PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: September 14, 2014

E. Crabbe
Google, Inc.
J. Medved
Cisco Systems, Inc.
I. Minei
Google, Inc.
R. Varga
Pantheon Technologies SRO
X. Zhang
D. Dhody
Huawei Technologies

March 13, 2014

Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE
draft-ietf-pce-stateful-sync-optimizations-00

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 reliable state synchronization mechanism between the PCE and the network, PCE and path computation clients (PCCs), and between cooperating PCEs. The basic mechanism for state synchronization is part of the stateful PCE specification. This draft presents motivations for optimizations to the base state synchronization procedure and specifies the required Path Computation Element Communication Protocol (PCEP) extensions.

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].

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 September 14, 2014.

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 Clients (PCCs) requests.

[I-D.ietf-pce-stateful-pce] 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 reliable state synchronization mechanism between the PCE and the network, PCE and PCC, and between cooperating PCEs. [I-D.ietf-pce-stateful-pce] describes the basic mechanism for state synchronization. This draft specifies optimizations for state synchronization and the correspoding PCEP extensions.

2. Terminology

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

This document uses the following terms defined in [I-D.ietf-pce-stateful-pce] : Delegation, Redelegation Timeout Interval, LSP State Report, LSP Update Request, LSP State Database.

Within this document, when describing PCE-PCE communications, the requesting PCE fills the role of a PCC. This provides a saving in documentation without loss of function.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

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]). [I-D.ietf-pce-stateful-pce] 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 state database, using a version tracking field called the LSP State Database Version Number.

The LSP State Database 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 state database. The LSP State Database 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 node should send back a PCErr with Error-type 20 Error-value 6 'Received an invalid LSP DB Version Number', and close the PCEP session. Operations that trigger a change to the local LSP state database 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.

State synchronization avoidance is advertised on a PCEP session during session startup using the INCLUDE-DB-VERSION (IDB) bit in the capabilities TLV (see Section 6). The peer may move in the network, either physically or logically, which may cause its connectivity details and transport-level identity (such as IP address) to change. To ensure that a PCEP peer can recognize a previously connected peer even in face of such mobility, each PCEP peer includes the SPEAKER-ENTITY-ID TLV described in Section 3.3.2 in the OPEN message.

If both PCEP speakers set the IDB 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 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 state synchronization avoidance 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 SHOULD ignore it were to receive one.

If a PCE's LSP state database 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 State Database Version Number received on an LSP State Report from the PCC in the previous PCEP session. If a PCC's LSP State Database 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 State Database Version Number. If a PCEP speaker's LSP state database did not survive the restart of a PCEP session, 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. Otherwise, the PCC MUST perform state synchronization to the stateful PCE. If the PCC attempts to skip state synchronization (i.e., the SYNC Flag = 0 on the first LSP State Report from the PCC as per [I-D.ietf-pce-stateful-pce]), the PCE MUST send back a PCErr with Error-type 20 Error-value 2 'LSP Database 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 database 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 database, the PCE MUST update the LSP database 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 Path Computation State Report (PCRpt) message with an LSP object containing a PLSP-ID of 0 and with the SYNC flag set to 0 ([I-D.ietf-pce-stateful-pce]). 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 database 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 State Database 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.

If state synchronization avoidance is enabled, a PCC MUST increment its LSP State Database Version Number when the 'Redelegation Timeout

Interval' timer expires (see [I-D.ietf-pce-stateful-pce]) for the use of the Redelegation Timeout Interval).

Figure 1 shows an example sequence where the state synchronization is skipped.

