Stateful PCE (Path Computation Element) and its corresponding protocol extensions provide a mechanism that enables PCE to do stateful control of MPLS Traffic Engineering Label Switched Paths (TE LSP). Stateful PCE supports manipulating the existing LSP’s state and attributes (e.g., bandwidth and route) and also creating totally new LSPs in the network.

In the current MPLS TE network using RSVP-TE, LSPs are often controlled by "make-before-break (M-B-B)" signaling by headend for the purpose of LSP restoration and reoptimization. In most cases, it is an essential operation to reroute LSP traffic without any data disruption.

This document specifies the procedure of applying stateful PCE’s control to make-before-break RSVP-TE signaling. In this document, two types of restoration/reoptimization procedures are defined, implicit mode and explicit mode. This document also specifies the usage and handling of stateful PCEP (PCE Communication Protocol) messages, expected behavior of PCC as RSVP-TE headend and several extensions of additional PCEP objects.

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."
Table of Contents

1. Introduction ................................................. 4
2. Conventions used in this document ............................ 4
3. Terminology .................................................... 4
4. Motivation ...................................................... 5
5. Make-Before-Break LSP procedures ............................ 5
   5.1. Implicit Make-Before-Break Mode ......................... 6
   5.2. Explicit Make-Before-Break Mode ......................... 8
       5.2.1. Establish new Trial LSP ............................. 9
       5.2.2. Switchover Data Traffic triggered by a PCUpd message ................ 10
       5.2.3. Tear Down old LSP ................................. 12
6. Objects and TLV Formats ....................................... 12
   6.1. Trial LSP TLV in LSP Objects ............................. 12
   6.2. DATA-CONTROL TLV in LSP Objects ....................... 12
7. IANA Considerations .......................................... 13
   7.1. PCEP TLV Indicators ...................................... 13
   7.2. PCEP Error Objects ....................................... 14
8. Security Considerations ....................................... 14
9. Acknowledgments ............................................... 14
10. References .................................................... 14
    10.1. Normative References ................................. 14
    10.2. Informative References ............................... 15
Authors’ Addresses .............................................. 15
1. Introduction

[I-D.ietf-pce-stateful-pce] describes the stateful Path Computation Elements (PCE). Stateful PCE defines the extensions to PCEP to enable stateful control of LSPs between and across PCEP sessions, and it also describes mechanisms to effect LSP state synchronization between PCCs and PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions.

Today, however, there is no detailed procedure specified as to how to restore and reoptimize one particular MPLS-TE LSP using stateful PCE. In today’s MPLS RSVP-TE mechanism, make-before-break (M-B-B) is a widely common scheme supported by headend LER in order to assure no traffic disruption during restoration and reoptimization. Hence it is naturally desirable for stateful PCE to control M-B-B based signaling and forwarding process.

This document specifies the definite procedures of applying stateful PCE’s control to M-B-B method. In this document, two types of restoration/reoptimization procedures are defined, Implicit mode and explicit mode. This document also specifies the usage and handling of stateful PCEP (PCE Communication Protocol) messages, expected behavior of PCC as RSVP-TE headend and several extensions of additional objects.

2. Conventions used in this document

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].

3. Terminology

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

This document uses the following terms defined in [RFC3209]: make-before-break.

This document uses the following terms defined in [RFC4426] and [RFC4427]: recovery, protection, restoration.

According to their definition the term "recovery" is generically used to denote both protection and restoration; the specific terms "protection" and "restoration" are used only when differentiation is required. The subtle distinction between protection and restoration
is made based on the resource allocation done during the recovery period. Hence the protection allocates LSP resource in advance of a failure, while the restoration allocates LSP after a failure occur.

4. Motivation

As for current MPLS mechanism, make-before-break (M-B-B) concept is outlined in [RFC3209], which allows adaptive and smooth RSVP-TE LSP rerouting that does not disrupt traffic or adversely impact network operations while rerouting is in progress. M-B-B is applicable for reoptimizing LSP’s route and resources for several use cases, for example, to adopt better path for reversion after failure, to change traversing node/links for planned maintenance, to change bandwidth of LSPs. M-B-B is also used for global restoration scenario in case of failure, which is effective if operators do not want to reserve both working and standby LSPs’ bandwidth in advance. In real deployment, it can also be operated with local protection scheme FRR (Fast ReRoute).

Since M-B-B operational scheme is universally common in MPLS network today, it is naturally much desirable to utilize it under the architecture of stateful PCE.

