              Optimizations of Label Switched Path State Synchronization
                          Procedures for a Stateful PCE

Abstract

   A stateful Path Computation Element (PCE) has access to not only the
   information disseminated by the network's Interior Gateway Protocol
   (IGP) but also the set of active paths and their reserved resources
   for its computation. The additional Label Switched Path (LSP) state
   information allows the PCE to compute constrained paths while
   considering individual LSPs and their interactions. This requires a
   State Synchronization mechanism between the PCE and the network, the
   PCE and Path Computation Clients (PCCs), and cooperating PCEs. The
   basic mechanism for State Synchronization is part of the stateful PCE
   specification. This document presents motivations for optimizations
   to the base State Synchronization procedure and specifies the
   required Path Computation Element Communication Protocol (PCEP)
   extensions.

Status of This Memo

   This is an Internet Standards Track document.

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

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   https://www.rfc-editor.org/info/rfc8232.

Table of Contents

1. Introduction

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

   [RFC8231] describes a set of extensions to PCEP to provide stateful
   control. A stateful PCE has access to not only the information
   carried by the network's Interior Gateway Protocol (IGP) but also the
   set of active paths and their reserved resources for its
   computations. The additional state allows the PCE to compute
   constrained paths while considering individual LSPs and their
   interactions. This requires a State Synchronization mechanism
   between the PCE and the network, the PCE and the PCC, and cooperating
   PCEs. [RFC8231] describes the basic mechanism for State
   Synchronization. This document specifies following optimizations for
   State Synchronization and the corresponding PCEP procedures and
   extensions:

   o State Synchronization Avoidance: To skip State Synchronization if
      the state has survived and not changed during session restart.
      (See Section 3.)

   o Incremental State Synchronization: To do incremental (delta) State
      Synchronization when possible. (See Section 4.)

   o PCE-Triggered Initial Synchronization: To let PCE control the
      timing of the initial State Synchronization. (See Section 5.)

   o PCE-Triggered Resynchronization: To let PCE resynchronize the
      state for sanity check. (See Section 6.)

   Support for each of the synchronization optimization capabilities is
   advertised during the PCEP initialization phase. See Section 7 for
   the new flags defined in this document. The handling of each flag is
   described in the relevant section.

1.1. Requirements Language

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and
   "OPTIONAL" in this document are to be interpreted as described in BCP
   14 [RFC2119] [RFC8174] when, and only when, they appear in all
   capitals, as shown here.

2. Terminology

   This document uses the following terms defined in [RFC5440]: PCC,
   PCE, and PCEP Peer.

   This document uses the following terms defined in [RFC8051]: Stateful
   PCE, Delegation, and LSP State Database (LSP-DB).

   This document uses the following terms defined in [RFC8231]:
   Redelegation Timeout Interval, LSP State Report, and LSP Update
   Request.

   Within this document, when describing PCE-PCE communications, the
   requesting PCE fills the role of a PCC as usual.

3. State Synchronization Avoidance

3.1. Motivation

   The purpose of State Synchronization is to provide a
   checkpoint-in-time state replica of a PCC's LSP state in a stateful
   PCE. State Synchronization is performed immediately after the
   initialization phase [RFC5440]. [RFC8231] describes the basic
   mechanism for State Synchronization.

   State Synchronization is not always necessary following a PCEP
   session restart. If the state of both PCEP peers did not change, the
   synchronization phase may be skipped. This can result in significant
   savings in both control-plane data exchanges and the time it takes
   for the stateful PCE to become fully operational.

3.2. State Synchronization Avoidance Procedure

   State Synchronization MAY be skipped following a PCEP session restart
   if the state of both PCEP peers did not change during the period
   prior to session re-initialization. To be able to make this
   determination, state must be exchanged and maintained by both PCE and
   PCC during normal operation. This is accomplished by keeping track
   of the changes to the LSP-DB, using a version tracking field called
   the LSP-DB Version Number.

   The INCLUDE-DB-VERSION (S) bit in the STATEFUL-PCE-CAPABILITY TLV
   (Section 7) is advertised on a PCEP session during session startup to
   indicate that the LSP-DB Version Number is to be included when the
   LSPs are reported to the PCE. The LSP-DB Version Number, carried in
   LSP-DB-VERSION TLV (see Section 3.3.1), is owned by a PCC, and it
   MUST be incremented by 1 for each successive change in the PCC's LSP-
   DB. The LSP-DB Version Number MUST start at 1 and may wrap around.

Values 0 and 0xFFFFFFFFFFFFFFFF are reserved. If either of the two
values are used during LSP State (re)Synchronization, the PCE speaker
receiving this value MUST send back a PCEP Error (PCErr) with Error-
type=20 and Error-value=6 'Received an invalid LSP-DB Version
Number', and close the PCEP session. Operations that trigger a
change to the local LSP-DB include a change in the LSP operational
state, delegation of an LSP, removal or setup of an LSP, or change in
any of the LSP attributes that would trigger a report to the PCE.

If the include LSP-DB version capability is enabled, a PCC MUST
increment its LSP-DB Version Number when the 'Redelegation Timeout
Interval' timer expires (see [RFC8231] for the use of the
Redelegation Timeout Interval).

If both PCEP speakers set the S flag in the OPEN object's
STATEFUL-PCE-CAPABILITY TLV to 1, the PCC MUST include the LSP-DB-
VERSION TLV in each LSP object of the Path Computation LSP State
Report (PCRpt) message. If the LSP-DB-VERSION TLV is missing in a
PCRpt message, the PCE will generate an error with Error-type=6
(Mandatory Object missing) and Error-value=12 'LSP-DB-VERSION TLV
missing', and close the session. If the include LSP-DB version
capability has not been enabled on a PCEP session, the PCC SHOULD NOT
include the LSP-DB-VERSION TLV in the LSP Object, and the PCE MUST
ignore it, were it to receive one.

If a PCE's LSP-DB survived the restart of a PCEP session, the PCE
will include the LSP-DB-VERSION TLV in its OPEN object, and the TLV
will contain the last LSP-DB Version Number received on an LSP State
Report from the PCC in the previous PCEP session. If a PCC's LSP-DB
survived the restart of a PCEP session, the PCC will include the LSP-
DB-VERSION TLV in its OPEN object, and the TLV will contain the
latest LSP-DB Version Number. If a PCEP speaker's LSP-DB did not
survive the restart of a PCEP session or at startup when the database
is empty, the PCEP speaker MUST NOT include the LSP-DB-VERSION TLV in
the OPEN object.

