PCE Working Group D. Dhody
Internet-Draft Huawei Technologies
Intended status: Standards Track December 28, 2016
Expires: July 1, 2017

Realtime Synchronization between Redundant Stateful PCEs.
draft-dhody-pce-stateful-pce-lspdb-realtime-sync-01

Abstract

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

The stateful PCE further extentds PCEP to enable stateful control of MPLS-TE and GMPLS Label Switched Paths (LSPs) via PCEP and maintaining of these LSPs at the stateful PCE. This document describes the mechanisms of realtime LSP Database (LSP-DB) synchronization between stateful PCEs.

Table of Contents

1. Introduction

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

[I-D.ietf-pce-stateful-pce] specifies a set of extensions to PCEP to enable stateful control of LSPs in compliance with [RFC4655]. It includes mechanisms for LSP state synchronization between a PCC and a PCE, i.e., all stateful PCEs synchronize their LSP states from the network. It further describe the handling of redundant stateful PCEs, where all PCEs receive the state from the network (PCCs). When the primary PCE fails, another PCE can take over.

Apart from the synchronization from the network, it is also useful if there is realtime synchronization mechanism between the stateful PCEs. As stateful PCE make changes to its delegated LSPs, these changes (pending LSPs and the sticky resources) can be synchronized immediately to the other PCEs. Further PCE may also synchronize any status change of its delegated LSPs to other PCEs. Note that some synchronization issues are identified in [RFC7399].

It should be noted that in some deployments the PCE function is part of the central controller architecture with multiple instances of PCE for load balancing and backup which uses proprietary mechanics to maintain consistent state between these PCE instance. In such deployment PCEP MAY not used as a database synchronization mechanism.

This document specifies the mechanisms of realtime LSP-DB synchronization between redundant stateful PCEs via PCEP.

1.1. Requirements Language

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

2. Terminology

The terminology is as per [RFC5440] and [I-D.ietf-pce-stateful-pce].

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

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

3. Architectural Considerations

Distributed computation model ([RFC4655]) refers to a domain or network that may include multiple PCEs where computation of paths is shared among the PCEs, this is further clarified in [RFC7399].

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

Primary or Backup PCE: A backup PCE exists to perform functions in the network, only in the event of a failure of the primary PCE. In this case, all LSPs to be delegated are under primary stateful PCE control while other PCEs in the domain act as backup.

Load-Balanced 'Backup' PCE: Load-Balanced PCEs share the computation load at all times, as well as act backup to each other. One PCE MAY serve a set of PCCs as the primary computation server, and only addresses requests from other PCCs in the event of the failure of some other PCE. Delegated LSPs are thus distributed among stateful PCEs.

In either case it is beneficial for the PCE to synchronize changes of its delegated LSPs to the other PCEs in realtime. This should include -

- o Any update made by the PCE to its delegated LSP.

- o Any status change learned from the network.

Note that the state synchronization as per [I-D.ietf-pce-stateful-pce] and [I-D.ietf-pce-stateful-sync-optimizations]remains unchanged. This include initial state synchronization as well as LSP state reports. The mechanism described in this document are in addition to those already present in [I-D.ietf-pce-stateful-pce].

4. Functions to Support LSP-DB Synchronization

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

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

- o LSP state synchronization (C-E)

- o LSP update request (E-C)

- o LSP state report (C-E)
- o LSP control delegation (C-E,E-C)
- o Stateful PCE discovery

This document extends some of these functions to support realtime LSP-DB synchronization. These are initiated from a PCE towards another PCE (E-E).

Capability negotiation (E-E): both the PCEs must announce during PCEP session establishment that they support PCEP Stateful PCE extensions defined in [I-D.ietf-pce-stateful-pce]. It should also declare whether it has realtime synchronization capability between PCEs. This is done via Open message.

LSP state report (E-E): a PCE sends an LSP state report to a PCE whenever the state of an delegated LSP changes. This is usually triggered on receiving the state report from the PCC. This is done via PCRpt message.