The basic procedure of the Make-Before-Break method is outlined as follows:

1. Establish a new LSP
2. Transfer data traffic from old LSP onto the new LSP
3. Tear down the old LSP

In M-B-B, it is an important behavior that headend node handles the sequence of data traffic switchover. The headend is able to "make" one or more new LSPs for a particular Tunnel (i.e., it is allowed to signal multiple RSVP sessions with different LSP-IDs that share a common Tunnel IDs), and the headend will switch the traffic upon only one (or some) of those LSPs. In some use cases about stateful PCE, it is expected that operators can watch and control when the data is switched over and which LSPs are used. Therefore, this document covers such a procedure and related message extensions.

5. Make-Before-Break LSP procedures

There are possibly two modes introduced for Make-Before-Break procedure under stateful PCE. The first one is "implicit M-B-B mode", where the operation is triggered by a PC Update Request (PCUpd) message from a PCE, and a PCC handles whole Make-Before-Break steps
(signaling and transferring data traffic) for itself. This mode utilizes the existing messages as defined in
[I-D.ietf-pce-stateful-pce].

The second one is "explicit M-B-B mode", where the operation is triggered by a PCUpd message with TRIAL LSP TLV, which is defined in Section 6.1. A PCE also controls timing and sequence of each granular step that a PCC takes. This procedure additionally uses a new extended TLV that is defined in Section 6.2.

Both types of procedure require at least two LSPs residing in a single MPLS-TE tunnel, working LSP and trial LSPs. An ingress node is currently transporting data traffic on the working LSP, and then it establishes one or more trial LSPs. As per [RFC3209] Section 2.5, "LSP ID" of a restoration LSP, which is newly signaled, differs from that of a working LSP. In this document, LSP ID of a working LSP describes "old" and that of a trial LSP describes "new" as a simple example.

Implicit mode has high affinity with most existing MPLS edge node implementations which perform entire steps of M-B-B automatically at once. This mode is particularly applicable for migration scenario for the existing deployment where service providers want their recovery operation be delegated to centralized PCE.

Explicit mode is much more flexible than Implicit mode since it allows PCEs to manage each LSP step-by-step. Explicit mode is applicable to several new use cases that require split control of signaling and data switchover. For example, if end-to-end data path is created by connecting multiple individual LSPs across different segments (e.g., LSP stitching), in reoptimization scenario, data flowing cannot be started unless all signaling of all LSPs is completed. Similarly, there is a case under Software Defined Network (SDN) applications, where MPLS domain is connected to other non-MPLS domains, and the end-to-end data switchover timing should be carefully coordinated with various different methods of path/flow setup in each domain.

PCC and PCE can distinguish which mode, implicit mode or explicit mode, is to be performed by checking the type of PCEP messages that are exchanged. The implementation MAY support both modes, but for each restoration/reoptimization operation, either one of them SHOULD be exclusively selected.

5.1. Implicit Make-Before-Break Mode

This specifies the detailed procedure of M-B-B LSP restoration and reoptimization using existing messages which are defined in
This procedure is based on the current existing messages/TLVs and no extended TLV is used. Once a PCC receives PCUpd message from a PCE, the PCC automatically executes the implicit M-B-B procedure as described in [I-D.ietf-pce-stateful-pce] Section 6.2.

First, a PCUpd message is sent from a PCE to trigger M-B-B procedure. Once a PCC received the PCUpd message, the PCC starts signaling a new restoration LSP and it sends back to the PCE a PCRpt message with LSP-IDENTIFIERS TLV in the LSP Object.

Second, once a restoration LSP is successfully established, a PCC transfers data traffic from working LSP to restoration LSP. If the restoration LSP failed in setup, the PCC notifies the PCE the result in a PCRpt message and it MAY wait for a next instruction from the PCE.

Finally, when a PCC successfully transferred data traffic to restoration LSP, the PCC tears down the (previous) working LSP by RSVP-TE signaling, then the PCC MUST send a PCRpt message. That PCRpt message MUST carry a LSP Object with LSP-IDENTIFIERS TLV which indicates the value of RSVP-TE signaling the PCC has just torn down.

Following Figure 1 illustrates the example of implicit M-B-B procedure, in following conditions.

- working LSP : ERO=a-b, Tunnel ID=T1, LSP ID=old
- restoration LSP : ERO=a-c-b, Tunnel ID=T1, LSP ID=new
5.2.  Explicit Make-Before-Break Mode

Comparing to the implicit M-B-B mode, explicit M-B-B mode allows a
PCE to control timing and sequence of subsequent make-before-break
steps as follows.

First, the PCE initiates PCC’s signaling of a new LSP by sending a
LSP Update Request (PCUpd) message with TRIAL-LSP TLV that are defined
in this document. Second, the PCE instructs the PCC to transfer data
traffic from old LSP to new LSP by sending a PCUpd message with
TRIAL-LSP TLV and DATA-CONTROL TLV that are defined in [I-D.tanaka-
pce-stateful-pce-data-ctrl]. Third, the PCE MAY instruct the PCC to

tear down the old LSP by sending a PCUpd message indicating LSP removal.

