

CCAMP Working Group
Internet Draft
Intended status: Standard Track
Expires: April 14, 2013

Zafar Ali
George Swallow
Clarence Filsfils
Siva Sivabalan
Stefano Previdi
Cisco Systems
Kenji Kumaki
KDDI Corporation
October 15, 2012

Additional Objective Functions and Metric Types in Path
Computation Element Communication Protocol (PCEP)
draft-ali-pce-additional-of-and-metric-00.txt

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 April 14, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Abstract

Network performance criteria such as latency and latency variation are becoming critical to data path selection, especially for networks used by financial institutions. This draft defines additional objective functions and metrics types related to latency and latency variation in Path Computation Element Communication Protocol (PCEP).

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 RFC 2119 [RFC2119].

Table of Contents

Copyright Notice.....	1
1. Introduction.....	3
2. PCEP extensions.....	3
2.1. New Metric Object Types.....	3
2.1.1. P2P Latency Metric.....	4
2.1.2. P2P Latency Variation Metric.....	4
2.1.3. P2MP Latency Metric.....	4
2.1.4. P2MP Latency Variation Metric.....	5
2.2. Handling of New Metric Object Types.....	5
2.3. New Objective Functions.....	5
2.3.1. Minimum Latency Path Objective Function.....	6
2.3.2. Minimum Latency Variation Path Objective Function.....	6
2.4. Handling of New Objective Functions.....	6
3. Security Considerations.....	6

4. IANA Considerations.....	6
5. References.....	7
5.1. Normative References.....	7
5.2. Informative References.....	7

1. Introduction

As noted in [OSPF-TE-METRIC] and [ISIS-TE-METRIC], in certain networks such as financial information networks (e.g. stock market data providers), performance criteria (e.g. latency, latency variation) are becoming critical to data path selection along with other metrics. Such networks may require selection of a path that minimizes end-to-end latency and/or end-to-end latency variation. Or a path may need to be found that optimizes some other metric, but is subjected to a latency and/or latency variation bound(s).

The METRIC object defined in [RFC5440] allows a PCC to specify a bounded acceptable path cost and/or optimization metric. While [RFC5440], [RFC5541] and [RFC6006] define various Metric Types, these RFCs do not address latency and latency variation metrics. This document extends [RFC 5540] with four new Metric Types namely Point-to-Point (P2P) latency metric, P2P latency variation metric, Point-to-Multipoint (P2MP) latency metric and P2MP latency variation metric.

[RFC5541] defines a framework to extend the PCEP to allow a PCE to indicate the set of objective functions it supports. [RFC5541] also define procedure so that a PCC can indicate in a path computation request the required objective function, and a PCE can report in a path computation reply the objective function that was used for path computation. While [RFC5541] and [RFC6006] define various objective functions, these documents do not define objective functions for optimizing network performance criteria such as latency and latency variation. This document extends the [RFC5541] with two new objective functions namely Minimum Latency Path (MLP) OF and Minimum Latency Variation Path (MLVP) OF.

2. PCEP extensions

This section defines PCEP extensions for requirements outlined in Section 1.

2.1. New Metric Object Types

This document defines the following four additional types for the <METRIC> object defined in [RFC5440}. For explanation of

these metrics, the following terminology is used and expanded along the way.

- A network comprises of a set of N links $\{L_i, (i=1\dots N)\}$.
- A path P of a P2P LSP is a list of K links $\{L_{pi}, (i=1\dots K)\}$.

2.1.1. P2P Latency Metric

Link delay metric is defined in [OSPF-TE-METRIC] and [ISIS-TE-METRIC]. P2P latency metric type of <METRIC> object in PCEP encodes the sum of the link delay metric of all links along a P2P Path. Specifically, extending on the above mentioned terminology:

- Link delay metric of link L is denoted $D(L)$.
- P2P latency metric for the Path $P = \text{Sum } \{D(L_{pi}), (i=1\dots K)\}$.

Value for P2P latency metric type is to be assigned by IANA (suggested value: 11).

2.1.2. P2P Latency Variation Metric

Link delay variation metric is defined in [OSPF-TE-METRIC] and [ISIS-TE-METRIC]. P2P latency variation metric type of <METRIC> object in PCEP encodes a function of the link delay variation metric of all links along a P2P Path. Specifically, extending on the above mentioned terminology:

- Latency variation of link L is denoted $DV(L)$.
- P2P latency variation metric for the Path $P = \text{Function } \{DV(L_{pi}), (i=1\dots K)\}$.

Specification of the "Function" used to drive latency variation metric of a path from latency variation metrics of individual links along the path is beyond the scope of this document.

Value for P2P latency variation metric is to be assigned by IANA (suggested value: 12).

2.1.3. P2MP Latency Metric

P2MP latency metric type of <METRIC> object in PCEP encodes the path latency metric for destination that observes the worst

latency metric among all destination of the P2MP tree.
Specifically, extending on the above mentioned terminology:

- A P2MP Tree T comprises of a set of M destinations {Dest_j, (j=1...M)}
- P2P latency metric of the Path to destination Dest_j is denoted by LM(Dest_j).
- P2MP latency metric for the P2MP tree T = Maximum {LM(Dest_j), (j=1...M)}.

Value for P2MP latency metric is to be assigned by IANA
(suggested value: 13).

2.1.4. P2MP Latency Variation Metric

P2MP latency variation metric type of <METRIC> object in PCEP encodes the path latency variation metric for destination that observes the worst latency variation metric among all destination of the P2MP tree. Specifically, extending on the above mentioned terminology:

- A P2MP Tree T comprises of a set of M destinations {Dest_j, (j=1...M)}
- P2P latency variation metric of the Path to destination Dest_j is denoted by LVM(Dest_j).
- P2MP latency variation metric for the P2MP tree T = Maximum {LVM(Dest_j), (j=1...M)}.

Value for P2MP latency variation metric is to be assigned by IANA
(suggested value: 14).

2.2. Handling of New Metric Object Types

This document does not propose any changes to handling of Metric object. Specifically, the new metric types defined in this document are handled in the same fashion as metric types defined in [RFC5440].

2.3. New Objective Functions

This document extends the [RFC 5541] with two new objective functions namely Minimum Latency Path (MLP) OF and Minimum Latency Variation Path (MLVP) OF. The objective function code for each of the new objective function is also defined.

2.3.1. Minimum Latency Path Objective Function

Minimum Latency Path (MLP) OF is defined as an objective function where a path is computed such that latency of the path is minimized.

Objective function code for MLP OF is to be assigned by IANA (suggested value: 9).

2.3.2. Minimum Latency Variation Path Objective Function

Minimum Latency Variation Path (MLVP) OF is defined as an objective function where a path is computed such that latency variation in the path is minimized.

Objective function code for MLVP OF is to be assigned by IANA (suggested value: 10).

2.4. Handling of New Objective Functions

This document does not propose any changes to handling of <OF> object. Specifically, the new OF types defined in this document are handled in the same fashion as OF types defined in [RFC5541].

3. Security Considerations

This document does not introduce any additional security issues beyond those identified in [RFC5440], [RFC5541] and [RFC6006].

4. IANA Considerations

This document defines the following four additional types for the <METRIC> object defined in [RFC5440].

