

PCE Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: August 17, 2014

Y. Tanaka  
Y. Kamite  
NTT Communications  
D. Dhody  
Huawei Technologies  
Feb 13, 2014

**Make-Before-Break MPLS-TE LSP restoration and reoptimization procedure  
using Stateful PCE  
draft-tanaka-pce-stateful-pce-mbb-03**

**Abstract**

Stateful Path Computation Element (PCE) and its corresponding protocol extensions provide a mechanism that enables PCE to do stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP). Stateful PCE supports manipulating of the existing LSP's state and attributes (e.g., bandwidth and path) via delegation and also instantiation of new LSPs in the network via PCE Initiation.

In the current MPLS TE network using Resource ReSerVation Protocol (RSVP-TE), LSPs are often controlled by Make-before-break (M-B-B) signaling by the 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 necessary 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."

This Internet-Draft will expire on August 17, 2014.

#### Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](http://trustee.ietf.org/bcp78) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.



## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">4</a>
<a href="#">2.</a>	Conventions used in this document . . . . .	<a href="#">4</a>
<a href="#">3.</a>	Terminology . . . . .	<a href="#">4</a>
<a href="#">4.</a>	Motivation . . . . .	<a href="#">5</a>
<a href="#">5.</a>	Make-Before-Break LSP procedures . . . . .	<a href="#">6</a>
<a href="#">5.1.</a>	Implicit Make-Before-Break Mode . . . . .	<a href="#">7</a>
<a href="#">5.2.</a>	Explicit Make-Before-Break Mode . . . . .	<a href="#">8</a>
<a href="#">5.2.1.</a>	Initiate Association Group for old LSP . . . . .	<a href="#">9</a>
<a href="#">5.2.2.</a>	Establish new Trial LSP . . . . .	<a href="#">10</a>
5.2.3.	Switchover Data Traffic triggered by a PCUpd message . . . . .	<a href="#">11</a>
<a href="#">6.</a>	Objects and TLV Formats . . . . .	<a href="#">13</a>
<a href="#">6.1.</a>	Trial LSP TLV in LSP Objects . . . . .	<a href="#">13</a>
<a href="#">7.</a>	IANA Considerations . . . . .	<a href="#">13</a>
<a href="#">7.1.</a>	PCEP TLV Indicators . . . . .	<a href="#">13</a>
<a href="#">7.2.</a>	PCEP Error Objects . . . . .	<a href="#">13</a>
<a href="#">8.</a>	Operational Considerations . . . . .	<a href="#">14</a>
<a href="#">8.1.</a>	Operation in multiple PCEs . . . . .	<a href="#">14</a>
<a href="#">9.</a>	Security Considerations . . . . .	<a href="#">14</a>
<a href="#">10.</a>	Acknowledgments . . . . .	<a href="#">14</a>
<a href="#">11.</a>	References . . . . .	<a href="#">15</a>
<a href="#">11.1.</a>	Normative References . . . . .	<a href="#">15</a>
<a href="#">11.2.</a>	Informative References . . . . .	<a href="#">15</a>
	Authors' Addresses . . . . .	<a href="#">15</a>



## **1. Introduction**

[I-D.ietf-pce-stateful-pce] describes the stateful Path Computation Elements (PCE) and defines the extensions to PCEP to enable stateful control of LSPs between and across PCEP sessions, further 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 for restoration and reoptimization of 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](#)[[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 (M-B-B), Path State Block (PSB).

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 resource 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 etc. 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 LSP's bandwidth in advance. Once failure occur, LSP becomes down, however PSB (Path State Block) of a headend node remains and keep resources intact. Using M-B-B, the headend node is able to resignals working LSP while the PSB remains until new restoration LSP is successfully established. 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 (Release old PSB)

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 Update Request(PCUpd) message from a PCE, and a PCC handles whole Make-Before-Break steps (signaling, transferring data traffic and teardown) by 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 the M-B-B steps that a PCC takes. This procedure additionally uses a new extended TLV that is defined in [\[I-D.tanaka-pce-stateful-pce-data-ctrl\]](#).

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 RSVP-TE. Note that it is also used for LSP-ID in LSP Identifiers TLVs in PCEP messages, and it differs from PLSP-ID ([\[I-D.ietf-pce-stateful-pce\]](#)). 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/reoptimization 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 signaling of all LSPs is completed. Similarly, there is a case under Software Defined Networking (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 (presence of certain TLV) 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 [\[I-D.ietf-pce-stateful-pce\]](#). This procedure is based on the current existing messages/TLVs and no extensions are required. 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 receiving the PCUpd message, the PCC starts signaling a new restoration/reoptimization LSP and it replies back to the PCE a PCRpt message with LSP-IDENTIFIERS TLV (with new LSP-ID) in the LSP Object to notify the result of signaling. If the new LSP failed to setup, the PCC sends to the PCE the detail of the result in a PCErr or PCRpt message with the same SRP (Stateful PCE Request Parameters) object as that of the PCUpd message and it MAY wait for a next instruction from the PCE.

