                LSP-DB Synchronization between Stateful PCEs
                 draft-palle-pce-stateful-pce-lspdb-sync-03

Abstract

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

   [STATEFUL-PCE] specifies a set of extensions to PCEP to enable
   stateful control of MPLS-TE and GMPLS Label Switched Paths (LSPs) via
   PCEP and maintaining of these LSPs at the stateful PCE. This
   document describes the mechanisms of LSP Database (LSP-DB)
   synchronization between stateful PCEs.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

[RFC5440] describes the Path Computation Element Protocol (PCEP) as
the communication between a Path Computation Client (PCC) and a Path
Computation Element (PCE), or between PCEs, enabling computation of
Multiprotocol Label Switching (MPLS) for Traffic Engineering Label
Switched Paths (TE LSPs).

[STATEFUL-PCE] specifies a set of extensions to PCEP to enable
stateful control of LSPs in compliance with [RFC4655]. It includes
mechanisms for LSP state synchronization between a PCC and a PCE,
i.e., all stateful PCEs synchronize their LSP states from the
network.

When multiple stateful PCEs are operating in the network, they could
be either Primary/Backup or Loadbalanced. In a scenario where the
network operator has deployed backup stateful PCE(s) with only
purpose to be used in the event of failure of the primary PCE, it
makes sense that in such a deployment, PCE should have as latest LSP
states as possible. Further as stateful PCE make changes to the
delegated LSPs, these changes (pending LSPs and sticky resources)
needs to be synchronized to other PCEs as soon as possible.
[PCE-QUESTIONS] describes the synchronization issues when PCEs are
synchronized from the network (PCC) only.

This document specifies the mechanisms of LSP-DB synchronization
between stateful PCEs (in the same domain) to be able to get the
latest LSP state.

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 [RFC2119].

2. Terminology

The terminology is as per [RFC5440] and [STATEFUL-PCE].

LSP-DB: A database of LSPs that are active in the network as
   maintained by a stateful PCE.

Sticky Resources: The temporarily assigned resources that are
   allocated to a pending LSP and are provisionally blocked.

3. Motivation and Use

   Distributed computation model ([RFC4655]) refers to a domain or
   network that may include multiple PCEs where computation of paths is
   shared among the PCEs, this is further clarified in [PCE-QUESTIONS].

   When multiple stateful PCEs are operating in the network, they could
   be either -

   Primary or Backup PCE: A backup PCE exists to perform functions in
      the network, only in the event of a failure of the primary PCE.
      In this case, all LSPs to be delegated are under primary stateful
      PCE control while other PCEs in the domain act as backup. The
      backup PCE should have the same view of LSP-DB as primary stateful
      PCE. The LSP-DB of a backup PCE can be synchronized via the
      primary stateful PCE or collected from multiple network nodes
      (PCC). In case of latter only, the backup PCE may face
      synchronization issues as described in [PCE-QUESTIONS]. Thus it
      is suggested that backup PCE can be synchronized via the primary
      stateful PCE, this mechanism is described in Section 5.1. Note
      that backup PCE MAY use synchronization from network as a
      mechanism to cross-check the LSP-DB.

   Load-Balanced 'Backup' PCE: Load-Balanced PCEs share the computation
      load all the time as well as act backup to each other. One PCE
      MAY serve a set of PCCs as the primary computation server, and
      only addresses requests from other PCCs in the event of the
      failure of some other PCE. Delegated LSPs are thus distributed
      among stateful PCEs. It is suggested that in this case each load-
      balanced stateful PCE should build their LSP-DB independently from
      the network (PCCs) (via mechanism described in [STATEFUL-PCE])
      during initial LSP state synchronization and not from other
      stateful PCEs. But it is important that these load-balanced
      stateful PCEs needs to be synchronized to have a similar view of
      pending LSPs and sticky resources, this mechanism is described in
      Section 5.2.

4. Functions to Support LSP-DB Synchronization

   [STATEFUL-PCE] specifies new functions to support a stateful PCE. It
   also specifies that a function can be initiated either from a PCC
   towards a PCE (C-E) or from a PCE towards a PCC (E-C).

   o Capability negotiation (E-C,C-E)

   o LSP state synchronization (C-E)

   o LSP update request (E-C)

o LSP state report (C-E)

o LSP control delegation (C-E,E-C)

o Stateful PCE discovery via [STATEFUL-PCE-DISC]

This document extends some of these functions to support LSP-DB
synchronization. Some are initiated either from a PCE towards
another PCE (E-E) or specifically from primary to backup PCE (PE-BE).