Value	Description
-----	-----
TBA (suggest value: 11)	P2P latency metric
TBA (suggest value: 12)	P2P latency variation metric
TBA (suggest value: 13)	P2MP latency metric
TBA (suggest value: 14)	P2MP latency variation metric

This document defines the following two objective functions codes for the <OF> object defined in [RFC5541].

Value	Description
-----	-----
TBA (suggest value: 9)	Minimum Latency Path (MLP) OF
TBA (suggest value: 10)	Minimum Latency Variation Path (MLVP) OF

5. References

5.1. Normative References

- [RFC5440] Vasseur, JP., Ed., and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, June 2009.
- [DRAFT-OSPF-TE-METRIC] S. Giacalone, D. Ward, J. Drake, A. Atlas, S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", draft-ietf-ospf-te-metric-extensions, work in progress.
- [DRAFT-ISIS-TE-METRIC] S. Previdi, S. Giacalone, D. Ward, J. Drake, A. Atlas, C. Filsfils, "IS-IS Traffic Engineering (TE) Metric Extensions", draft-previdi-isis-te-metric-extensions, work in progress.

5.2. Informative References

- [RFC6006] Zhao, Q., Ed., King, D., Ed., Verhaeghe, F., Takeda, T., Ali, Z., and J. Meuric, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 6006, September 2010.

Authors' Addresses

Zafar Ali
Cisco Systems
Email: zali@cisco.com

George Swallow
Cisco Systems
swallow@cisco.com

Clarence Filsfils
Cisco Systems
cfilsfil@cisco.com

Siva Sivabalan
Cisco Systems
msiva@cisco.com

Stefano Previdi
Cisco Systems
sprevidi@cisco.com

Kenji Kumaki
KDDI Corporation
Email: ke-kumaki@kddi.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 12, 2013

E. Crabbe
Google, Inc.
I. Minei
Juniper Networks, Inc.
S. Sivabalan
Cisco Systems, Inc.
R. Varga
Pantheon Technologies SRO
October 9, 2012

PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model
draft-crabbe-pce-pce-initiated-lsp-00

Abstract

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

The extensions described in [I-D.ietf-pce-stateful-pce] provide stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP) via PCEP, for a model where the PCC delegates control over one or more locally configured LSPs to the PCE. This document describes the creation and deletion of PCE-initiated LSPs under the stateful PCE model.

Requirements Language

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

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 April 12, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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

1. Introduction	4
2. Terminology	4
3. Architectural Overview	4
3.1. Motivation	5
3.2. Operation overview	5
4. Support of PCE-initiated LSPs	6
4.1. Stateful PCE Capability TLV	6
5. PCE-initiated LSP creation	7
5.1. The LSP Create Message	7
6. LSP delegation and cleanup	9
6.1. LSP delegation procedures	9
6.2. LSP cleanup procedures	9
6.2.1. LSP-CLEANUP TLV	10
7. IANA considerations	10
7.1. PCEP-Error Object	11
7.2. PCEP TLV Type Indicators	11
8. Security Considerations	11
8.1. Malicious PCE	11
9. Acknowledgements	12
10. References	12
10.1. Normative References	12
10.2. Informative References	13
Authors' Addresses	13

1. Introduction

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

Stateful pce [I-D.ietf-pce-stateful-pce] specifies a set of extensions to PCEP to enable stateful control of TE LSPs between and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP state synchronization between PCCs and PCEs, delegation of control of LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions and focuses on a model where LSPs are configured on the PCC and control over them is delegated to the PCE.

This document describes the setup and teardown of PCE-initiated LSPs under the stateful PCE model, without the need for local configuration on the PCC, thus allowing for a dynamic network that is centrally controlled and deployed.

2. 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, Delegation, Delegation Timeout Interval, LSP State Report, LSP Update Request.

The following terms are defined in this document:

PCE-initiated LSP: LSP that is instantiated as a result of a request from the PCE.

LSP cleanup timer: PCE-defined timer for cleanup of PCE-initiated LSPs that are no longer delegated to a PCE.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

3. Architectural Overview

3.1. Motivation

[I-D.ietf-pce-stateful-pce] provides stateful control over LSPs that are locally configured on the PCC. This model relies on the LER taking an active role in delegating locally configured LSPs to the PCE, and is well suited in environments where the LSP placement is fairly static. However, in environments where the LSP placement needs to change in response to application demands, it is useful to support dynamic creation and tear down of LSPs. The ability for a PCE to trigger the creation of LSPs on demand can make possible agile software-driven network operation, and can be seamlessly integrated into a controller-based network architecture, where intelligence in the controller can determine when and where to set up paths.

A possible use case is one of a software-driven network, where applications request network resources and paths from the network infrastructure. For example, an application can request a path with certain constraints between two LSRs by contacting the PCE. The PCE can compute a path satisfying the constraints, and instruct the head end LSR to create and signal it. When the path is no longer required by the application, the PCE can request its teardown.

Another use case is that of demand engineering, where a PCE with visibility into both the network state and the demand matrix can anticipate and optimize how traffic is distributed across the infrastructure. Such optimizations may require creating new paths across the infrastructure.

3.2. Operation overview

A PCC indicates its ability to support PCE provisioned dynamic LSPs during the PCEP Initialization Phase via a new flag in the STATEFUL-PCE-CAPABILITY TLV (see details in Section 4.1).

The decision when to create a PCE-initiated LSP is out of the scope of this document. To instantiate an LSP, the PCE sends a new message, the LSP Create Request (PCCreate) message to the PCC. The LSP Create Request MUST include the END-POINTS and LSPA objects, and the LSPA object MUST include the SYMBOLIC-PATH-NAME TLV. The PCC creates the LSP using the attributes communicated by the PCE, and local values for the unspecified parameters. It assigns a unique LSP-ID for the LSP and automatically delegates the LSP to the PCE. It then generates an LSP State Report (PCRpt) for the LSP, carrying the LSP-ID and the delegation bit. The PCE may update the attributes of the LSP via subsequent PCUpd messages.

Subsequent LSP State Report and LSP Update Request for the LSP will carry the PCC-assigned LSP-ID, which uniquely identifies the LSP.

The LSPA Object included in these messages MUST carry the SYMBOLIC-PATH-NAME TLV which will be used to correlate between the PCC-assigned LSP-ID and the LSP. See details in Section 5.

Removal of PCE-initiated LSPs is done by the PCE by setting the R flag in the LSP Object in the PCUpd message. Upon receiving the PCUpd message with the R Flag set, the PCC deletes the LSP. See details in Section 5.

Once instantiated, a PCRpt is generated for the LSP, with the delegation bit set. After this, the delegation procedures for PCE-initiated LSPs are the same as for PCC initiated LSPs. Upon session failure, PCE-initiated LSPs are not immediately removed, in order to avoid LSP flap and service interruption. However, to allow for network cleanup without manual intervention, such "orphan" PCE-initiated LSPs must be either adopted by a different PCE or cleaned up within a time interval. This time is negotiated between PCE and PCC at session initialization time. See details in Section 6.