If both PCEP speakers include the LSP-DB-VERSION TLV in the OPEN
object and the TLV values match, the PCC MAY skip State
Synchronization, and the PCE does not wait for the end-of-
synchronization marker [RFC8231]. Otherwise, the PCC MUST perform
full State Synchronization (see [RFC8231]) or incremental State
Synchronization (see Section 4 if this capability is advertised) to
the stateful PCE. In other words, if the incremental State
Synchronization capability is not advertised by the peers, based on
the LSP-DB Version Number match, either the State Synchronization is
skipped or a full State Synchronization is performed. If the PCC
attempts to skip State Synchronization, by setting the SYNC flag to 0
and PLSP-ID to a non-zero value on the first LSP State Report from

the PCC as per [RFC8231], the PCE MUST send back a PCErr with Error-
type=20 and Error-value=2 'LSP-DB version mismatch', and close the
PCEP session.

If State Synchronization is required, then prior to completing the
initialization phase, the PCE MUST mark any LSPs in the LSP-DB that
were previously reported by the PCC as stale. When the PCC reports
an LSP during State Synchronization, if the LSP already exists in the
LSP-DB, the PCE MUST update the LSP-DB and clear the stale marker
from the LSP. When it has finished State Synchronization, the PCC
MUST immediately send an end-of-synchronization marker. The end-of-
synchronization marker is a PCRpt message with an LSP object
containing a PLSP-ID of 0 and with the SYNC flag set to 0 [RFC8231].
The LSP-DB-VERSION TLV MUST be included in this PCRpt message. On
receiving this state report, the PCE MUST purge any LSPs from the
LSP-DB that are still marked as stale.

Note that a PCE/PCC MAY force State Synchronization by not including
the LSP-DB-VERSION TLV in its OPEN object.

Since a PCE does not make changes to the LSP-DB Version Number, a PCC
should never encounter this TLV in a message from the PCE (other than
the OPEN message). A PCC SHOULD ignore the LSP-DB-VERSION TLV, were
it to receive one from a PCE.

Figure 1 shows an example sequence where the State Synchronization is
skipped.

```
           +-+-+                       +-+-+
           |PCC|                       |PCE|
           +-+-+                       +-+-+
             |                           |
             |--Open--,                  |
             |  DBv=42 \      ,---Open--|
             |    S=1    \   /   DBv=42  |
             |            \ /     S=1    |
             |             /\            |
             |            /   '-------->| (OK to skip sync)
   (Skip sync) |<--------'              |
             |             .            |
             |             .            |
             |             .            |
             |                           |
             |--PCRpt,DBv=43,SYNC=0-->| (Regular
             |                           |  LSP State Report)
             |--PCRpt,DBv=44,SYNC=0-->| (Regular
             |                           |  LSP State Report)
             |--PCRpt,DBv=45,SYNC=0-->|
             |                           |
```
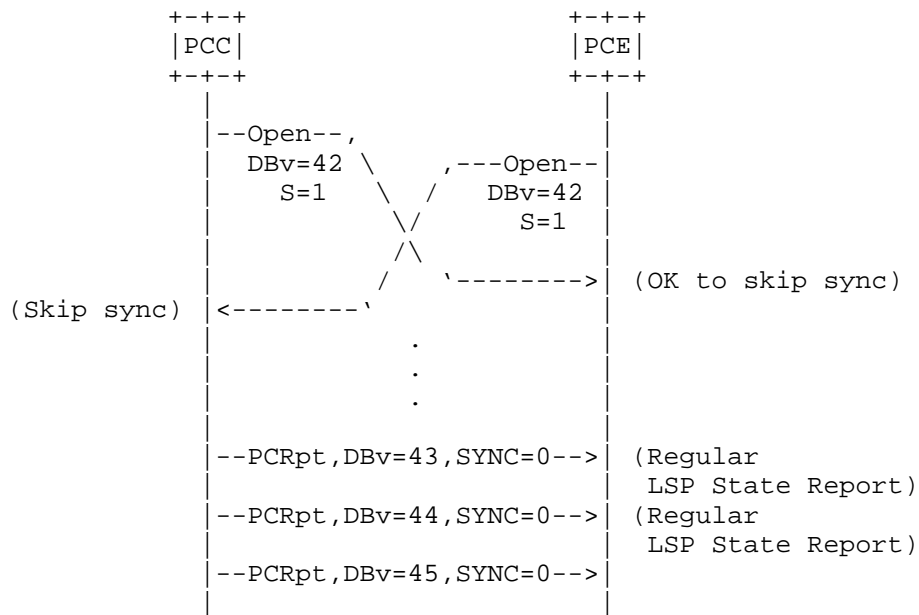
               Figure 1: State Synchronization Skipped

