodes of the domains (the Core Tree) should be optimal.
3. It should be possible to limit the number of entry points to a domain.
4. It should be possible to force the branches for all leaves within a domain to be in that domain.

7. P2MP Path Computation Procedures

The following sections describe the Core Tree based procedures to satisfy the requirements specified in the previous section. A core tree based solution provides an optimal inter-domain P2MP TE LSP.

7.1. Core Trees

A Core Tree is defined as a node tree, with nodes from the domains corresponding to the domain tree PCE topology, which satisfies the following conditions:

- o The root of the core tree is the ingress LSR in the root domain;
- o The leaves of the core tree are the entry nodes in the leaf domains;

- o The transit and branch nodes of the core tree are from the entry and exit nodes from the transit and branch domains.

For example, consider the Domain Tree from the figure below, representing a domain tree of 5 domains, and part of the resulting Core Tree which satisfies the aforementioned conditions.

RN: Root Node
 EN: Entry Border Node (domain, index)
 XN: Exit Border Node (domain, index)

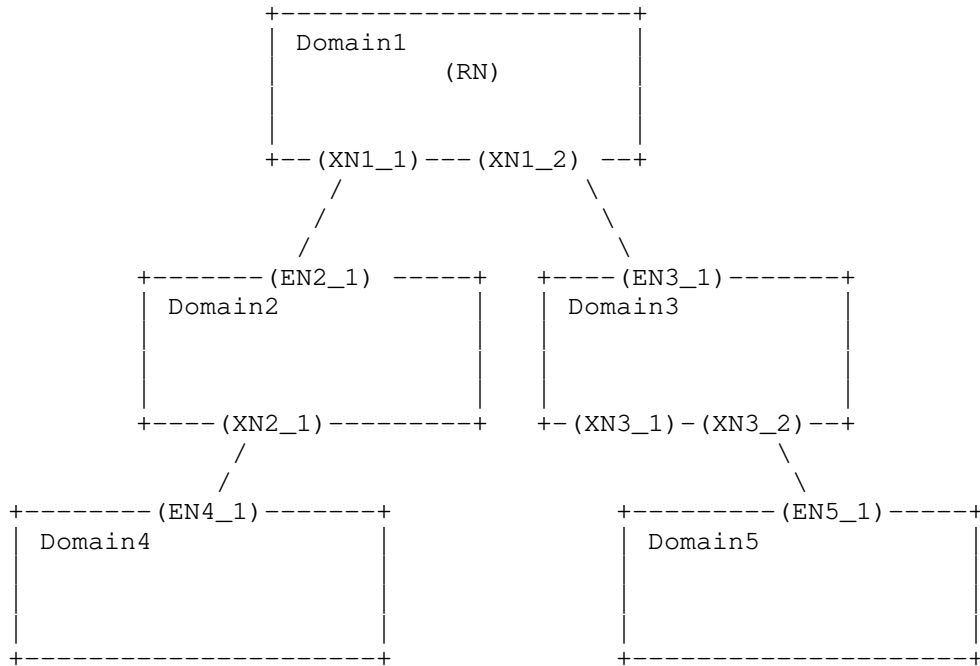


Figure 2: Domain Tree Example

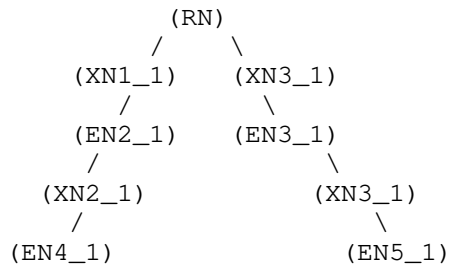


Figure 3: Core Tree

7.2. Core Tree Computation Procedures

Computing the complete P2MP LSP path tree is done in two phases:

The algorithms to compute the optimal large core tree are outside scope of this document. The following extended BRPC based procedure can be used to compute the core tree.

BRPC Based Core Tree Path Computation Procedure:

1. Using the BRPC procedures to compute the VSPT(i) for each leaf BN(i), $i=1$ to n , where n is the total number of entry nodes for all the leaf domains. In each VSPT(i), there are a number of $P(i)$ paths.
2. When the root PCE has computed all the VSPT(i), $i=1$ to n , take one path from each VSPT and form a set of paths, we call it a PathSet(j), $j=1$ to M , where $M=P(1) \times P(2) \dots \times P(n)$;
3. For each PathSet(j), there are n S2L (Source to Leaf BN) paths and form these n paths into a Core Tree(j);
4. There will be M number of Core Trees computed from step3. Apply the OF to each of these M Core Trees and find the optimal Core Tree.

Note that the application of BRPC in the aforementioned procedure differs from the typical one since paths returned from a downstream PCE are not necessary pruned from the solution set by intermediate PCEs.

The reason for this is that if the PCE in a downstream domain does the pruning and returns the single optimal sub-path to its parent PCE, BRPC insures that the ingress PCE will get all the best optimal sub-paths for each LN (Leaf Border Nodes), but the combination of these single optimal sub-paths into a P2MP tree is not necessarily optimal even each S2L (Source-to-Leaf) sub-path is optimal.

Without trimming, the ingress PCE will get all the possible S2L sub-paths set for LN, and eventually by looking through all the combinations, and taking one sub-path from each set to built one p2mp tree it finds the optimal tree.

The proposed method may present a scalability problem for the dynamic computation of the Core Tree (by iterative checking of all combinations of the solution space), specially with dense/meshed

domains. Considering a domain sequence D1, D2, D3, D4, where the Leaf border node is at domain D4, PCE(4) will return 1 path. PCE(3) will return N paths, where N is $E(3) \times X(3)$, where $E(k) \times X(k)$ denotes the number of entry nodes times the number of exit nodes for that domain. PCE(2) will return M paths, where $M = E(2) \times X(2) \times N = E(2) \times X(2) \times E(3) \times X(3) \times 1$, etc. Generally speaking the number of potential paths at the ingress PCE $Q = \prod E(k) \times X(k)$.

Consequently, it is expected that the Core Path will be typically computed offline, without precluding the use of dynamic, online mechanisms such as the one presented here, in which case it SHOULD be possible to configure transit PCEs to control the number of paths sent upstream during BRPC (trading trimming for optimality at the point of trimming and downwards).

7.3. Sub Tree Computation Procedures

Once the core tree is built, the grafting of all the leaf nodes from each domain to the core tree can be achieved by a number of algorithms. One algorithm for doing this phase is that the root PCE will send the request with C bit set for the path computation to the destination(s) directly to the PCE where the destination(s) belong(s) along with the core tree computed from the phase 1.

This approach requires that the root PCE manage a potentially large number of adjacencies (either in persistent or non-persistent mode), including PCEP adjacencies to PCEs that are not within neighboring domains.

A first alternative would involve establishing PCEP adjacencies that correspond to the PCE domain tree. This would require that branch PCEs forward requests and responses from the root PCE towards the leaf PCEs and vice-versa.

Finally, another alternative would use a hierarchical PCE (H-PCE) architecture. The "hierarchically" parent would request sub tree path computations.

The algorithms to compute the optimal large sub tree are outside scope of this document. In the case that the number of destinations and the number of BNs within a domain are not big, the incremental procedure based on p2p path computation using the OSPF can be used.

7.4. PCEP Protocol Extensions

7.4.1. The Extension of RP Object

The extended format of the RP object body to include the C bit is as follows:

The C bit is added in the flag bits field of the RP object to signal the receiver of the message that the request/reply is for inter-domain P2MP Core Tree or not.