4. Support of PCE-initiated LSPs

A PCC indicates its ability to support PCE provisioned dynamic LSPs 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 I (LSP-INSTANTIATION-CAPABILITY) flag is introduced to indicate support for instantiation of PCE-initiated LSPs. A PCE wishing to initiate LSPs, can do so 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 I flag. A PCE or PCC that advertise support of LSP initiation MUST also advertise a cleanup time for the removal of such LSPs. The cleanup time is advertised via a new TLV in the Open Object, the LSP-CLEANUP TLV, discussed in Section 6, and the value is negotiated to the lower one advertised on a session.

4.1. Stateful PCE Capability TLV

The format of the STATEFUL-PCE-CAPABILITY TLV is shown in the following figure:

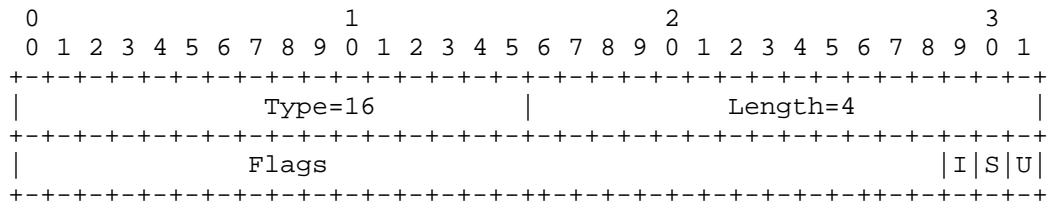


Figure 1: STATEFUL-PCE-CAPABILITY TLV format

The type of the TLV is defined in [I-D.ietf-pce-stateful-pce] and it has a fixed length of 4 octets.

The value comprises a single field - Flags (32 bits). The U and S bits are defined in [I-D.ietf-pce-stateful-pce].

If set to 1 by a PCC, the I Flag indicates that the PCC allows instantiation of an LSP by a PCE. If set to 1 by a PCE, the I flag indicates that the PCE will attempt to instantiate LSPs. The LSP-INSTANTIATION-CAPABILITY flag must be set by both PCC and PCE in order to support PCE-initiated LSP instantiation.

Unassigned bits are considered reserved. They MUST be set to 0 on transmission and MUST be ignored on receipt.

5. PCE-initiated LSP creation

To create a PCE-initiated LSP, a PCE sends a PCCreate message to a PCC, which include a set of objects and TLVs describing the LSP to be instantiated. The message format, the objects and TLVs are discussed separately below.

5.1. The LSP Create Message

A Path Computation LSP Create message (also referred to as PCCreate message) is a PCEP message sent by a PCE to a PCC to trigger an LSP instantiation. The Message-Type field of the PCEP common header for the PCCreate message is set to [TBD].

The PCCreate message MUST include the END-POINTS and the LSPA objects. In the LSPA object, it MUST include the SYMBOLIC-PATH-NAME TLV for the LSP. The PCCreate message MAY include other attributes for the LSP. If specified, the PCC MUST use them for the LSP instantiation, otherwise it MUST use its locally configured values. The error messages will be specified in a future version of this document.

The format of a PCCreate message is as follows:

```
<PCCreate Message> ::= <Common Header>
                        <lsp-instantiation-list>
```

Where:

```
<lsp-instantiation-list> ::= <lsp-instantiation-request>[<lsp-instantiation-1
ist>]
```

```
<lsp-instantiation-request> ::= <END-POINTS>
                                <LSPA>
                                [<ERO>]
                                [<BANDWIDTH>]
                                [<metric-list>]
```

Where:

```
<metric-list> ::= <METRIC>[<metric-list>]
```

The END-POINTS Object contains the source and destination addresses for provisioning the PCE-initiated LSP. If the END-POINTS Object is missing, the PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=3 (END-POINTS Object missing).

The LSPA Object MUST include the SYMBOLIC-PATH-NAME TLV, which will be used to correlate between the PCC-assigned LSP-ID and the LSP. The symbolic name used for provisioning PCE-initiated LSPs must not have conflict with the LSP name of any existing LSP in the PCC. (Existing LSPs may be either statically configured, or initiated by another PCE). If there is conflict with the LSP name, the PCC MUST send a PCErr message with Error-type=TBD (Invalid Parameter) and Error-value=TBD (Bad Symbolic Path Name). The only exception to this rule is for LSPs for which the LSP-cleanup timer is running (see Section 6).

PCE-initiated removal of a PCE-initiated LSP is done by setting the R (remove) flag in the LSP Object in the PCUpd request from the PCE. The definition of the R bit is updated as follows:

R (Remove - 1 bit): On PCRpt messages the R Flag indicates that the LSP has been removed from the PCC. Upon receiving a PCRpt message with the R Flag set to 1, the PCE SHOULD remove all state related to the LSP from its database. In PCUpd messages the R flag indicates that the PCE wishes to disable the LSP. Upon receiving the PCUpd message with the R Flag set for a PCE-initiated LSP, the PCC tears

down the LSP and removes its state.

A PCC SHOULD be able to place a limit on either the number of LSPs or the percentage of resources that are allocated to honor PCE-initiated LSP requests. As soon as that limit is reached, the PCC MUST send a PCErr message of type 19 (Invalid Operation) and value TBD "PCE-initiated limit reached" and is free to drop any incoming PCUpd messages without additional processing.

A PCC SHOULD relay to the PCE errors it encounters in the setup of PCE-initiated LSP. The error codes and error processing will be detailed in a future version of this document.

6. LSP delegation and cleanup

6.1. LSP delegation procedures

PCE-initiated LSPs are automatically delegated by the PCC to the PCE upon instantiation. The PCC MUST delegate the LSP to the PCE by setting the delegation bit to 1 in the PCRpt that includes the assigned LSP-Id. All subsequent messages from the PCC must have the delegation bit set to 1. The PCC cannot revoke the delegation for PCE-initiated LSPs for an active PCEP session. Sending a PCRpt message with the delegation bit set to 0 results in a PCErr message of type 19 (Invalid Operation) and value TBD "Delegation for PCE-initiated LSP cannot be revoked".

A PCE MAY return a delegation to the PCC, to allow for LSP transfer between PCEs. Doing so MUST trigger the LSP cleanup timer described in Section 6.2.

Control over PCE-initiated LSPs reverts to the PCC at the expiration of the delegation timeout. To obtain control of a PCE-initiated LSP, a PCE (either the original or one of its backups) sends a PCCreate message specifying the endpoints and symbolic name (the same process used when initiating an LSP from the PCE). See more in the next section.

6.2. LSP cleanup procedures

The LSP cleanup timer ensures that a PCE crash does not result in automatic and immediate disruption for the services using PCE-initiated LSPs. PCE-initiated LSPs are not be removed immediately upon PCE failure. Instead, they are cleaned up on the expiration of this timer. This allows for network cleanup without manual intervention. The LSP cleanup timer is advertised in the session open message via a mandatory TLV for sessions where PCE-initiated

LSPs are supported. The timer is started upon PCEP session failure and is stopped when the LSP is delegated to a PCE. Both PCE and PCC advertise a value for this timer, and the timer value is negotiated to the lower value of the two.

