Network Working Group   F. Zhang, Ed.
Internet-Draft   Q. Zhao
Intended status: Standards Track   Huawei
Expires: August 29, 2013   O. Gonzalez de Dios, Ed.
                                                      Telefonica I+D
                                                         R. Casellas
                                                                CTTC
                                                             D. King
                                                  Old Dog Consulting
                                                   February 25, 2013

Extensions to Path Computation Element Communication Protocol (PCEP) for
           Hierarchical Path Computation Elements (PCE)
                draft-zhang-pce-hierarchy-extensions-03

Abstract

   The Hierarchical Path Computation Element (H-PCE) architecture,
   defined in the companion framework document [RFC6805], provides a
   mechanism to allow the optimum sequence of domains to be selected,
   and the optimum end-to-end path to be derived through the use of a
   hierarchical relationship between domains.

   This document defines the Path Computation Element Protocol (PCEP)
   extensions for the purpose of implementing Hierarchical PCE
   procedures which are described in the aforementioned 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 August 29, 2013.

Copyright Notice

   Copyright (c) 2013 IETF Trust and the persons identified as the

Table of Contents

1. Introduction

   [RFC6805] describes a Hierarchical PCE (H-PCE) architecture which can
   be used for computing end-to-end paths for inter-domain MPLS Traffic
   Engineering (TE) and GMPLS Label Switched Paths (LSPs).

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

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

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

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

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

   o The parent PCE sends the intra-domain path computation requests
      (between the domain border nodes) to the child PCEs which are
      responsible for the domains along the domain path;

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

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

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

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

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

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

1.1. Scope

The following functions are out of scope of this document.

o Finding end point addresses;

o Parent Traffic Engineering Dabase (TED) methods;

o Domain connectivity;

The document also uses a number of [editor notes] to describe options and alternative solutions. These options and notes will be removed before publication once agreement is reached.

1.2. Terminology

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

1.3. Requirements Language

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

2. Requirements for H-PCE

This section compiles the set of requirements of the PCEP protocol to support the H-PCE architecture and procedures.

[RFC6805] identifies high-level requirements of PCEP extensions required to support the hierarchical PCE model.

2.1. PCEP Requests

The PCReq messages are used by a PCC or PCE to make a path computation request to a PCE. In order to achieve the full functionality of the H-PCE procedures, the PCReq message needs to include:

o Qualification of PCE Requests.

o Multi-domain Objective Functions (OF).

o Multi-domain Metrics.

2.1.1. Qualification of PCEP Requests

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

   o It MUST be possible for a child PCE to indicate that a 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 a per-domain (PD) [RFC5152] or a backward
      recursive path computation (BRPC) [RFC5441].

   o As stated in [RFC6805], section 4.5, if a PCC knows the egress
      domain, it can supply this information as the path computation
      request. It SHOULD be possible to specify the destination domain
      information in a PCEP request, if it is known.

2.1.2. Multi-domain Objective Functions

   For inter-domain path computation, there are two new objective
   functions which are defined in section 1.3.1 and 4.1 of [RFC6805]:

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

   o Disallow domain re-entry.[Editor's note: Disallow domain re-entry
      may not be an objective function, but an option in the request].

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

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

   For the reasons described previously, new OF codes need to be defined
   for the new inter-domain objective functions. Then the PCE can
   notify its new inter-domain objective functions to the PCC by
   carrying them in the OF-list TLV which is carried in the OPEN object.
   The PCC can specify which objective function code to use, which is
   carried in the OF object when requesting a PCE to compute an inter-
   domain path.

The proposed solution may need to differentiate between the OF code
that is requested at the parent level, and the OF code that is
requested at the intra-domain (child domain).

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

2.1.3. Multi-domain Metrics

For inter-domain path computation, there are several path metrics of
interest [Editor's note: Current framework only mentions metric
objectives. The metric itself should be also defined]:

o Domain count (number of domains crossed).

o Border Node count.

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

2.2. Parent PCE Capability Discovery

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

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

2.3. PCE Domain and PCE ID Discovery

A PCE domain is a single domain with an associated PCE. Although 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.

3. **PCEP Extensions (Encoding)**

3.1. **OPEN object**

3.1.1. **OF Codes**

   There are two new OF codes defined here for H-PCE:

   o MTD

      * Name: Minimize the number of Transit Domains.

      * Objective Function Code: (to be assigned by IANA, recommended
         12).

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

   o MBN

      * Name: Minimize the number of border nodes.

      * Objective Function Code: (to be assigned by IANA, recommended
         13).

      * Description: Find a path P such that it passes through the
         least number of border nodes.

   o DDR

      * Name: Disallow Domain Re-entry (DDR)

      * Objective Function Code: (to be assigned by IANA, recommended
         14)

      * 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 would signal that 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 would signal that the PCE can be used as a parent

PCE by the peer PCE.