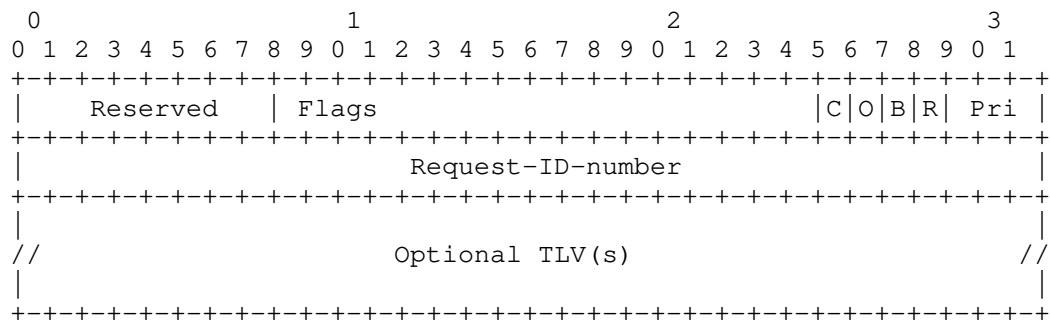


Figure 4: RP Object Body Format

The following flag is added in this draft:

C bit (P2MP Core Tree bit - 1 bit):

0: This indicates that this is normal PCReq/PCRep for P2MP.

1: This indicates that this is PCReq or PCRep message for inter-domain Core Tree P2MP. When the C bit is set, then the request message should have the Core Tree passed along with the destinations which and then graphed to the tree.

7.4.2. The PCE Sequence Object

The PCE Sequence Object is added to the existing PCE protocol. A list of this objects will represent the PCE topology tree. A list of Sequence Objects can be exchanged between PCEs during the PCE capability exchange or on the first path computation request message between PCEs. In this case, the request message format needs to be changed to include the list of PCE Sequence Objects for the PCE inter-domain P2MP calculation request.

Each PCE Sequence can be obtained from the domain sequence for a

specific path. All the PCE sequences for all the paths of P2MP inter-domain form the PCE Topology Tree of the P2MP LSP.

The format of the new PCE Sequence Object for IPv4 (Object-Type 3) is as follows:

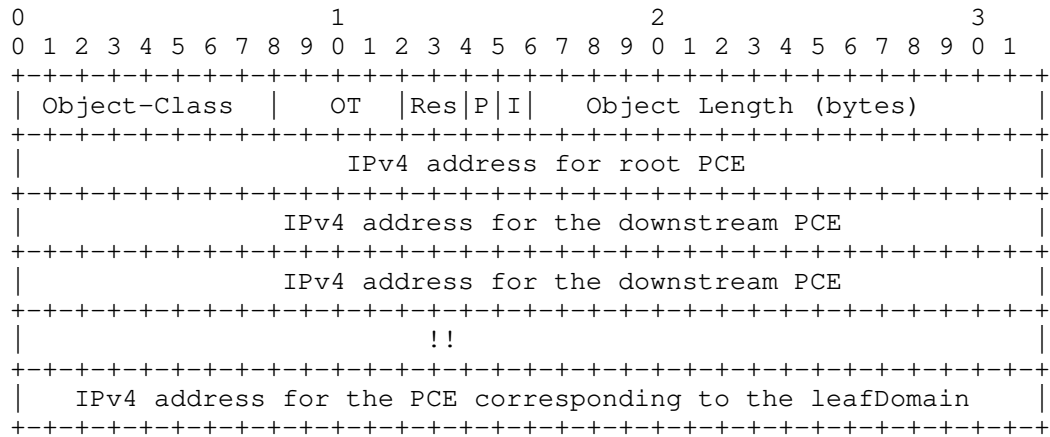


Figure 5: The New PCE Sequence Object Body Format for IPv4

The format of the new PCE Sequence Object for IPv6 (Object-Type 3) is as follows:

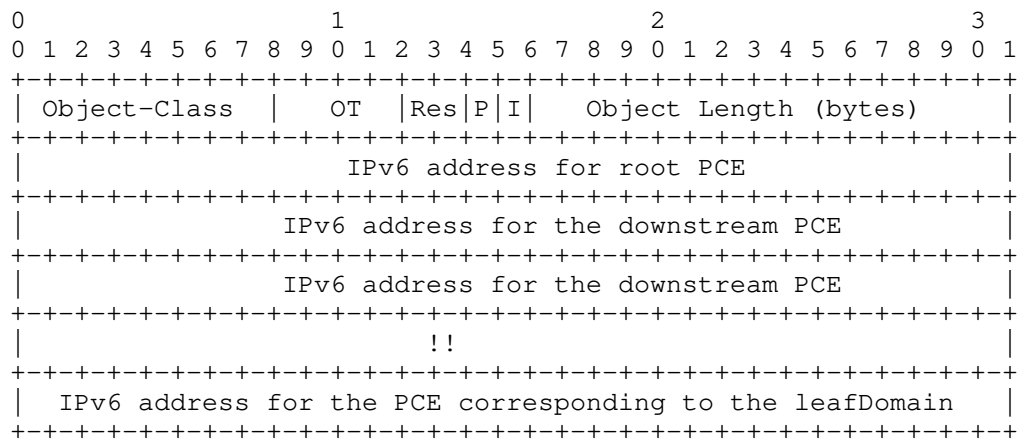


Figure 6: The New PCE Sequence Object Body Format for IPv6

7.5. Relationship with Hierarchical PCE

The actual grafting of subtrees into the Multi-Domain tree needs to be carried out by the source node. This means that the source node needs to get the computed sub-paths from all the involved domains. This requires that the source node either has a PCEP session with all the PCEs, or PCEP messages are routed via the PCEP sessions. This may mean an excessive number of sessions or an added complexity in implementations.

Alternatively, one may use an architecture based on the concept of hierarchical PCE [H-PCE]. The parent PCE would be responsible to request Intra-domain subtrees to the PCEs, combine them and return the overall P2MP tree.

7.6. Parallelism

In order to minimize latency in path computation in multi-domain networks, intra-domain path segments and intra-domain sub-trees SHOULD be computed in parallel when possible. The proposed procedures in this draft present opportunities for parallelism:

1. The BRPC procedure for each leaf node can be launched in parallel by the ingress/root PCE if the dynamic computation of the Core Tree is enabled.
2. Intra-domain P2MP paths can also be computed in parallel by the PCEs once the entry and exit nodes within a domain are known

One of the potential issues of parallelism is that the ingress PCE would require a potentially high number of PCEP adjacencies to "remote" PCEs and that may not be desirable, but a given PCE would only receive requests for the destinations that are in its domain (+ the core nodes), without PCEs forwarding requests.

8. Manageability Considerations

[RFC5862] describes various manageability requirements in support of P2MP path computation when applying PCEP. This section describes how manageability requirements mentioned in [RFC5862] are supported in the context of PCEP extensions specified in this document.

Note that [RFC5440] describes various manageability considerations in PCEP, and most of manageability requirements mentioned in [PCE-P2MP

P2MP] are already covered there.

9. Control of Function and Policy

In addition to configuration parameters listed in [RFC5440], the following parameters MAY be required.

- o P2MP path computations enabled or disabled.
- o Advertisement of P2MP path computation capability enabled or disabled (discovery protocol, capability exchange).

10. Information and Data Models

As described in [RFC5862], MIB objects MUST be supported for PCEP extensions specified in this document.

11. Liveness Detection and Monitoring

There are no additional considerations beyond those expressed in [RFC5440], since [RFC5862] does not address any additional requirements.

12. Verifying Correct Operation

There are no additional considerations beyond those expressed in [RFC5440], since [RFC5862] does not address any additional requirements.

13. Requirements on Other Protocols and Functional Components

As described in [RFC5862], the PCE MUST obtain information about the P2MP signaling and branching capabilities of each LSR in the network.

Protocol extensions specified in this document do not provide such capability. Other mechanisms MUST be present.

14. Impact on Network Operation

It is expected that use of PCEP extensions specified in this document will not have a significant impact on network operations.

15. Security Considerations

As described in [RFC5862], P2MP path computation requests are more CPU-intensive and also use more link bandwidth. Therefore, it may be more vulnerable to denial of service attacks. Therefore, it is more important that implementations conform to security requirements of [RFC5440], and the implementer utilize those security features.