6.2.1. LSP-CLEANUP TLV

The LSP-CLEANUP TLV is advertised in the Open Object and is mandatory when the I flag is set in the STATEFUL-PCE-CAPABILITY TLV. The LSP-CLEANUP TLV contains the time in seconds that the PCC has to wait before cleaning up any PCE-initiated LSPs belonging to a particular PCEP session when a PCEP session terminates. Both PCE and PCC advertise a value for the cleanup time, and the cleanup timer is set to the lower of the two. The timer is triggered on PCEP session failure and reset when the LSP is delegated to a PCE.

Failure to include the mandatory LSP-CLEANUP TLV in the Open Object when the I flag is set MUST trigger PCerr of type 6 (Mandatory Object missing) and value TBD (LSP-CLEANUP TLV missing).

The format of the LSP-CLEANUP TLV is shown in the following figure:

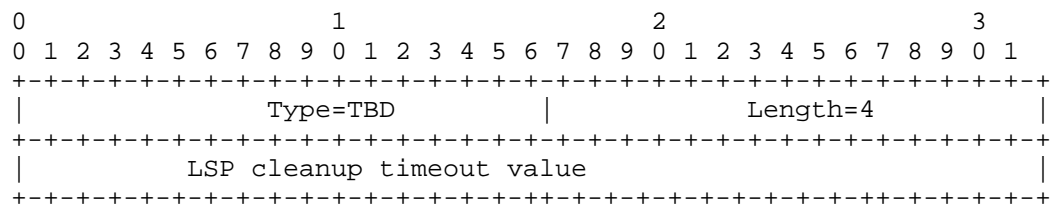


Figure 2: LSP-CLEANUP TLV format

The type of the TLV is TBD and it has a fixed length of 4 octets.

The value comprises a single field, the LSP cleanup timeout value.

The time in seconds to wait before cleaning up PCE-initiated LSPs. Zero means immediate removal. The value 0xFFFFFFFF is reserved.

A PCE may take control of the dynamic LSPs for which the LSP cleanup timer is running by sending an PCCreate request for the LSP. In this case, the "Bad Symbolic Path Name" error MUST NOT be generated, the LSP MUST be delegated and the cleanup timer MUST be stopped.

7. IANA considerations

7.1. PCEP-Error Object

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

Error-Type	Meaning
6	Mandatory Object missing Error-value=8: LSP cleanup TLV missing
19	Invalid operation Error-value=TBD: PCE-initiated LSP limit reached Error-value=TBD: Delegation for PCE-initiated LSP cannot be revoked

7.2. PCEP TLV Type Indicators

This document defines the following new PCEP TLVs:

Value	Meaning	Reference
???	LSP cleanup	This document

8. Security Considerations

The security considerations described in [I-D.ietf-pce-stateful-pce] apply to the extensions described in this document. Additional considerations related to a malicious PCE are introduced.

8.1. Malicious PCE

The LSP instantiation mechanism described in this document allows a PCE to generate state on the PCC and throughout the network. As a result, it introduces a new attack vector: an attacker may flood the PCC with LSP instantiation requests and consume network and LSR resources, either by spoofing messages or by compromising the PCE itself.

A PCC can protect itself from such an attack by imposing a limit on either the number of LSPs or the percentage of resources that are allocated to honor PCE-initiated LSP requests. As soon as that limit is reached, the PCC MUST send a PCErr message of type 19 (Invalid Operation) and value TBD "PCE-initiated LSP limit reached" (XXX TBD add to the IANA section) and is free to drop any incoming PCUpd messages without additional processing.

Rapid flaps triggered by the PCE can also be an attack vector. This will be discussed in a future version of this document.

9. Acknowledgements

We would like to thank Jan Medved, Ambrose Kwong and Raveendra Trovi for their contributions to this document.

10. References

10.1. Normative References

- [I-D.ietf-pce-stateful-pce]
Crabbe, E., Medved, J., Varga, R., and I. Minei, "PCEP Extensions for Stateful PCE",
draft-ietf-pce-stateful-pce-01 (work in progress),
July 2012.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [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.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC5088] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax

Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.

10.2. Informative References

- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC3346] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D., Christian, B., and W. Lai, "Applicability Statement for Traffic Engineering with MPLS", RFC 3346, August 2002.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, December 2008.
- [RFC5557] Lee, Y., Le Roux, J.L., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, July 2009.

Authors' Addresses

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: edc@google.com

Ina Minei
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: ina@juniper.net

Siva Sivabalan
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: msiva@cisco.com

Robert Varga
Pantheon Technologies SRO
Mlynske Nivy 56
Bratislava 821 05
Slovakia

Email: robert.varga@pantheon.sk

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 13, 2014

E. Crabbe
Google, Inc.
I. Minei
Juniper Networks, Inc.
S. Sivabalan
Cisco Systems, Inc.
R. Varga
Pantheon Technologies SRO
October 10, 2013

PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model
draft-crabbe-pce-pce-initiated-lsp-03

Abstract

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

The extensions described in [I-D.ietf-pce-stateful-pce] provide stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP) via PCEP, for a model where the PCC delegates control over one or more locally configured LSPs to the PCE. This document describes the creation and deletion of PCE-initiated LSPs under the stateful PCE model.

Requirements Language

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

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 April 13, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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

1. Introduction	4
2. Terminology	4
3. Architectural Overview	4
3.1. Motivation	4
3.2. Operation overview	5
4. Support of PCE-initiated LSPs	6
4.1. Stateful PCE Capability TLV	7
5. PCE-initiated LSP instantiation and deletion	7
5.1. The LSP Initiate Message	7
5.2. The R flag in the SRP Object	8
5.3. LSP instantiation	9
5.3.1. The Create flag	11
5.4. LSP deletion	11
6. LSP delegation and cleanup	12
7. Implementation status	12
8. IANA considerations	13
8.1. PCEP Messages	13
8.2. LSP Object	13
8.3. PCEP-Error Object	14
9. Security Considerations	14
9.1. Malicious PCE	14
9.2. Malicious PCC	15
10. Acknowledgements	15
11. References	15
11.1. Normative References	15
11.2. Informative References	16
Authors' Addresses	17

1. Introduction

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

Stateful pce [I-D.ietf-pce-stateful-pce] specifies a set of extensions to PCEP to enable stateful control of TE LSPs between and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP state synchronization between PCCs and PCEs, delegation of control of LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions and focuses on a model where LSPs are configured on the PCC and control over them is delegated to the PCE.

This document describes the setup, maintenance and teardown of PCE-initiated LSPs under the stateful PCE model, without the need for local configuration on the PCC, thus allowing for a dynamic network that is centrally controlled and deployed.

2. 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, Delegation, Redelegation Timeout, State Timeout Interval LSP State Report, LSP Update Request.

The following terms are defined in this document:

PCE-initiated LSP: LSP that is instantiated as a result of a request from the PCE.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

3. Architectural Overview

3.1. Motivation

[I-D.ietf-pce-stateful-pce] provides stateful control over LSPs that are locally configured on the PCC. This model relies on the LER taking an active role in delegating locally configured LSPs to the

PCE, and is well suited in environments where the LSP placement is fairly static. However, in environments where the LSP placement needs to change in response to application demands, it is useful to support dynamic creation and tear down of LSPs. The ability for a PCE to trigger the creation of LSPs on demand can make possible agile software-driven network operation, and can be seamlessly integrated into a controller-based network architecture, where intelligence in the controller can determine when and where to set up paths.