Capability negotiation (E-E): both the PCEs must announce during
   PCEP session establishment that they support PCEP Stateful PCE
   extensions defined in [STATEFUL-PCE]. It should also declare
   whether it has primary or backup stateful PCE capability. This is
   done via Open message.

LSP state synchronization (PE-BE): after the session between the
   stateful PCEs is initialized, the backup PCE must learn the state
   of LSPs from the primary PCE. This is done via PCRpt message.

LSP update request (E-E): When a PCE requests modification of
   attributes of a delegated LSP, this information should also be
   sent to other PCEs. This is done via PCUpd message. This is
   needed to synchronize the pending LSPs and sticky resources.

Stateful PCE discovery: PCE can advertise its primary or backup
   capability via IGP.

5. Architectural Overview

LSP-DB synchronization function is defined in section 5.4 of
[STATEFUL-PCE] between PCC and PCEs. This document extends the LSP
state synchronization between stateful PCEs.

5.1. LSP-DB Synchronization between Primary and Backup Stateful PCEs

As shown in Figure 1, PCE1 is the primary stateful PCE and PCE2 is
the backup stateful PCE. PCC1 and PCC2 synchronize the LSP-DB with
the primary stateful PCE1 after session initialization phase. And
primary stateful PCE1 synchronizes LSP-DB with its backup stateful
PCE2 after session initialization phase. This is LSP state
synchronization as described in Section 4 and uses PCRpt message.

PCC1 and PCC2 delegates LSP1 and LSP2 to the primary PCE1. Whenever
there is an update in LSP, PCE1 sends a PCUpd message to
corresponding PCC and also to backup PCE2. This is LSP update
request as described in Section 4 and uses PCUpd message. This makes
sure that the pending LSP changes and sticky resources are backed up.

