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

Y. Tanaka
Y. Kamite
NTT Communications
I. Minei
Google
D. Dhody
Huawei Technologies

Feb 13, 2014

Stateful PCE Extensions for Data Plane Switchover and Balancing
draft-tanaka-pce-stateful-pce-data-ctrl-02

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). One application that stateful PCE can realize is data traffic reoptimization among the LSPs. Data traffic traversed in a LSP can be switched to another PCE-initiated LSP. Moreover, data traffic can be balanced to multiple PCE-initiated LSPs and may also be policed based on a signaling bandwidth of a PCE-Initiated LSP using stateful PCE.

This document specifies the extensions to Path Computation Element Protocol (PCEP) that allow a stateful PCE to do switchover, balancing and policing of data traffic with PCE-initiated LSPs. This document also specifies the extensions, usage and handling of stateful PCEP messages and the expected behavior of PCC as the RSVP-TE headend.

Status of this Memo

Table of Contents

1. Introduction

[I-D.ietf-pce-stateful-pce] describes the stateful Path Computation Elements (PCE) procedures 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. A PCE can update LSP settings (such as bandwidth, priority, path) using an update message (called PCUpd).

[I-D.ietf-pce-pce-initiated-lsp] defines the extensions to PCEP to allow a PCE to instantiate new LSPs (called PCE-Initiated LSPs). Before these extensions, the LSP ingress point had to be preconfigured at the head end Label Edge Router (LER), the LSP control automatically delegated to the initiating stateful PCE and then its parameters (e.g., bandwidth, priority, path) could be modified via a PCUpd message. The extensions for PCE-initiated LSPs eliminate the need for preconfiguration, and allow more flexible operations on the network. Stateful-PCE with LSP instantiation is attracting attention as an enabler for Software Defined Networking (SDN) operation of MPLS networks.

In SDN, it is highly expected to support intelligent and interactive control of the amount of network traffic by means of a logically-centralized controller. Optimizing the path and bandwidth of MPLS-TE LSP by using stateful PCE is a leading use case of SDN applications. A PCE is able to calculate an optimized route from the topology and bandwidth information in the Traffic Engineering Database (TED) and the LSP state database (LSPDB) and it can integrate with a controller that takes into account additional information such as historical trends and service orders to trigger some PCE actions. For example, when data traffic on a LSP increases the bandwidth utilization and if there is no capacity left in the currently signaled path (i.e., no remaining bandwidth of links), a PCE is able to update the existing LSP’s parameters (PCE-updated LSP) or create a totally new LSP (PCE-initiated LSP).

The former method is oriented for keeping the existing instance of LSP tunnel. Meanwhile, the latter method is oriented for adding a new instance of a LSP tunnel.

Specifically regarding the latter method, PCE-initiated LSP, there are some operational scenarios in the network: one is that PCE instantiate a new LSP that have alternate route with increased-bandwidth LSP and performs switchover from old LSP. Another is that PCE creates one or more additional LSPs and performs load balancing of data traffic. Today, however, there is no detailed procedure

specified as to how to control data traffic switching from an old LSP to new PCE-Initiated LSP(s).

For another example, when data traffic on a LSP increases its bandwidth utilization and if there is strict traffic contract, a PCE is able to force a PCC not to exceed the contract bandwidth.

This document specifies the procedures that a stateful PCE can use to control data traffic switchover, load balancing with multiple PCE-Initiated LSPs and policing activation/deactivation. This document also specifies the usage and handling of stateful PCEP messages and the expected behavior of PCC as an RSVP-TE headend.

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 [I-D.ietf-pce-stateful-pce]: Stateful PCE, LSP State Request, LSP Update Request.

This document uses the following terms defined in [I-D.ietf-pce-pce-initiated-lsp]: LSP Initiate Message.

4. PCEP Operation for Data Switchover and Balancing

There are two typical operations for explaining the functionality of data switchover and balancing.

- o Whole data switchover, where a PCC switches all data traffic from one LSP tunnel to another.
- o Load balancing of multi-instance LSP tunnels with different paths, where a PCC (headend) balances data traffic among two or more tunnels (ex fifty percent each, for two instances).

Both operational cases are completed by the messaging over a single protocol, PCEP, keeping this as a simple and straightforward solution for MPLS networks.

A PCEP speaker indicates its ability to support PCE control over the data switchover and balancing during the PCEP Initialization phase. The Open Object in the Open message contains the "Stateful PCE Capability" TLV, defined in [I-D.ietf-pce-stateful-pce]. A new flag, the W (LSP-DATASWITCHOVER-BALANCE-CAPABILITY) flag is introduced. A PCE can control the data switchover and loadbalancing only for PCCs that advertised this capability and a PCC will follow the procedures described in this document only on sessions where the PCE advertised the W flag. (Refer Section 5.4)

Data switchover and balancing for an MPLS-TE LSP is available once a PCEP session is established and then a PCC delegates its LSPs to a PCE.

