LABEL-DB Synchronization Procedures for a PCE as a central controller (PCECC)
draft-palle-pce-controller-labeldb-sync-00

Abstract

[I-D.zhao-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller [I-D.zhao-teas-pce-control-function] where LSPs are calculated/setup/initiated and label forwarding entries are downloaded through a centralized PCE server to each network devices along the LSP path while leveraging the existing PCE technologies as much as possible.

Labels downloaded to forwarding entries requires a reliable synchronization mechanism between the path computation clients (PCCs) and the PCECC. This draft specify the label database synchronization mechanism for managing of label database (LABEL-DB) at node (PCC) aligning with LABEL-DB at PCECC on initial session UP or session flap and specifies the required Path Computation Element Communication Protocol (PCEP) extensions.

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 November 16, 2016.
1. Introduction

[I-D.zhao-pce-pcep-extension-for-pce-controller] specify the procedures and PCEP protocol extensions for using the PCE as the central controller [I-D.zhao-teas-pce-control-function] and user cases where LSPs are calculated/setup/initiated/downloaded through extending the existing PCE architectures and PCEP.

Labels downloaded to forwarding entries requires a reliable synchronization mechanism between the path computation clients (PCCs) and the PCECC. This draft specify the PCECC maintenance of label
database per session, and describes the label database (LABEL-DB) synchronization mechanism for managing of label database at node (PCC) aligning with label database at PCECC on initial session UP or session flap and specifies the required Path Computation Element Communication Protocol (PCEP) extensions.

This draft specifies the optimizations for LABEL-DB synchronization and the corresponding PCEP procedures and extensions.

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. LABEL-DB Synchronization

PCECC MUST maintains the LABEL-DB for each PCEP session separately. The purpose of LABEL-DB synchronization is to make sure that the PCECC’s view of LABEL-DB matches with the PCC’s LABEL-DB. The LABEL-DB synchronization MUST be performed from PCECC to PCC immediately after the LSP state synchronization. [I-D.ietf-pce-stateful-pce] describes the basic mechanism for LSP state synchronization. [I-D.ietf-pce-stateful-sync-optimizations] describes the optimizations for LSP state synchronization.

By default a Full LABEL-DB is performed from PCECC to PCC on Initial session UP or every session flap. See Section 3 for detail procedures.

But a Full LABEL-DB synchronization is not always necessary following a PCEP session restart and providing an Optimizations for LABEL-DB synchronization can result in significant savings in both control-plane data exchanges and the time it takes for the PCC to become fully operational.

Optimizations for LABEL-DB synchronization describes the need that both PCEP speakers support label database version capability and maintain label database version for each session. See Section 4 for detail procedures.

3. Full LABEL-DB Synchronization procedure

During Full LABEL-DB Synchronization, a PCECC first takes a snapshot of the label database for the session, then sends this snapshot to the PCC in a sequence of Label Update message (PCLabelUpd message defined in [I-D.zhao-pce-pcep-extension-for-pce-controller]). Each
PCLabelUpd message sent during LABEL-DB Synchronization has the SYNC Flag in the SRP Object (see Section 5.1) set to 1.

The end of synchronization marker is a PCLabelUpd message with the SYNC Flag set to 0 for SRP Object with Label equal to reserved value 0 in the LABEL object ([I-D.zhao-pce-pcep-extension-for-pce-controller]). If the PCECC has no label to synchronize, it will only send the end of synchronization marker.

A PCECC SHOULD NOT send PCUpd messages to a PCC before LABEL-DB Synchronization is complete.

Either the PCECC or the PCC MAY terminate the session using the PCEP session termination procedures during the LABEL-DB synchronization phase. If the session is terminated, the PCC MUST clean up label(s) it received from this PCECC. The session reestablishment MUST be re-attempted as per the procedures defined in [RFC5440], including use of a back-off timer.

The PCC does not send positive acknowledgements for properly received label database synchronization messages. It MUST respond with a PCErr message with Error-type TBD1 (Label Database Synchronization Error) and Error-value 1 (indicating an error in processing the PCLabelUpd) if it encounters a problem with the Label Update it received from the PCECC and it MUST terminate the session.