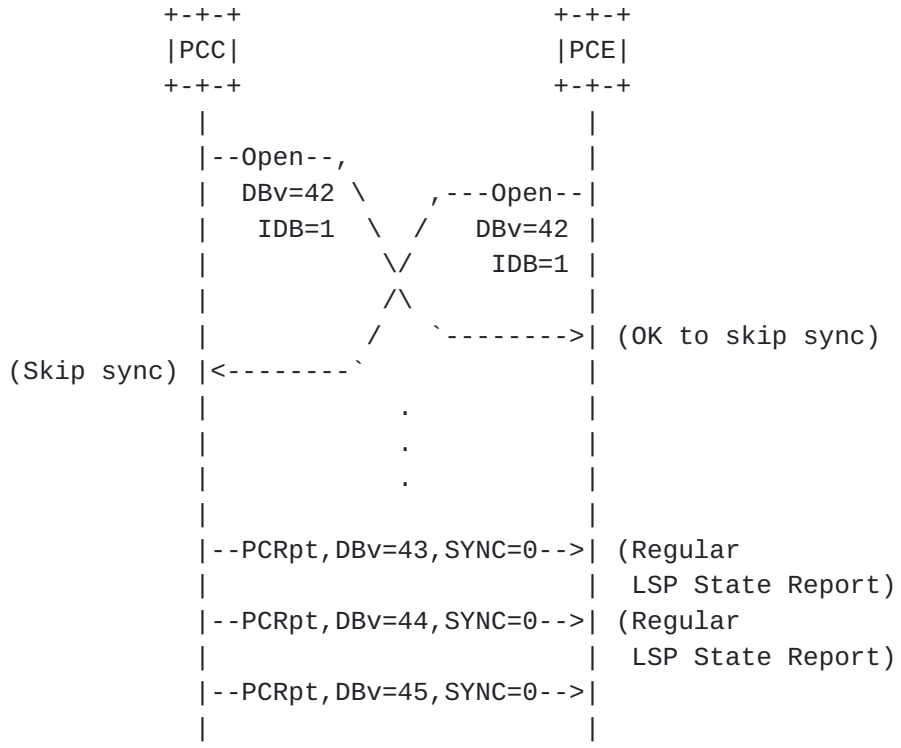

Figure 1: State Synchronization Skipped

Figure 2 shows an example sequence where the state synchronization is performed due to LSP state database 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.