First step is LSP instantiation. In this step, a PCE sends as many PCInitiate messages as PCE-Initiated LSP as per demand. Once the PCC receives them and successfully establishes PCE-Initiated LSPs, it sends PCRpt messages in reply to the PCInitiate messages and delegates the newly established LSP to the PCE. Message formats and behaviors of the PCC and the PCE are described in detail in [I-D.ietf-pce-pce-initiated-lsp].

Second step is LSP association. After the PCE-Initiated LSP successfully established and delegated the PCE sends a PCUpd message that contains the ASSOCIATION-GROUP TLV in the LSP Object in order to assemble the members of an association group of LSPs to take over the traffic. Once a PCC receives the PCUpd message with ASSOCIATION-GROUP TLV, the PCC sends back a PCRpt message that contains the ASSOCIATION-GROUP TLV with current operational status.

[Editor’s Note: The option of specifying the association at LSP instantiation time (as part of the PCInitiate message) will be evaluated in a future version of this document.]

Third step is executing the data switchover and/or load balancing. In this step, the PCE sends a single PCUpd message which updates the operational status of the LSP from "up and carrying traffic" to just "up". This Update request message for data plane switchover/ balancing execution MUST contain DATA-CONTROL TLV in LSP Object. The associated group of traffic origin and that of target to take over the traffic are listed in the DATA-CONTROL TLV. The PCC (LSP headend) load-balances between LSPs in the same association group based on their respective bandwidths. The switchover case is supported since there will be an association of a single LSP, so that LSP will get hundred percent of data traffic.

The PCC MUST send a PCRpt message to the PCE in order to notify of the result of the data switchover/balancing. The PCRpt message MUST

have the DATA-CONTROL TLV that indicates the actual assigned percentages of each member of association group after the execution of the data switchover/balancing operation. The LSP object in the PCRpt will have the reserved PLSP-ID of 0.

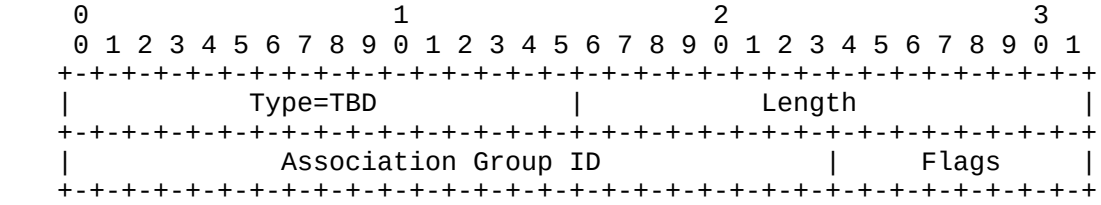
The final step is the deletion of old LSP. It is OPTIONAL to carry out this step. The PCE sends PCInitiate message requesting deletion of the LSP that does not carry data traffic anymore after data switchover/balancing execution. Once the PCC tears down the LSP, a PCRpt message MUST be sent from the PCC to the PCE in order to notify that the LSP is no longer used.

Note that, both RSVP-TE [RFC3209] Tunnel-ID and LSP-ID for PCE-Initiated LSP signaling is not allocated by a PCE. A PCC locally assigns those IDs that are related to RSVP-TE parameters. Therefore, the operations of data switchover and balancing specified in this document is the traffic control procedure across multiple RSVP-TE Tunnels (i.e., different Tunnel instances). Data switchover method across LSPs within a single RSVP-TE Tunnel, which is the switchover in the middle of make-before-break reoptimization, is covered by [I-D.tanaka-pce-stateful-pce-mbb].

5. TLVs in LSP Objects

5.1. ASSOCIATION-GROUP TLV in LSP Objects

This section defines ASSOCIATION-GROUP TLV in LSP Objects. An ASSOCIATION-GROUP TLV is used in the LSP Object in PCUpd messages when a PCE creates an association group of LSPs on a PCC. Further it is used in a LSP object in a PCRpt message to confirm the association.



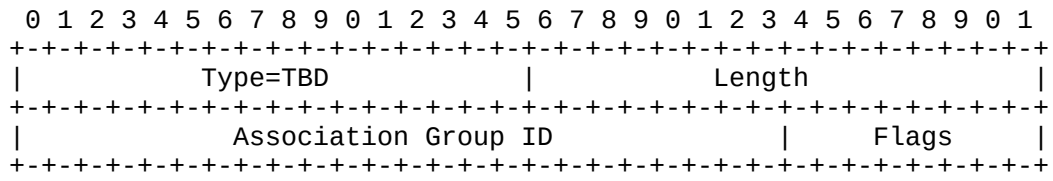


Figure 1: ASSOCIATION-GROUP TLV format

Flags and fields

Association Group ID - 24 bits: This field specifies a identifier of association group of LSPs. The IDs are assigned by a PCE. 0x000000 and 0xFFFFFF is reserved for special use.
Flags - 8 bit: None defined. MUST be set to zero.