   Figure 2 shows an example sequence where the State Synchronization is
   performed due to LSP-DB version mismatch during the PCEP session
   setup. Note that the same State Synchronization sequence would
   happen if either the PCC or the PCE would not include the LSP-DB-
   VERSION TLV in their respective Open messages.

```
                  +-+-+                        +-+-+
                  |PCC|                        |PCE|
                  +-+-+                        +-+-+
                    |                            |
                    |--Open--,                   |
                    |  DBv=46 \      ,---Open--|
                    |    S=1    \   /    DBv=42  |
                    |            \/       S=1    |
                    |            /\              |
                    |           /   '-------->| (Expect sync)
        (Do sync) |<--------'                 |
                    |                            |
                    |--PCRpt,DBv=46,SYNC=1-->| (Sync start)
                    |            .              |
                    |            .              |
                    |            .              |
                    |--PCRpt,DBv=46,SYNC=0-->| (Sync done)
                    |            .              | (Purge LSP state
                    |            .              |  if applicable)
                    |            .              |
                    |--PCRpt,DBv=47,SYNC=0-->| (Regular
                    |                            |  LSP State Report)
                    |--PCRpt,DBv=48,SYNC=0-->| (Regular
                    |                            |  LSP State Report)
                    |--PCRpt,DBv=49,SYNC=0-->|
                    |                            |
```
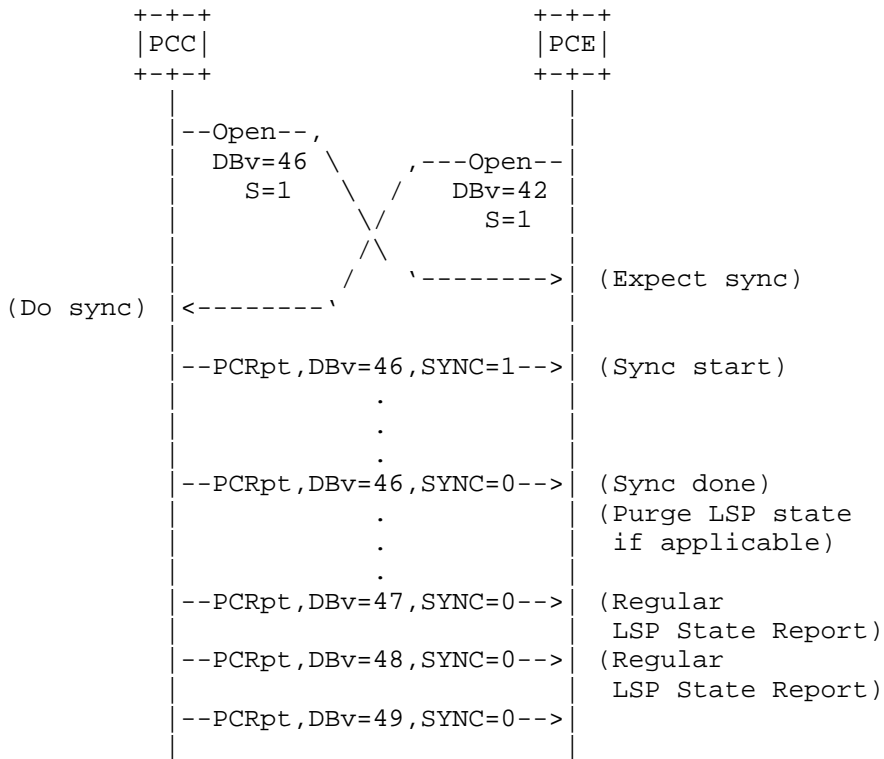
Figure 2: State Synchronization Performed

Figure 3 shows an example sequence where the State Synchronization is
skipped, but because one or both PCEP speakers set the S flag to 0,
the PCC does not send LSP-DB-VERSION TLVs in subsequent PCRpt
messages to the PCE. If the current PCEP session restarts, the PCEP
speakers will have to perform State Synchronization, since the PCE
does not know the PCC's latest LSP-DB Version Number information.

```
                     +-+-+                    +-+-+
                     |PCC|                    |PCE|
                     +-+-+                    +-+-+
                       |                        |
                       |--Open--,               |
                       |  DBv=42 \     ,---Open--|
                       |    S=0   \   /   DBv=42 |
                       |           \ /     S=0   |
                       |           /\            |
                       |          /   '-------->| (OK to skip sync)
         (Skip sync) |<--------'               |
                       |            .            |
                       |            .            |
                       |            .            |
                       |------PCRpt,SYNC=0----->| (Regular
                       |                         |  LSP State Report)
                       |------PCRpt,SYNC=0----->| (Regular
                       |                         |  LSP State Report)
                       |------PCRpt,SYNC=0----->|
                       |                        |
```

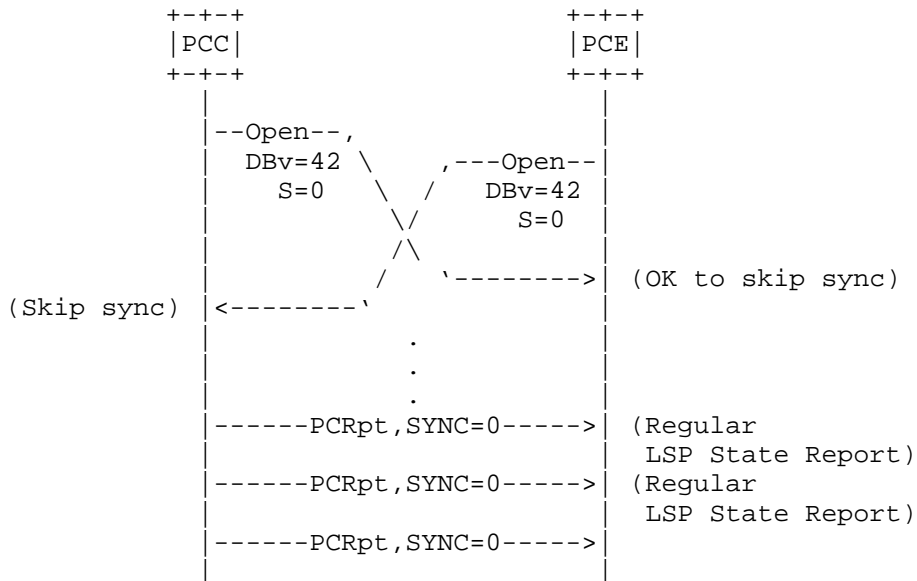   Figure 3: State Synchronization Skipped;
                 No LSP-DB-VERSION TLVs Sent from the PCC

3.2.1. IP Address Change during Session Re-establishment

   There could be a case during PCEP session re-establishment when the
   PCC's or PCE's IP address can change. This includes, but is not
   limited to, the following cases:

   o A PCC could use a physical interface IP address to connect to the
      PCE. In this case, if the line card that the PCC connects from
      changes, then the PCEP session goes down and comes back up again,
      with a different IP address associated with a new line card.

   o The PCC or PCE may move in the network, either physically or
      logically, which may cause its IP address to change. For example,
      the PCE may be deployed as a virtual network function (VNF), and
      another virtualized instance of the PCE may be populated with the
      original PCE instance's state, but it may be given a different IP
      address.

   To ensure that a PCEP peer can recognize a previously connected peer,
   each PCEP peer includes the SPEAKER-ENTITY-ID TLV described in
   Section 3.3.2 in the OPEN message.