If the PCECC encounters a problem which prevents it from completing the label transfer, it MUST send a PCErr message with Error-type TBD1 (Label Database Synchronization Error) and Error-value 2 (indicating an internal PCECC Error) to the PCC and terminate the session.

The successful LABEL-DB Synchronization sequence is shown in Figure 1.
The sequence where the PCC fails during the LABEL-DB Synchronization phase is shown in Figure 2.

Figure 1: Successful LABEL-DB synchronization

Figure 2: Failed LABEL-DB synchronization (PCC failure)
The sequence where the PCECC fails during the LABEL-DB Synchronization phase is shown in Figure 3.

```
+++-+-+                         +-+-+
|PCECC|                         |PCC|
+++-+-+                         +-+-+
  |               |               |
  |-----PCLabelUpd, SYNC=1------> (Sync start) |
  |               |               |
  |-----PCLabelUpd, SYNC=1------> |
  | .               |               |
  | .               |               |
  |---------- PCErr=? ---------> |
```

Figure 3: Failed LABEL-DB synchronization (PCECC failure)

4. Optimizations for LABEL-DB Synchronization

This section adds some of the optimization mechanisms for LABEL-DB synchronization. By default, the full LABEL-DB synchronization is performed.

4.1. LABEL-DB Synchronization Avoidance Procedure

The LABEL-DB synchronization MAY be skipped following a PCEP session restart if there is no change in the LABEL-DB of the session at PCECC, during the period prior to session re-initialization. To be able to make this determination, labels must be exchanged and maintained by both PCECC and PCC during normal operation. This is accomplished by keeping track of the changes to the label database, using a version tracking field called the Label Database Version Number.

The Label Database Version Number, carried in LABEL-DB-VERSION TLV (see Section 5.3), is owned by a PCECC and it MUST be incremented by 1 for each successive change in the PCECC’s label database. The Label 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 LABEL-DB synchronization, the PCC speaker receiving this node should send back a PCErr with Error-type TBD1 Error-value 3 ‘Received an invalid Label Database Version Number’, and close the PCEP session. Operations that trigger a change to the Label database include an addition or deletion of labels that would trigger a label update to the PCC.
LABEL-DB synchronization avoidance is advertised on a PCEP session during session startup using the INCLUDE-LABEL-DB-VERSION (I) bit in the PCECC capability TLV (see Section 5.2). The PCEP peer MAY include the SPEAKER-ENTITY-ID TLV described in [I-D.ietf-pce-stateful-sync-optimizations] in the OPEN message to identify the peer in case of IP address change.

If both PCEP speakers set the I flag in the OPEN object’s PCECC Capability TLV to 1, the PCECC MUST include the LABEL-DB-VERSION TLV in each LABEL object of the PCLabelUpd message. If the LABEL-DB-VERSION TLV is missing in a PCLabelUpd message, the PCC will generate an error with Error-Type 6 (mandatory object missing) and Error-Value TBD2 ‘LABEL-DB-VERSION TLV missing’ and close the session. If LABEL-DB synchronization avoidance has not been enabled on a PCEP session, the PCECC SHOULD NOT include the LABEL-DB-VERSION TLV in the LABEL Object and the PCC SHOULD ignore it were it to receive one.

If a PCC’s label database survived the restart of a PCEP session, the PCC will include the LABEL-DB-VERSION TLV in its OPEN object, and the TLV will contain the last Label Database Version Number received on an Label Update from the PCECC in the previous PCEP session. If a PCECC’s Label Database survived the restart of a PCEP session, the PCECC will include the LABEL-DB-VERSION TLV in its OPEN object and the TLV will contain the latest Label Database Version Number. If a PCEP speaker’s label database did not survive the restart of a PCEP session, the PCEP speaker MUST NOT include the LABEL-DB-VERSION TLV in the OPEN object.