An association group is a group of LSPs that is referenced by a single identifier, by both the PCE and PCC. This number is significant in the context of a single PCEP session. An association group may have one or more LSPs. Association groups with zero members are removed and the id can be reused. The PCE is the entity managing association, and this is considered PCE's state that will be cleaned up when the State Timeout Interval expires.

The status of the association group is active when the group is up and carrying data traffic. Otherwise, it is inactive when the group does not carry any data traffic. An LSP is able to associate with up to two association groups, unless both association groups are active at any given point in time. This is done to allow a new LSP to be instantiated and assigned with a new inactive association group, the existing LSP is also associated with this group. This allows switching to the new group.

To create a new association group on a PCC, a PCE sends a PCUpd message which contains the LSP Object(e.g. PLSP-ID=100) and ASSOCIATION-GROUP TLV (Association Group ID=10) in the LSP object. Next, a PCE sends the another PCUpd message with another LSP Object(e.g. PLSP-ID=200) and ASSOIATION-GROUP TLV(Association Group ID=10). As a result, the PCC and PCE both recognize that Association Group ID 10 represents PLSP-ID=100 and 200.

To remove a specific PLSP-ID from the association group, a PCE sends PCUpd message which contains the LSP Object(PLSP-ID=100) and ASSOCIATION-GROUP TLV (Association Group ID=0x0000). Then a PCC removes the PLSP-ID 100 from any inactive association group on the

PCC.

To flush all association groups on a PCC, a PCE sends a PCUpd message which contains the LSP Object(PLSP-ID=0x0000) and ASSOCIATION-GROUP TLV(Association Group ID=0x0000). Then a PCC flushes all association groups. A traffic handling behavior of a PCC when it flushes the active association group is left for a future version of this document.

To associate a PLSP-ID with the existing inactive association group, A PCE sends a PCUpd message which contains the PLSP-ID and the existing Association Group ID. A PCE is not allowed to add any PLSP-ID to the active association group in order to avoid rebalancing traffic without data-ctrl requests. If the PCUpd message contains a PLSP-ID and the active Association Group ID, the PCC MUST send out a PCErr with error value TBD to indicate an invalid operation.

When the LSP of the active association group is torn down by a reason of either network failure or administrative down-request from the PCE, a PCC MUST remove the PLSP-ID from the group and rebalance the traffic based on the respective bandwidths of the rest of LSPs. After rebalancing, The PCC MUST report the actual percentage to the PCE using PCRpt with DATA-REPORT TLV (Section 5.3).

Note that a PCE is able to associate not only PCE-Initiated LSP but also existing LSP(i.e., PCE-updated LSP) with any association group on a PCC.

The definition of PCRpt messages when a PCC creates/removes/flushes an association group will be clarified in the future version of this draft. Redundant stateful PCE section needs the PCRpt in order to sync the association group IDs and actual percentages of balancing.

5.2. DATA-CONTROL TLV in LSP Objects

This section defines DATA-CONTROL TLV in LSP Objects. A DATA-CONTROL TLV is used in the LSP Object in PCUpd messages when a PCE makes a PCC to execute traffic switchover or load balancing. It is also mandatory in a LSP object in a PCRpt message with DATA-REPORT TLV to notify the results of execution.

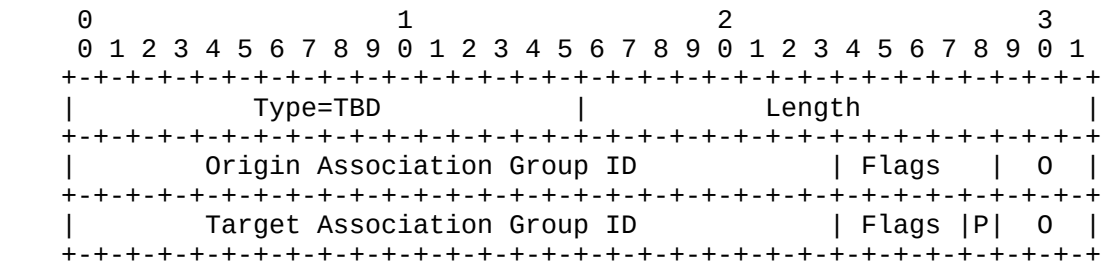


Figure 2: DATA-CONTROL TLV format

Flags and fields

Origin Association Group ID - 24 bits: data traffic origin
Target Association Group ID - 24 bits: for taking over whole data traffic from origin.
P (Policing - 1 bit: This flag is used when a PCE makes a PCC apply traffic policer. If this flag is set 1, traffic exceeding the bandwidth of the LSP is discarded on the PCC after traffic switchover execution. Otherwise, the PCC does not apply any traffic policer and traffic on a target association group will not be discarded.
O (Operational - 3 bits): This flag represents the requested operational status for each Origin Association Group ID and Target Association Group ID by a PCE when this TLV is used in a PCUpd message. It is also used as a status report in a PCRpt message. The meanings of the values are defined in [I-D.ietf-pce-stateful-pce].

An LSP Object in a PCUpd message MUST have DATA-CONTROL TLV when a PCE operates data switchover and balancing on a PCC. DATA-CONTROL TLV is sub-TLV of an LSP Object and is used in both a PCUpd and a PCRpt message.