This TLV is used during the State Synchronization procedure to
identify the PCEP session as a re-establishment of a previous session
that went down. Then State Synchronization optimizations such as
state sync avoidance can be applied to this session. Note that this
usage is only applicable within the State Timeout Interval [RFC8231].
After the State Timeout Interval expires, all state associated with
the PCEP session is removed, which includes the SPEAKER-ENTITY-ID
received. Note that the PCEP session initialization [RFC5440]
procedure remains unchanged.

3.3. PCEP Extensions

A new INCLUDE-DB-VERSION (S) bit is added in the stateful
capabilities TLV (see Section 7 for details).

3.3.1. LSP-DB Version Number TLV

The LSP-DB Version Number (LSP-DB-VERSION) TLV is an optional TLV
that MAY be included in the OPEN object and the LSP object.

This TLV is included in the LSP object in the PCRpt message to
indicate the LSP-DB version at the PCC. This TLV SHOULD NOT be
included in other PCEP messages (Path Computation Update Request
(PCUpd), Path Computation Request (PCReq), and Path Computation Reply
(PCRep)) and MUST be ignored if received.

The format of the LSP-DB-VERSION TLV is shown in the following
figure:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|            Type=23            |           Length=8            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                    LSP-DB Version Number                      |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
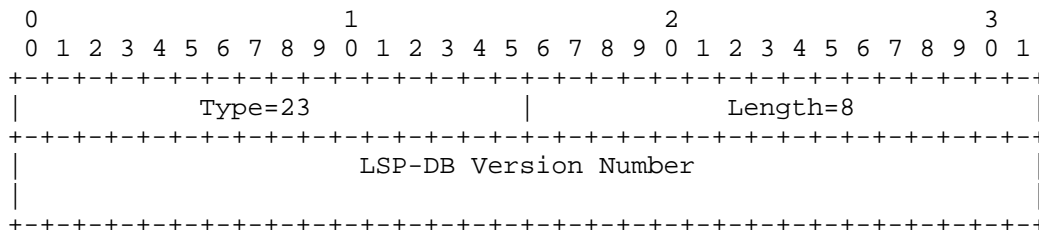
Figure 4: LSP-DB-VERSION TLV Format

The type of the TLV is 23, and it has a fixed length of 8 octets.
The value contains a 64-bit unsigned integer, carried in network byte
order, representing the LSP-DB Version Number.

3.3.2. Speaker Entity Identifier TLV

   The Speaker Entity Identifier TLV (SPEAKER-ENTITY-ID) is an optional
   TLV that MAY be included in the OPEN object when a PCEP speaker
   wishes to determine if State Synchronization can be skipped when a
   PCEP session is restarted. It contains a unique identifier for the
   node that does not change during the lifetime of the PCEP speaker.
   It identifies the PCEP speaker to its peers even if the speaker's IP
   address is changed.

   In case of a remote peer IP address change, a PCEP speaker would
   learn the Speaker Entity Identifier on receiving the open message,
   but it MAY have already sent its open message without realizing that
   it is a known PCEP peer. In such a case, either a full
   synchronization is done or the PCEP session is terminated. This may
   be a local policy decision. The new IP address is associated with
   the Speaker Entity Identifier for the future either way. In the
   latter case when the PCEP session is re-established, it would be
   correctly associated with the Speaker Entity Identifier and not be
   considered as an unknown peer.

   The format of the SPEAKER-ENTITY-ID TLV is shown in the following
   figure:

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |            Type=24            |         Length (variable)     |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                                                               |
   //                 Speaker Entity Identifier                   //
   |                                                               |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
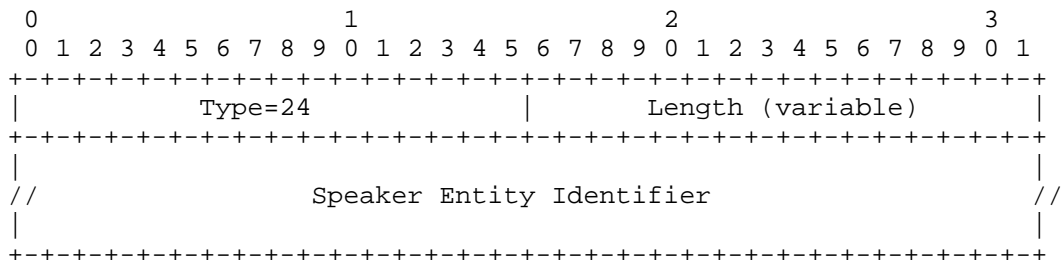
                 Figure 5: SPEAKER-ENTITY-ID TLV Format

   The type of the TLV is 24, and it has a variable length, which MUST
   be greater than 0. The value is padded to a 4-octet alignment. The
   padding is not included in the Length field. The value contains the
   Speaker Entity Identifier (an identifier of the PCEP speaker
   transmitting this TLV). This identifier is required to be unique
   within its scope of visibility, which is usually limited to a single
   domain. It MAY be configured by the operator. Alternatively, it can
   be derived automatically from a suitably stable unique identifier,

   such as a Media Access Control (MAC) address, serial number, Traffic
   Engineering Router ID, or similar. In the case of inter-domain
   connections, the speaker SHOULD prefix its usual identifier with the
   domain identifier of its residence, such as an Autonomous System
   number, an IGP area identifier, or similar to make sure it remains
   unique.

   The relationship between this identifier and entities in the Traffic
   Engineering database is intentionally left undefined.

   From a manageability point of view, a PCE or PCC implementation
   SHOULD allow the operator to configure this Speaker Entity
   Identifier.

   If a PCEP speaker receives the SPEAKER-ENTITY-ID on a new PCEP
   session, that matches with an existing alive PCEP session, the PCEP
   speaker MUST send a PCErr with Error-type=20 and Error-value=7
   'Received an invalid Speaker Entity Identifier', and close the PCEP
   session.

4. Incremental State Synchronization

   [RFC8231] describes the LSP State Synchronization mechanism between
   PCCs and stateful PCEs. During the State Synchronization, a PCC
   sends the information of all its LSPs (i.e., the full LSP-DB) to the
   stateful PCE. In order to reduce the State Synchronization overhead
   when there is a small number of LSP state changes in the network
   between the PCEP session restart, this section defines a mechanism
   for incremental (Delta) LSP-DB synchronization.