16. IANA Considerations

A new flag of the RP object (specified in [RFC5440]) is defined in this document.

TBD.

17. Acknowledgements

The authors would like to thank Adrian Farrel, Dan Tappan and Olufemi Komolafe for their valuable comments on this draft.

18. References

18.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC6006] Takeda, T., Chaitou M., Le Roux, J.L., Ali Z., Zhao, Q., King, D., "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC6006, September 2010.

18.2. Informative References

- [RFC4461] Yasukawa, S., "Signaling Requirements for Point-to-Multipoint Traffic-Engineered MPLS Label Switched Paths (LSPs)", RFC 4461, April 2006.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

- [RFC4875] Aggarwal, R., Papadimitriou, D., and S. Yasukawa, "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, May 2007.
- [RFC5152] Vasseur, JP., Ayyangar, A., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.
- [RFC5376] Bitar, N., Zhang, R., and K. Kumaki, "Inter-AS Requirements for the Path Computation Element Communication Protocol (PCECP)", RFC 5376, November 2008.
- [RFC5441] Roux, J., Vasseur, J., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5441, June 2009.
- [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, April 2009.
- [RFC5671] Yasukawa, S. and A. Farrel, "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering (TE)", RFC 5671, August 2009.
- [RFC5862] Yasukawa, S. and A. Farrel, "PCC-PCE Communication Requirements for Point to Multipoint Multiprotocol Label Switching Traffic Engineering (MPLS-TE)", RFC 5862, June 2010.
- [H-PCE] King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS & GMPLS", July 2010.

Authors' Addresses

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

Email: qzhao@huawei.com

Zafar Ali
Cisco Systems
US

Email: zali@cisco.com

Tarek Saad
Cisco Systems
US

Email: tsaad@cisco.com

Siva Sivabalan
Cisco Systems
Canada

Email: msiva@cisco.com

Daniel King
Old Dog Consulting
UK

Email: daniel@olddog.co.uk

Ramon Casellas
CTTC - Centre Tecnologic de Telecomunicacions de Catalunya
Spain

Email: ramon.casellas@cttc.es

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 21, 2011

Q. Zhao
D. Dhody
U. Palle
Huawei Technology
D. King
Old Dog Consulting
September 21, 2010

Management Information Base for the PCE Communications Protocol (PCEP)
When Requesting Point-to-Multipoint Services
draft-zhao-pce-pcep-p2mp-mib-01

Abstract

This memo defines an experimental portion of the Management Information Base for use with network management protocols in the Internet community. In particular, it describes managed objects for modeling of the Path Computation Element communication Protocol (PCEP) for communications between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between two PCEs when point-to-multipoint services are requested.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on March 16, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Requirements Language

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 BCP 14, RFC 2119 [RFC2119].

Table of Contents

| | |
|---|----|
| 1. Introduction | 3 |
| 2. Terminology | 3 |
| 3. The Internet-Standard Management Framework | 4 |
| 4. PCEP P2MP MIB Module Architecture | 4 |
| 5. Example of the PCEP P2MP MIB module usage | 4 |
| 6. Object definitions | 5 |
| 6.1. PCE-PCEP-P2MP-DRAFT-MIB | 5 |
| 6.2. Objects for inclusion in module PCE-PCEP-DRAFT-MIB | 18 |
| 7. IANA Considerations | 19 |
| 8. Security Considerations | 19 |
| 9. References | 20 |
| 9.1. Normative References | 20 |
| 9.2. Informative References | 21 |

1. Introduction

The Path Computation Element (PCE) defined in [RFC4655] is an entity that is capable of computing a network path or route based on a network graph, and applying computational constraints. A Path Computation Client (PCC) may make requests to a PCE for paths to be computed.

A P2MP LSP is comprised of multiple source-to-leaf (S2L) sub-LSPs. These S2L sub-LSPs are set up between ingress and egress LSRs and are appropriately combined by the branch LSRs using computation results from the PCE to determine the path of a P2MP TE LSP.

The PCE communication protocol (PCEP) is designed as a communication protocol between PCCs and PCEs for point-to-point (P2P) path computations and is defined in [RFC5440]. [PCE-PCEP-P2MP-EXT] explains how to extend the PCEP protocol for P2MP scenario.

[PCE-PCEP-DRAFT-MIB] defines a portion of the Management Information Base (MIB) for use with network management protocols in the Internet community for P2P path computations.

This memo defines an experimental portion of the Management Information Base for use with network management protocols in the Internet community. In particular, it describes managed objects for modeling of Path Computation Element communication Protocol (PCEP) [RFC5440] for communications between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between two PCEs in P2MP scenarios.

Some objects maybe moved to [PCE-PCEP-DRAFT-MIB] after consensus with the authors and working group, these are defined in section 6.2.

2. Terminology

The following terminology is used in this document.

Domain: Any collection of network elements within a common sphere of address management or path computational responsibility. Examples of domains include Interior Gateway Protocol (IGP) areas and Autonomous Systems (ASs).

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

PCC: Path Computation Client: any client application requesting a path computation to be performed by a Path Computation Element.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

P2MP: Point-to-Multipoint

P2P: Point-to-Point

3. The Internet-Standard Management Framework

For a detailed overview of the documents that describe the current Internet-Standard Management Framework, please refer to section 7 of RFC 3410 [RFC3410].

Managed objects are accessed via a virtual information store, termed the Management Information Base or MIB. MIB objects are generally accessed through the Simple Network Management Protocol (SNMP). Objects in the MIB are defined using the mechanisms defined in the Structure of Management Information (SMI). This memo specifies a MIB module that is compliant to the SMIV2, which is described in STD 58, RFC 2578 [RFC2578] and STD 58, RFC 2580 [RFC2580].

4. PCEP P2MP MIB Module Architecture

The PCEP P2MP MIB is just an extension of the existing architecture defined in [PCE-PCEP-DRAFT-MIB] by adding additional objects which are either common to P2P and P2MP or which are specific to P2MP. All these new objects are added into the two new tables (pcePcepExtSessionTable and pcePcepExtClientTable) defined in this new MIB module. The relationship among the two new tables to the two existing tables in [PCE-PCEP-DRAFT-MIB] are shown in the following figure:

```
pcePcepSessionTable <----- pcePcepExtSessionTable
pcePcepClientTable  <----- pcePcepExtClientTable
```

An arrow in the figure above shows that the MIB table pointed from contains a reference to the MIB table pointed to.

5. Example of the PCEP P2MP MIB module usage

In this section we provide an example (pcePcepExtClientTable 1) of using the MIB objects described in Section 6 (Object definitions) to monitor. While this example is not meant to illustrate every

permutation of the MIB, it is intended as an aid to understanding some of the key concepts. It is meant to be read after going through the MIB itself.

```
pcePcepExtClientTable 1 of the PCE-PCEP-P2MP-DRAFT-MIB module :
{
    pcePcepClientP2mpCapabilityStatus    enable(1),
    pcePcepClientOverloadStatus          resumed(2),
    pcePcepClientOverloadDuration        (10),
}
```

6. Object definitions

6.1. PCE-PCEP-P2MP-DRAFT-MIB

This MIB module makes references to the following documents.

[RFC2578], [RFC2580], [RFC3411], [RFC2863], [RFC3813], [PCE-PCEP-DRAFT-MIB].