An operation of data switchover/balancing is the action of transferring traffic from an origin association group to a target association group. A PCUpd message with reserved LSP Object (PLSP-ID=0x00000) and DATA-CONTROL TLV (a set of an origin and a target association group) MUST triggers data switchover/balancing execution.

Traffic policer is able to be applied in both traffic switchover case and load-balancing case.

5.3. DATA-REPORT TLV in LSP Objects

This section defines DATA-REPORT TLV in LSP Objects. A DATA-REPORT TLV is used in the LSP Object in PCRpt message to notify the results of execution with the DATA-CONTROL TLV.

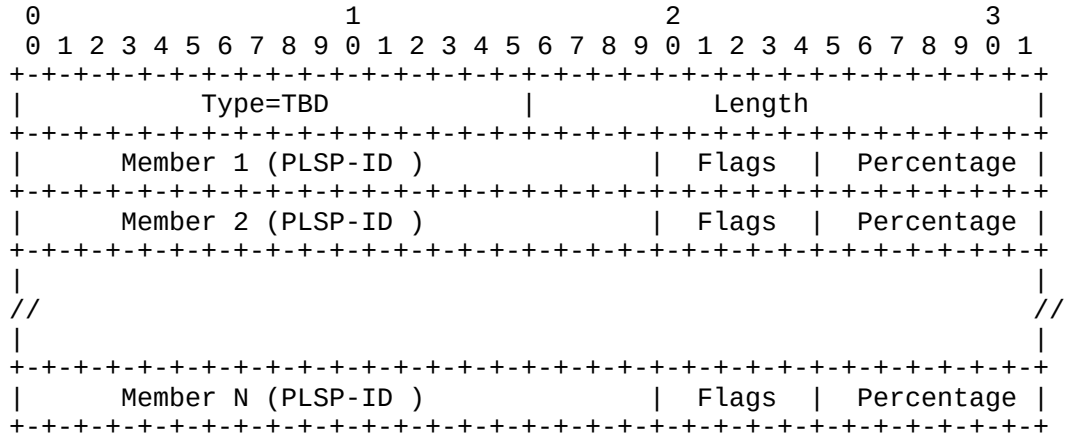


Figure 3: DATA-REPORT TLV format

Flags and fields

Member(PLSP-ID) - 20 bit: This TLV is only used in a PCRpt message and represents actual percentages of load balancing per respective PLSP-ID after load balancing execution. Member field fills PLSP-ID that is member of target association group. As per [I-D.ietf-pce-stateful-pce].

Flags - 5 bit: None defined. MUST be set to zero.

Percentage - 7 bits: This field specifies actual percentage of load balancing as a closest integer, with 100% as the max allowed value.

A PCC replies to a PCE a PCRpt message as an acknowledgment of data switchover/balancing result. The PCRpt message MUST have reserved LSP Object(PLSP-ID=0x00000) and DATA-CONTROL TLV with DATA-REPORT TLV inside.

The PCC load-balances between LSPs in the same association group based on their respective bandwidths.If one of the LSPs goes down by network failure, the traffic would load-balance correctly over the others. If a PCE updates the bandwidth of the LSP, the traffic would

rebalance after a PCC completes the signaling. If one of the LSPs is signaled with zero bandwidth, no traffic would be transferred to the LSP. If all LSPs of the association group are signaled with zero bandwidth, the traffic would load-balance equally. In switchover case, the hundred percent traffic will be transferred to the LSP even if the LSP is zero bandwidth.

The traffic on the existing LSP is able to load-balance over both the existing LSP itself and new PCE-Initiated LSPs, by means of that the existing LSP belongs to both the origin association group and that of target.

5.4. Advertising Support of Data Switchover and Balancing

New flags are defined for the STATEFUL-PCE-CAPABILITY TLV defined in [I-D.ietf-pce-stateful-pce].

W (LSP-DATASWITCHOVER-BALANCE-CAPABILITY - 1 bit): if set to 1 by a PCEP speaker, it indicates that the PCEP speaker allows data switchover and balancing.

6. Operation Examples

For easy understanding this section introduces typical operation examples of data switchover/balancing.

6.1. Data switchover operation (100:0 => 0:100)

A PCE instructs a PCC to switchover 100% traffic from association group ID 1 to association group ID 2. A PCE sends single PCUpd message containing the mandatory LSP Objects with DATA-CONTROL TLV.

Expected PCUpd, PCRpt messages to create association group and to trigger data switchover follow.