4.1. Motivation

   According to [RFC8231], if a PCE restarts and its LSP-DB survived,
   PCCs with a mismatched LSP-DB Version Number will send all their LSPs
   information (full LSP-DB) to the stateful PCE, even if only a small
   number of LSPs underwent state change. It can take a long time and
   consume large communication channel bandwidth.

   Figure 6 shows an example of LSP State Synchronization.

```
                         +-----+
                         | PCE |
                         +-----+
                        /
                       /
                      /
                     /
          +------+            +------+
          | PCC1 |------------| PCC2 |
          +------+            +------+
             |                   |
             |                   |
          +------+            +------+
          | PCC3 |------------| PCC4 |
          +------+            +------+
```
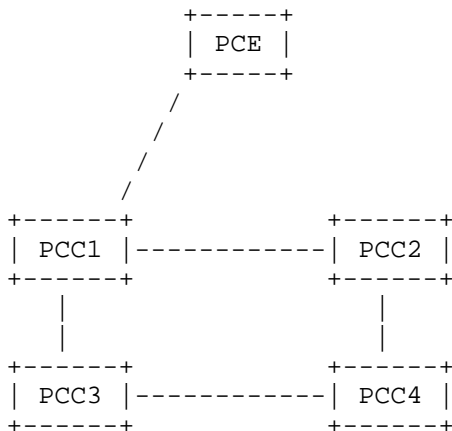
                    Figure 6: Topology Example

   Assume that there are 320 LSPs in the network, with each PCC having
   80 LSPs. During the time when the PCEP session is down, 20 LSPs of
   each PCC (i.e., 80 LSPs in total), are changed. Hence, when the PCEP
   session restarts, the stateful PCE needs to synchronize 320 LSPs with
   all PCCs. But actually, 240 LSPs stay the same. If performing full
   LSP State Synchronization, it can take a long time to carry out the
   synchronization of all LSPs. It is especially true when only a low
   bandwidth communication channel is available (e.g., in-band control
   channel for optical transport networks), and there is a substantial
   number of LSPs in the network. Another disadvantage of full LSP
   synchronization is that it is a waste of communication bandwidth to
   perform full LSP synchronization given the fact that the number of
   LSP changes can be small during the time when the PCEP session is
   down.

   An incremental (Delta) LSP-DB State Synchronization is described in
   this section, where only the LSPs that underwent state change are
   synchronized between the session restart. This may include
   new/modified/deleted LSPs.

4.2. Incremental Synchronization Procedure

   [RFC8231] describes State Synchronization and Section 3 of this
   document describes State Synchronization avoidance by using
   LSP-DB-VERSION TLV in its OPEN object. This section extends this
   idea to only synchronize the delta (changes) in case of version
   mismatch.

If both PCEP speakers include the LSP-DB-VERSION TLV in the OPEN
object and the LSP-DB-VERSION TLV values match, the PCC MAY skip
State Synchronization. Otherwise, the PCC MUST perform State
Synchronization. Incremental State Synchronization capability is
advertised on a PCEP session during session startup using the
DELTA-LSP-SYNC-CAPABILITY (D) bit in the capabilities TLV (see
Section 7). Instead of dumping full LSP-DB to the stateful PCE
again, the PCC synchronizes the delta (changes) as described in
Figure 7 when the D and S flags are set to 1 by both the PCC and PCE.
Other combinations of D and S flags set by the PCC and PCE result in
full LSP-DB synchronization procedures as described in [RFC8231]. By
setting the D flag to zero in the OPEN message, a PCEP speaker can
skip the incremental synchronization optimization, resulting in a
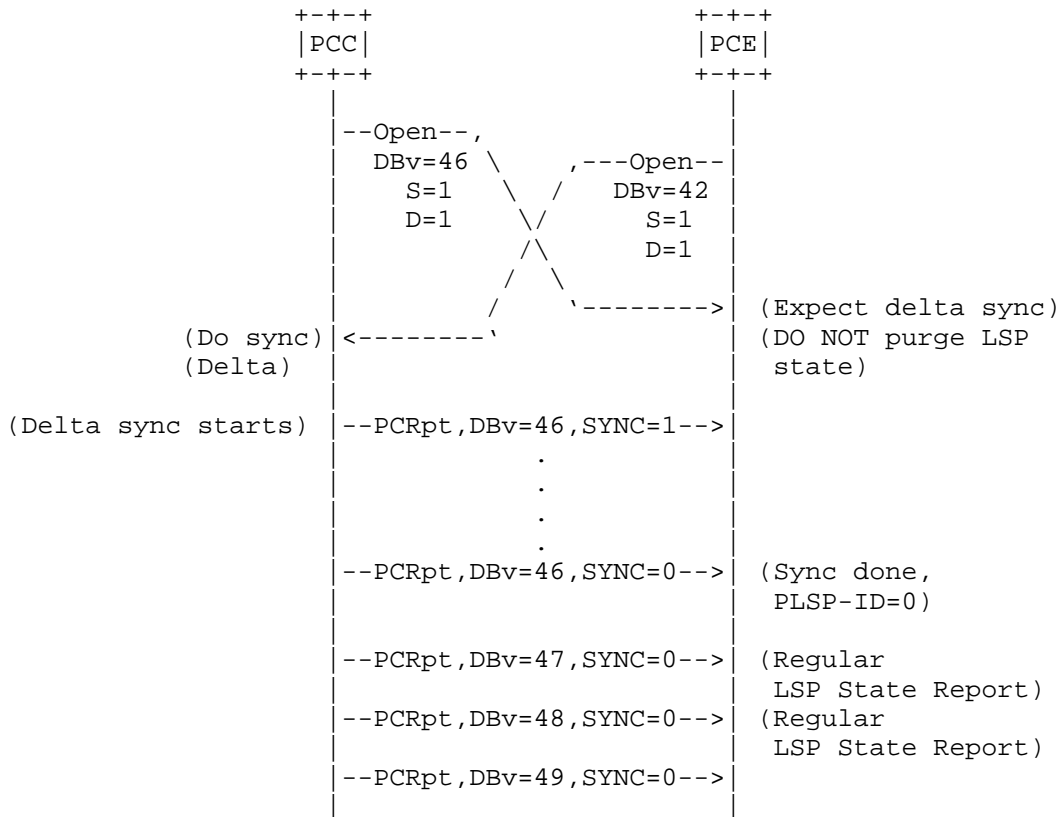full LSP-DB synchronization.

```
                    +-+-+                      +-+-+
                    |PCC|                      |PCE|
                    +-+-+                      +-+-+
                      |                          |
                      |--Open--,                 |
                      |  DBv=46 \      ,---Open--|
                      |    S=1   \    /   DBv=42 |
                      |    D=1    \  /      S=1   |
                      |           \/       D=1   |
                      |           /\             |
                      |          /  \            |
                      |         /    '-------->| (Expect delta sync)
          (Do sync)|<--------'              | (DO NOT purge LSP
            (Delta) |                          |  state)
                      |                          |
  (Delta sync starts) |--PCRpt,DBv=46,SYNC=1-->|
                      |            .             |
                      |            .             |
                      |            .             |
                      |            .             |
                      |--PCRpt,DBv=46,SYNC=0-->| (Sync done,
                      |                          |  PLSP-ID=0)
                      |                          |
                      |--PCRpt,DBv=47,SYNC=0-->| (Regular
                      |                          |  LSP State Report)
                      |--PCRpt,DBv=48,SYNC=0-->| (Regular
                      |                          |  LSP State Report)
                      |--PCRpt,DBv=49,SYNC=0-->|
                      |                          |
```