PCE-PCEP-P2MP-DRAFT-MIB DEFINITIONS ::= BEGIN

```
IMPORTS
    MODULE-IDENTITY, OBJECT-TYPE,
    Unsigned32,
    Counter32,
    experimental
        FROM SNMPv2-SMI                -- [RFC2578]

    pcePcepClientPcepId, pcePcepClientIndex,
    pcePcepPeerPcepId
        FROM PCE-PCEP-DRAFT-MIB

    MODULE-COMPLIANCE,
    OBJECT-GROUP,
        FROM SNMPv2-CONF;              -- [RFC2580]
```

pcePcepP2mpDraftMIB MODULE-IDENTITY

LAST-UPDATED "201009151200Z" --Sep 15, 2010

ORGANIZATION "Path Computation Element (PCE) Working Group"

CONTACT-INFO "

Quintin Zhao
Dhruv Dhody
Udayasree Palle
Huawei Technology
Daniel King
OldDog Consulting

EMail: qzhao@huawei.com

EMail: dhruvd@huawei.com

EMail: udayasreepalle@huawei.com

EMail: daniel@oldog.co.uk

Email comments directly to the PCE WG Mailing List at pce@ietf.org

WG-URL: <http://www.ietf.org/html.charters/pce-charter.html>

"

DESCRIPTION

"This extended MIB module defines a collection of objects for managing PCE communication protocol(PCEP) when point-to-multipoint services are requested"

-- Revision history

REVISION

"201009151200Z" -- 15 Sep 2010 12:00:00 EST

DESCRIPTION

"

Changes from -00 draft :

1. Removed pathkey objects as these objects to be made as a new MIB module for pathkey. As per section 6.2 of [RFC5520].
2. Rearrangement of the sections for better understanding
3. Addition of STATUS (optional or mandatory) in the definitions
4. Addition of section 6.2 to gather all objects which may be moved to [PCE-PCEP-DRAFT-MIB]"

REVISION

"201007051200Z" -- July 05 2010 12:00:00 EST

DESCRIPTION

"draft-00 version"

::= { experimental 9999 } --

```

pcePcepExtMIBObjects OBJECT IDENTIFIER ::= { pcePcepExtDraftMIB 0 }
pcePcepExtConformance OBJECT IDENTIFIER ::= { pcePcepExtDraftMIB 1 }
pcePcepExtClientObjects OBJECT IDENTIFIER ::= { pcePcepExtMIBObjects
1 }

--

-- PCE Extended Client Objects

--

pcePcepClientVersionnumber OBJECT-TYPE
    SYNTAX      Unsigned32
    MAX-ACCESS  read-only
    STATUS      optional
    DESCRIPTION
        "The current version number of the PCEP protocol is 1."
    ::= { pcePcepExtClientObjects 1 }

pcePcepExtClientTable OBJECT-TYPE
    SYNTAX      SEQUENCE OF PcePcepClientEntry
    MAX-ACCESS  not-accessible
    STATUS      current
    DESCRIPTION
        "This table contains information about the
        PCEP Client."
    ::= { pcePcepExtClientObjects 2 }

pcePcepExtClientEntry OBJECT-TYPE
    SYNTAX      PcePcepClientEntry
    MAX-ACCESS  not-accessible
    STATUS      current
    DESCRIPTION
        "An entry in this table represents a PCEP client.
        An entry can be created by a network administrator
        or by an SNMP agent as instructed by PCEP."

    INDEX      { pcePcepClientPcepId,
                  pcePcepClientIndex,
                  pcePcepPeerPcepId }

    ::= { pcePcepExtClientTable 1 }

PcePcepExtClientEntry ::= SEQUENCE {
    pcePcepClientP2mpCapabilityStatus    INTEGER,
    pcePcepClientOverloadStatus          INTEGER,
    pcePcepClientOverloadDuration        Unsigned32
}

```

```
pcePcepClientP2mpCapabilityStatus OBJECT-TYPE
    SYNTAX      INTEGER {
                    enable (1),
                    disable(2)
                }
    MAX-ACCESS   read-only
    STATUS       mandatory
    DESCRIPTION
        "The P2MP capability status of this PCEP client."
    ::= { pcePcepExtClientEntry 1 }

pcePcepClientOverloadStatus OBJECT-TYPE
    SYNTAX      INTEGER {
                    overloaded(1),
                    resumed(2)
                }
    MAX-ACCESS   read-only
    STATUS       optional
    DESCRIPTION
        "The Overload status of this PCE client."
    ::= { pcePcepExtClientEntry 2 }

pcePcepClientOverloadDuration OBJECT-TYPE
    SYNTAX      Unsigned32
    UNITS        "seconds"
    MAX-ACCESS   read-only
    STATUS       optional
    DESCRIPTION
        "The period of time during which no further request should
        be sent to the PCE client. Once this period of time has
        elapsed, the PCE client should no longer be considered in
        a congested state."
    ::= { pcePcepExtClientEntry 3 }

pcePcepExtSessionObjects OBJECT IDENTIFIER ::= { pcePcepExtMIBObjects
2 }

--

-- The PCEP Ext Sessions Table

--
```

```
pcePcepExtSessionTable OBJECT-TYPE
    SYNTAX      SEQUENCE OF pcePcepExtSessionEntry
    MAX-ACCESS  not-accessible
    STATUS      current
    DESCRIPTION
        "A table of extended sessions characteristics between
        PCEP clients. Each row in this table represents a
        single session."
    ::= { pcePcepExtSessionObjects 1 }

pcePcepExtSessionEntry OBJECT-TYPE
    SYNTAX      pcePcepExtSessionEntry
    MAX-ACCESS  not-accessible
    STATUS      current
    DESCRIPTION
        "An entry in this table represents information on a
        single session between two PCEP clients.
        The information contained in a row is read-only."
    ::= { pcePcepExtSessionTable 1 }
```



```

PcePcepExtSessionEntry ::= SEQUENCE {
    pcePcepSessionP2mpPCReqMessagesSent      Unsigned32,
    pcePcepSessionP2mpPCRepMessagesSent      Unsigned32,
    pcePcepSessionP2mpPCReqMessagesReceived  Unsigned32,
    pcePcepSessionP2mpPCRepMessagesReceived  Unsigned32,
    pcePcepSessionP2mpAddLeaves              Unsigned32,
    pcePcepSessionP2mpRemoveLeaves           Unsigned32,
    pcePcepSessionP2mpModifyLeaves           Unsigned32,
    pcePcepSessionP2mpUnchangedLeaves        Unsigned32,
    pcePcepSessionTotalMessagesSent          Unsigned32,
    pcePcepSessionOpenMessagesSent          Unsigned32,
    pcePcepSessionKeepaliveMessagesSent     Unsigned32,
    pcePcepSessionPCNtfMessagesSent         Unsigned32,
    pcePcepSessionPCErrMessagesSent         Unsigned32,
    pcePcepSessionTotalMessagesReceived     Unsigned32,
    pcePcepSessionOpenMessagesReceived      Unsigned32,
    pcePcepSessionKeepaliveMessagesReceived Unsigned32,
    pcePcepSessionPCNtfMessagesReceived     Unsigned32,
    pcePcepSessionPCErrMessagesReceived     Unsigned32,
    pcePcepSessionIntraDomainRequest        Unsigned32,
    pcePcepSessionInterDomainRequest        Unsigned32,
    pcePcepSessionSuccessComps              Unsigned32,
    pcePcepSessionNoReply                   Unsigned32,
    pcePcepSessionSynchronization           Unsigned32,
    pcePcepSessionReoptimization            Unsigned32,
    pcePcepSessionFragmentation             Unsigned32,
    pcePcepSessionP2pPCReqMessagesSent      Unsigned32,
    pcePcepSessionP2pPCRepMessagesSent      Unsigned32,
    pcePcepSessionP2pPCReqMessagesReceived  Unsigned32,
    pcePcepSessionP2pPCRepMessagesReceived  Unsigned32
}

```

pcePcepSessionP2mpPCReqMessagesSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS mandatory

DESCRIPTION

"The number of P2MP Request messages sent on this session."

::= { pcePcepExtSessionEntry 1 }

pcePcepSessionP2mpPCRepMessagesSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS mandatory

DESCRIPTION

"The number of P2MP Reply messages sent on this session."