3.1.3. Domain-ID TLV

The type of Domain-ID TLV is to be assigned by IANA (recommended 7).
The length is variable. The format of this TLV is defined below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|          Domain Type          |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Domain ID                             |
//                                                             //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: Domain-ID TLV

Domain Type (8 bits): Indicates the domain type. Two types of domain
are currently defined:

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 (variable): Indicates an IGP Area ID or AS number. It can
be 2 bytes, 4 bytes or 8 bytes long depending on the domain
identifier used.

[Editor's note: draft-dhody-pce-pcep-domain-sequence, section 3.2
deals with the encoding of domain sequences, using ERO-subobjects.
Work is ongoing to define domain identifiers for OSPF-TE areas, IS-IS
area (which are variable sized), 2-byte and 4-byte AS number, and any
other domain that may be defined in the future. It uses RSVP-TE
subobject discriminators, rather than new type 1/ type 2. A domain
sequence may be encoded as a route object. The "VALUE" part of the
TLV could follow common RSVP-TE subobject format:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|0|   Type      |   Length      |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        AS Id (4 bytes)                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+


 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|   Type      |   Length      | AS Id (2 bytes)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
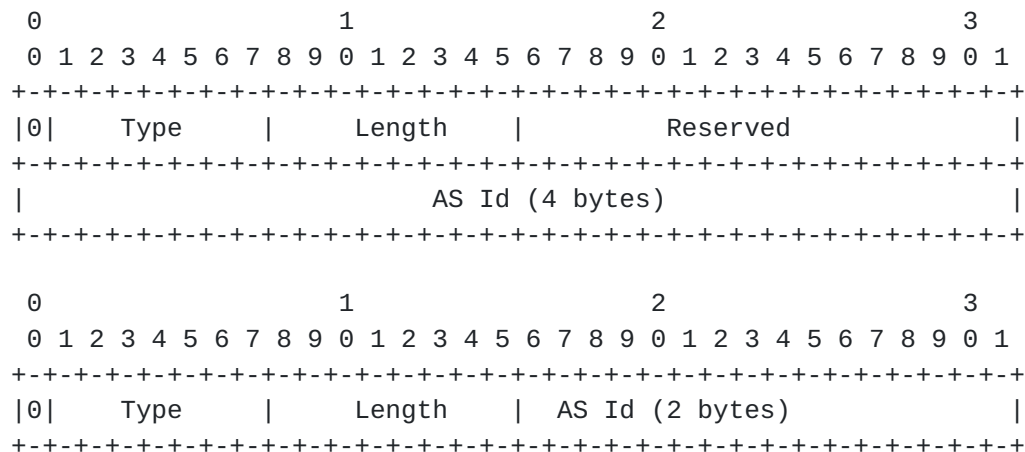
Figure 2: Alternative Domain-ID TLV

3.1.4. PCE-ID TLV

The type of PCE-ID TLV is to be assigned by IANA (recommended 8).
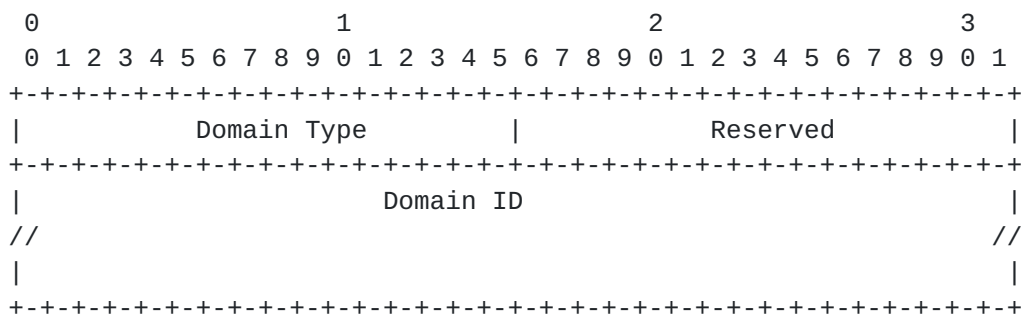The length is variable. The format of this TLV is defined below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Address Type       |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                      PCE IP Address                         //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
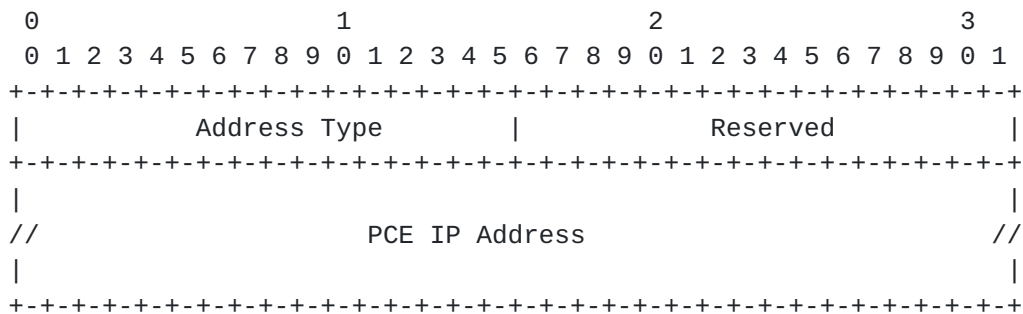
Figure 3: PCE-ID TLV

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.