If both PCEP speakers include the LABEL-DB-VERSION TLV in the OPEN Object and the TLV values match, the PCECC MAY skip LABEL-DB synchronization. Otherwise, the PCECC MUST perform full LABEL-DB synchronization (see Section 3) or incremental LABEL-DB synchronization (see Section 4.2) to the PCC. Incase, the PCECC attempts to skip LABEL-DB synchronization, by setting the SYNC Flag to 0 on the first Label Update from the PCECC, the PCC MUST send back a PCErr with Error-type TBD1 (Label Database Synchronization Error) and Error-value 4(Label Database Version mismatch), and close the PCEP session.

If LABEL-DB synchronization is required, then prior to completing the initialization phase, the PCC MUST mark any labels in the label database that were previously updated by the PCECC as stale. When the PCECC updates a label during LABEL-DB synchronization, if the label already exists in the label database, the PCC MUST update the label database and clear the stale marker from the label. When it has finished LABEL-DB synchronization, the PCECC MUST immediately send an end of synchronization marker. The end of synchronization marker is a Path Computation Label Update (PCLabelUpd) message with a
SRP object containing the SYNC flag set to 0 (see Section 5.1) and Label as 0 in the LABEL object. The LABEL-DB-VERSION TLV MUST be included in this PCLabelUpd message. On receiving this Label Update, the PCC MUST purge any labels from the label database that are still marked as stale.

Note that a PCECC/PCC MAY force LABEL-DB synchronization by not including the LABEL-DB-VERSION TLV in its OPEN object.

Figure 4 shows an example sequence where the LABEL-DB synchronization is skipped.

```
+-----+    +-----+
|PCECC|    |PCC|
+-----+    +-----+

,----Open---
|       |
| DBv=35 |
|--Open--, |
| I=1   |
| DBv=35 |
| I=1   |
|       |

(Skip sync)

<--------' (OK to skip sync)

--PCLabelUpd, DBv=36, SYNC=0--> (Regular Label Update)
--PCLabelUpd, DBv=37, SYNC=0--> (Regular Label Update)
--PCLabelUpd, DBv=38, SYNC=0--> |

Figure 4: LABEL-DB synchronization Skipped
```

Figure 5 shows an example sequence where the LABEL-DB synchronization is performed due to label database version mismatch during the PCEP session setup. Note that the same LABEL-DB synchronization sequence would happen if either the PCC or the PCECC would not include the LABEL-DB-VERSION TLV in their respective Open messages.
Figure 5: LABEL-DB synchronization Performed

Figure 6 shows an example sequence where the LABEL-DB synchronization is skipped, but because one or both PCEP speakers set the I Flag to 0, the PCECC does not send LABEL-DB-VERSION TLVs in subsequent PCLabelUpd messages to the PCC. If the current PCEP session restarts, the PCEP speakers will have to perform full LABEL-DB synchronization, since the PCC does not know the PCECC’s latest Label Database Version Number information.
4.2. Incremental LABEL-DB Synchronization Procedure

If a PCC restarts and its label database survived, PCECC with mismatched Label Database Version Number will send all their Labels information (full LABEL-DB) to the PCC, even if only a small number of changes happened. It can take a long time and consume large communication channel bandwidth.

This section extends the idea to only synchronize the delta (changes) in case of Label Database Version Number of both PCEP peers is non-zero and mismatch.

If both PCEP speakers include the LABEL-DB-VERSION TLV in the OPEN object and the LABEL-DB-VERSION TLV values match, the PCECC MAY skip LABEL-DB synchronization. Otherwise, the PCECC MUST perform LABEL-DB synchronization. Incremental label database synchronization capability is advertised on a PCEP session during session startup using the DELTA-LABEL-SYNC-CAPABILITY (D) bit in the capabilities TLV (see Section 5.2). Instead of dumping full LABEL-DB to the PCC again, the PCECC synchronizes the delta (changes) as described in Figure 7 when D flag and I flag is set to 1 by both PCC and PCECC. Other combinations of D and I flags setting by PCC and PCECC result in full LABEL-DB synchronization procedure as described in Section 3.
The PCECC MAY force a full LABEL-DB synchronization by setting the D flag to zero in the OPEN message.