   The PCC sends a PCRpt message to the primary PCE, indicating the
   LSP's status, the primary PCE further synchronizes the state with
   backup PCEs via PCRpt message.

```
   +----+                +----+                +----+                +----+
   |PCC1|                |PCC2|                |PCE1|                |PCE2|
   +-+--+                +-+--+                +-+--+                +-+--+
     |                     |                     |                     |
     |---- LSP SYNC ---+---------------->|                     |
     |                     |---- LSP SYNC --->|                     |
     |                     |                     |---- LSP SYNC----->|
     |                     |---- LSP SYNC ----+----------------->|
     |---- LSP SYNC ---+----------------+----------------->|
     |                     |                     |                     |
     |-- PCRpt,lsp1,D -+---------------->|                     |
     |<---------------+----PCUpd,lsp1 ---|                     |
     |                     |                     |--- PCUpd,lsp1 --->|
     |-- PCRpt,lsp1,up-+---------------->|                     |
     |-- PCRpt,lsp1,up-+----------------+----------------->|
     |                     |                     |----PCRpt,lsp1,up->|
     |                     |                     |                     |
     |                     |-- PCRpt,lsp2,D ->|                     |
     |                     |<---PCUpd,lsp2 ---|                     |
     |                     |                     |--- PCUpd,lsp2---->|
     |                     |-- PCRpt,lsp2,up->|                     |
     |                     |-- PCRpt,lsp2,up--+----------------->|
     |                     |                     |----PCRpt,lsp2,up->|
     |                     |                     |                     |
```
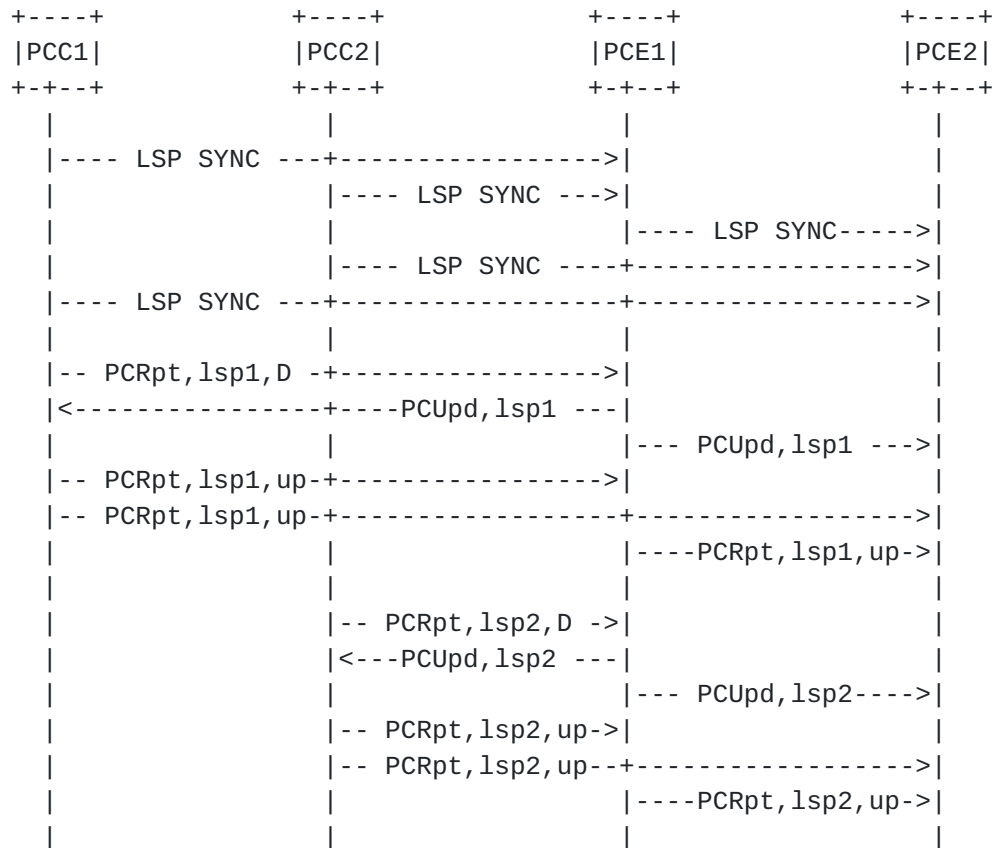
       Figure 1: LSP-DB synchronization between primary and backup stateful
                                     PCEs

   In this case LSP state synchronization is done via primary stateful
   PCE. The backup PCE MAY choose to cross-check the LSP-DB with the
   state learned from the network (PCCs).

   The backup PCE is used only in case the primary PCE fails. At the
   time of failure of primary PCE (PCE1), the backup PCE (PCE2) act as a
   primary. In case of multiple backup PCEs, a selection mechanism
   (e.g. least IP address among backup PCEs) may be used. When PCE1
   recovers from failure, the acting primary PCE (PCE2) should backup
   using the mechanism as described in this section and restart all its
   PCEP sessions, thus making sure all PCEP speakers now considers PCE1
   as primary.

   [STATEFUL-LSPSYNC-OPT] describes LSP state synchronization
   optimizations - PCE triggered synchronization, state synchronization

avoidance, and incremental state synchronization. A Backup PCE
should trigger the state synchronization with primary PCE first in
order to get the full LSP-DB, likewise when the original primary PCE
gets up, it can trigger the state synchronization with the current
stateful PCE first. Further state synchronization avoidance and
incremental state synchronization can optimize the state
synchronization process making sure that the PCEs have the latest
LSP-DB as quickly as possible.

5.2. LSP-DB Synchronization between Load-Balanced Stateful PCEs

As shown in Figure 2, PCE1 and PCE2 are load-balanced stateful PCEs
and share the computation load as well as act as backup to each
other. PCC1 and PCC2 synchronize their LSP-DB with both PCEs after
session initialization phase as per [STATEFUL-PCE]. In this case,
state synchronization does not happen between PCE1 and PCE2 as they
synchronize the LSP-DB with the network (PCCs).