              Figure 7: Incremental Synchronization Procedure

   As per Section 3, the LSP-DB Version Number is incremented each time
   a change is made to the PCC's local LSP-DB. Each LSP is associated
   with the DB version at the time of its state change. This is needed
   to determine which LSP and what information needs to be synchronized
   in incremental State Synchronization. The incremental state sync is
   done from the last LSP-DB version received by the PCE to the latest
   DB version at the PCC. Note that the LSP-DB Version Number can wrap
   around, in which case the incremental state sync would also wrap till
   the latest LSP-DB Version Number at the PCC.

   In order to carry out incremental State Synchronization, it is not
   necessary for a PCC to store a complete history of LSP-DB change for
   all time, but remember the LSP state changes (including LSP
   modification, setup, and deletion) that the PCE did not get to
   process during the session down. Note that, a PCC would be unaware
   that a particular LSP report has been processed by the PCE before the
   session to the PCE went down. So a PCC implementation MAY choose to

   store the LSP-DB Version Number with each LSP at the time its status
   changed, so that when a session is re-established, an incremental
   synchronization can be attempted based on the PCE's last LSP-DB
   Version Number. For an LSP that is deleted at the PCC, the PCC
   implementation would need to remember the deleted LSP in some way to
   make sure this could be reported as part of incremental
   synchronization later. The PCC would discard this information based
   on a local policy or when it determines that this information is no
   longer needed with sufficient confidence. In the example shown in
   Figure 7, the PCC needs to store the LSP state changes that happened
   between DB Versions 43 to 46 and synchronize these changes, when
   performing incremental LSP state update.

   If a PCC finds out it does not have sufficient information to
   complete incremental synchronization after advertising incremental
   LSP State Synchronization capability, it MUST send a PCErr with
   Error-type=20 and Error-value=5 'A PCC indicates to a PCE that it can
   not complete the State Synchronization' (defined in [RFC8231]), and
   terminate the session. The PCC SHOULD re-establish the session with
   the D bit set to 0 in the OPEN message.

   The other procedures and error checks remain unchanged from the full
   State Synchronization [RFC8231].

5. PCE-Triggered Initial Synchronization

5.1. Motivation

   In networks such as optical transport networks, the control channel
   between network nodes can be realized through in-band overhead, thus
   it has limited bandwidth. With a stateful PCE connected to the
   network via one network node, it is desirable to control the timing
   of PCC State Synchronization so as not to overload the low
   communication channel available in the network during the initial
   synchronization (be it incremental or full) when the session
   restarts, when there is a comparatively large amount of control
   information needing to be synchronized between the stateful PCE and
   the network. The method proposed, i.e., allowing PCE to trigger the
   State Synchronization, is similar to the function proposed in
   Section 6 but is used in different scenarios and for different
   purposes.

5.2. PCE-Triggered Initial State Synchronization Procedure

   Support of PCE-triggered initial State Synchronization is advertised
   during session startup using the TRIGGERED-INITIAL-SYNC (F) bit in
   the STATEFUL-PCE-CAPABILITY TLV (see Section 7).

   In order to allow a stateful PCE to control the LSP-DB
   synchronization after establishing a PCEP session, both PCEP speakers
   MUST set the F bit to 1 in the OPEN message. If the LSP-DB-VERSION
   TLV is included by both PCEP speakers and the TLV value matches, the
   State Synchronization can be skipped as described in Section 3.2. If
   the TLV is not included or the LSP-DB Version is mismatched, the PCE
   can trigger the State Synchronization process by sending a PCUpd
   message with PLSP-ID = 0 and SYNC = 1. The PCUpd message SHOULD
   include an empty Explicit Route Object (ERO) (with no ERO sub-object
   and object length of 4) as its intended path and SHOULD NOT include
   the optional objects for its attributes for any parameter update.
   The PCC MUST ignore such an update when the SYNC flag is set. If the
   TRIGGERED-INITIAL-SYNC capability is not advertised by a PCE and the
   PCC receives a PCUpd with the SYNC flag set to 1, the PCC MUST send a
   PCErr with the SRP-ID-number of the PCUpd, Error-type=20, and
   Error-value=4 'Attempt to trigger a synchronization when the PCE
   triggered synchronization capability has not been advertised' (see
   Section 8.1). If the TRIGGERED-INITIAL-SYNC capability is advertised
   by a PCE and the PCC, the PCC MUST NOT trigger State Synchronization
   on its own. If the PCE receives a PCRpt message before the PCE has
   triggered the State Synchronization, the PCE MUST send a PCErr with
   Error-type=20 and Error-value=3 'Attempt to trigger synchronization
   before PCE trigger' (see Section 8.1).