[Editor's note: [RFC5886] already defines the PCE-ID object. If a
semantically equivalent PCE-ID TLV is needed (to avoid modifying
message grammars to include the object), it can align with the PCEP
object: in any case, the length (4 / 16 bytes) can be used to know
whether it is an IPv4 or an IPv6 PCE, the address type is not
needed.]

3.2. RP object

3.2.1. RP Object Flags

The following flags are defined:

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 3.1.3. 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.]

3.3. Metric Object

There are two new metrics defined in this document for H-PCE:

o Domain count (number of domains crossed).

o Border Node Count (number of border nodes crossed).

3.4. 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 19):

```
+------------+----------------------------------------------------+
| Error-Type | Meaning                                            |
+------------+----------------------------------------------------+
| 19         | H-PCE error Error-value=1: parent PCE capability   |
|            | cannot be provided                                 |
+------------+----------------------------------------------------+
```

H-PCE error table

3.5. NO-PATH Object

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

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

o Bit 23(to be assigned by IANA): When set, the parent PCE indicates
   that destination domain unknown;

o Bit 22(to be assigned by IANA): When set, the parent PCE indicates
   un-responsive child PCE(s);

o Bit 21(to be assigned by IANA): When set, the parent PCE indicates
   no available resource available in one or more domain(s).

4. H-PCE Procedures

4.1. OPEN Procedure between Child PCE and Parent PCE

If a child PCE wants to use the peer PCE as a parent, it 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.

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

   When a specific child PCE sends a PCReq to a peer PCE that requires
   parental activity and 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 (IANA) and error-value (1) in the PCEP-ERROR
   object.

4.2. Procedure to obtain Domain Sequence

   If a child PCE only wants to get the domain sequence for a multi-
   domain path computation from a parent PCE, it can set the Domain Path
   Request bit in the 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 a to-be-defined IGP area
   type). Otherwise it sends a PCReq message which carries the NO-PATH
   object to the child PCE.

5. Error Handling

   A PCE that is capable of acting as a parent PCE might not be
   configured or willing to act as the parent for a specific child PCE.
   This fact could be determined when the child sends a PCReq that
   requires parental activity (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.

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

   o A child PCE cannot find a suitable path to the egress;

   o The parent PCE do not hear from a child PCE for a specified time;

   o The objective functions specified in the path request cannot be
      met.

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

6. Manageability Considerations

   TBD.

7. IANA Considerations

As per [RFC5226], IANA is requested to create/update the following
registries

7.1. Objective Function (OF) codes

Value	Meaning	Reference
12	MBN	This document
13	MTD	This document
14	DDR	This document

7.2. OPEN Object Flags

Bit Number	Meaning	Reference
0	Parent PCE Request	This document
1	Parent PCE Indication	This document

7.3. RP Object Flags

Bit Number	Meaning	Reference
17	Domain Path Request	This document

7.4. PCEP TLVs

Value	Meaning	Reference
x	Interdomain Link TLV	This document (section Section 3.3.2)
x	Interdomain Node TLV	This document (section Section 3.3.3)

7.5. PCEP PCEP-ERROR types

```
          Type      Value                 Meaning
     H-PCE Error 19   1   parent PCE capability cannot be provided
                      2                    TBD
                      3                    TBD
```

7.6. New No-Path Reasons

```
   +------------+-----------------------------------+--------------+
   | Bit Number |               Name                |  Reference   |
   +------------+-----------------------------------+--------------+
   |     23     |     destination domain unknown    | This document |
   |     22     |     un-responsive child PCE(s)    | This document |
   |     21     |    no resource available in some  | This document |
   |            |              domains              |              |
   +------------+-----------------------------------+--------------+
```

8. Security Considerations

   To be added.

9. Contributing Authors

      Xian Zhang
      Huawei
      zhang.xian@huawei.com

10. Acknowledgments

   To be added.

11. 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.

   [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
              Element (PCE)-Based Architecture", RFC 4655, August 2006.

   [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.

   [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an
              IANA Considerations Section in RFCs", BCP 26, RFC 5226,
              May 2008.

   [RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in
              Support of Inter-Autonomous System (AS) MPLS and GMPLS
              Traffic Engineering", RFC 5316, December 2008.

   [RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in
              Support of Inter-Autonomous System (AS) MPLS and GMPLS
              Traffic Engineering", RFC 5392, January 2009.

   [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.

   [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.

   [RFC6805] King, D. and A. Farrel, "The Application of the Path
              Computation Element Architecture to the Determination of a
              Sequence of Domains in MPLS and GMPLS", RFC 6805,
              November 2012.

Authors' Addresses

   Fatai Zhang (editor)
   Huawei
   Huawei Base, Bantian, Longgang District
   Shenzhen,   518129
   China

   Phone: +86-755-28972912
   Email: zhangfatai@huawei.com

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

Phone:
Email: qzhao@huawei.com

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

Phone: +34913128832
Email: ogondio@tid.es

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

Phone: +34 93 645 29 00
Email: ramon.casellas@cttc.es

Daniel King
Old Dog Consulting
UK

Phone:
Email: daniel@olddog.co.uk