PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 7, 2018

Y. Tanaka
Y. Kamite
NTT Communications
D. Dhody
R. Palleti
Huawei Technologies
January 3, 2018

Make-Before-Break MPLS-TE LSP restoration and reoptimization procedure using Stateful PCE
draft-tanaka-pce-stateful-pce-mbb-05

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 procedures.

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.

Copyright Notice

Table of Contents

1. Introduction 3
2. Conventions used in this document 3
3. Terminology 3
4. Motivation 4
5. Make-Before-Break LSP procedures 5
5.1. Implicit Make-Before-Break Mode 6
5.2. Explicit Make-Before-Break Mode 7
5.2.1. Initiate Association Group for old LSP 8
5.2.2. Establish new Trial LSP 9
5.2.3. Switchover Data Traffic triggered by a PCUpd message 11
6. Protocol extension 12
6.1. Association group 13
6.2. Trial LSP TLV in ASSOCIATION Objects 13
6.3. Optional TLVs 13
7. Security Considerations 14
8. IANA Considerations 14
8.1. PCEP TLV Indicators 14
8.2. Association Object Type Indicator 14
9. Operational Considerations 14
9.1. Operation in multiple PCEs 14
10. Acknowledgments 15
11. References 15
11.1. Normative References 15
11.2. Informative References 16

Authors' Addresses 16

1. Introduction

[RFC5440] describes the Path Computation Element Protocol PCEP. PCEP enables the communication between a Path Computation Client (PCC) and a Path Control Element (PCE), or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalzied MPLS (GMPLS) for Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] describes the stateful Path Computation Elements (PCE) and defines the extensions to PCEP to enable stateful control of LSPs within 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 Label Edge Router (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 of the M-B-B procedures. 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", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

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 to only one (or some) of those LSPs. In some use cases about stateful PCE, it is expected that controller/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 and procedures as defined in [RFC8231] .

The second one is "explicit M-B-B mode", where the operation is triggered by a PCUpd message with a new TRIAL LSP TLV (defined in Section 6.2). A PCE also controls timing and sequence of the M-B-B steps that a PCC takes. This procedure uses ASSOCIATION Object that is defined in [I-D.ietf-pce-association-group].

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 ([RFC8231]). 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 a centralized PCE.

Explicit mode is much more flexible than Implicit mode since it allows PCEs to manage each step of the M-B-B. 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 presence of ASSOCIATION and certain TLV in the PCEP messages. The implementation MAY support both modes, but for each restoration/reoptimization operation, either one of them SHOULD be exclusively applied.

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 [RFC8231] . 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 [RFC8231] Section 6.2.

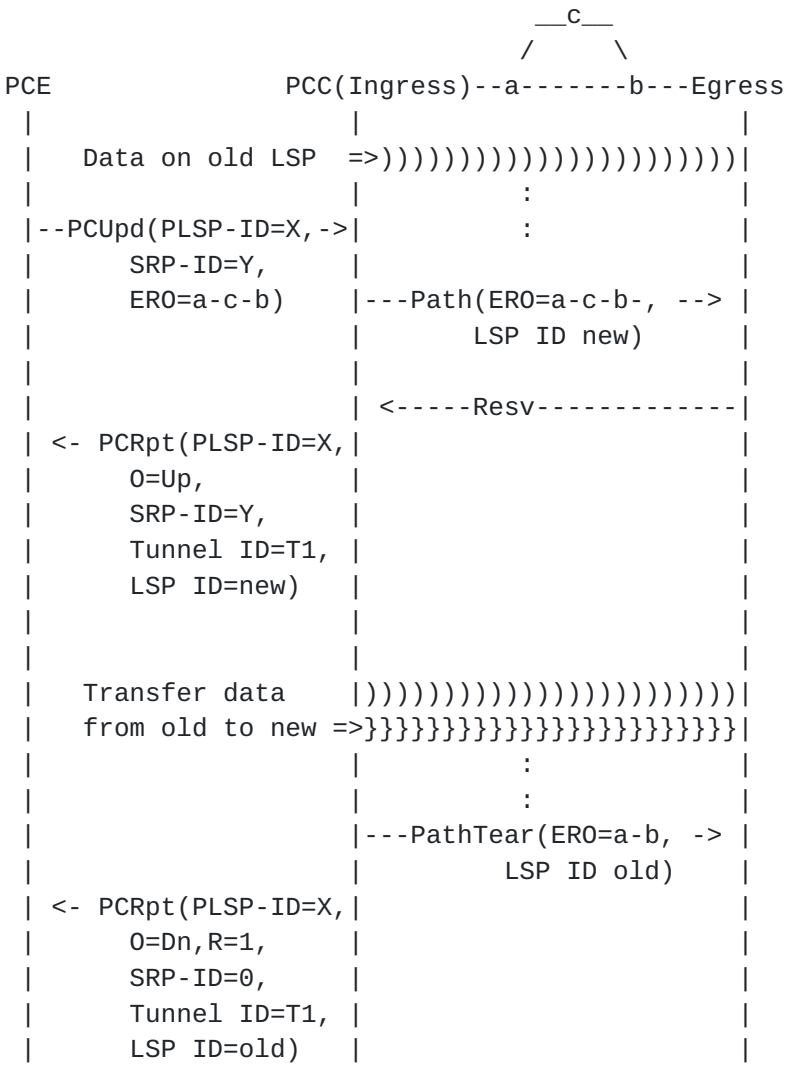
First, A PCUpd message is sent from a PCE to trigger M-B-B procedure. Once receiving the PCUpd message, the PCC starts signaling a new restoration/reoptimization LSP and it replies back to the PCE a PCRpt message with LSP-IDENTIFIERS TLV (with new LSP-ID) in the LSP Object to notify the result of signaling. 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 [RFC8231], 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 per [I-D.ietf-pce-association-group], LSPs are associated with other LSPs with which they interact by adding them to a common association group. In this draft, this grouping is used to define associations between a set of LSPs. This document define one new association type called "Explicit MBB Association Type" of value TBD1.