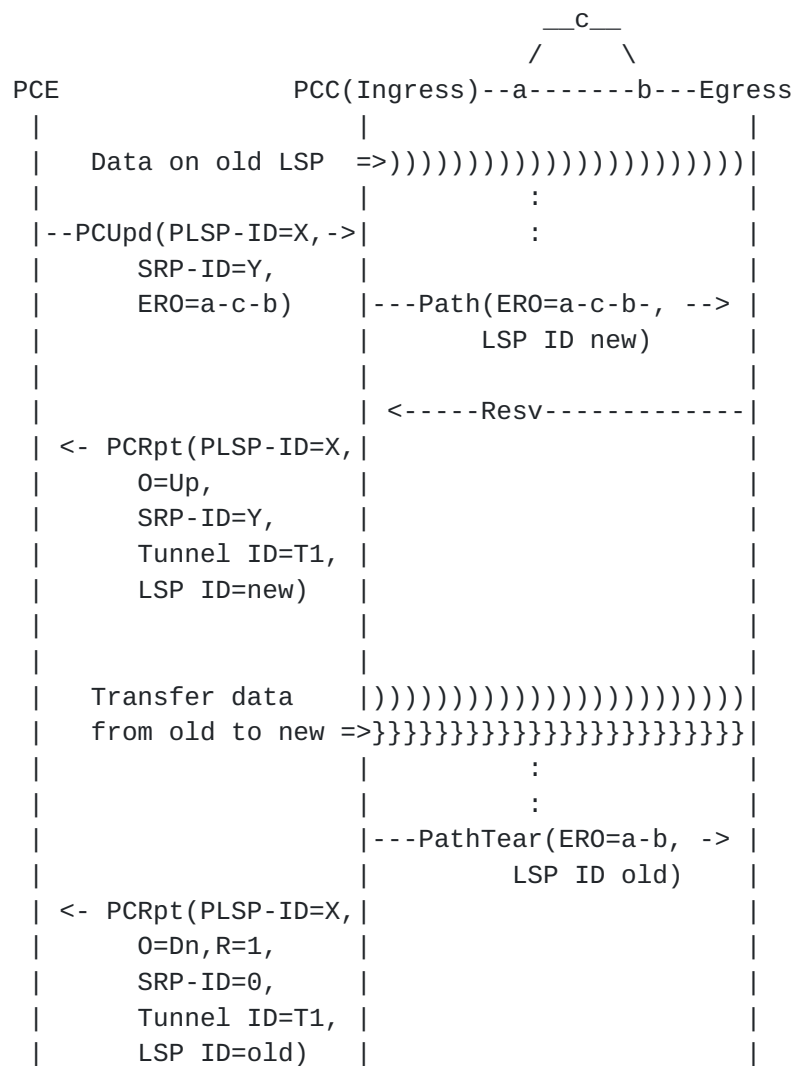
Second, once a new LSP is successfully established, a PCC transfers data traffic from working LSP to new LSP automatically.

Finally, when a PCC successfully transferred data traffic to the new LSP, the PCC tears down the (previous) working LSP by RSVP-TE signaling, then the PCC MUST send another PCRpt message. That PCRpt message MUST carry a LSP Object with LSP-IDENTIFIERS TLV (with old LSP-ID) which indicates the value of RSVP-TE signaling the PCC has just torn down. As per [\[I-D.ietf-pce-stateful-pce\]](#), the message has to have SRP-ID set to 0x00000000.

Following Figure 1 illustrates the example of implicit M-B-B procedure, in following conditions. Tunnel ID and LSP ID are included in an LSP Identifiers TLV in a LSP Object.

working LSP : ERO=a-b, Tunnel ID=T1, LSP ID=old, PLSP-ID=X  
restoration LSP : ERO=a-c-b, Tunnel ID=T1, LSP ID=new, PLSP-ID=X





O flag = Operational flag in LSP object.

R flag = Remove flag in LSP object.

Figure 1: Implicit Make-Before-Break Procedure

## 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.

Prior to start of explicit M-B-B mode, the PCE initiates Association Group for working LSP by sending a PCUpd message with both TRIAL-LSP TLV and ASSOCIATION-GROUP TLV (defined in [\[I-D.tanaka-pce-stateful-pce-data-ctrl\]](#)) in the LSP Object. This is



a pre-requisite.