::= { pcePcepExtSessionEntry 2 }

```
pcePcepSessionP2mpPCReqMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of P2MP Request messages received on this
        session."
    ::= { pcePcepExtSessionEntry 3 }

pcePcepSessionP2mpPCRepMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of P2MP Reply messages received on this
        session."
    ::= { pcePcepExtSessionEntry 4 }

pcePcepSessionP2mpAddLeaves OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of leaves to be Added (Type1) for the
        total P2MP requests (PCReq message) received by
        the PCE."
    ::= { pcePcepExtSessionEntry 5 }

pcePcepSessionP2mpRemoveLeaves OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of leaves to be Removed (Type2) for the
        total P2MP requests (PCReq message) received by the
        PCE."
    ::= { pcePcepExtSessionEntry 6 }

pcePcepSessionP2mpModifyLeaves OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of leaves to be Modified (Type3) for the
        total P2MP requests (PCReq message) received by the
        PCE."
    ::= { pcePcepExtSessionEntry 7 }
```

```
pcePcepSessionP2mpUnchangedLeaves OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of leaves not to be changed (Type4) for
        the total P2MP requests (PCReq message) received
        by the PCE."
    ::= { pcePcepExtSessionEntry 8 }

pcePcepSessionTotalMessagesSent OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The total number of PCEP messages sent on this
        session."
    ::= { pcePcepExtSessionEntry 9 }

pcePcepSessionOpenMessagesSent OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of Open messages sent on this session."
    ::= { pcePcepExtSessionEntry 10 }

pcePcepSessionKeepaliveMessagesSent OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of Keepalive messages sent on this session."
    ::= { pcePcepExtSessionEntry 11 }

pcePcepSessionPCNtfMessagesSent OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of PCNtf messages sent on this session."
    ::= { pcePcepExtSessionEntry 12 }
```

```
pcePcepSessionPCErrMessagesSent OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of PCErr messages sent on this session."
    ::= { pcePcepExtSessionEntry 13 }

pcePcepSessionTotalMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The total number of PCEP messages received on this
        session."
    ::= { pcePcepExtSessionEntry 14 }

pcePcepSessionOpenMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of Open messages received on this
        session."
    ::= { pcePcepExtSessionEntry 15 }

pcePcepSessionKeepaliveMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of Keepalive messages received on this
        session."
    ::= { pcePcepExtSessionEntry 16 }

pcePcepSessionPCNtfMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of PCNtf messages received on this
        session."
    ::= { pcePcepExtSessionEntry 17 }
```

```
pcePcepSessionPCErrMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of PCErr messages received on this
        session."
    ::= { pcePcepExtSessionEntry 18 }

pcePcepSessionIntraDomainRequest OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of requests sent for the Intra-Domain
        path computation."
    ::= { pcePcepExtSessionEntry 19 }

pcePcepSessionInterDomainRequest OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of requests sent for the Inter-Domain path
        computation."
    ::= { pcePcepExtSessionEntry 20 }

pcePcepSessionSuccessComps OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of requests which had successful
        computations. In case of PCC-PCE session, it is core
        computation value and in case of PCE-PCE session, it
        is transit computation value."
    ::= { pcePcepExtSessionEntry 21 }

pcePcepSessionNoReply OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        " The number of requests which had not been replied
        either success or failure."
    ::= { pcePcepExtSessionEntry 22 }
```

```
pcePcepSessionSynchronization OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of synchronized path computation requests
        that can be either dependent or independent."
    ::= { pcePcepExtSessionEntry 23 }

pcePcepSessionReoptimization OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of requests for Reoptimization."
    ::= { pcePcepExtSessionEntry 24 }

pcePcepSessionFragmentation OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS optional
    DESCRIPTION
        "The number of packets of a PCReq / PCRep
        message which had been fragmented."
    ::= { pcePcepExtSessionEntry 25 }

pcePcepSessionP2pPCReqMessagesSent OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of P2P Request messages sent on this
        session."
    ::= { pcePcepExtSessionEntry 26 }

pcePcepSessionP2pPCRepMessagesSent OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of P2P Reply messages sent on this session."
    ::= { pcePcepExtSessionEntry 27 }
```

```

pcePcepSessionP2pPCReqMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of P2P PCReq messages received on this
        session."
    ::= { pcePcepExtSessionEntry 28 }

pcePcepSessionP2pPCRepMessagesReceived OBJECT-TYPE
    SYNTAX Counter32
    MAX-ACCESS read-only
    STATUS mandatory
    DESCRIPTION
        "The number of P2P PCRep messages received on this
        session."
    ::= { pcePcepExtSessionEntry 29 }

--*****
-- Module Conformance Statement
--*****

pcePcepExtGroups
    OBJECT IDENTIFIER ::= { pcePcepExtConformance 1 }

pcePcepExtCompliances
    OBJECT IDENTIFIER ::= { pcePcepExtConformance 2 }

--
-- Full Compliance
--

pcePcepExtModuleFullCompliance MODULE-COMPLIANCE
    STATUS current
    DESCRIPTION
        "The Module is implemented with support
        for read-create and read-write. In other
        words, both monitoring and configuration
        are available when using this MODULE-COMPLIANCE."

    MODULE -- this module
        MANDATORY-GROUPS { pcePcepExtGeneralGroup,
                           }

    ::= { pcePcepExtCompliances 1 }

```

```
--
-- Read-Only Compliance
--

pcePcepExtModuleReadOnlyCompliance MODULE-COMPLIANCE
    STATUS current
    DESCRIPTION
        "The Module is implemented with support
        for read-only.  In other words, only monitoring
        is available by implementing this MODULE-COMPLIANCE."

    MODULE -- this module
        MANDATORY-GROUPS { pcePcepExtGeneralGroup,
                           }
        ::= { pcePcepExtCompliances 2 }

-- units of conformance
```



```

pcePcepExtGeneralGroup OBJECT-GROUP
    OBJECTS {

        pcePcepClientP2mpCapabilityStatus,
        pcePcepSessionP2mpPCReqMessagesSent,
        pcePcepSessionP2mpPCRepMessagesSent,
        pcePcepSessionP2mpPCReqMessagesReceived,
        pcePcepSessionP2mpPCRepMessagesReceived,
        pcePcepSessionP2mpAddLeaves,
        pcePcepSessionP2mpRemoveLeaves,
        pcePcepSessionP2mpModifyLeaves,
        pcePcepSessionP2mpUnchangedLeaves,
        pcePcepSessionTotalMessagesSent,
        pcePcepSessionOpenMessagesSent,
        pcePcepSessionKeepaliveMessagesSent,
        pcePcepSessionPCNtfMessagesSent,
        pcePcepSessionPCErrMessagesSent,
        pcePcepSessionTotalMessagesReceived,
        pcePcepSessionOpenMessagesReceived,
        pcePcepSessionKeepaliveMessagesReceived,
        pcePcepSessionPCNtfMessagesReceived,
        pcePcepSessionPCErrMessagesReceived,
        pcePcepSessionP2pPCReqMessagesSent,
        pcePcepSessionP2pPCRepMessagesSent,
        pcePcepSessionP2pPCReqMessagesReceived,
        pcePcepSessionP2pPCRepMessagesReceived
    }
    STATUS      current
    DESCRIPTION
        "Objects that apply to all PCEP P2MP MIB implementations."

    ::= { pcePcepExtGroups 1 }

END

```

6.2. Objects for inclusion in module PCE-PCEP-DRAFT-MIB

Following are the objects maybe moved to [PCE-PCEP-DRAFT-MIB] after consensus with the authors and working group.

pcePcepClientVersionnumber,
pcePcepClientP2mpCapabilityStatus,
pcePcepClientOverloadStatus,
pcePcepClientOverloadDuration,
pcePcepSessionTotalMessagesSent,
pcePcepSessionOpenMessagesSent,
pcePcepSessionKeepaliveMessagesSent,
pcePcepSessionPCNtfMessagesSent,
pcePcepSessionPCErrMessagesSent,
pcePcepSessionTotalMessagesReceived,
pcePcepSessionOpenMessagesReceived,
pcePcepSessionKeepaliveMessagesReceived,
pcePcepSessionPCNtfMessagesReceived,
pcePcepSessionPCErrMessagesReceived,
pcePcepSessionIntraDomainRequest,
pcePcepSessionInterDomainRequest,
pcePcepSessionSuccessComps,
pcePcepSessionNoReply,
pcePcepSessionSynchronization,
pcePcepSessionReoptimization,
pcePcepSessionFragmentation,
pcePcepSessionP2pPCReqMessagesSent,
pcePcepSessionP2pPCRepMessagesSent,
pcePcepSessionP2pPCReqMessagesReceived,
pcePcepSessionP2pPCRepMessagesReceived

7. IANA Considerations

TBD

8. Security Considerations

The readable objects in the PCE-PCEP-DRAFT-MIB module (i.e., those with MAX-ACCESS other than not-accessible) may be considered sensitive in some environments since, collectively, they provide information about the amount and frequency of path computation requests and responses within the network and can reveal some aspects of their configuration.