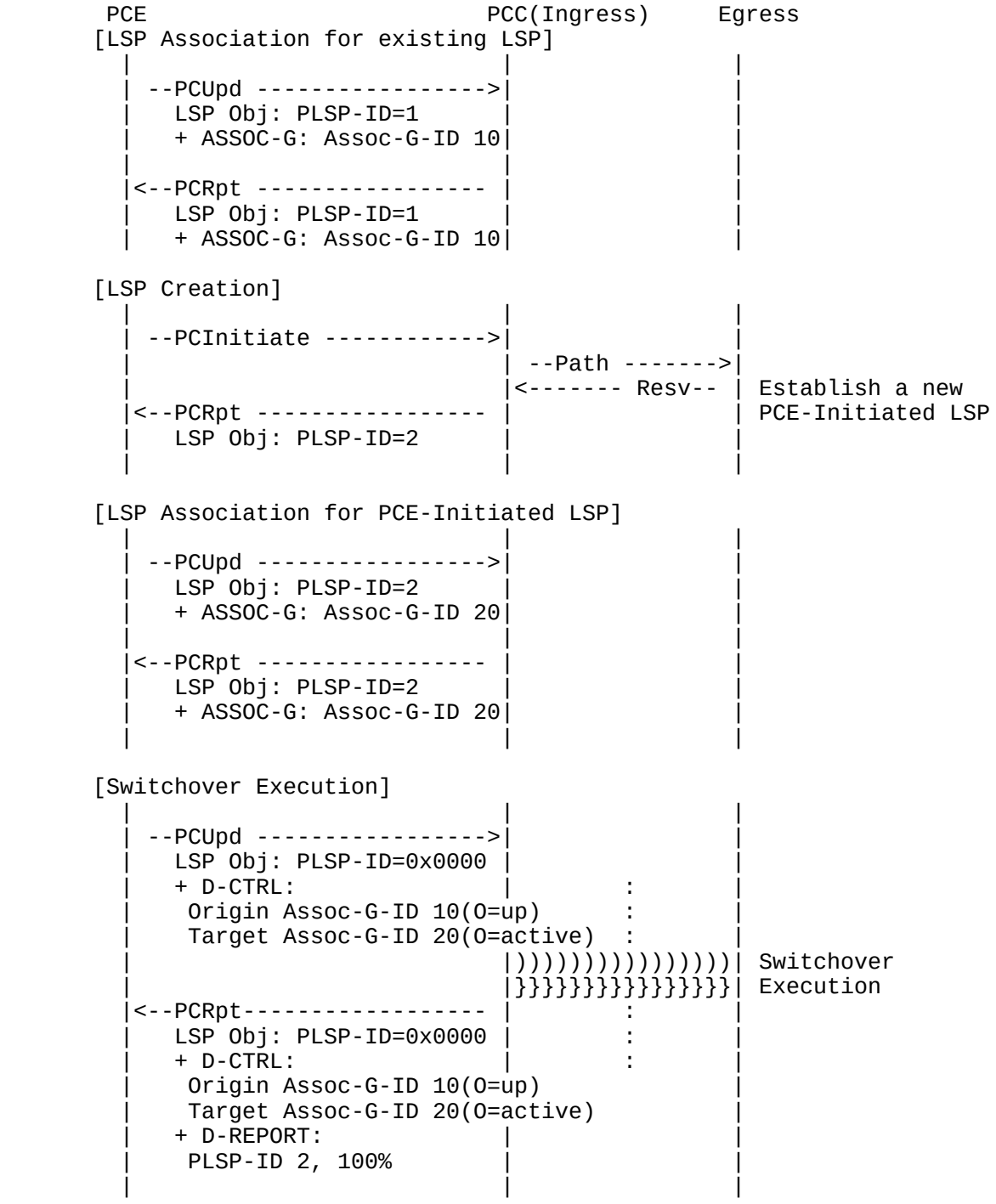
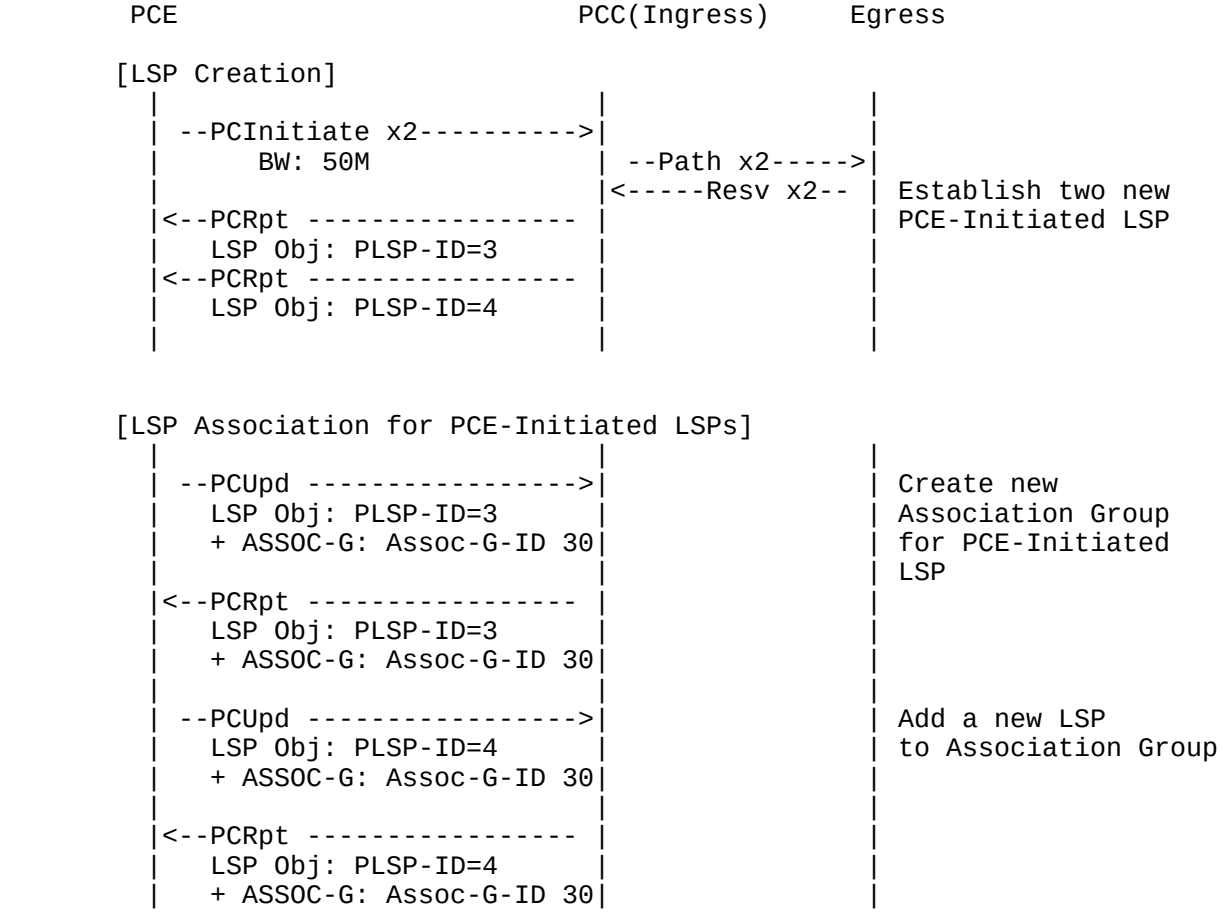