PCC1 delegates LSP1 to PCE1. Whenever there is an update in LSP1,
PCE1 sends the PCUpd message to PCC1 and other stateful PCEs (PCE2).
Similarly, PCC2 delegates LSP2 to PCE2. Whenever there is an update
in LSP2, PCE2 sends the PCUpd message to PCC2 and other stateful PCEs
(PCE1). This is LSP update request as described in Section 4 and it
makes sure that the pending LSP changes and sticky resources are
synchronized. The PCC sends an PCRpt message to the all load-
balanced PCEs as per [STATEFUL-PCE], indicating the LSP's status.

Note that the PCUpd message are exchanged between load-balanced PCEs
for pending LSP changes and sticky resources. And the status of the
LSPs are received from the network (PCC) via PCRpt message as
described in [STATEFUL-PCE] as well as from the PCE.

```
   +----+              +----+              +----+              +----+
   |PCC1|              |PCC2|              |PCE1|              |PCE2|
   +-+--+              +-+--+              +-+--+              +-+--+
     |                   |                   |                   |
     |---- LSP SYNC ---+---------------->|                   |
     |---- LSP SYNC ---+----------------+----------------->|
     |                 |---- LSP SYNC --->|                 |
     |                 |---- LSP SYNC ----+----------------->|
     |                   |                   |                   |
     |-- PCRpt,lsp1,D -+---------------->|                   |
     |<---------------+----PCUpd, lsp1---|                   |
     |                 |                 |--- PCUpd, lsp1--->|
     |-- PCRpt,lsp1,up-+---------------->|                   |
     |-- PCRpt,lsp1,up-+----------------+----------------->|
     |                 |                 |----PCRpt,lsp1,up->|
     |                 |                 |                   |
     |                 |                 |                   |
     |                 |-- PCRpt,lsp2,D --+----------------->|
     |                 |<---PCUpd, lsp2---|------------------|
     |                 |                 |<--- PCUpd, lsp2 --|
     |                 |-- PCRpt,lsp2,up->|                 |
     |                 |-- PCRpt,lsp2,up--+----------------->|
     |                 |                 |<---PCRpt,lsp1,up--|
     |                 |                 |                   |
```
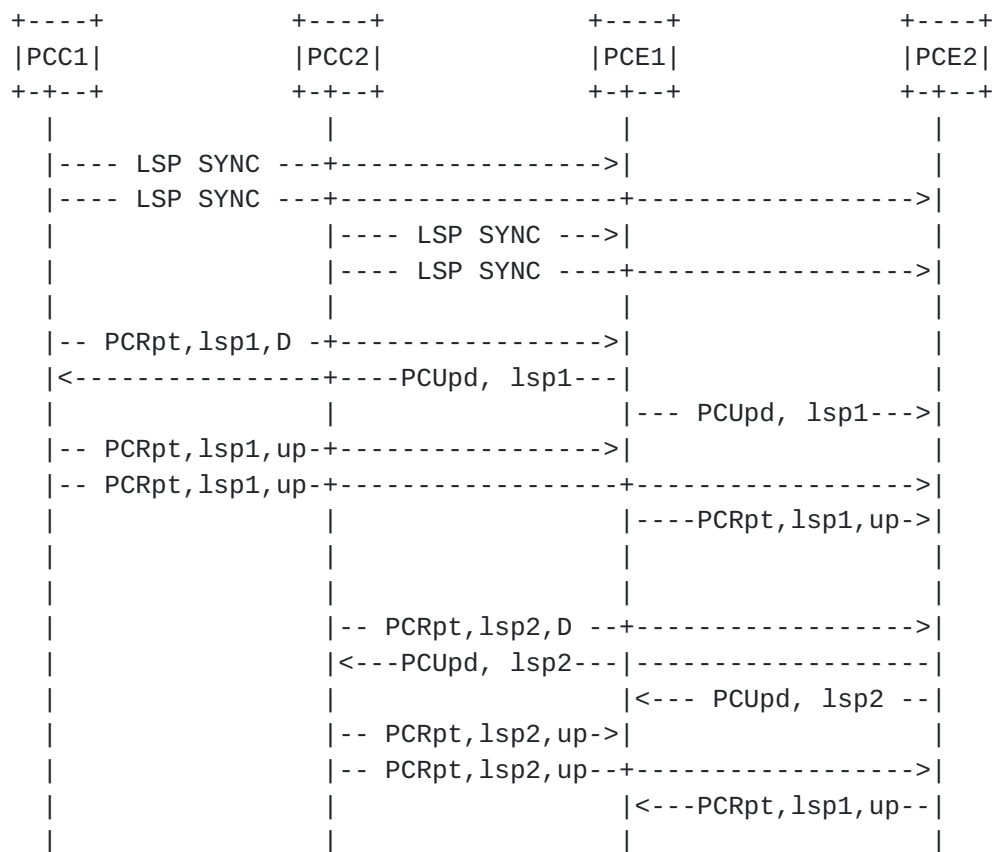
   Figure 2: LSP-DB synchronization between load-balanced stateful PCEs