In such environments it is important to control also GET and NOTIFY access to these objects and possibly even to encrypt their values when sending them over the network via SNMP.

SNMP versions prior to SNMPv3 did not include adequate security. Even if the network itself is secure (for example by using IPsec), even then, there is no control as to who on the secure network is allowed to access and GET/SET (read/change/create/delete) the objects in this MIB module.

It is RECOMMENDED that implementers consider the security features as provided by the SNMPv3 framework (see [RFC3410], section 8), including full support for the SNMPv3 cryptographic mechanisms (for authentication and privacy).

Further, deployment of SNMP versions prior to SNMPv3 is NOT RECOMMENDED. Instead, it is RECOMMENDED to deploy SNMPv3 and to enable cryptographic security. It is then a customer/operator responsibility to ensure that the SNMP entity giving access to an instance of this MIB module is properly configured to give access to the objects only to those principals (users) that have legitimate rights to indeed GET or SET (change/create/delete) them.

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2578] McCloghrie, k., Perkins, D., Schoenwaelder, J., Case, J., Rose, M., and S. Waldbusser, "Structure of Management Information Version 2 (SMIv2)", April 1999.
- [RFC2580] McCloghrie, k., Perkins, D., Schoenwaelder, J., Case, J., Rose, M., and S. Waldbusser, "Conformance Statements for SMIv2", April 1999.
- [RFC2863] McCloghrie, k. and F. Kastenholz, "The Interfaces Group MIB", June 2000.
- [RFC3411] Harrington, D., Presuhn, R., and B. Wijnen, "An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks", December 2002.
- [RFC3813] Srinivasan, C., Viswanathan, A., and T. Nadeau, "MPLS Multiprotocol Label Switching (MPLS) Label Switch Router Management Information Base", June 2004.
- [RFC5440] Ayyangar, A ., Farrel, A ., Oki, E., Atlas, A., Dolganow, A., Ikejiri, Y., Kumaki, K., Vasseur, J., and J. Roux, "Path Computation Element (PCE) communication Protocol (PCEP)", March 2009.

9.2. Informative References

- [PCE-PCEP-DRAFT-MIB] Kiran Koushik, A S., Stephan, E., Zhao, Q., and D. King, "PCE communication protocol(PCEP) Management Information Base", July 2010.
- [PCE-PCEP-P2MP-EXT] Zhao, Q. and D. King, "Introduction and Applicability Statements for Internet-Standard Management Framework", May 2010.
- [RFC3410] Case, J ., Mundy, R., Partain, D., and B. Stewart, "Introduction and Applicability Statements for Internet-Standard Management Framework", December 2002.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC5520, April 2009.

Authors' Addresses

Quintin Zhao
Huawei Technology
125 Nagog Technology Park
Acton, MA 01719
US
EMail: qzhao@huawei.com

Dhruv Dhody
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA
EMail: dhruvd@huawei.com

Udayasree Palle
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA
EMail: Udayasreepalle@huawei.com

Daniel King
Old Dog Consulting
UK
EMail: daniel@olddog.co.uk

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 22, 2011

YL. Zhao
J. Zhang
BUPT
RJ. Jing
China Telecom Beijing Research
Institute
DJ. Wang
XH. Fu
ZTE Corporation
October 19, 2010

Protocol Extension Requirement for Cooperation between PCE and
Distributed Routing Controller in GMPLS Networks
draft-zhaoyl-pce-dre-01

Abstract

Path Computation Element (PCE) and distributed routing controller in GMPLS networks have different advantages of path computation respectively. PCE is suitable for the path computation in multi-layer and multi-domain networks, especially in multi-constraints environment. While distributed routing controller is good at the path computation in parallel and distributed network control in the local domain. A cooperative path computation architecture named Dual Routing Engine (DRE) is proposed, which is based on the two path computation engines and can combine the advantages of centralized and distributed. The corresponding PCE communication protocol extension and other protocol requirements for cooperation between PCE and distributed routing controller are listed in this document.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 22, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

| | |
|--|----|
| 1. Introduction | 4 |
| 1.1. Conventions Used in This Document | 4 |
| 2. Terminologies | 4 |
| 3. General Assumptions | 5 |
| 4. Cooperative Architecture based on PCE and Distributed Routing Controller | 6 |
| 4.1. DRC | 7 |
| 4.2. PCE | 7 |
| 5. Different Application Scenarios | 8 |
| 5.1. Cross layers and cross domains | 8 |
| 5.2. Independent coexistence | 8 |
| 5.3. Security backup | 8 |
| 5.4. Policy-enabled | 8 |
| 5.5. Constraint-based | 9 |
| 5.6. Service-oriented | 9 |
| 5.7. Cooperation between network level and node level | 9 |
| 6. PCEP Extension Requirements | 10 |
| 6.1. PCEP extension requirement for the communication between RES and PCE | 10 |
| 6.2. PCEP extension requirement for the communication between RES and DRC | 11 |
| 7. Other Protocol Extension Requirements | 11 |
| 7.1. OSPF-TE extension requirement for the cooperative architecture | 11 |
| 7.2. RSVP-TE extension requirement for the cooperative architecture | 12 |
| 8. Discussions | 12 |
| 9. Security Considerations | 12 |
| 10. Acknowledgments | 12 |
| 11. References | 13 |

| | |
|--|----|
| 11.1. Normative References | 13 |
| 11.2. Informative References | 13 |
| Authors' Addresses | 13 |

1. Introduction

Path Computation Element (PCE) is proposed to complete the constraint-based shortest path computation in Multi-protocol Label Switching (MPLS) and Generalized MPLS (GMPLS) multi-layer and multi-domain networks [RFC4665]. Then a series of Request for Comments (RFCs) related to PCE technology, such as requirements for PCE discovery, PCE Communication Protocol (PCEP), a backward recursive PCE-based computation procedure (BRPC) and so on, have been standardized. Studies prove that PCE is suitable for cross-layer and cross-domain design of networks, capable of end-to-end path computation under multiple constraints [1-2] and potentially applied to resource allocation and routing optimization [3-4]. While traditional distributed routing controller (DRC) in GMPLS/ASON control plane is good at the path computation in parallel and distributed network control in the local domain. On the other hand, as the huge bandwidth requirement emerges, capacity of Tbit/s transmission links and Pbit/s switching are necessary for next generation optical networks. According to the current photonics technology level, the node architecture will be very complicated and of large power consumption. Then how to configure the switching architecture in the node will be very important for the entire network performance. DRC can also be used for the management and configuration of resource in the internal node. Of course, both PCE and DRC have corresponding different disadvantages respectively.

In order to optimize the performance of optical networks under different application scenarios, PCE and distributed routing controller need to cooperate with each other. A cooperative path computation architecture named Dual Routing Engine (DRE) is proposed, which includes two path computation engines, i.e. PCE and distributed routing controller, and can combine their advantages. Several different application scenarios are described in this document. PCE communication protocol extension and other protocol extension requirements for cooperation between PCE and distributed routing controller are listed here.