The following subsections specify each Make-Before-Break steps in detail.

5.2.1. Establish new Trial LSP

As a first step of M-B-B procedure, a PCC establishes a new LSP for restoration once PCC receives a PCUpd message with TRIAL-LSP TLV from a PCE. We call this newly established LSPs for restoration "trial LSP". A trial LSP is signaled the same RSVP-TE Tunnel ID but different LSP ID from active working LSP, and both the active working LSP and new trial LSPs MUST be signaled with Shared Explicit style as describes in [RFC3209]. TRIAL-LSP TLV triggers explicit mode M-B-B. A PCE do not have to assign RSVP-TE LSP ID for trial LSP signaling, however it MAY specify RSVP-TE LSP ID that the PCC is going to establish.

When a new trial LSP was signaled successfully, the PCC sends a PCRpt message toward the PCE to notify the result. The PCRpt message from the PCC MUST have the LSP object with LSP-IDENTIFIERS TLV that indicates RSVP-TE Tunnel ID and LSP ID the PCC has just established.

If a new trial LSP failed to be established by some reason of RSVP-TE signaling, the PCC MUST send a PCRpt message carrying LSP-IDENTIFIERS TLV and RSVP-ERROR-SPEC TLV as defined in [I-D.ietf-pce-stateful-pce] Section 7.3.4. to the PCE.

A PCC SHOULD accept multiple PCUpd messages with TRIAL-LSP TLV in a LSP Object. And a PCC SHOULD establish as many trial lsps as the number of PCUpd messages it receives.

Figure 2 illustrates an example, working LSP(PLSP-ID P1, Tunnel ID T1, LSP-ID old, ERO Ingress-a-b-Egress), trial LSP(Tunnel ID T1, LSP-ID new, ERO Ingress-a-c-b-Egress).
5.2.2. Switchover Data Traffic triggered by a PCUpd message

As a second step, PCC(Ingress) transfers data traffic from a working LSP to a trial LSP. To specify desired LSP for transferring data traffic, a PCUpd message from a PCE MUST have a TRIAL-LSP TLV and a DATA-CONTROL TLV in a LSP Object.

A TRIAL-LSP TLV specifies desired RSVP-TE LSP-ID a PCC starts using for transferring data traffic. And a DATA-CONTROL TLV triggers data traffic switch over. Both TLVs MUST be assembled into a single LSP Object.

Once the PCC receives the PCUpd message with TRIAL-LSP TLV and DATA-CONTROL TLV in the LSP Object, the PCC MUST start transfer data traffic to new trial LSP immediately. (See Figure 3)

In DATA-CONTROL TLV, Origin(O) bit, which represents traffic origin, SHOULD set to 1, Continue(C) bit SHOULD set to 0, and Percentage(P) bit SHOULD set to 100% in order to perform whole data traffic switchover.

If the TRIAL-LSP TLV in the PCUpd message specifies invalid LSP,
PCErr MUST be sent out from the PCC to the PCE. The error message with Error-Type-19 (Invalid Operation) and Error-Value[TBD](See Section 7.2).

```
\[ \]
PCE                  PCC(Ingress)--a-------b---Egress
|                                    |                                    |
\|--PCUpd            -------\                         |                                    |
\|LSP Object         |]]]]]]]]]]]]]]]]]]]]]]]]]]]]\            | data on old LSP |
\|   PLSP ID=P1      |]]]]]]]]]]]]]]]]]]]]]]]]]]]]\            | data on new LSP |
\|     +TRIAL-LSP TLV:                                    |                                    |
\|       LSP-ID=new                                                |
\|     +DATA-CTRL TLV:                                         |
\|       Origin=1,                                              |
\|       Continue=0,                                           |
\|       Percent=100                                           |
\|                                    |                                    |
\|--PCRpt            -------\                         |                                    |
\|LSP Object         |]]]]]]]]]]]]]]]]]]]]]]]]]]]]\            |                                    |
\|   PLSP ID=P1, Tunnel ID=T1, LSP-ID=new,                    |
\|   +DATA-CTRL TLV:                                         |
\|     Origin=1,                                              |
\|     Continue=0,                                           |
\|     Percent=100                                           |
\|                                    |                                    |
\|--PCRpt (R=1, ---\                                           |
\|   PLSP ID=P1, Tunnel ID=T1, LSP-ID=old)                     |
```

R flag = Remove flag in LSP object.

Figure 3: Transfer data traffic from old LSP to new LSP
5.2.3. Tear Down old LSP

As a final step of Make-Before-Break procedure, the PCC tears down the working LSP and the other trial lsps which the data traffic is no longer used.