PCE
PCC(Ingress)
Egress
[LSP Association for existing LSP]
--PCUpd ----------->
LSP Obj: PLSP-ID=1
+ ASSOC-G: Assoc-G-ID 10
<--PCRpt -----------
LSP Obj: PLSP-ID=1
+ ASSOC-G: Assoc-G-ID 10
[LSP Creation]
--PCInitiate ------------>
--Path ------->
<------- Resv--
Establish a new
PCE-Initiated LSP
<--PCRpt -----------
LSP Obj: PLSP-ID=2
[LSP Association for PCE-Initiated LSP]
--PCUpd ----------->
LSP Obj: PLSP-ID=2
+ ASSOC-G: Assoc-G-ID 20
<--PCRpt -----------
LSP Obj: PLSP-ID=2
+ ASSOC-G: Assoc-G-ID 20
[Switchover Execution]
--PCUpd ----------->
LSP Obj: PLSP-ID=0x0000
+ D-CTRL:
Origin Assoc-G-ID 10(O=up)
Target Assoc-G-ID 20(O=active)
Switchover
Execution
<--PCRpt-----------
LSP Obj: PLSP-ID=0x0000
+ D-CTRL:
Origin Assoc-G-ID 10(O=up)
Target Assoc-G-ID 20(O=active)
+ D-REPORT:
PLSP-ID 2, 100%

Figure 4: Switchover Operation Example

6.2. Load balancing operation (100:0 => 50:50)

The scenario is one where the starting state is a single LSP (of bandwidth 100 M) is carrying the traffic. To enable better bin-packing, the PCE may want to create two smaller LSPs instead, each of 50M, and load balance the traffic over them. To accomplish this, two association groups are used, the first (say association group ID 10) contains the LSP carrying the traffic, and the second (say association group ID 30) contains the two new smaller LSPs. Expected PCUpd, PCRpt messages to create association group and to trigger load-balance follow (The instantiation of the original LSP of bandwidth 100M and its association into group ID 10 is not shown)