   In this way, the PCE can control the sequence of LSP synchronization
   among all the PCCs that are re-establishing PCEP sessions with it.
   When the capability of PCE control is enabled, only after a PCC
   receives this message, it will start sending information to the PCE.
   This PCE-triggering capability can be applied to both full and
   incremental State Synchronization. If applied to the latter, the
   PCCs only send information that PCE does not possess, which is
   inferred from the LSP-DB version information exchanged in the OPEN
   message (see Section 4.2 for a detailed procedure).

   Once the initial State Synchronization is triggered by the PCE, the
   procedures and error checks remain unchanged [RFC8231].

   If a PCC implementation that does not implement this extension should
   not receive a PCUpd message to trigger State Synchronization as per
   the capability advertisement, but if it were to receive it, it will
   behave as per [RFC8231].

6. PCE-Triggered Resynchronization

6.1. Motivation

   The accuracy of the computations performed by the PCE is tied to the
   accuracy of the view the PCE has on the state of the LSPs.
   Therefore, it can be beneficial to be able to resynchronize this
   state even after the session has been established. The PCE may use
   this approach to continuously sanity check its state against the
   network or to recover from error conditions without having to tear
   down sessions.

6.2. PCE-Triggered State Resynchronization Procedure

   Support of PCE-triggered state resynchronization is advertised by
   both PCEP speakers during session startup using the TRIGGERED-RESYNC
   (T) bit in the STATEFUL-PCE-CAPABILITY TLV (see Section 7). The PCE
   can choose to resynchronize its entire LSP-DB or a single LSP.

   To trigger resynchronization for an LSP, the PCE sends a Path
   Computation State Update (PCUpd) for the LSP, with the SYNC flag in
   the LSP object set to 1. The PCE SHOULD NOT include any parameter
   updates for the LSP, and the PCC MUST ignore such an update when the
   SYNC flag is set. The PCC MUST respond with a PCRpt message with the
   LSP state, SYNC flag set to 0 and MUST include the SRP-ID-number of
   the PCUpd message that triggered the resynchronization. If the PCC
   cannot find the LSP in its database, PCC MUST also set the R (remove)
   flag [RFC8231] in the LSP object in the PCRpt message.

   The PCE can also trigger resynchronization of the entire LSP-DB. The
   PCE MUST first mark all LSPs in the LSP-DB that were previously
   reported by the PCC as stale, and then send a PCUpd with an LSP
   object containing a PLSP-ID of 0 and with the SYNC flag set to 1.
   The PCUpd message MUST include an empty ERO (with no ERO sub-object
   and object length of 4) as its intended path and SHOULD NOT include
   the optional objects for its attributes for any parameter update.
   The PCC MUST ignore such update if the SYNC flag is set. This PCUpd
   message is the trigger for the PCC to enter the synchronization phase
   as described in [RFC8231] and start sending PCRpt messages. After
   the receipt of the end-of-synchronization marker, the PCE will purge
   LSPs that were not refreshed. The SRP-ID-number of the PCUpd that
   triggered the resynchronization SHOULD be included in each of the
   PCRpt messages. If the PCC cannot resynchronize the entire LSP-DB,
   the PCC MUST respond with a PCErr message with Error-type=20 and
   Error-value=5 'cannot complete the State Synchronization' [RFC8231],
   and it MAY terminate the session. The PCE MUST remove the stale mark
   for the LSPs that were previously reported by the PCC. Based on the
   local policy, the PCE MAY reattempt synchronization at a later time.

If the TRIGGERED-RESYNC capability is not advertised by a PCE and the PCC receives a PCUpd with the SYNC flag set to 1, it MUST send a PCErr with the SRP-ID-number of the PCUpd, Error-type=20, and Error-value=4 'Attempt to trigger a synchronization when the PCE triggered synchronization capability has not been advertised' (see Section 8.1).

Once the state resynchronization is triggered by the PCE, the procedures and error checks remain unchanged from the full state synchronization [RFC8231]. This would also include the PCE triggering multiple state resynchronization requests while synchronization is in progress.

If a PCC implementation that does not implement this extension should not receive a PCUpd message to trigger resynchronization as per the capability advertisement, but if it were to receive it, it will behave as per [RFC8231].

7. Advertising Support of Synchronization Optimizations

   Support for each of the optimizations described in this document requires advertising the corresponding capabilities during session establishment time.

   The STATEFUL-PCE-CAPABILITY TLV is defined in [RFC8231]. This document defines the following new flags in the STATEFUL-PCE-CAPABILITY TLV:

Bit	Description
30	S bit (INCLUDE-DB-VERSION)
27	D bit (DELTA-LSP-SYNC-CAPABILITY)
26	F bit (TRIGGERED-INITIAL-SYNC)
28	T bit (TRIGGERED-RESYNC)

   If the S bit (INCLUDE-DB-VERSION) is set to 1 by both PCEP speakers, the PCC will include the LSP-DB-VERSION TLV in each LSP object. See Section 3.2 for details.

   If the D bit (DELTA-LSP-SYNC-CAPABILITY) is set to 1 by a PCEP speaker, it indicates that the PCEP speaker allows incremental (delta) State Synchronization. See Section 4.2 for details.

   If the F bit (TRIGGERED-INITIAL-SYNC) is set to 1 by both PCEP speakers, the PCE SHOULD trigger initial (first) State Synchronization. See Section 5.2 for details.

If the T bit (TRIGGERED-RESYNC) is set to 1 by both PCEP speakers,
the PCE can trigger resynchronization of LSPs at any point in the
life of the session. See Section 6.2 for details.

See Section 8.3 for IANA allocations.

8. IANA Considerations

IANA has allocated code points for the protocol elements defined in
this document.

8.1. PCEP-Error Object

IANA has allocated the following values in the "PCEP-ERROR Object
Error Types and Values" registry.

```
   Error-Type   Meaning                              Reference
   ------------------------------------------------------------
        6        Mandatory Object missing             [RFC5440]

                 Error-value
                 12: LSP-DB-VERSION TLV missing       This document

       20        LSP State Synchronization Error      [RFC8231]

                 Error-value
                 2: LSP-DB version mismatch.          This document

                 3: Attempt to trigger               This document
                 synchronization before PCE
                 trigger.

                 4: Attempt to trigger a             This document
                 synchronization when the
                 PCE triggered synchronization
                 capability has not been
                 advertised.

                 6: Received an invalid              This document
                 LSP-DB Version Number.

                 7: Received an invalid              This document
                 Speaker Entity Identifier.
```

8.2. PCEP TLV Type Indicators

   IANA has allocated the following values in the "PCEP TLV Type
   Indicators" registry.