The PCC SHOULD tear down the old working LSP and other trial LSPs immediately once the data traffic successfully switched over (See Figure 3). In OPTIONAL, a PCC tears down old lsp separately.

6. Objects and TLV Formats

6.1. Trial LSP TLV in LSP Objects

This document defines a new TLV named TRIAL-LSP TLV.

```
+---------------+---------------+---------------+---------------+
|               |               |               |               |
|   0 1 2 3 4   |   1 2 3 4 5   |   6 7 8 9 0  |   1 2 3 4 5   |
+---------------+---------------+---------------+---------------+
| Type=TBD      | Length        |
+---------------+---------------+
| MUST be Zero  | LSP-ID        |
+---------------+---------------+
```

Figure 4: TRIAL-LSP TLV format

TRIAL-LSP TLV is sub-TLV of the LSP Object and is used in a PCUpd message especially to perform explicit mode M-B-B. A PCC signals a trial LSP once it receives a PCUpd in which LSP object has a TRIAL-LSP TLV(LSP-ID=0). It MUST set RSVP LSP-ID in LSP-ID field of TRIAL-LSP TLV in order to notify a PCC of desired trial LSP to be carried data traffic.

LSP-ID: This field fills the same value of RSVP-TE LSP-ID that is used in signaling. LSP-ID MUST be zero in a PCUpd message when a PCE requests a PCC to signal new trial LSP. LSP-ID MUST be non-zero when a PCE sends a PCUpd message to trigger traffic switchover execution.

6.2. DATA-CONTROL TLV in LSP Objects

DATA-CONTROL TLV in LSP Objects follows for easy reference of [I-D.tanaka-pce-stateful-pce-data-ctrl].
DATA-CONTROL TLV format

Flags and fields

O (traffic Origin - 1 bit): "traffic Origin(O) = 1" indicates this is an active LSP (i.e., carrying traffic now) whose traffic is to be switched over or to be load-balanced. A PCE uses this bit to specify traffic origin that it wants to manipulate. On the other hand, a PCC uses this bit in PCRpt message to notify a PCE that switching traffic succeeded and carrying data traffic.

C (Continue - 1 bit): If this flag set to 1, it indicates the next LSP Object encoded in the PCUpd has also DATA-CONTROL TLV. If this flag set to 0, it indicates no more LSP Object continues and load balancing calculation is completed, and then the PCC MUST perform switching traffic or load-balancing.

Percentage - 8 bit (0B11111111 is reserved): This field specifies ratio of switching traffic as an unsigned char. The sum of this field across subsequent LSP Object has to be hundred percent. The value must be less than or equal to 100% (0B01100100) (e.g., If you want to set 50%, this field should be set to 0B00110010). If no traffic goes through the corresponding LSP, this field should be set to 0%. 0% LSP MUST be deleted immediately after switchover. The special value 0B11111111 indicates traffic 0%, but the LSP MUST remain after switchover.

7. IANA Considerations

7.1. PCEP TLV Indicators

This document defines the following new PCEP TLVs:

<table>
<thead>
<tr>
<th>Value</th>
<th>Meaning</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBD</td>
<td>DATA-CONTROL</td>
<td>This document</td>
</tr>
</tbody>
</table>
7.2. PCEP Error Objects

This document defines new Error-Type and Error-Value for the following new error conditions:

<table>
<thead>
<tr>
<th>Error-Type</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Mandatory Object missing</td>
</tr>
<tr>
<td></td>
<td>Error-value=TBD: LSP Identifiers TLV missing</td>
</tr>
<tr>
<td>19</td>
<td>Invalid operation</td>
</tr>
<tr>
<td></td>
<td>Error-value=TBD: Percentage is not hundred. for explicit mode</td>
</tr>
<tr>
<td></td>
<td>Error-value=TBD: Specified LSP-ID is not existing. for explicit mode</td>
</tr>
<tr>
<td></td>
<td>Error-value=TBD: Specified LSP-ID is not operational. for explicit mode</td>
</tr>
</tbody>
</table>

8. Security Considerations

TBD

9. Acknowledgments

Many thanks to Ina Minei, Adrian Farrel, Yimin Shen, Xian Zhang and their develop team for their ideas and feedback in documentation.

10. References

10.1. 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-05 (work in progress), July 2013.


10.2. Informative References


Authors’ Addresses

Yosuke Tanaka
NTT Communications Corporation
Granpark Tower
3-4-1 Shibaura, Minato-ku
Tokyo 108-8118
Japan

Email: yosuke.tanaka@ntt.com

Yuji Kamite
NTT Communications Corporation
Granpark Tower
3-4-1 Shibaura, Minato-ku
Tokyo 108-8118
Japan

Email: y.kamite@ntt.com