1.1. 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].

2. Terminologies

LSR: Label Switching Router.

LSP: Label Switched Path.

PCC: Path Computation Client. Any client application requesting a path computation to be performed by the Path Computation Element.

PCE (Path Computation Element): an entity (component, application or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

PCEP: PCE Communication Protocols, the communication protocol between PCC and PCE.

DRC: Distributed Routing Controller, the module that can complete path computation in the local domain or the internal node.

DRE: Dual Routing Engine, including PCE and DRC.

TED: Traffic Engineering Database.

RID: Resource Information Databases.

3. General Assumptions

PCE and distributed routing controller are two path computation engines which have been standardized by IETF working group. They can both complete the path computation in GMPLS networks. In order to show how the cooperative architecture based on PCE and distributed routing controller works, we make some assumptions as follows.

- o Each GMPLS-based control node is equipped with a distributed routing controller which can complete the path computation in the local domain, even the path computation and resource configuration in the internal node, and each domain is equipped with no less than one PCE which cannot only complete the intra-domain path computation, but also complete the inter-domain path computation.
- o The topology and TE information are updated as soon as any change occurs in the network, and the information kept at PCE and distributed routing controller are synchronized by OSPF-TE.
- o PCE and distributed routing controller can be selected arbitrary according to the local routing strategy.
- o Constraints and strategies can be considered by PCE during the process of path computation including the intra-domain path and inter-domain path.

- o An end-to-end path which crosses domains or crosses layers can be completed by several PCEs or several DRCs or several PCEs and DRCs. For example, the section of an end-to-end path in one domain may be computed by PCE, but the section of this path in another domain may be computed by DRC.

4. Cooperative Architecture based on PCE and Distributed Routing Controller

As shown in Fig. 1, cooperative architecture consists of Path Computation Element (PCE) and distributed routing controller (DRC). Both of them maintain Traffic Engineering Databases (TEDs) to keep network topology and other status information within the scope of their respective functions. There is a function module named Routing Engine Selector (RES) at each control node, which is responsible for managing the switchover process between PCE and DRC and choosing the right one while a LSP setup request arrives. At the same time, RES poses the function of PCC and can be implemented in Connection Controller (CC) module of GMPLS-based control plane.

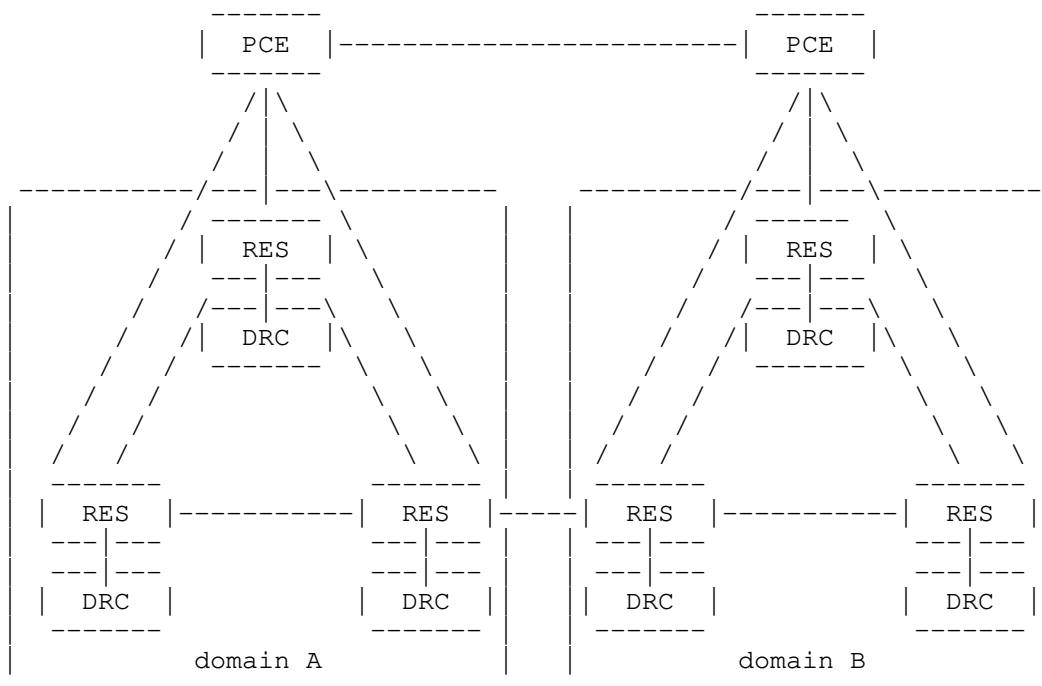


Fig.1 Cooperative architecture in multi-layer and multi-domain

networks

4.1. DRC

DRC offers general routing functions based on GMPLS/ASON control plane including link state advertisement, topology update and path computation. A typical DRC contains two essential modules which are topology analysis module and path computer module. The former floods network topology and resource utilization information to maintain a TED in each node. The latter responds to a routing request from Connection Controller (CC), finds the optimized path solution and returns it to CC. Except the general routing functions, DRC can also complete the resource allocation and configuration, which refers to internal resource allocation, ports interconnection, and switch fabric configuration in the internal node. This function is getting more and more important with the structure of the internal node getting more and more complex. Details of DRC operation are out of this document.

4.2. PCE

PCE has the advantage of centralized path computation especially in multi-layer and multi-domain networks, especially in multi-constraints environment. There is a Client/Server model between RES and PCE. RES sends a TE-LSP computation request to PCE. PCE performs path computation and returns the results back to RES.

Firstly RES makes choice of the best PCE based on a certain policy. It sends performance query requests to multiple related PCEs, each of which evaluates itself individually and then feeds back to the RES. RES gathers performance status information from PCEs and decides which one is feasible. Then a combination of RES and PCE is founded. Secondly PCE serves to deal with the path computation requests coming from RES through message parser. If a request carries path information, it will be directed towards path computer. If it carries policy information from RES, it is first interpreted by policy parser along with local policy settings and then loaded into path computer. Finally path computer implements multi-constraint routing on the basis of the path information, policy information and TE information. If path computation is successful the details of the whole route are returned to RES. If it only gets an incomplete route a new application for path computation request will be sent to the next hop RES. Then a PCE or DRC will be selected to complement the incomplete route.

5. Different Application Scenarios

Cooperation between PCE and DRC is one of the key issues for the cooperative architecture, which helps to achieve fast and exact routing due to the advantages of both PCE and DRC. This section will list several typical application scenarios of cooperation between PCE and DRC.

5.1. Cross layers and cross domains

It is necessary for PCE to collaborate with DRC when the requirements for path computation occur in multi-layer and multi-domain GMPLS networks. PCE could be shared among several domains or layers and make the best use of the inter-domain and inter-layer network resources, so it is more suitable for inter-domain or inter-layer path computation. DRC runs usually to fulfill intra-domain or intra-layer routing in contrast to PCE.

5.2. Independent coexistence

Both PCE and DRC are working independently under this scenario. While one customer applies a TE-LSP computation RES could select PCE or DRC arbitrarily. Of course only one path computation engine can be selected at each time. If a lot of applications for path computation arrive simultaneously, the burst computing load may be also balanced between them. The changed topology and resource status information have to be maintained in PCE and DRC. So it is difficult for the management of information synchronization on both sides.

5.3. Security backup

The cooperation mechanisms among different engines make it possible that PCE and DRC backup each other to enhance the routing security and reliability, since they both satisfy the demands of path computation from customers. When the working engine (e.g. PCE) fails, computation tasks could be switched to another reserved engine (e.g. DRC) as soon as possible. In such a case both PCE and DRC have to maintain the accordant network topology and resource status information.

5.4. Policy-enabled

In order to compute the optimal path in consideration of traffic engineering, different policies which mean series of rules and actions from management plane or control plane are involved. PCE is obviously more suitable for policy-enabled path computation framework than DRC. Tab.1 lists some typical policy application instances that may be exerted to the cooperative path computation architecture.