A possible use case is one of a software-driven network, where applications request network resources and paths from the network infrastructure. For example, an application can request a path with certain constraints between two LSRs by contacting the PCE. The PCE can compute a path satisfying the constraints, and instruct the head end LSR to instantiate and signal it. When the path is no longer required by the application, the PCE can request its teardown.

Another use case is one of dynamically adjusting aggregate bandwidth between two points in the network using multiple LSPs. This functionality is very similar to auto-bandwidth, but allows for providing the desired capacity through multiple LSPs. This approach overcomes two of the limitations auto-bandwidth can experience: 1) growing the capacity between the endpoints beyond the capacity of individual links in the path and 2) achieving good bin-packing through use of several small LSPs instead of a single large one. The number of LSPs varies based on the demand, and LSPs are created and deleted dynamically to satisfy the bandwidth requirements.

Another use case is that of demand engineering, where a PCE with visibility into both the network state and the demand matrix can anticipate and optimize how traffic is distributed across the infrastructure. Such optimizations may require creating new paths across the infrastructure.

3.2. Operation overview

A PCC or PCE indicates its ability to support PCE provisioned dynamic LSPs during the PCEP Initialization Phase via a new flag in the STATEFUL-PCE-CAPABILITY TLV (see details in Section 4.1).

The decision when to instantiate or delete a PCE-initiated LSP is out of the scope of this document. To instantiate or delete an LSP, the PCE sends a new message, the Path Computation LSP Initiate Request (PCInitiate) message to the PCC. The LSP Initiate Request MUST include the SRP and LSP objects, and the LSP object MUST include the Symbolic Path Name TLV and MUST have a PLSP-ID of 0.

For an instantiation operation, the PCE MUST include the ERO and END-

POINTS object and may include various attributes as per [RFC5440]. The PCC creates the LSP using the attributes communicated by the PCE, and local values for the unspecified parameters. It assigns a unique PLSP-ID for the LSP and automatically delegates the LSP to the PCE. It also generates an LSP State Report (PCRpt) for the LSP, carrying the newly assigned PLSP-ID and indicating the delegation via the Delegate flag in the LSP object. In addition to the Delegate flag, the PCC also sets the Create flag in the LSP object (see Section 5.3.1), to indicate that the LSP was created as a result of a PCInitiate message. This PCRpt message MUST include the SRP object, with the SRP-id-number used in the SRP object of the PCInitiate message. The PCE may update the attributes of the LSP via subsequent PCUpd messages. Subsequent LSP State Report and LSP Update Request for the LSP will carry the PCC-assigned PLSP-ID, which uniquely identifies the LSP. See details in Section 5.3.

Once instantiated, the delegation procedures for PCE-initiated LSPs are the same as for PCC initiated LSPs as described in [I-D.ietf-pce-stateful-pce]. This applies to the case of a PCE failure as well. In order to allow for network cleanup without manual intervention, the PCC SHOULD support removal of PCE-initiated LSPs as one of the behaviors applied on expiration of the State Timeout Interval [I-D.ietf-pce-stateful-pce]. The behavior SHOULD be picked based on local policy, and can result either in LSP removal, or into reverting to operator-defined default parameters. See details in Section 6. A PCE may return a delegation to the PCC in order to facilitate re-delegation of its LSPs to an alternate PCE.

To indicate a delete operation, the PCE MUST use the R flag in the SRP object in a PCUpd message. As a result of the deletion request, the PCC MUST remove all state related to the LSP, and send a PCRpt with the R flag set in the LSP object for the removed state. See details in Section 5.4.

4. Support of PCE-initiated LSPs

A PCC indicates its ability to support PCE provisioned dynamic LSPs 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 I (LSP-INSTANTIATION-CAPABILITY) flag is introduced to indicate support for instantiation of PCE-initiated LSPs. A PCE can initiate LSPs 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 I flag.

4.1. Stateful PCE Capability TLV

The format of the STATEFUL-PCE-CAPABILITY TLV is shown in the following figure:

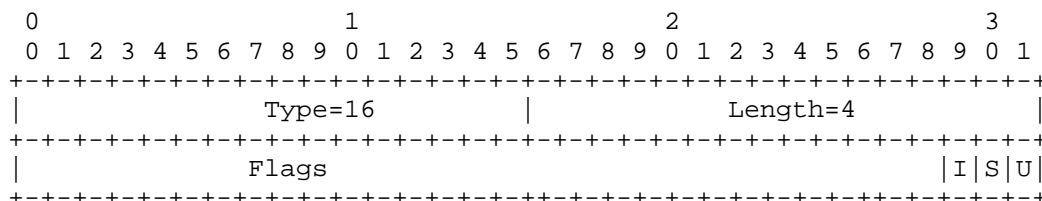


Figure 1: STATEFUL-PCE-CAPABILITY TLV format

The type of the TLV is defined in [I-D.ietf-pce-stateful-pce] and it has a fixed length of 4 octets.

The value comprises a single field - Flags (32 bits). The U and S bits are defined in [I-D.ietf-pce-stateful-pce].

I (LSP-INSTANTIATION-CAPABILITY - 1 bit): If set to 1 by a PCC, the I Flag indicates that the PCC allows instantiation of an LSP by a PCE. If set to 1 by a PCE, the I flag indicates that the PCE will attempt to instantiate LSPs. The LSP-INSTANTIATION-CAPABILITY flag must be set by both PCC and PCE in order to support PCE-initiated LSP instantiation.

Unassigned bits are considered reserved. They MUST be set to 0 on transmission and MUST be ignored on receipt.

5. PCE-initiated LSP instantiation and deletion

To initiate an LSP, a PCE sends a PCInitiate message to a PCC. The message format, objects and TLVs are discussed separately below for the creation and the deletion cases.

5.1. The LSP Initiate Message

A Path Computation LSP Initiate Message (also referred to as PCInitiate message) is a PCEP message sent by a PCE to a PCC to trigger LSP instantiation or deletion. The Message-Type field of the PCEP common header for the PCInitiate message is set to [TBD]. The PCInitiate message MUST include the SRP and the LSP objects, and may contain other objects, as discussed later in this section. If either the SRP or the LSP object is missing, the PCC MUST send a PCErr as described in [I-D.ietf-pce-stateful-pce]. LSP instantiation is done

by sending an LSP Initiate Message with an LSP object with the reserved PLSP-ID 0. LSP deletion is done by sending an LSP Initiate Message with an LSP object carrying the PLSP-ID of the LSP to be removed and an SRP object with the R flag set (see Section 5.2).

The format of a PCInitiate message for LSP instantiation is as follows:

```
<PCInitiate Message> ::= <Common Header>
                           <PCE-initiated-lsp-list>
```

Where:

```
<PCE-initiated-lsp-list> ::= <PCE-initiated-lsp-request>[<PCE-initiated-lsp-list>]
```

```
<PCE-initiated-lsp-request> ::= (<PCE-initiated-lsp-instantiation>|<PCE-initiated-lsp-deletion>)
```

```
<PCE-initiated-lsp-instantiation> ::= <SRP>
                                       <LSP>
                                       <END-POINTS>
                                       <ERO>
                                       [<attribute-list>]
```

```
<PCE-initiated-lsp-deletion> ::= <SRP>
                                   <LSP>
```

Where:

<attribute-list> is defined in [RFC5440] and extended by PCEP extensions.