         Value   Meaning   Reference
         ------------------------ ---------------- -------------
         23   LSP-DB-VERSION   This document
         24   SPEAKER-ENTITY-ID This document

8.3. STATEFUL-PCE-CAPABILITY TLV

   The STATEFUL-PCE-CAPABILITY TLV is defined in [RFC8231]. The
   "STATEFUL-PCE-CAPABILITY TLV Flag Field" registry has been created to
   manage the flags in the TLV. IANA has allocated the following values
   in this registry.

         Bit   Description   Reference
         ------------------------ ------------------------ -------------
         26   TRIGGERED-INITIAL-SYNC   This document
         27   DELTA-LSP-SYNC-CAPABILITY This document
         28   TRIGGERED-RESYNC   This document
         30   INCLUDE-DB-VERSION   This document

9. Manageability Considerations

   All manageability requirements and considerations listed in [RFC5440]
   and [RFC8231] apply to PCEP protocol extensions defined in this
   document. In addition, requirements and considerations listed in
   this section apply.

9.1. Control of Function and Policy

   A PCE or PCC implementation MUST allow configuring the State
   Synchronization optimization capabilities as described in this
   document. The implementation SHOULD also allow the operator to
   configure the Speaker Entity Identifier (Section 3.3.2). Further,
   the operator SHOULD be to be allowed to trigger the resynchronization
   procedures as per Section 6.2.

9.2. Information and Data Models

   An implementation SHOULD allow the operator to view the stateful
   capabilities advertised by each peer and the current synchronization
   status with each peer. To serve this purpose, the PCEP YANG module
   [PCEP-YANG] can be extended to include advertised stateful
   capabilities and synchronization status.

9.3. 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].

9.4. Verify Correct Operations

   Mechanisms defined in this document do not imply any new operation
   verification requirements in addition to those already listed in
   [RFC5440] and [RFC8231].

9.5. Requirements on Other Protocols

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

9.6. Impact on Network Operations

   Mechanisms defined in [RFC5440] and [RFC8231] also apply to PCEP
   extensions defined in this document.

   The State Synchronization optimizations described in this document
   can result in a reduction of the amount of data exchanged and the
   time taken for a stateful PCE to be fully operational when a PCEP
   session is re-established. The ability to trigger resynchronization
   by the PCE can be utilized by the operator to sanity check its state
   and recover from any mismatch in state without tearing down the
   session.

10. Security Considerations

   The security considerations listed in [RFC8231] apply to this
   document as well. However, this document also introduces some new
   attack vectors. An attacker could spoof the SPEAKER-ENTITY-ID and
   pretend to be another PCEP speaker. An attacker may flood the PCC
   with triggered resynchronization requests at a rate that exceeds the
   PCC's ability to process them by either spoofing messages or
   compromising the PCE itself. The PCC can respond with a PCErr
   message as described in Section 6.2 and terminate the session. Thus,
   securing the PCEP session using Transport Layer Security (TLS)
   [PCEPS], as per the recommendations and best current practices in
   [RFC7525], is RECOMMENDED. An administrator could also expose the
   Speaker Entity Identifier as part of the certificate, for the peer
   identity verification.

11. References

11.1. Normative References

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

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

   [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC
              2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174,
              May 2017, <https://www.rfc-editor.org/info/rfc8174>.

   [RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path
              Computation Element Communication Protocol (PCEP)
              Extensions for Stateful PCE", RFC 8231,
              DOI 10.17487/RFC8231, September 2017,
              <http://www.rfc-editor.org/info/rfc8231>.

11.2. Informative References

   [PCEP-YANG]
              Dhody, D., Hardwick, J., Beeram, V., and j.
              jefftant@gmail.com, "A YANG Data Model for Path
              Computation Element Communications Protocol (PCEP)", Work
              in Progress, draft-ietf-pce-pcep-yang-05, July 2017.

   [PCEPS]   Lopez, D., Dios, O., Wu, Q., and D. Dhody, "Secure
              Transport for PCEP", Work in Progress,
              draft-ietf-pce-pceps-18, September 2017.

   [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre,
              "Recommendations for Secure Use of Transport Layer
              Security (TLS) and Datagram Transport Layer Security
              (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May
              2015, <https://www.rfc-editor.org/info/rfc7525>.

   [RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a
              Stateful Path Computation Element (PCE)", RFC 8051,
              DOI 10.17487/RFC8051, January 2017,
              <https://www.rfc-editor.org/info/rfc8051>.

Acknowledgments

Contributors

   Gang Xie
   Huawei Technologies
   F3-5-B R&D Center, Huawei Industrial Base, Bantian, Longgang District
   Shenzhen, Guangdong, 518129
   China
   Email: xiegang09@huawei.com

Authors' Addresses

   Edward Crabbe
   Oracle
   Email: edward.crabbe@gmail.com

   Ina Minei
   Google, Inc.
   1600 Amphitheatre Parkway
   Mountain View, CA 94043
   United States of America
   Email: inaminei@google.com

   Jan Medved
   Cisco Systems, Inc.
   170 West Tasman Dr.
   San Jose, CA 95134
   United States of America
   Email: jmedved@cisco.com

   Robert Varga
   Pantheon Technologies SRO
   Mlynske Nivy 56
   Bratislava 821 05
   Slovakia
   Email: robert.varga@pantheon.tech

   Xian Zhang
   Huawei Technologies
   F3-5-B R&D Center, Huawei Industrial Base, Bantian, Longgang District
   Shenzhen, Guangdong 518129
   China
   Email: zhang.xian@huawei.com

   Dhruv Dhody
   Huawei Technologies
   Divyashree Techno Park, Whitefield
   Bangalore, Karnataka 560066
   India
   Email: dhruv.ietf@gmail.com