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

Stateful PCE discovery: PCE can advertise its realtime synchronization capability between PCEs via IGP.

5. Operations

5.1. Relatime LSP-DB Synchronization between redundant Stateful PCEs

PCE (including redundant stateful PCEs) learn LSP state from the PCCs. Apart from that, for each LSP delegated to a stateful PCE -

- o When it sends an LSP Update (PCUpd message) to the PCC for the delegated LSP, it also sends an LSP update to other stateful PCEs.
- o When it receives an LSP report (LSRpt message) from the PCC for the delegated LSP, it also sends an LSP report to other stateful PCEs.

Thus a PCE may learn LSP state from both the PCC as well as the PCE to which LSP is delegated.

In Figure 1, PCE1 is the primary stateful PCE and PCE2 is the backup stateful PCE (all LSPs are delegated to PCE1). PCC1 and PCC2

synchronize the LSP-DB with PCE1 and PCE2 after session initialization phase.

PCC1 and PCC2 delegates LSP1 and LSP2 to the primary PCE1. Whenever there is an update in LSP, PCE1 sends a PCUpd message to corresponding PCC and also to backup PCE2. This is LSP update request as described in Section 4 and uses PCUpd message. This makes sure that the pending LSP changes and sticky resources are backed up. The PCC sends a PCRpt message to the primary PCE, indicating the LSP's status, the primary PCE further synchronizes the state with backup PCEs via PCRpt message.

```
+----+            +----+             +----+              +----+
|PCC1|            |PCC2|             |PCE1|              |PCE2|
+-+--+            +-+--+             +-+--+              +-+--+
  |                 |                  |                   |
  |---- LSP SYNC ---+----------------->|                   |
  |                 |---- LSP SYNC --->|                   |
  |                 |                  |                   |
  |                 |---- LSP SYNC ----+------------------>|
  |---- LSP SYNC ---+------------------+------------------>|
  |                 |                  |                   |
  |-- PCRpt,lsp1,D -+----------------->|                   |
  |<----------------+----PCUpd,lsp1 ---|                   |
  |                 |                  |--- PCUpd,lsp1 --->|
  |-- PCRpt,lsp1,up-+----------------->|                   |
  |-- PCRpt,lsp1,up-+------------------+------------------>|
  |                 |                  |----PCRpt,lsp1,up->|
  |                 |                  |                   |
  |                 |-- PCRpt,lsp2,D ->|                   |
  |                 |<---PCUpd,lsp2 ---|                   |
  |                 |                  |--- PCUpd,lsp2---->|
  |                 |-- PCRpt,lsp2,up->|                   |
  |                 |-- PCRpt,lsp2,up--+------------------>|
  |                 |                  |----PCRpt,lsp2,up->|
  |                 |                  |                   |
```

Figure 1: Relatime LSP-DB synchronization between primary and backup stateful PCEs

The backup PCE is used only in case the primary PCE fails. At the time of failure of primary PCE (PCE1), the backup PCE (PCE2) act as a primary.

In Figure 2, PCE1 and PCE2 are load-balanced stateful PCEs and share the computation load as well as act as backup to each other. PCC1

and PCC2 synchronize their LSP-DB with both PCEs after session initialization phase as per [I-D.ietf-pce-stateful-pce].

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

```
+----+            +----+            +----+              +----+
|PCC1|            |PCC2|            |PCE1|              |PCE2|
+-+--+            +-+--+            +-+--+              +-+--+
  |                 |                 |                   |
  |---- LSP SYNC ---+---------------->|                   |
  |---- LSP SYNC ---+-----------------+------------------>|
  |                 |---- LSP SYNC -->|                   |
  |                 |---- LSP SYNC ----+----------------->|
  |                 |                 |                   |
  |-- PCRpt,lsp1,D -+---------------->|                   |
  |                 |-- PCRpt,lsp2,D --+----------------->|
  |                 |                 |                   |
  |                 |                 |                   |
  |<----------------+----PCUpd, lsp1---|                  |
  |                 |                 |--- PCUpd, lsp1--->|
  |-- PCRpt,lsp1,up-+---------------->|                   |
  |-- PCRpt,lsp1,up-+-----------------+------------------>|
  |                 |                 |----PCRpt,lsp1,up->|
  |                 |                 |                   |
  |                 |                 |                   |

  |                 |<---PCUpd, lsp2---|------------------|
  |                 |                 |<--- PCUpd, lsp2 --|
  |                 |-- PCRpt,lsp2,up--+----------------->|
  |                 |                 |<---PCRpt,lsp1,up--|
  |                 |-- PCRpt,lsp2,up->|                  |
  |                 |                 |                   |
```