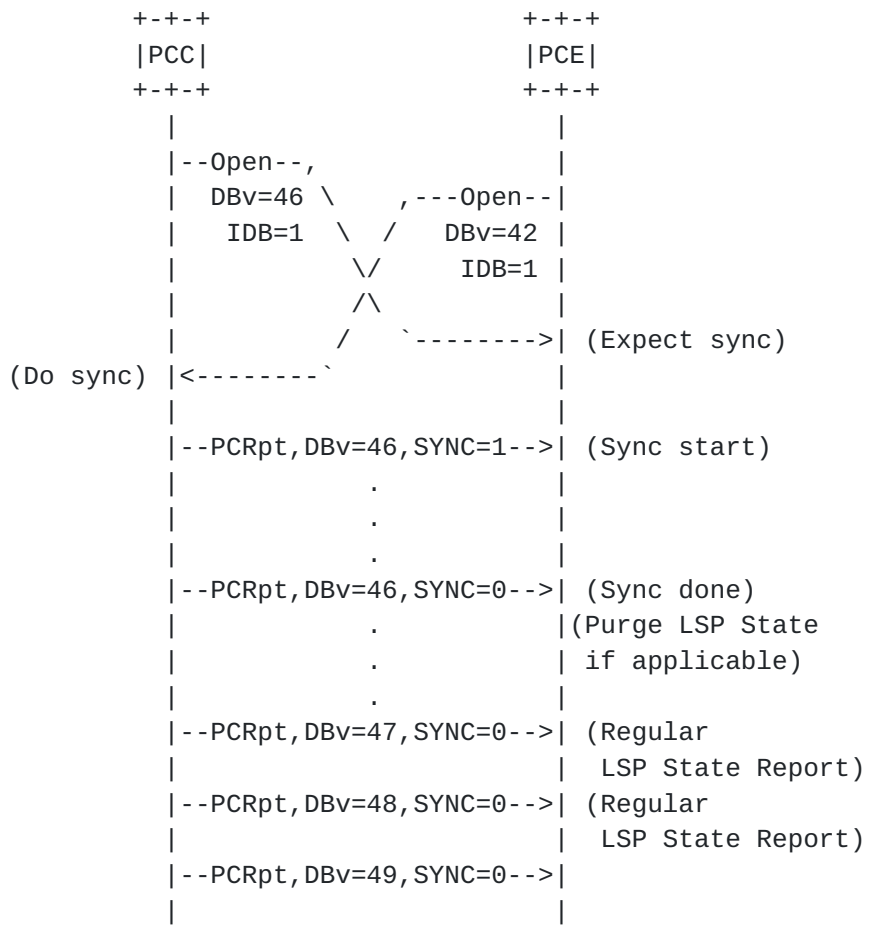

Figure 2: State Synchronization Performed

Figure 3 shows an example sequence where the state synchronization is skipped, but because one or both PCEP speakers set the IDB 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 State Database Version Number information.

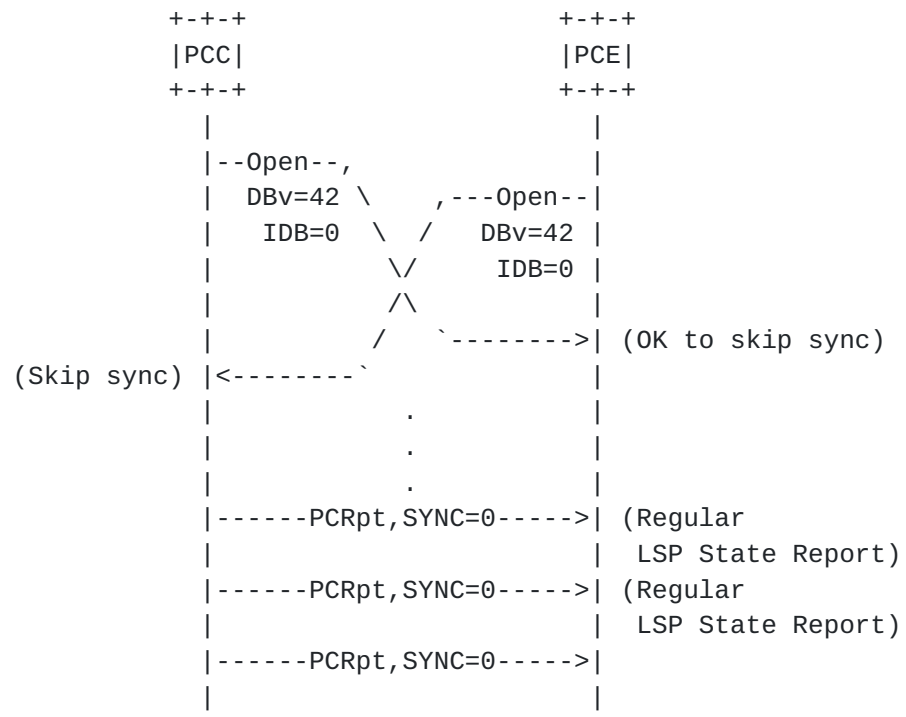

Figure 3: State Synchronization Skipped, no LSP-DB-VERSION TLVs sent from PCC

3.3. PCEP Extensions

3.3.1. LSP State Database Version Number TLV

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

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

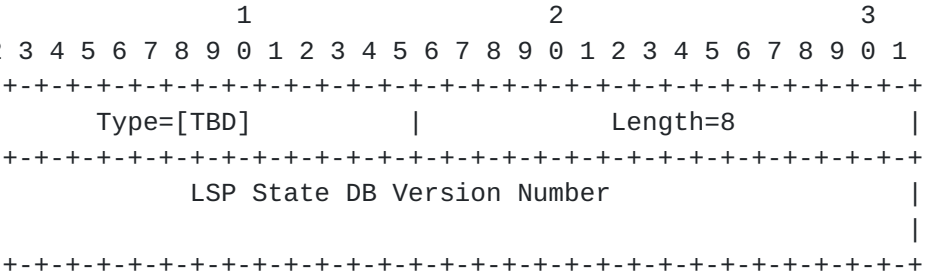

Figure 4: LSP-DB-VERSION TLV format

The type of the TLV is [TBD] and it has a fixed length of 8 octets. The value contains a 64-bit unsigned integer, representing the LSP State 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.