Effective combinations of the above scenarios as well as possible new scenarios could occur in the real networks.

| Policy application scenarios | Description |
|------------------------------|--|
| Policy configured paths | To centrally administer configured paths |
| Provider selection policy | To be applied in multi-provider topologies |
| Policy based constraints | To provide constraints in a path computation request |
| Advanced load balancing | To balance the traffic load for the whole network |

Policy-enabled path computation instances

Table 1

5.5. Constraint-based

Constraint-based path computation is a basic function especially for TE-LSP establishment. Available bandwidth, diversity, Shared Risk Link Group (SRLG), optical impairments, wavelength continuity and other constraints are likely to be considered. However, it is difficult to compute an optimal path with these constraints under the condition of the general GMPLS/ASON routing architecture. The centralized operation manner makes PCE easy to fulfill constraint-based path computation.

5.6. Service-oriented

PCEs can collaborate to finish constraint-based path computation without sharing TE information with each other, which are particularly useful when end-to-end constraints have to be taken into account because of protection and path diversity. PCEs should play an important role of service-oriented applications such as Layer 1 Virtual Private Network (L1 VPN), Bandwidth on Demand (BoD) and so on. Based on GMPLS/ASON architecture, the advantages of PCE and service plane can be combined to implement the framework of service-oriented application.

5.7. Cooperation between network level and node level

In this application scenario, PCE maintains Traffic Engineering Databases (TEDs) to keep network topology, and DRC maintains Resource

Information Databases (RIDs) to keep node internal topology and resource status. Both of them maintain the status information within the scope of their functions. Routing Engine Selector (RES) is responsible for managing the switchover process between PCE and DRC and choosing the right one when a LSP setup request arrives.

The interconnection between different nodes is general routing problem, which can be solved by PCE framework effectively, especially in multi-domain, multi-layer and multi-constraints scenarios, while the interconnection within the node is the problem of resource allocation and configuration, which refers to internal resource allocation, ports interconnection, and switch fabric configuration. Through the cooperation of PCE and DRC, we can make resource configuration more effectively and improve the performance of the entire optical networks.

6. PCEP Extension Requirements

As an extension of PCC, RES can not only complete the communication with PCE and DRC, but also complete the cooperation between PCE with DRC. There are some PCEP extension requirements for the cooperative path computation architecture and the procedure.

6.1. PCEP extension requirement for the communication between RES and PCE

PCEP is the communication protocol between RES and PCE. However, there is some extension requirement for PCEP in the cooperative path computation architecture.

Firstly, the path computation request messages from RES to PCE should be added an identification which appoints different application scenarios, and the corresponding data structure should be defined according to different identifications. For example, in the policy-enabled scenario, a policy object is necessary to be defined in the message sent from RES to PCE, and PCE should be able to parse the different policies and conduct the corresponding operations during the procedure of path computation.

Secondly, in the reply message from PCE to RES, some indication information should be contained except the routes information, such as the inter-domain loose path or the intra-domain path, the complete end-to-end path or section path, some failure information and path computation engine switchover requests.

6.2. PCEP extension requirement for the communication between RES and DRC

DRC can complete the path computation in the local domain in distributed method, which is usually implemented in the OSPF-TE module of GMPLS-based control plane. After the path computation, DRC returns the computation results to RES (CC) including the detailed routes and some failure or indication information.

As another important application, DRC can complete the path computation, resource management and configuration in the internal node with the node structure getting more and more complex. In this application scenario, DRC has the function of routing and resource allocation.

In all the application scenarios, DRC has the analogous functions with PCE and some different extension functions, such as the functions above. So there is requirement for application scenarios identification in the communication message between RES and DRC. The communication protocol between RES and DRC is necessary to be standardized as the function of control node getting more and more complex. Because RES and DRC are implemented in the same control node and DRC has similar functions with PCE, then a simplified PCEP can be used as the communication protocol between RES and DRC, which can include the path computation request, identification and reply messages only.

7. Other Protocol Extension Requirements

7.1. OSPF-TE extension requirement for the cooperative architecture

In the cooperative path computation architecture, the topology and TE information should be kept and synchronized at each control node and PCE in the local domain, and should also be updated as soon as any change occurs. Meanwhile, all the inter-domain links and TE information should be kept and synchronized at each PCE. So there is extension requirement for OSPF-TE to guarantee the information at each node synchronized in time.

Except the normal topology and TE information, some other constraints information may be necessary to be contained in the OSPF-TE protocol flooding in the entire networks, such as the physical impairment information. Then there is an extension requirement for the OSPF-TE protocol to enable the synchronization of all the constraint information at each node, which is of great value for the path computation and resource allocation.

7.2. RSVP-TE extension requirement for the cooperative architecture

In different cooperation modes between PCE and DRC, an end-to-end path may be computed by several PCE and DRC, and then be setup by signaling from the source node to the destination node. However, sometimes an end-to-end path which crosses domains or layers may be computed and setup section by section. Signaling should be triggered when a section path is gained. The current RSVP-TE protocol is necessary to be extended to support this application scenario.

Furthermore, as the scheme of path and resource allocation in the internal node is gained, the resource reservation and action of switches are to be conducted by some protocol as the control node is getting more and more complex. RSVP-TE is the optimal option for this function, and to be extended.

8. Discussions

According to the development of GMPLS networks, the cooperative architecture can be introduced in two steps. First, PCE and DRC can cooperate with each other to complete the path computation of the entire networks at network level. Second, PCE and DRC can cooperate with each other to complete the path computation at network level and resource computation and allocation at node level with the network size increasing and node structure getting complex.

9. Security Considerations

The cooperation between PCE and DRC can enhance the security of networks because they can backup each other. However, because the information is kept at both PCE and each DRC, especially the exchange of information across domain boundaries is necessary in the multi-domain operation, there is some security and confidentiality risk, which can inherit the security requirement defined [RFC5440] and [RFC5376].

10. Acknowledgments

The RFC text was produced using Marshall Rose's xml2rfc tool.

11. References

11.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFC's to Indicate Requirement Levels", RFC 2119, March 1997.
- [RFC4665] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

11.2. Informative References

- [1] Nishioka, Itaru., "End-to-End path routing with PCEs in multi-domain GMPLS networks", July 2008.
- [2] Jabbari, Bijan., "On Constraints for Path Computation in Multi-Layer Switched Networks", August 2007.
- [3] Giorgetti, A. and F. Paolucci, "Routing and Wavelength Assignment in PCE-based Wavelength Switched Optical Networks", September 2008.
- [4] Hayashi, Michiaki., "Advance Reservation-Based Network Resource Manger for Optical Networks", February 2008.

Authors' Addresses

Yongli Zhao
BUPT
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613811761857
Email: yonglizhao@bupt.edu.cn
URI: <http://www.bupt.edu.cn/>

Jie Zhang
BUPT
No.10,Xitucheng Road,Haidian District
Beijing 100876
P.R.China

Phone: +8613911060930
Email: lgr24@bupt.edu.cn
URI: <http://www.bupt.edu.cn/>

Ruiquan Jing
China Telecom Beijing Research Institute
118 Xizhimenwai Avenue
Beijing 100035
P.R.China

Phone: +86-10-58552000
Email: jingrq@ctbri.com.cn
URI: <http://www.ctbri.com.cn/>

Dajiang Wang
ZTE Corporation
No.16, Huayuan Road, Haidian District
Beijing 100191
P.R.China

Phone: +8613811795408
Email: wang.dajiang@zte.com.cn
URI: <http://www.zte.com.cn/>

Xihua Fu
ZTE Corporation
West District, ZTE Plaza, No.10, Tangyan South Road, Gaoxin District
Xi'an 710065
P.R.China

Phone: +8613798412242
Email: fu.xihua@zte.com.cn
URI: <http://www.zte.com.cn/>