Figure 7: Incremental Synchronization Procedure

As per Section 4.1, the Label Database Version Number is incremented each time a change is made to the PCECC’s label database. Each label is associated with the DB version at the time of its addition. This is needed to determine which label and what information needs to be synchronized in incremental LABEL-DB synchronization.

It is not necessary for a PCECC to store a complete history of label database change, but rather remember the labels (including label addition and deletion) that happened between the PCEP session(s) restart in order to carry out incremental LABEL-DB synchronization. After the synchronization procedure finishes, the PCECC can dump this history information. In the example shown in Figure 7, the PCECC needs to store the label changes that happened between DB Version 35...
to 39 and synchronizes these changes only when performing incremental label update. So a PCECC needs to remember at least the label changes that happened after an existing PCEP session with a PCC goes down to have any chance of doing incremental synchronization when the session is re-established.

If a PCECC finds out it does not have sufficient information to complete incremental synchronization after advertising incremental LABEL-DB synchronization capability, it MUST send a PCErr with Error-Type TBD1 and Error-Value 5 ’A PCECC indicates to a PCC that it can not complete the LABEL-DB synchronization’ and terminate the session. The PCECC 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 LABEL-DB synchronization defined in Section 3.

5. PCEP Extensions

5.1. Extension of SRP object

SRP object is defined in [I-D.ietf-pce-stateful-pce] and extended in [I-D.ietf-pce-pce-initiated-lsp]. This draft defines a new ’SYNC’ flag (S bit) to specify the LABEL-DB synchronization operation.

The format of the SRP object is shown Figure 8:

```
+-----------------+-----------------+-----------------+-----------------+
| Flags           | S | R | SR-ID-number |
+-----------------+-----------------+-----------------+-----------------+
| Optional TLVs   |
+-----------------+-----------------+-----------------+-----------------+
```

Figure 8: SRP Object format

S (SYNC - 1 bit): The S Flag MUST be set to 1 on each PCLabelUpd sent from a PCECC during LABEL-DB Synchronization. The S Flag MUST be set to 0 in other messages sent from the PCECC.
5.2. Extension of PCECC Capability TLV

PCECC Capability TLV is defined in [I-D.zhao-pce-pcep-extension-for-pce-controller]. This draft defines a new ‘INCLUDE-LABEL-DB-VERSION’ flag (I bit) to specify the label database version capability and ‘DELTA-LABEL-SYNC-CAPABILITY’ to specify the incremental label database synchronization capability.

The format of the PCECC Capability TLV is shown Figure 9:

```
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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Type            |            Length=4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                             Flags                       |D|I|S|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

**Figure 9: PCECC Capability TLV**

I (INCLUDE-LABEL-DB-VERSION – 1 bit): if set to 1 by both PCEP Speakers, the PCECC will include the LABEL-DB-VERSION TLV in each LABEL Object.

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

5.3. New LABEL-DB-VERSION TLV

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

The format of the LABEL-DB-VERSION TLV is shown in the following figure:
The type of the TLV is [TBD3] and it has a fixed length of 8 octets. The value contains a 64-bit unsigned integer, representing the Label Database Version Number.

6. Manageability Considerations

TBD

7. Security Considerations

TBD

8. IANA Considerations

TBD

9. Acknowledgements

This document borrows some of the structure and text from [I-D.ietf-pce-stateful-sync-optimizations], and would like to thanks the authors and contributors of the document.

10. References

10.1. Normative References


[I-D.ietf-pce-stateful-pce]

[I-D.zhao-pce-pcep-extension-for-pce-controller]

10.2. Informative References

[I-D.zhao-teas-pce-control-function]

[I-D.ietf-pce-pce-initiated-lsp]

[I-D.ietf-pce-stateful-sync-optimizations]

Authors’ Addresses

Udayasree Palle
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka  560066
India

EMail: udayasree.palle@huawei.com