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

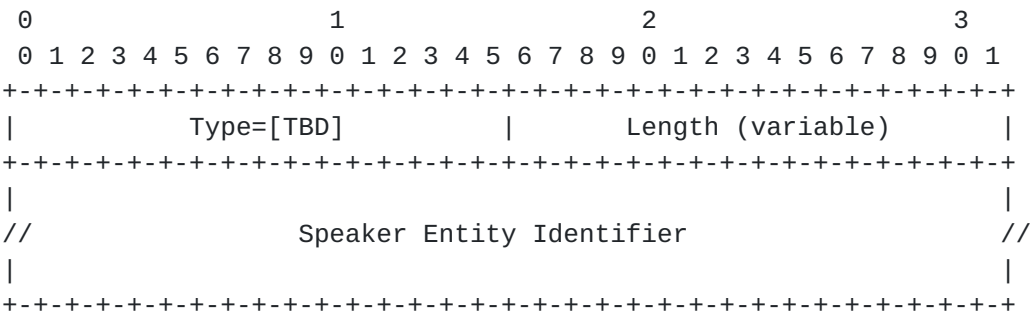

Figure 5: SPEAKER-ENTITY-ID TLV format

The type of the TLV is [TBD] and it has a a variable length, which MUST be greater than 0 and padded to 4-octet alignment (, and padding is not included in the Length field). The value contains the entity identifier of the 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 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 Autonomous System number, IGP area identifier, or similar.

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.

4. PCE-triggered State Synchronization

4.1. 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.

4.2. PCE-triggered State Synchronization Procedure

Support of PCE-triggered state synchronization is advertised during session startup using the TRIGGERED-SYNC (T) bit in the STATEFUL-PCE-CAPABILITY TLV (see Section 6). The PCE can choose to resynchronize its entire LSP database or a single LSP.

To trigger resynchronization for an LSP, the PCE MUST first mark the LSP as stale and then send a Path Computation State Update (PCUpd) for it, 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 SHOULD ignore such updates if the SYNC flag is set. The PCC MUST respond with a PCRpt message and SHOULD include the SRP-ID-number of the PCUpd that triggered the resynchronization.

The PCE can also trigger resynchronization of the entire LSP database. The PCE MUST first mark all LSPs in the LSP database 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. This PCUpd message is the trigger for the PCC to enter the synchronization phase as described in [I-D.ietf-pce-stateful-pce] and start sending PCRpt messages. After the receipt of the end-of-synchronization marker, the PCE will purge LSPs which 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 TRIGGERED-SYNC capability is not advertised 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.(see Section 7.1)

5. Incremental State Synchronization

[I-D.ietf-pce-stateful-pce] describes the LSP state synchronization mechanism between PCCs and stateful PCEs. During the state synchronization, a PCC sends the information of all its LSPs (full

LSP-DB) to the stateful PCE. In order to save the state synchronization overhead when there is a small number of LSP state change in the network between PCEP session restart as well as avoiding overloading a PCE during state (re-)synchronization phase, this section proposes a mechanism for incremental (Delta) LSP Database (LSP-DB) synchronization as well as allowing PCE to control the timing of the LSP-DB synchronization process during incremental syncronization.

5.1. Motivation

According to [I-D.ietf-pce-stateful-pce] , if a PCE restarts and its LSP-DB survived, PCCs with mismatched LSP State Database 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. Moreover, the stateful PCE can get overloaded with all the PCC performing full synchronization with it at the same time.

Figure 6 shows an example of LSP state synchronization.