Figure 2: Relatime LSP-DB synchronization between load-balanced stateful PCEs

At the time of failure of one of the PCEs (say PCE1), the other PCE (PCE2) may take up the load.

5.2. Other Considerations

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

- This document does not tackle the issue about TED synchronization which is described in detail in [RFC7399].

6. PCEP Messages

[Editor's Note: There are ongoing discussions to come up with a singular extention for inter-stateful-PCE communications. This section will be updated based on the outcome of the discussion.]

6.1. The PCRpt Message

The format of PCRpt message is defined in [I-D.ietf-pce-stateful-pce]. It specifies the PCRpt message is sent from PCC to PCE in reporting the LSP state. This document extends the usage of PCRpt message between redundant stateful PCEs for realtime LSP synchronization as described in Section 5.1. A unique PLSP-ID needs to be generated at the PCE and should also carry the PCC generated PLSP-ID along in a REALTIME-SYNC TLV in the LSP object.

6.2. The PCUpd Message

The format of PCUpd Message is defined in [I-D.ietf-pce-stateful-pce]. It specifies the PCUpd message is sent from PCE to PCC to request changes in LSP attributes. This document extends the usage of PCUpd message between stateful PCEs for realtime LSP synchronization as described in Section 5.1. Whenever there is a PCUpd message sent from PCE to PCC, PCE should also send it to other PCEs along with the PCC generated PLSP-ID in a REALTIME-SYNC TLV in the LSP object.

7. TLVs

7.1. Stateful PCE Capability TLV

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

R (REALTIME-SYNC-PCE - 1 bit): if set to 1 by PCE, the PCE has the capability for realtime synchronization between PCEs. In case of PCC, this bit has no meaning and is simply ignored.

7.2. Speaker Entity Identifier TLV

[I-D.ietf-pce-stateful-sync-optimizations] describes 'Speaker Entity Identifier TLV' to be included in OPEN object. This document uses the same TLV in the LSP object for realtime sync between redundant stateful PCEs.

For a PCE that supports realtime sync (REALTIME-SYNC-PCE R flag in Stateful PCE Capability TLV), the PCC MUST include 'Speaker Entity Identifier TLV' in the OPEN message. Note a PCC may have to bring down the current session and include the TLV in the subsequent open message.

Any realtime state synchronization (PCRpt or PCUpd message between PCEs) MUST include 'Speaker Entity Identifier TLV' in LSP object with the PCC's speaker identity. If the TLV is missing, the PCE will generate an error with error-type 6 (mandatory object missing) and error-value TBD1 (Speaker Entity Identifier TLV missing) and close the session.

The format of Speaker Entity Identifier is defined in [I-D.ietf-pce-stateful-sync-optimizations].

7.3. REALTIME-SYNC TLV

PCC uses the PLSP-ID in LSP object to uniquely identify an LSP. For PCE to PCE realtime sync, another unique PLSP-ID needs to be generated at the PCE and should also carry the PCC's generated PLSP-ID in the REALTIME-SYNC TLV. This way redundant PCE can correlate the LSP from the state received from the PCCs.

This TLV MUST be encoded in the PCRpt and PCUpd message between redundant stateful PCEs. If the TLV is missing, the PCE will generate an error with error-type 6 (mandatory object missing) and error-value TBD2 (REALTIME-SYNC TLV missing) and close the session.