Prior to start of explicit M-B-B mode, PCE makes an association group for the working LSP by including the Association Object (defined in [I-D.ietf-pce-association-group]) with "Explicit MBB Association Type". This allows the PCEs to identify the LSP belong to a Make-Before-Break association group. PCE may include the TRIAL-LSP TLV that is defined in this document with D(Data Switchover) and T(Trial LSP) flags set to 0 in Association Object. This is a pre-requisite for the explicit M-B-B.

First step of the explicit M-B-B, the PCE triggers signaling of a new LSP at the PCC by sending a PCUpd message with T flag in TRIAL-LSP TLV set to 1, in the ASSOCIATION Object. The PCC sends a PCRpt message back to the PCE to notify the result of the 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 D flag in TRIAL-LSP TLV set to 1, in the ASSOCIATION Object. 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.

[Editor's Note - 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.]

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, PCE makes an association group for working LSP by sending PCUpd message that contains ASSOCIATION object with TRIAL-LSP TLV with both D and T flags set to zero. TRIAL-LSP TLV is optional in the ASSOCIATION object at this step.

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).

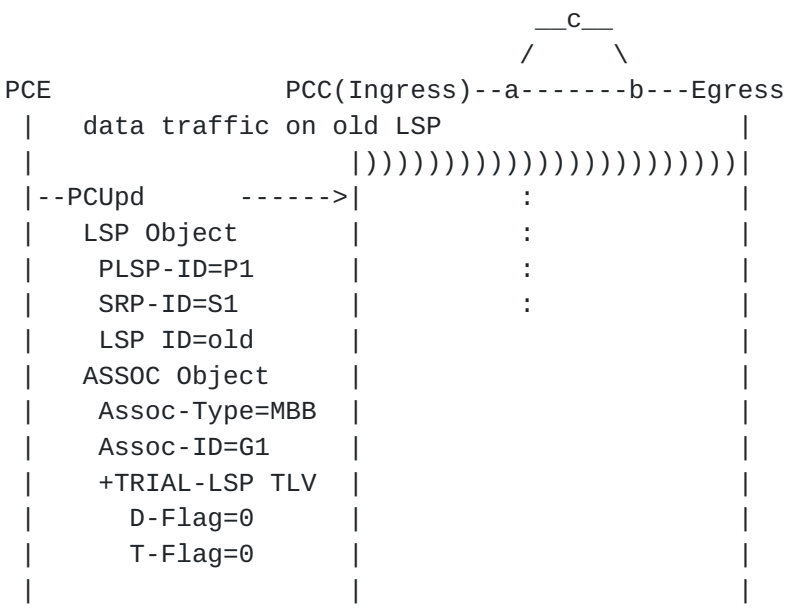


Figure 2: Initiate Associate Group for old LSP

5.2.2. 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 T flag (in TRIAL-LSP TLV) set to 1, in a ASSOCIATION Object 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].