The SRP object is used to correlate between initiation requests sent by the PCE and the error reports and state reports sent by the PCC. Every request from the PCE receives a new SRP-ID-number. This number is unique per PCEP session and is incremented each time an operation (initiation, update, etc) is requested from the PCE. The value of the SRP-ID-number MUST be echoed back by the PCC in PCErr and PCrpt messages to allow for correlation between requests made by the PCE and errors or state reports generated by the PCC. Details of the SRP object and its use can be found in [I-D.ietf-pce-stateful-pce].

5.2. The R flag in the SRP Object

The format of the SRP object is shown Figure 2:

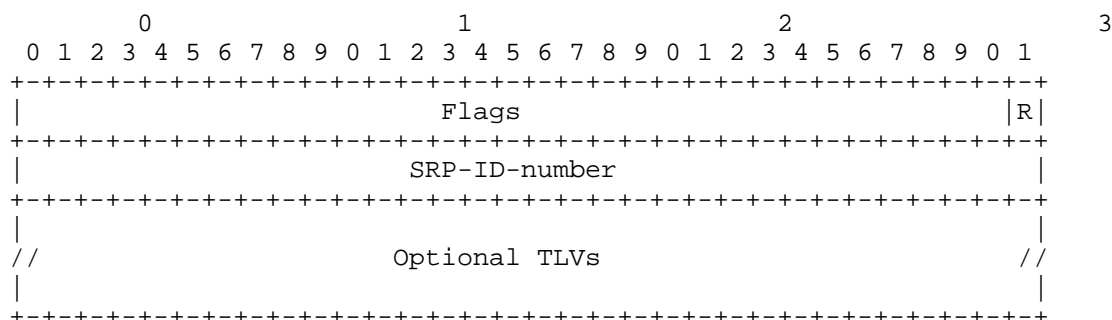


Figure 2: The SRP Object format

The type object is defined in [I-D.ietf-pce-stateful-pce].

A new flag is defined to indicate a delete operation initiated by the PCE:

R (LSP-REMOVE - 1 bit): If set to 1, it indicates a removal request initiated by the PCE.

5.3. LSP instantiation

LSP instantiation is done by sending an LSP Initiate Message with an LSP object with the reserved PLSP-ID 0. The LSP is set up using RSVP-TE, extensions for other setup methods are outside the scope of this draft.

Receipt of a PCInitiate Message with a non-zero PLSP-ID and the R flag in the SRP object set to zero results in a PCErr message of type 19 (Invalid Operation) and value 8 (non-zero PLSP-ID in LSP initiation request).

The END-POINTS Object is mandatory for an instantiation request of an RSVP-signaled LSP. It contains the source and destination addresses for provisioning the LSP. If the END-POINTS Object is missing, the PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=3 (END-POINTS Object missing).

The ERO Object is mandatory for an instantiation request. It contains the ERO for the LSP. If the ERO Object is missing, the PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=9 (ERO Object missing).

The LSP Object MUST include the SYMBOLIC-PATH-NAME TLV, which will be used to correlate between the PCC-assigned PLSP-ID and the LSP. If

the TLV is missing, the PCC MUST send a PCErr message with Error-type=6(Mandatory object missing) and Error-value=14 (SYMBOLIC-PATH-NAME TLV missing). The symbolic name used for provisioning PCE-initiated LSPs must not have conflict with the LSP name of any existing LSP in the PCC. (Existing LSPs may be either statically configured, or initiated by another PCE). If there is conflict with the LSP name, the PCC MUST send a PCErr message with Error-type=23 (Bad Parameter value) and Error-value=1 (SYMBOLIC-PATH-NAME in use). The only exception to this rule is for LSPs for which the State timeout timer is running (see Section 6).

The PCE MAY include various attributes as per [RFC5440]. The PCC MUST use these values in the LSP instantiation, and local values for unspecified parameters. After the LSP setup, the PCC MUST send a PCRpt to the PCE, reflecting these values. The SRP object in the PCRpt message MUST echo the value of the PCInitiate message that triggered the setup. LSPs that were instantiated as a result of a PCInitiate message MUST have the C flag set in the LSP object.

If the PCC determines that the LSP parameters proposed in the PCInitiate message are unacceptable, it MUST trigger a PCErr with error-type=TBD (PCE instantiation error) and error-value=1 (Unacceptable instantiation parameters). If the PCC encounters an internal error during the processing of the PCInitiate message, it MUST trigger a PCErr with error-type=TBD (PCE instantiation error) and error-value=2 (Internal error).

A PCC MUST relay to the PCE errors it encounters in the setup of PCE-initiated LSP by sending a PCErr with error-type=TBD (PCE instantiation error) and error-value=3 (RSVP signaling error). The PCErr MUST echo the SRP-id-number of the PCInitiate message. The PCEP-ERROR object SHOULD include the RSVP Error Spec TLV (if an ERROR SPEC was returned to the PCC by a downstream node). After the LSP is set up, errors in RSVP signaling are reported in PCRpt messages, as described in [I-D.ietf-pce-stateful-pce].

A PCC SHOULD be able to place a limit on either the number of LSPs or the percentage of resources that are allocated to honor PCE-initiated LSP requests. As soon as that limit is reached, the PCC MUST send a PCErr message of type 19 (Invalid Operation) and value TBD "PCE-initiated limit reached" and is free to drop any incoming PCInitiate messages without additional processing.

Similarly, the PCE SHOULD be able to place a limit on either the number of LSP initiation requests pending for a particular PCC, or on the time it waits for a response (positive or negative) to a PCInitiate request from a PCC and MAY take further action (such as closing the session or removing all its LSPs) if this limit is

reached.

On successful completion of the LSP instantiation, the PCC assigns a PLSP-ID, and immediately delegates the LSP to the PCE by sending a PCRpt with the Delegate flag set. The PCRpt MUST include the SRP-ID-number of the PCInitiate request that triggered its creation. PCE-initiated LSPs are identified with the Create flag in the LSP Object.

5.3.1. The Create flag

The LSP object is defined in [I-D.ietf-pce-stateful-pce] and included here for easy reference.

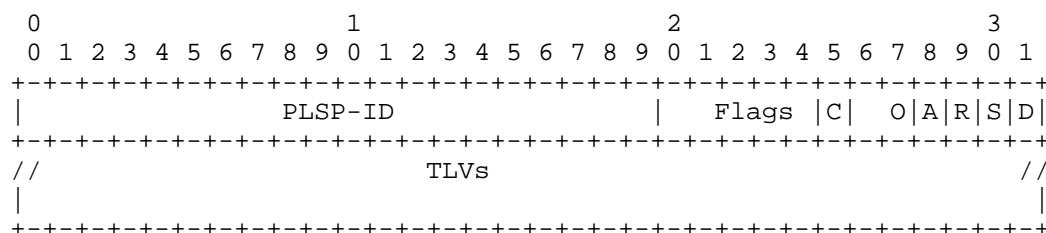


Figure 3: The LSP Object format

A new flag, the Create (C) flag is introduced. On a PCRpt message, the C Flag set to 1 indicates that this LSP was created via a PCInitiate message. The C Flag MUST be set to 1 on each PCRpt message for the duration of existence of the LSP. The Create flag allows PCEs to be aware of which LSPs were PCE-initiated (a state that would otherwise only be known by the PCC and the PCE that initiated them).