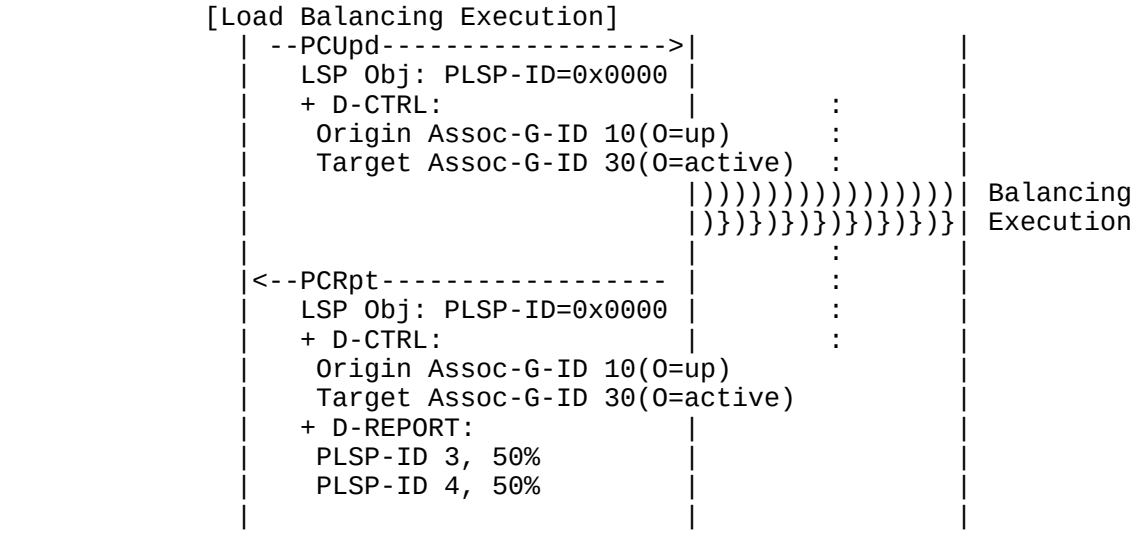
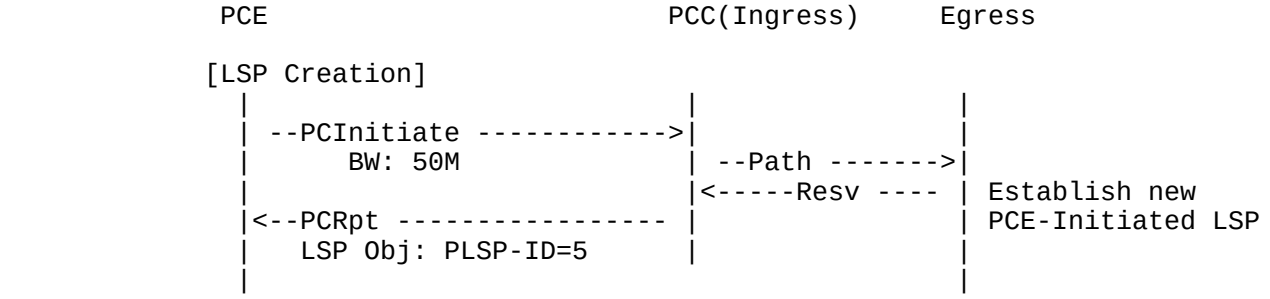


Figure 5: Load-Balance Operation Example

6.3. Load balancing operation (100:0 => 66:33)

The scenario is one where the starting state is a single LSP (of bandwidth 100 M) is carrying the traffic. But as the data traffic load increases another 50 M is required. The PCE may want to create another LSP of 50 M, and load balance the traffic over the existing and new LSP. To accomplish this, two association groups are used, the first (say association group ID 10) contains the LSP carrying the traffic, and the second (say association group ID 40) contains the new initiated LSP as well as the original LSP. Expected PCUpd, PCRpt messages to create association group and to trigger load-balance follow (The instantiation of the original LSP of bandwidth 100M and its association into group ID 10 is not shown)