   At the time of failure of one of the PCEs (say PCE1), the other PCE
   (PCE2) may take up the load. When PCE1 recovers from failure, the
   load can be redistributed again among the PCEs.

5.3. Other Considerations

   o This document does not tackle the issue about TED synchronization
      which is described in detail in [PCE-QUESTIONS].

   o The computation mechanism and how PCE chooses to handle the sticky
      resources during computation is out of scope of this document.

6. PCEP Messages

6.1. The PCRpt Message

   The format of PCRpt message is defined in [STATEFUL-PCE]. It
   specifies the PCRpt message is sent from PCC to PCE in reporting the
   LSP state. This document extends the usage of PCRpt message between

   primary and backup stateful PCEs for LSP synchronization as described
   in Section 5.1.

6.2. The PCUpd Message

   The format of PCUpd Message is defined in [STATEFUL-PCE]. It
   specifies the PCUpd message is sent from PCE to PCC to request
   changes in LSP attributes. This document extends the usage of PCUpd
   message between stateful PCEs for LSP synchronization of pending LSPs
   and sticky resources as described in Section 5.2. Whenever there is
   a PCUpd message sent from PCE to PCC, PCE should also send it to
   other PCEs (backup or load-balanced).

7. TLVs

7.1. Stateful PCE Capability TLV

   As per [STATEFUL-PCE], STATEFUL-PCE-CAPABILITY TLV can be used in the
   OPEN object for stateful PCE capability negotiation. A stateful PCE
   must announce during PCEP session establishment that they support
   PCEP stateful PCE extensions defined in [STATEFUL-PCE]. A new flag
   is added -

   B (BACKUP - 1 bit): if set to 1 by PCE, the PCE should act as a
      backup. It MAY become an 'acting primary PCE' only in case of
      failure or unavailability of primary PCE. In case of PCC, this
      bit has no meaning and is simply ignored.

7.2. PCE Redundancy Group Identifier TLV

   [STATEFUL-PCE] defines a PREDUNDANCY-GROUP-ID TLV which is an unique
   identifier of a PCC and carried in OPEN object, [STATEFUL-PCE] also
   specifies PLSP-ID in LSP object and SYMBOLIC-PATH-NAME TLV which is
   used to identify the originating PCC.

   To uniquely identify LSP across stateful PCEs, PREDUNDANCY-GROUP-ID
   TLV MUST be encoded along with LSP object when PCRpt message is sent
   from primary to backup stateful PCE. This way the backup stateful
   PCE will also learn the unique identifier for the PCC that does not
   change.

   The existing PREDUNDANCY-GROUP-ID TLV MAYBE encoded in LSP object's
   optional TLV to identify the originating PCC.

7.3. PCE-CAP-FLAGS sub-TLV

[RFC5088] and [RFC5089] describe the mechanism to advertise the PCE Discovery information via OSPF and IS-IS respectively along with processing rules for the sub-TLVs. [STATEFUL-PCE-DISC] further enhances the optional PCE-CAP-FLAGS sub-TLV used to advertise PCE stateful capabilities.

Further a new bit is added -

```
     Bit        Capabilities

     TBD        Backup Stateful PCE
```

If this bit is set to 1, the PCE should act as a backup. It MAY become an 'acting primary PCE' only in case of failure or unavailability of primary PCE.

8. Other Considerations

[STATEFUL-PCE-INTERDOMAIN] describes general considerations for the deployment of stateful PCE(s) in inter-domain scenarios including inter-area and inter-AS. It further mentions an alternative approach for state synchronisation of inter-domain LSP to transit and egress domain PCE, where each PCE may synchronise the state with other PCEs in other domain. A mechanism similar to LSP-DB backup described in this document may be utilized for this purpose.