First step of the explicit M-B-B, the PCE triggers PCC's signaling of a new LSP by sending a PCUpd message with TRIAL-LSP TLV that is defined in this document, and it creates a new Association Group for the new LSP by ASSOCIATION-GROUP TLV. The PCC sends back to the PCE a PCRpt message to notify the result of signaling of the new LSP.

Second, the PCE instructs the PCC to transfer data traffic from old LSP to new LSP by sending a PCUpd message with DATA-CONTROL TLV (defined in [[I-D.tanaka-pce-stateful-pce-data-ctrl](#)]). The PCC automatically tears down the (previous) working LSP once the traffic switchover successfully is executed. Then it sends back to the PCE a PCRpt message to notify the result of the switchover.

The operator may want to separate the second step into traffic switchover and tearing down old LSP. It is further study about the separate operation of third step.

f The following subsections specify each Explicit Make-Before-Break step in detail.

#### **5.2.1. Initiate Association Group for old LSP**

As a pre-requisite before starting explicit M-B-B is to initiate association group for working LSP. The PCE sends to the PCC a PCUpd message that contain both TRIAL-LSP TLV and ASSOCIATION-GROUP TLV in a LSP object to identify the start of explicit M-B-B procedure. [Editor's Note - this need further study].

RSVP-TE LSP-ID that the PCE knows from LSP Identifiers TLVs in a PCRpt message MUST be set to LSP-ID of TRIAL-LSP TLV. The PCE assigns ASSOCIATION-GROUP ID in DATA-CONTROL TLV that is unique in the PCEP session.

Figure 2 illustrates an example of working LSP(PLSP-ID P1, Tunnel ID T1, LSP-ID old, Association Group ID G1 and ERO Ingress-a-b-Egress).









Simultaneously, the PCE creates a new Association Group for the new trial LSP in this step by attaching ASSOCIATION-GROUP TLV in the LSP Object. The PCE assigns a new ASSOCIATION-GROUP-ID, which MUST be different from that of working LSP and be unique in the PCEP session.

Figure 3 illustrates a example, working LSP(PLSP-ID P1, Tunnel ID T1, LSP-ID old, ERO Ingress-a-b-Egress), trial LSP(PLSP-ID P1, Tunnel ID T1, LSP-ID new, ERO Ingress-a-c-b-Egress). And a new ASSOCIATION-GROUP-ID G2 for the new trial LSP.

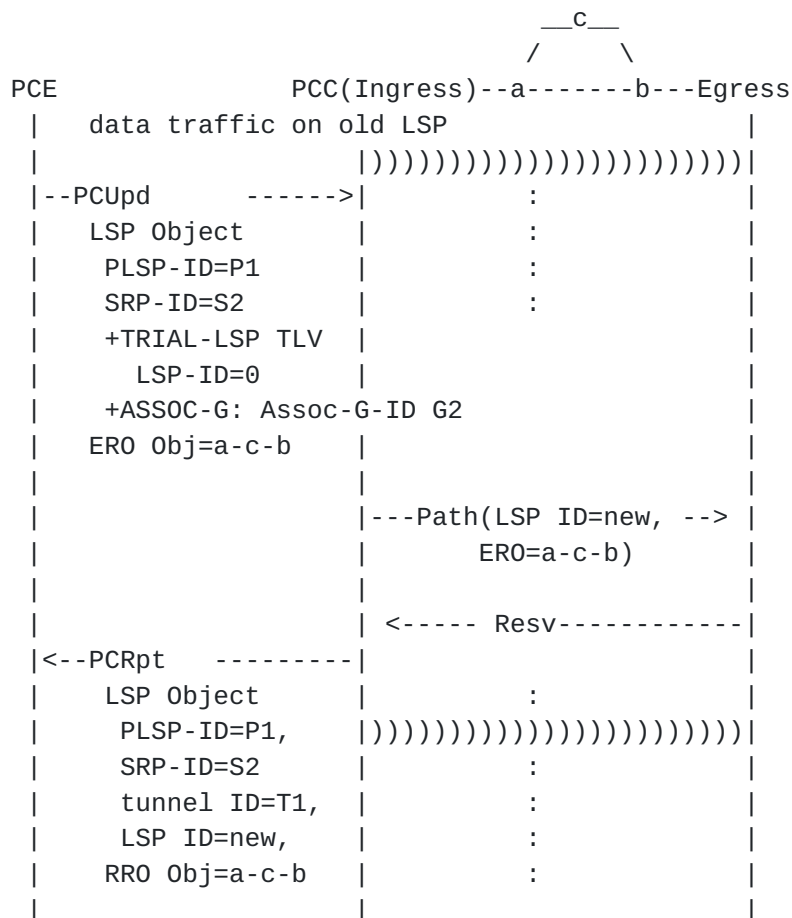


Figure 3: Establish new LSP

### 5.2.3. Switchover Data Traffic triggered by a PCUpd message

As a second step, the 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 DATA-CONTROL TLV in a LSP Object.