5.4. LSP deletion

PCE-initiated removal of a PCE-initiated LSP is done by setting the R (remove) flag in the SRP Object in the PCInitiate message from the PCE. The LSP is identified by the PLSP-ID in the LSP object. If the PLSP-ID is unknown, the PCC MUST generate a PCErr with error type 19, error value 3, "Unknown PLSP-ID". A PLSP-ID of zero removes all LSPs that were initiated by the PCE. If the PLSP-ID specified in the PCInitiate message is not delegated to the PCE, the PCC MUST send a PCErr message indicating "LSP is not delegated" (Error code 19, error value 1 ([I-D.ietf-pce-stateful-pce])). If the PLSP-ID specified in the PCInitiate message was not created by the PCE, the PCC MUST send a PCErr message indicating "LSP is not PCE initiated" (Error code 19, error value TBD). Following the removal of the LSP, the PCC MUST send a PCRpt as described in [I-D.ietf-pce-stateful-pce]. The SRP object in the PCRpt MUST include the SRP-ID-number from the

PCInitiate message that triggered the removal. The R flag in the SRP object SHOULD be set.

6. LSP delegation and cleanup

PCE-initiated LSPs are automatically delegated by the PCC to the PCE upon instantiation. The PCC MUST delegate the LSP to the PCE by setting the delegation bit to 1 in the PCRpt that includes the assigned PLSP-ID. All subsequent messages from the PCC must have the delegation bit set to 1. The PCC cannot revoke the delegation for PCE-initiated LSPs for an active PCEP session. Sending a PCRpt message with the delegation bit set to 0 results in a PCErr message of type 19 (Invalid Operation) and value TBD "Delegation for PCE-initiated LSP cannot be revoked". The PCE MAY further react by closing the session.

A PCE MAY return a delegation to the PCC, to allow for LSP transfer between PCEs. Doing so MUST trigger the State Timeout Interval timer ([I-D.ietf-pce-stateful-pce]).

In case of PCEP session failure, control over PCE-initiated LSPs reverts to the PCC at the expiration of the redelegation timeout. To obtain control of a PCE-initiated LSP, a PCE (either the original or one of its backups) sends a PCInitiate message, including just the SRP and LSP objects, and carrying the PLSP-ID of the LSP it wants to take control of. Receipt of a PCInitiate message with a non-zero PLSP-ID normally results in the generation of a PCErr. If the State Timeout timer is running, the PCC MUST NOT generate an error and redelegate the LSP to the PCE. The State Timeout timer is stopped upon the redelegation. After obtaining control of the LSP, the PCE may remove it using the procedures described in this document.

The State Timeout timer ensures that a PCE crash does not result in automatic and immediate disruption for the services using PCE-initiated LSPs. PCE-initiated LSPs are not be removed immediately upon PCE failure. Instead, they are cleaned up on the expiration of this timer. This allows for network cleanup without manual intervention. The PCC SHOULD support removal of PCE-initiated LSPs as one of the behaviors applied on expiration of the State Timeout Interval [I-D.ietf-pce-stateful-pce]. The behavior SHOULD be picked based on local policy, and can result either in LSP removal, or into reverting to operator-defined default parameters.

7. Implementation status

This section to be removed by the RFC editor.

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in RFC 6982. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to RFC 6982, "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

Two vendors are implementing the extensions described in this draft and have included the functionality in releases that will be shipping in the near future. An additional entity is working on implementing these extensions in the scope of research projects.

8. IANA considerations

8.1. PCEP Messages

This document defines the following new PCEP messages:

Value	Meaning	Reference
12	Initiate	This document

8.2. LSP Object

The following values are defined in this document for the Flags field in the LSP Object.

Bit	Description	Reference
24	Create	This document

8.3. PCEP-Error Object

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

Error-Type	Meaning
6	Mandatory Object missing Error-value=13: LSP cleanup TLV missing Error-value=14: SYMBOLIC-PATH-NAME TLV missing
19	Invalid operation Error-value=6: PCE-initiated LSP limit reached Error-value=7: Delegation for PCE-initiated LSP cannot be revoked Error-value=8: Non-zero PLSP-ID in LSP initiation request
23	Bad parameter value Error-value=1: SYMBOLIC-PATH-NAME in use
24	LSP instantiation error Error-value=1: Unacceptable instantiation parameters Error-value=2: Internal error Error-value=3: RSVP signaling error

9. Security Considerations

The security considerations described in [I-D.ietf-pce-stateful-pce] apply to the extensions described in this document. Additional considerations related to a malicious PCE are introduced.

9.1. Malicious PCE

The LSP instantiation mechanism described in this document allows a PCE to generate state on the PCC and throughout the network. As a result, it introduces a new attack vector: an attacker may flood the PCC with LSP instantiation requests and consume network and LSR resources, either by spoofing messages or by compromising the PCE itself.

A PCC can protect itself from such an attack by imposing a limit on either the number of LSPs or the percentage of resources that are allocated to honor PCE-initiated LSP requests. As soon as that limit is reached, the PCC MUST send a PCErr message of type 19 (Invalid Operation) and value 3 "PCE-initiated LSP limit reached" and is free to drop any incoming PCInitiate messages for LSP instantiation without additional processing.

Rapid flaps triggered by the PCE can also be an attack vector. This will be discussed in a future version of this document.

9.2. Malicious PCC

The LSP instantiation mechanism described in this document requires the PCE to keep state for LSPs that it instantiates and relies on the PCC responding (with either a state report or an error message) to requests for LSP instantiation. A malicious PCC or one that reached the limit of the number of PCE-initiated LSPs, can ignore PCE requests and consume PCE resources. A PCE can protect itself by imposing a limit on the number of requests pending, or by setting a timeout and it MAY take further action such as closing the session or removing all the LSPs it initiated.

10. Acknowledgements

We would like to thank Jan Medved, Ambrose Kwong, Ramon Casellas, Dhruv Dhody, and Raveendra Trovi for their contributions to this document.

11. References

11.1. 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-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.
- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [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.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC5088] Le Roux, J.L., Vasseur, J.P., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.

- [RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.

11.2. Informative References

- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC3346] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D., Christian, B., and W. Lai, "Applicability Statement for Traffic Engineering with MPLS", RFC 3346, August 2002.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, December 2008.
- [RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path

Computation Element Communication Protocol (PCEP)
Requirements and Protocol Extensions in Support of Global
Concurrent Optimization", RFC 5557, July 2009.

[RFC6982] Sheffer, Y. and A. Farrel, "Improving Awareness of Running
Code: The Implementation Status Section", RFC 6982,
July 2013.

Authors' Addresses

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: edc@google.com

Ina Minei
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: ina@juniper.net

Siva Sivabalan
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: msiva@cisco.com

Robert Varga
Pantheon Technologies SRO
Mlynske Nivy 56
Bratislava 821 05
Slovakia

Email: robert.varga@pantheon.sk

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 18, 2013

E. Crabbe
Google, Inc.
J. Medved
Cisco Systems, Inc.
I. Minei
Juniper Networks, Inc.
R. Varga
Pantheon Technologies SRO
October 15, 2012

Stateful PCE extensions for MPLS-TE LSPs
draft-crabbe-pce-stateful-pce-mpls-te-00

Abstract

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

[I-D.ietf-pce-stateful-pce] describes a set of extensions to PCEP to provide stateful control. This document describes the objects and TLVs to be used with these PCEP extensions to control Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP) via a stateful PCE.