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 to the PCE a PCRpt message carrying LSP-IDENTIFIERS TLV and RSVP-ERROR-SPEC TLV as defined in [RFC8231] Section 7.3.4.

A PCC SHOULD accept multiple PCUpd messages with TRIAL-LSP TLV in a ASSOCIATION Object. And a PCC SHOULD establish as many trial lsps as the number of PCUpd messages it receives. A PCC may also choose to implement a limit on the number of such PCUpd message.

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).

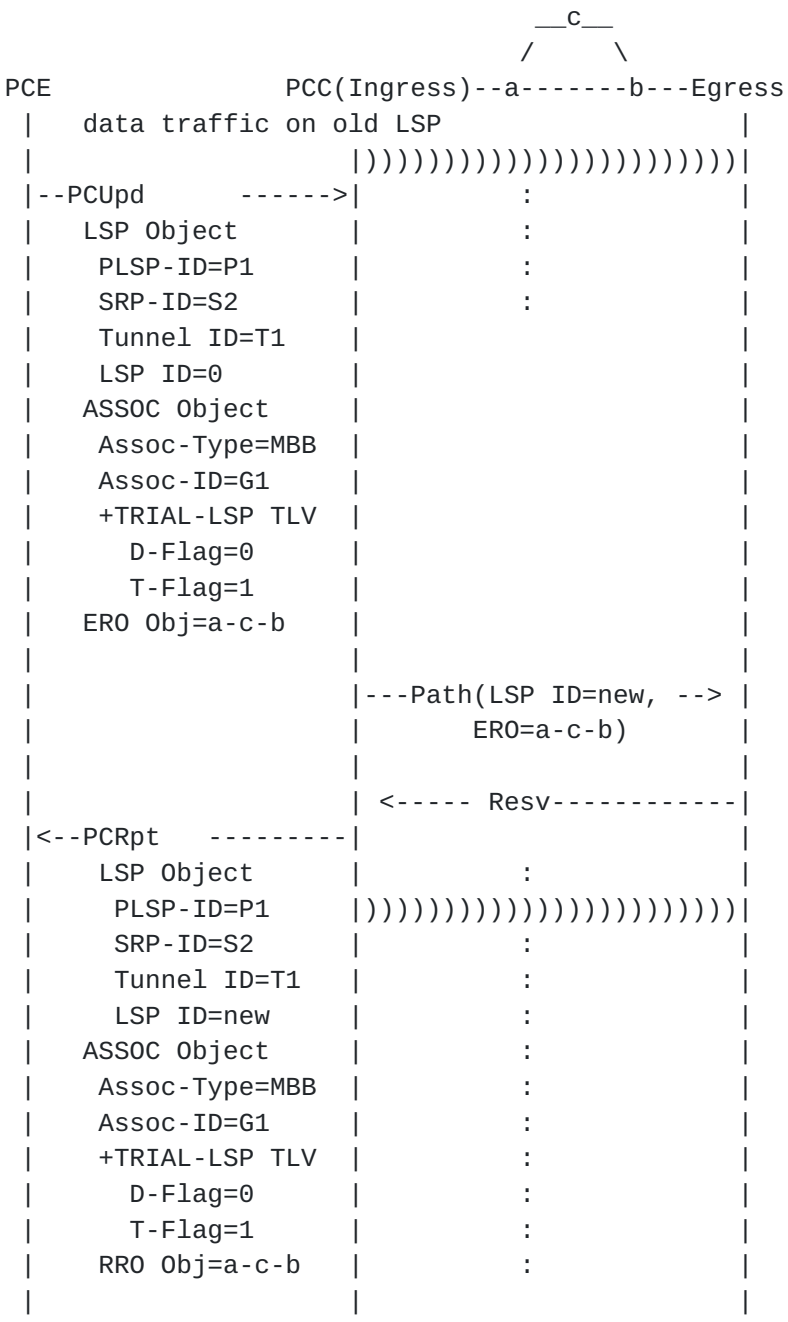


Figure 3: Establish new LSP

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 TRIAL-LSP TLV set D flag to 1, in a ASSOCIATION Object.

Data switchover happens from old LSP to new trial LSP, once PCC receives a PCUpd message with D flag in TRIAL-LSP TLV set to 1 in the ASSOCIATION object from a PCE.