9. Security Considerations

This document does not introduce any new security concerns besides those in [STATEFUL-PCE].

10. Manageability Considerations

10.1. Control of Function and Policy

A PCE may be deployed to act only as a backup (Section 5.1), an operator SHOULD be able to configure a PCE as backup.

10.2. Information and Data Models

[PCEP-MIB] describes the PCEP MIB, there are no new MIB Objects for this document.

10.3. 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 [RFC5440].

10.4. Verify Correct Operations

   Mechanisms defined in this document do not imply any new operation
   verification requirements in addition to those already listed in
   [RFC5440].

10.5. Requirements On Other Protocols

   Mechanisms defined in this document do not imply any new requirements
   on other protocols.

10.6. Impact On Network Operations

   Mechanisms defined in this document do not have any impact on network
   operations in addition to those already listed in [RFC5440].

11. IANA Considerations

11.1. STATEFUL-PCE-CAPABILITY TLV

   As discussed in Section 7.1, a new STATEFUL-PCE-CAPABILITY TLV Flag
   Field has been defined. IANA has made the following allocation from
   the PCEP "STATEFUL-PCE-CAPABILITY TLV Flag Field" sub-registry:

         Bit   Description   Reference

         TBD   BACKUP   [This I.D.]

11.2. PCE-CAP-FLAGS sub-TLV

   As discussed in Section 7.1, a new bit is added, IANA is requested to
   allocate a new bit in "PCE Capability Flags" registry for backup
   stateful PCE capability as follows:

         Bit   Description   Reference

         TBD   BACKUP   [This I.D.]

12. Acknowledgments

Thanks to Adrian Farrel and Daniel King for writing [PCE-QUESTIONS].

We would like to thank Avantika Kumar for her useful comments and suggestions.

13. References

13.1. Normative References

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

13.2. Informative References

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

[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.

[STATEFUL-PCE]
            Crabbe, E., Medved, J., Minei, I., and R. Varga,, "PCEP
            Extensions for Stateful PCE (draft-ietf-pce-stateful-
            pce)", October 2013.

[PCE-QUESTIONS]
            Farrel, A. and D. King, "Unanswered Questions in the Path
            Computation Element Architecture (draft-ietf-pce-
            questions)", October 2013.

[STATEFUL-PCE-DISC]
            Sivabalan, S., Medved, J., and X. Zhang, "IGP Extensions
            for Stateful PCE Discovery (draft-sivabalan-pce-disco-
            stateful-01)", January 2014.

   [STATEFUL-LSPSYNC-OPT]
             Crabbe, E., Medved, J., Minei, I., Varga,, R., Zhang, X.,
             and D. Dhody, "Optimizations of Label Switched Path State
             Synchronization Procedures for a Stateful PCE (draft-
             minei-pce-stateful-sync-optimizations)", December 2013.

   [PCEP-MIB]
             Kiran Koushik, A., Stephan, E., Zhao, Q., King, D., and J.
             Hardwick, "PCE communication protocol(PCEP) Management
             Information Base [draft-ietf-pce-pcep-mib]", January 2014.

   [STATEFUL-PCE-INTERDOMAIN]
             Dhody, D. and X. Zhang, "Stateful Path Computation Element
             (PCE) Inter-domain Considerations", draft-dhody-pce-
             stateful-pce-interdomain-00 (work in progress), July 2014.

Appendix A. Contributor Addresses

      Young Lee
      Huawei
      1700 Alma Drive, Suite 100
      Plano, TX 75075
      US

      Phone: +1 972 509 5599 x2240
      Fax: +1 469 229 5397
      EMail: leeyoung@huawei.com

Authors' Addresses

      Udayasree Palle
      Huawei Technologies
      Leela Palace
      Bangalore, Karnataka 560008
      INDIA

      EMail: udayasree.palle@huawei.com

      Dhruv Dhody
      Huawei Technologies
      Leela Palace
      Bangalore, Karnataka 560008
      INDIA

      EMail: dhruv.ietf@gmail.com

      Xian Zhang
      Huawei Technologies
      Bantian, Longgang District
      Shenzhen 518129
      P.R.China

      EMail: zhang.xian@huawei.com