Figure 4: Transfer data traffic from old LSP to new LSP



## 6. Objects and TLV Formats

### 6.1. Trial LSP TLV in LSP Objects

This document defines a new TLV named TRIAL-LSP TLV which can be optionally carried in the LSP object.

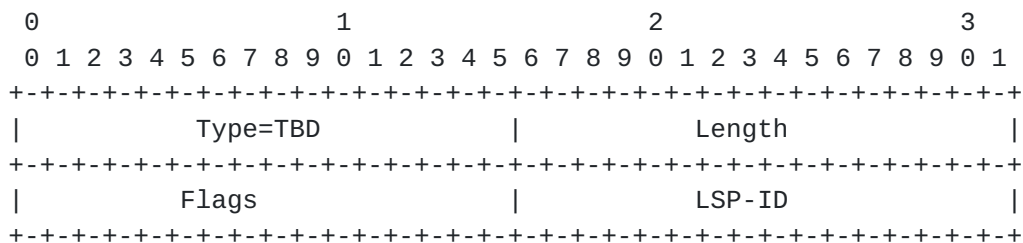


Figure 5: TRIAL-LSP TLV format

TRIAL-LSP TLV is an optional 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=0x0000).

**LSP-ID:** This field MUST be zero in a PCUpd message when a PCE requests a PCC to signal new trial LSP. It MUST be non-zero and fill in the RSVP-TE LSP ID when a PCE sends a PCUpd message to initiate or to create Association Groups for a working/trial LSP.

**Flags:** None defined. MUST be set to zero.

## 7. IANA Considerations

### 7.1. PCEP TLV Indicators

This document defines the following new PCEP TLVs:

Value	Meaning	Reference
TBD	TRIAL-LSP TLV	This document

### 7.2. PCEP Error Objects

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





Error-Type	Meaning
6	Mandatory Object missing Error-value=TBD: LSP Identifiers TLV missing
19	Invalid operation Error-value=TBD: Specified ASSOCIATION-GROUP-ID is not existing for explicit mode Error-value=TBD: Specified LSP-ID is not existing. for explicit mode

## **8. Operational Considerations**

### **8.1. Operation in multiple PCEs**

In addition to basic operations under multiple PCEs as described in [[I-D.ietf-pce-stateful-pce](#)], a PCC supports both types of M-B-B operations.

Implicit mode M-B-B requires only one PCUpd message to trigger M-B-B process, therefore a PCC accepts a message from a primary PCE whom the PCC delegates the LSPs to. An attempt to update parameters of a non-delegated LSP results in the PCC sending a PCErr message as defined in [[I-D.ietf-pce-stateful-pce](#)].

Explicit mode M-B-B requires at least three PCUpd messages(1. for trial-LSP signaling, 2. for new Association-Group creation, 3. for traffic switchover) to trigger each subsequent step. All steps MUST be taken by one primary PCE because state synchronization of trial-LSPs between the primary and backup PCE may be complex. If the PCC revokes LSP delegations after a Redlegation Timeout Interval, the PCC MUST tear down all trial-LSPs and redelegate a working LSP to alternate PCE. An attempt to trigger either step of explicit mode M-B-B of a non-delegated LSP results in the PCC sending the same PCErr as implicit mode M-B-B.

## **9. Security Considerations**

TBD

## **10. Acknowledgments**

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



## **11. References**

### **11.1. Normative References**

- [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-00](#) (work in progress), December 2013.
- [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-07](#) (work in progress), October 2013.
- [I-D.tanaka-pce-stateful-pce-data-ctrl]  
Tanaka, Y., Kamite, Y., and I. Minei, "Stateful PCE Extensions for Data Plane Switchover and Balancing", [draft-tanaka-pce-stateful-pce-data-ctrl-01](#) (work in progress), October 2013.
- [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.

### **11.2. Informative References**

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", [RFC 3209](#), December 2001.
- [RFC4426] Lang, J., Rajagopalan, B., and D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Recovery Functional Specification", [RFC 4426](#), March 2006.
- [RFC4427] Mannie, E. and D. Papadimitriou, "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", [RFC 4427](#), March 2006.



Authors' Addresses

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

Email: [yosuke.tanaka@ntt.com](mailto: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](mailto:y.kamite@ntt.com)

Dhruv Dhody  
Huawei Technologies  
Leela Palace  
Bangalore, Karnataka 560008  
INDIA

Email: [dhruv.ietf@gmail.com](mailto:dhruv.ietf@gmail.com)