The PCC SHOULD tear down the old working LSP and other trial LSPs which the data traffic is no longer used immediately once the data traffic successfully switched over (See Figure 4).

[Editor's Note - Another option would be, a PCC tears down old lsp separately using mechanism in [RFC8281] for PCE-Initiated LSPs.]

The PCC sends to the PCE a PCRpt message to notify the removal of both old LSP and other trial LSPs, which SRP-ID is set to 0x00000000.

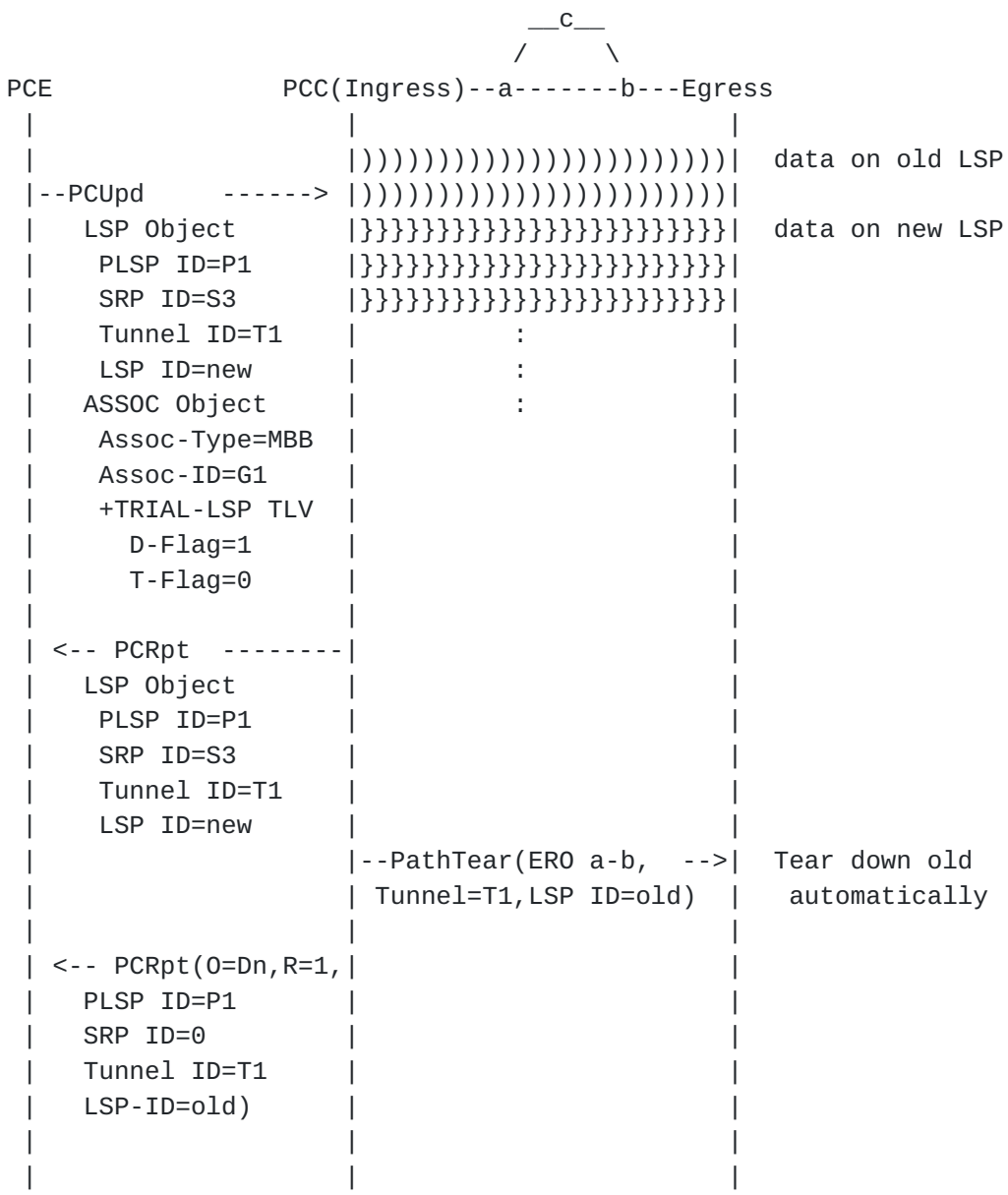


Figure 4: Transfer data traffic from old LSP to new LSP

6. Protocol extension

6.1. Association group

As per [I-D.ietf-pce-association-group], LSPs are associated with other LSPs with which they interact by adding them to a common association group. The Association ID will be used to identify the MBB group a set of LSPs belongs to. This document defines a new Association type, based on the generic Association object -

o Association type = TBD1 ("Explicit MBB Association Type").