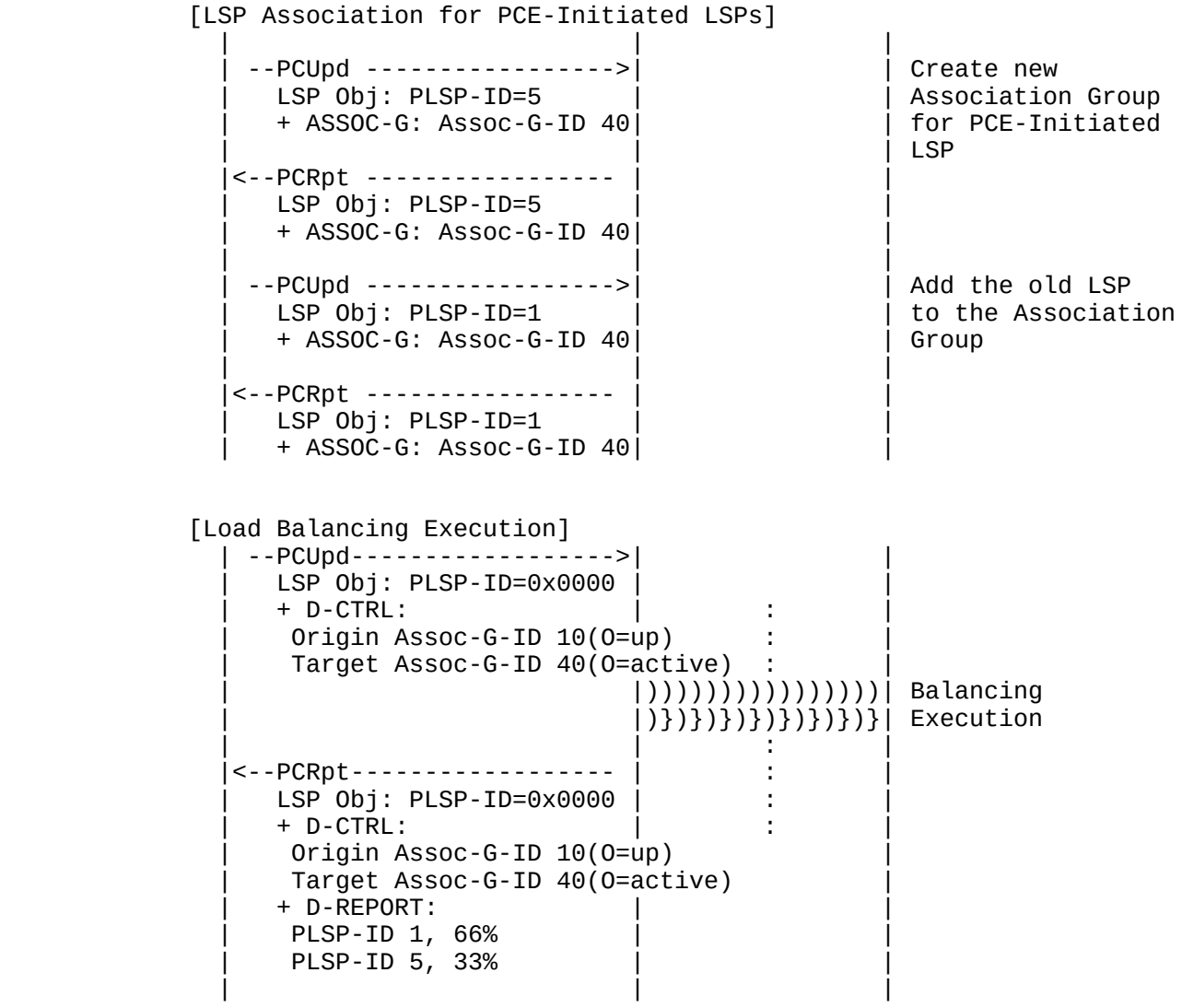


Figure 6: Load-Balance Operation Example

7. Redundant stateful PCEs

Association group IDs are unique within a PCEP session across the primary PCE and the PCC. A backup PCE has to synchronize the

association group IDs, PCE that created the association group and balancing percentages in advance of the failure on the primary PCE. One practical method to synchronize is a PCC replicates each PCRpt message for the backup PCEP session. A backup PCE is able to receive the association group IDs from ASSOCIATION-GROUP TLV and the result of balancing percentages from DATA-REPORT TLV.

8. Security Considerations

This document defines extensions to PCEP to control load balancing of traffic across multiple LSPs or to completely switch traffic from one LSP to another. The nature of these extensions results in more information being available for a hypothetical adversary and a number of additional attack surfaces which must be protected. As a general precaution, it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority

In addition to the security considerations and recommendations described in [I-D.ietf-pce-stateful-pce], the following also apply.

8.1. Malicious PCE

A malicious PCE may flap the traffic between several LSPs, creating shifting patterns in the network and excessive load on the PCC. A PCC may protect itself from such an attack by enforcing a limit on the number of data-control requests per unit of time and MAY take additional steps ranging from delegation revocation to closing the PCEP session.

8.2. Malicious PCC

Because the PCE keeps state regarding LSP associations for all the PCCs, it is RECOMMENDED that the PCE have a bound on the amount of state each PCC can occupy, and in the context of this draft, the number of associations on a PCC and the number of associations each LSP may be part of. Otherwise, a malicious PCC may create an unbounded number of associations. Additionally, a malicious PCC may purposely fail data-control messages in order to force the PCE to continuously resend them and create artificial load on the PCE. The PCE may protect itself from these situations by placing a limit on the number of failures and closing the PCEP session.

9. IANA Considerations

9.1. PCEP TLV Indicators

This document defines the following new PCEP TLVs:

Value	Meaning	Reference
TBD	DATA-CONTROL	This document
TBD	DATA-REPORT	This document

9.2. 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:	DATA-CONTROL TLV missing.
	Error-value=TBD:	DATA-REPORT TLV missing.
19	Invalid operation	
	Error-value=TBD:	No association group existing.
	Error-value=TBD:	No association group specified.
	Error-value=TBD:	No PLSP can be added to the active association group.

10. Acknowledgments

Many thanks to Adrian Farrel for their ideas and suggestions.

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.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC4872] Lang, J., Rekhter, Y., and D. Papadimitriou, "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.

[RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

11.2. Informative References

[I-D.tanaka-pce-stateful-pce-mbb] Tanaka, Y. and Y. Kamite, "Make-Before-Break MPLS-TE LSP restoration and reoptimization procedure using Stateful PCE", draft-tanaka-pce-stateful-pce-mbb-02 (work in progress), October 2013.

[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.

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

Ina Minei
Google
US

Email: inaminei@google.com

Dhruv Dhody
Huawei Technologies
Leela Palace
Bangalore, Karnataka 560008
INDIA

Email: dhruv.ietf@gmail.com