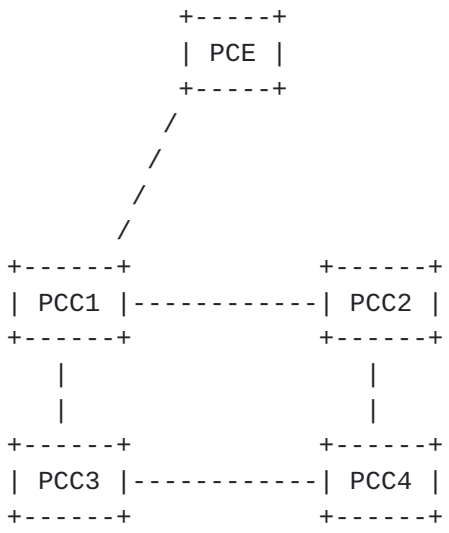

Figure 6: Topology Example

Assuming 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 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 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 PCEP session is down.

An incremental (Delta) LSP Database (LSP-DB) state synchronization is described in this section, where only the LSPs underwent state change are synchronized between the session restart. This may include new/ modified/deleted LSPs. Furthermore, to avoid overloading the PCE, the proposed method enable a stateful PCE to trigger the LSP synchronization (similar to Section 4).

PCEP extensions for stateful PCEs to perform LSP synchronization SHOULD allow:

- Incremental LSP state synchronization between session restarts. Note this does not exclude the need for a stateful PCE to request a full LSP DB synchronization.

- A stateful PCE to control the timing of PCC synchronizing its LSP state with the PCE during incremental synchronisation.

5.2. Incremental Synchronization Procedure

[I-D.ietf-pce-stateful-pce] describes state synchronization and Section 3 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 as well as to allow a stateful PCE to control the timing of this process.

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. Instead of dumping full LSP-DB to PCE again, the PCC synchronizes the delta (changes) as described in Figure 7 when DELTA-LSP-SYNC-CAPABILITY (D flag) is set to 1 by both PCC and PCE (see Section 6). Other combinations of D flag setting by PCC and PCE result in full LSP-DB synchronization procedure as described in [I-D.ietf-pce-stateful-pce]. If a PCC has to force full LSP DB synchronization due to reasons including but not limited: (1) local policy configured at the PCC; (2) no sufficient LSP state caches for incremental update, the PCC can set the D flag to 0. Note a PCC may have to bring down the current session and force a full LSPDB synchronization with D flag set to 0 in the subsequent open message.