The format of the REALTIME-SYNC 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=TBD3           |           Length=4            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           PCC's PLSP-ID               |     reserved          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 3: REALTIME-SYNC TLV format

The type of the TLV is to be assigned by IANA and it has a fixed length of 4 octets. The value contains the following fields:

PCC's PLSP-ID (20 bits): The PCC's original PLSP-ID as received in the PCRpt message from the PCC. This along with Speaker Entity Identifier TLV can be used to co-relate information received from the network (PCCs).

7.4. PCE-CAP-FLAGS sub-TLV

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

Further a new bit is added -

Bit	Capabilities
TBD4	Realtime Sync between PCEs

If this bit is set to 1, the PCE has the capability for realtime synchronization between PCEs.

8. Other Considerations

8.1. PCE Initiated LSP

[I-D.ietf-pce-pce-initiated-lsp] describes the setup and teardown of PCE-initiated LSPs under the active stateful PCE model. As the PCE sends PCInitiate message to PCC to create or delete LSP, the PCE should also send PCUpd message to other PCEs. For the initiation, the PCUpd message should have PCC's PLSP-ID as zero. The rest of the processing remains unchanged.

9. Security Considerations

This document does not introduce any new security concerns besides those in [I-D.ietf-pce-stateful-pce].

10. Manageability Considerations

10.1. Control of Function and Policy

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

10.2. Information and Data Models

[RFC7420] describes the PCEP MIB, there are no new MIB Objects for this document.

10.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

10.4. Verify Correct Operations

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

10.5. Requirements On Other Protocols

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

10.6. Impact On Network Operations

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

11. IANA Considerations

11.1. STATEFUL-PCE-CAPABILITY TLV

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

Bit	Description	Reference
TBD	REALTIME-SYNC-PCE	[This I.D.]

11.2. PCE-CAP-FLAGS sub-TLV

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

Bit	Description	Reference
TBD4	Realtime Sync between PCEs	[This I.D.]

11.3. REALTIME-SYNC TLV

This document defines the following new PCEP TLV:

Value	Meaning	Reference
TBD3	REALTIME-SYNC TLV	This document

11.4. PCEP-Error Object

IANA is requested to make the following allocation in the "PCEP-ERROR Object Error Types and Values" registry.

Error-Type	Meaning	Reference
6	Mandatory Object missing	[RFC5440]
	Error-Value= TBD2 Speaker Entity Identifier TLV missing	This document
	Error-Value= TBD3 REALTIME-SYNC TLV missing	This document

12. Acknowledgments

Thanks to Adrian Farrel and Daniel King for writing [RFC7399].

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

13. References

13.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://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, <http://www.rfc-editor.org/info/rfc5440>.

[I-D.ietf-pce-stateful-pce] Crabbe, E., Minei, I., Medved, J., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-18 (work in progress), December 2016.

[I-D.ietf-pce-stateful-sync-optimizations]
Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X., and D. Dhody, "Optimizations of Label Switched Path State Synchronization Procedures for a Stateful PCE", draft-ietf-pce-stateful-sync-optimizations-07 (work in progress), December 2016.

13.2. Informative References

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <http://www.rfc-editor.org/info/rfc4655>.

[RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, DOI 10.17487/RFC5088, January 2008, <http://www.rfc-editor.org/info/rfc5088>.

[RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, DOI 10.17487/RFC5089, January 2008, <http://www.rfc-editor.org/info/rfc5089>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <http://www.rfc-editor.org/info/rfc7399>.

[RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <http://www.rfc-editor.org/info/rfc7420>.

[I-D.ietf-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-ietf-pce-pce-initiated-lsp-07 (work in progress), July 2016.

Appendix A. Contributor Addresses

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

EMail: udayasree.palle@huawei.com

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

EMail: zhang.xian@huawei.com

Venugopal Reddy Kondreddy
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: venugopalreddyk@huawei.com

Author's Address

Dhruv Dhody
Huawei Technologies
Divyasree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

EMail: dhruv.ietf@gmail.com