6.2. Trial LSP TLV in ASSOCIATION Objects

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

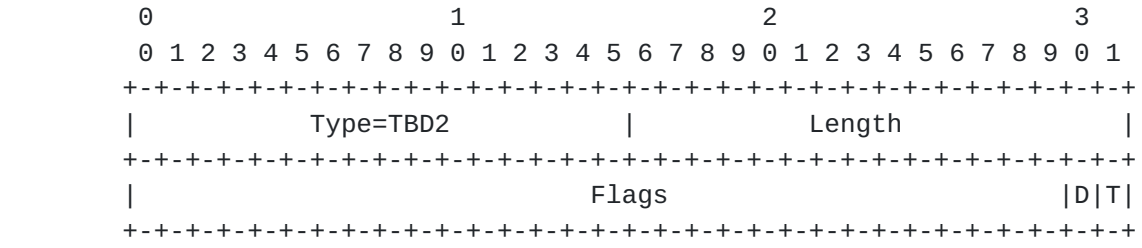


Figure 5: TRIAL-LSP TLV format

TRIAL-LSP TLV is an optional TLV of the ASSOCIATION 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 ASSOCIATION object has a TRIAL-LSP TLV.

T(Trial LSP - 1 bit): This field MUST be set to 1 in a PCUpd message when a PCE requests a PCC to signal new trial LSP. It MUST be zero for a working LSP.
D(Data switchover - 1 bit): This field MUST be set to 1 in a PCUpd message when a PCE requests a PCC to switchover data traffic for new trial LSP. It MUST be zero otherwise.
Flags: None defined. MUST be set to zero. Ignored on receipt.

6.3. Optional TLVs

The MBB association group MAY carry some optional TLVs including but not limited to:

o VENDOR-INFORMATION-TLV: Used to communicate arbitrary vendor specific behavioral information,, described in [RFC7470].

7. Security Considerations

This document defines one new type for association, which do not add any new security concerns beyond those discussed in [RFC5440], [RFC8231] and [I-D.ietf-pce-association-group] in itself.

8. IANA Considerations

8.1. PCEP TLV Indicators

This document defines the following new PCEP TLVs:

Value	Meaning	Reference
TBD2	TRIAL-LSP TLV	This document

8.2. Association Object Type Indicator

This document defines the following new association type originally defined in [I-D.ietf-pce-association-group].

Value	Name	Reference
TBD1	MBB Association Type	This document

9. Operational Considerations

9.1. Operation in multiple PCEs

In addition to basic operations under multiple PCEs as described in [RFC8231], 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 [RFC8231].

Explicit mode M-B-B requires at least three PCUpd messages(1. for new Association-Group creation, 2. for trial-LSP signaling, 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 Redelegation 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.

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-association-group] Minei, I., Crabbe, E., Sivabalan, S., Ananthakrishnan, H., Dhody, D., and Y. Tanaka, "PCEP Extensions for Establishing Relationships Between Sets of LSPs", draft-ietf-pce-association-group-04 (work in progress), August 2017.

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

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, DOI 10.17487/RFC8231, September 2017, <https://www.rfc-editor.org/info/rfc8231>.

[RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <https://www.rfc-editor.org/info/rfc8281>.

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, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[RFC4426] Lang, J., Ed., Rajagopalan, B., Ed., and D. Papadimitriou, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Recovery Functional Specification", RFC 4426, DOI 10.17487/RFC4426, March 2006, <https://www.rfc-editor.org/info/rfc4426>.

[RFC4427] Mannie, E., Ed. and D. Papadimitriou, Ed., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4427, DOI 10.17487/RFC4427, March 2006, <https://www.rfc-editor.org/info/rfc4427>.

[RFC7470] Zhang, F. and A. Farrel, "Conveying Vendor-Specific Constraints in the Path Computation Element Communication Protocol", RFC 7470, DOI 10.17487/RFC7470, March 2015, <https://www.rfc-editor.org/info/rfc7470>.

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

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

Email: dhruv.ietf@gmail.com

Ramanjaneya Reddy Palleti
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560066
India

Email: ramanjaneya.palleti@huawei.com