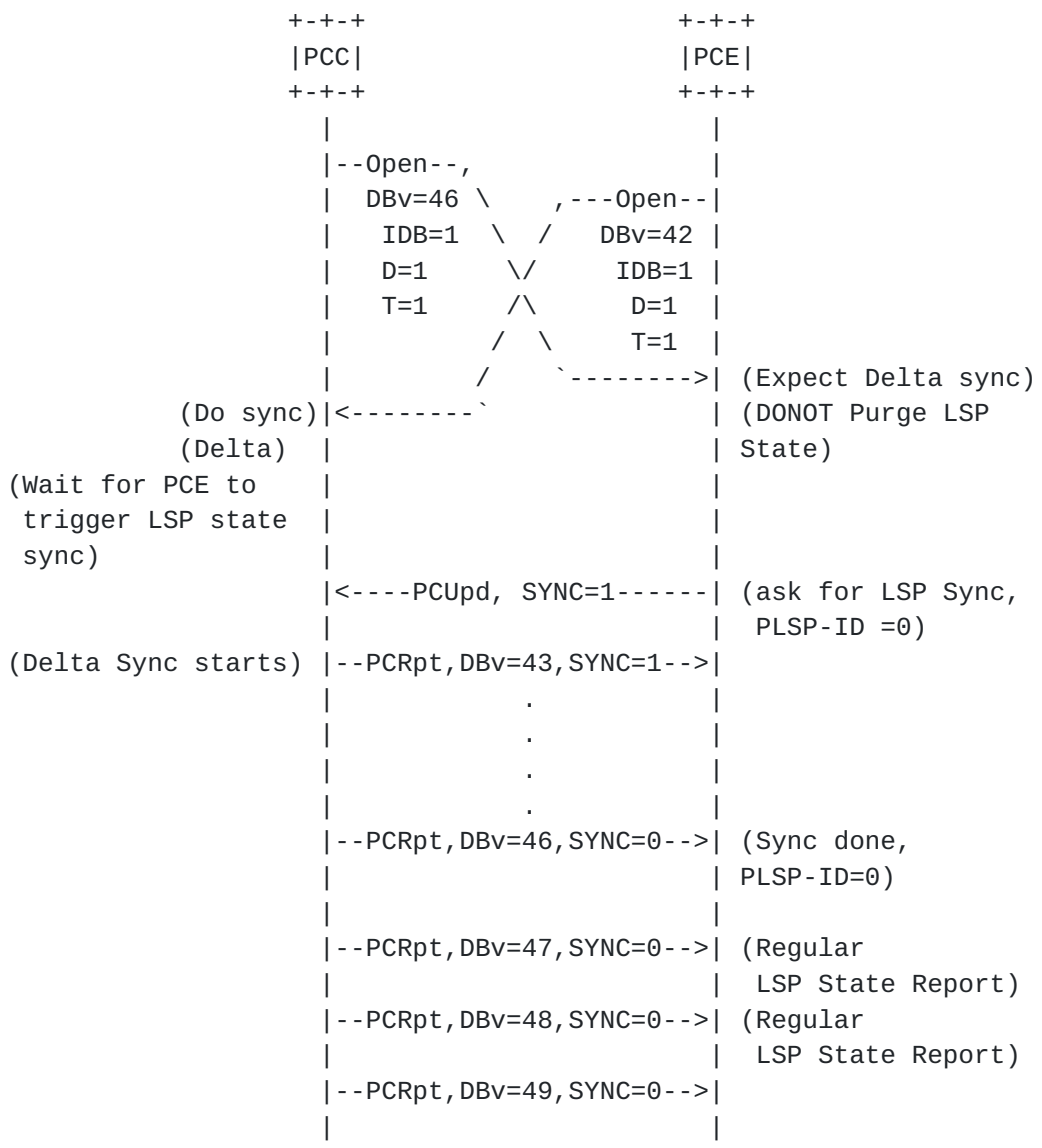

Figure 7: Incremental Synchronization Procedure

A stateful PCE MAY choose to control the LSP-DB synchronization process. To allow PCE to do so, PCEP speakers MUST set T bit to 1 to indicate this (as described in Section 4). If the LSP-DB Version is mis-matched, it can send a PCUpd message with PLSP-ID = 0 and SYNC = 1 in order to trigger the LSP-DB synchronization process. 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 that PCE does not possess, which is inferred from the LSP-DB version information exchanged in the OPEN message. Note that the PCE should not mark the

existing LSPs as stale for incremental state synchronisation procedure.

As per Section 3, the LSP State Database Version Number is incremented each time a change is made to the PCC's local LSP State Database. 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.

It is not necessary for a PCC to store a complete history of LSP Database change, but rather remember the LSP state changes (including LSP modification, setup and deletion) that happend between the PCEP session(s) restart in order to carry out incremental state synchronization. After the synchronization procedure finishes, the PCC can dump this history information. In the example shown in Figure 7, the PCC needs to store the LSP state changes that happend between DB Version 43 to 46 and synchronizes these changes only when performing incremental LSP state update. So a PCC needs to remember the LSP state changes that happened when an existing PCEP session to a stateful PCE goes down in the hope of doing incremental synchronisation when the session is re-established.

If a PCC finds out it does not have sufficient information to complete incremental synchronisation after advertising incremental LSP state synchronization capability, it MUST send a PCErr with error-type 20 and error-value 5(see Section 7.1) and terminate the session.

6. Advertising Support of Synchronization Optimizations

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

New flags are defined for the STATEFUL-PCE-CAPABILITY TLV defined in [I-D.ietf-pce-stateful-pce]. Its format is shown in the following figure:

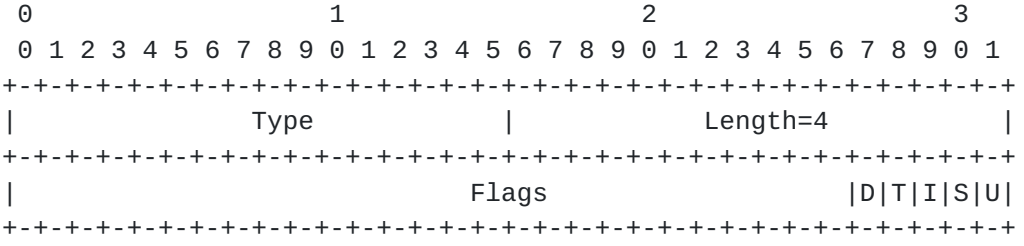

Figure 8: STATEFUL-PCE-CAPABILITY TLV Format

The value comprises a single field - Flags (32 bits):

U (LSP-UPDATE-CAPABILITY - 1 bit): defined in [I-D.ietf-pce-stateful-pce].

S (INCLUDE-DB-VERSION - 1 bit): if set to 1 by both PCEP Speakers, the PCC will include the LSP-DB-VERSION TLV in each LSP Object.

I (LSP-INSTANTIATION-CAPABILITY - 1 bit): defined in [I-D.crabbe-pce-pce-initiated-lsp].

T (TRIGGERED-SYNC - 1 bit): if set to 1 by both PCEP Speakers, the PCE can trigger synchronization of LSPs at any point in the life of the session.

D (DELTA-LSP-SYNC-CAPABILITY - 1 bit): if set to 1 by a PCEP speaker, it indicates that the PCEP speaker allows incremental state synchronization.

7. IANA Considerations

This document requests IANA actions to allocate code points for the protocol elements defined in this document. Values shown here are suggested for use by IANA.

7.1. PCEP-Error Object

This document defines new Error-Value values for the LSP state synchronization error defined in [I-D.ietf-pce-stateful-pce].

Error-Type	Meaning
6	Mandatory Object missing
	Error-value=12: LSP-DB-VERSION TLV missing
20	LSP State synchronization error
	Error-value=2: LSP Database version mismatch.
	Error-value=3: The LSP-DB-VERSION TLV Missing when state synchronization avoidance is enabled.
	Error-value=4: Attempt to trigger a synchronization when the TRIGGERED-SYNC capability has not been advertised.
	Error-value=5: No sufficient LSP change information for incremental LSP state synchronization.

Error-value=6: Received an invalid LSP DB Version Number

7.2. PCEP TLV Type Indicators

This document defines the following new PCEP TLVs:

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

7.3. STATEFUL-PCE-CAPABILITY TLV

The following values are defined in this document for the Flags field in the STATEFUL-PCE-CAPABILITY-TLV in the OPEN object:

Bit	Description	Reference
28	DELTA-LSP-SYNC-CAPABILITY	This document
29	TRIGGERED-SYNC	This document
30	INCLUDE-DB-VERSION	This document

8. Security Considerations

The security considerations listed in [I-D.ietf-pce-stateful-pce] apply to this document as well.

9. Acknowledgements

We would like to thank Young Lee for his contributions.

10. Contributors

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

11. References

11.1. Normative References

[I-D.ietf-pce-stateful-pce]
Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-08 (work in progress), February 2014.

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

[RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

[RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.

11.2. Informative References

[I-D.crabbe-pce-pce-initiated-lsp] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", draft-crabbe-pce-pce-initiated-lsp-03 (work in progress), October 2013.

Authors' Addresses

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: edc@google.com

Jan Medved
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: jmedved@cisco.com

Ina Minei
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: inaminei@google.com

Robert Varga
Pantheon Technologies SRO
Mlynske Nivy 56
Bratislava 821 05
Slovakia

Email: robert.varga@pantheon.sk

Xian Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Industrial Base, Bantian, Longgang District
Shenzhen, Guangdong 518129
P.R.China

Email: zhang.xian@huawei.com

Dhruv Dhody
Huawei Technologies
Leela Palace
Bangalore, Karnataka 560008
INDIA

Email: dhruv.ietf@gmail.com