Requirements Language

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

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 April 18, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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

1. Introduction	3
2. Terminology	3
3. Motivation	3
4. MPLS-TE specific descriptors used in PCEP Messages	3
4.1. MPLS-TE specific descriptors for the PCRpt Message	4
4.2. MPLS-TE specific descriptors for the PCUpd Message	4
4.3. MPLS-TE specific encoding for the PCReq Message for stateful PCE	6
4.4. MPLS-TE specific encoding for the PCRep Message for stateful PCE	7
5. Object and TLV Formats	8
5.1. LSP Identifiers TLVs	8
5.2. Tunnel ID TLV	11
5.3. LSP Update Error Code TLV	11
6. IANA Considerations	12
6.1. PCEP Objects	12
6.2. PCEP-Error Object	12
6.3. PCEP TLV Type Indicators	12
7. Security Considerations	13
8. Acknowledgements	13
9. References	13
9.1. Normative References	13
9.2. Informative References	14
Authors' Addresses	15

1. Introduction

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

[I-D.ietf-pce-stateful-pce] describes a set of extensions to PCEP to provide stateful control. This document describes the objects and TLVs to be used with these PCEP extensions to control Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP) via a stateful PCE.

2. Terminology

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

This document uses the following terms defined in [RFC4090]: MPLS TE Fast Reroute (FRR), FRR One-to-One Backup, FRR Facility Backup.

This document uses the following terms defined in [I-D.ietf-pce-stateful-pce] : Passive Stateful PCE, Active Stateful PCE, Delegation, Delegation Timeout Interval, LSP State Report, LSP Update Request, LSP Priority, LSP State Database, Revocation.

Within this document, when describing PCE-PCE communications, the requesting PCE fills the role of a PCC. This provides a saving in documentation without loss of function.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

3. Motivation

Several use cases for stateful PCE in an MPLS-TE network are included in [I-D.ietf-pce-stateful-pce].

4. MPLS-TE specific descriptors used in PCEP Messages

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable-length body made of a set of objects that can be either mandatory or optional. [I-D.ietf-pce-stateful-pce] describes the messages and objects needed in support of stateful PCE. The following sections contain MPLS-TE specific descriptors used in some of these messages.

4.1. MPLS-TE specific descriptors for the PCRpt Message

The format of the PCRpt message is defined in [I-D.ietf-pce-stateful-pce] as follows, and included here for easy reference:

```
<PCRpt Message> ::= <Common Header>
                        <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report>[<state-report-list>]
```

```
<state-report> ::= <LSP>
                    [<path-list>]
```

Where:

```
<path-list> ::= <path>[<path-list>]
```

For MPLS-TE LSPs, the path descriptor is defined as follows:

```
<path> ::= <ERO><attribute-list>
```

Where:

```
<attribute-list> ::= [<LSPA>]
                      [<BANDWIDTH>]
                      [<RRO>]
                      [<metric-list>]
```

```
<metric-list> ::= <METRIC>[<metric-list>]
```

The LSP State Report MAY contain a path descriptor for the primary path and one or more path descriptors for backup paths. A path descriptor MUST contain an ERO object as it was specified by a PCE or an operator. A path descriptor MUST contain the RRO object if a primary or secondary LSP is set up along the path in the network. A path descriptor MAY contain the LSPA, BANDWIDTH, and METRIC objects. The ERO, LSPA, BANDWIDTH, METRIC, and RRO objects are defined in [RFC5440].

4.2. MPLS-TE specific descriptors for the PCUpd Message

A Path Computation LSP Update Request message (also referred to as PCUpd message) is a PCEP message sent by a PCE to a PCC to update attributes of an LSP. A PCUpd message can carry more than one LSP Update Request. The Message-Type field of the PCEP common header for the PCUpd message is set to [TBD].

The format of the PCUpd message is defined in [I-D.ietf-pce-stateful-pce] and included here for easy reference:

```
<PCUpd Message> ::= <Common Header>
                        <update-request-list>
```

Where:

```
<update-request-list> ::= <update-request>[<update-request-list>]
```

```
<update-request> ::= <LSP>
                        [<path-list>]
```

Where:

```
<path-list> ::= <path>[<path-list>]
```

For MPLS-TE LSPs, the encoding of path descriptor is defined as follows:

```
<path> ::= <ERO><attribute-list>
```

Where:

```
<path> ::= <ERO><attribute-list>
```

Where:

```
<attribute-list> ::= [<LSPA>]
                        [<BANDWIDTH>]
                        [<metric-list>]
```

```
<metric-list> ::= <METRIC>[<metric-list>]
```

There is one mandatory object that MUST be included within each LSP Update Request in the PCUpd message: the LSP object (see [I-D.ietf-pce-stateful-pce]). If the LSP object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=[TBD] (LSP object missing).

The LSP Update Request MUST contain a path descriptor for the primary path, and MAY contain one or more path descriptors for backup paths. A path descriptor MUST contain an ERO object. A path descriptor MAY further contain the BANDWIDTH, IRO, and METRIC objects. The ERO, LSPA, BANDWIDTH, METRIC, and IRO objects are defined in [RFC5440].

Each LSP Update Request results in a separate LSP setup operation at a PCC. An LSP Update Request MUST contain all LSP parameters that a PCC wishes to set for the LSP. A PCC MAY set missing parameters from

locally configured defaults. If the LSP specified the Update Request is already up, it will be re-signaled. The PCC will use make-before-break whenever possible in the re-signaling operation.

A PCC MUST respond with an LSP State Report to each LSP Update Request to indicate the resulting state of the LSP in the network. A PCC MAY respond with multiple LSP State Reports to report LSP setup progress of a single LSP.

If the rate of PCUpd messages sent to a PCC for the same target LSP exceeds the rate at which the PCC can signal LSPs into the network, the PCC MAY perform state compression and only re-signal the last modification in its queue.

Note that a PCC MUST process all LSP Update Requests - for example, an LSP Update Request is sent when a PCE returns delegation or puts an LSP into non-operational state. The protocol relies on TCP for message-level flow control.

Note also that it's up to the PCE to handle inter-LSP dependencies; for example, if ordering of LSP set-ups is required, the PCE has to wait for an LSP State Report for a previous LSP before triggering the LSP setup of a next LSP.

4.3. MPLS-TE specific encoding for the PCReq Message for stateful PCE

A PCC MAY include the LSP object defined in [I-D.ietf-pce-stateful-pce] in the PCReq message if the stateful PCE capability has been negotiated on a PCEP session between the PCC and a PCE. The definition of the PCReq message (see [RFC5440], Section 6.4) is then extended as follows:

```
<PCReq Message>::= <Common Header>
                    [<svec-list>]
                    <request-list>
```

Where:

```
<svec-list>::=<SVEC>[<svec-list>]
<request-list>::=<request>[<request-list>]

<request>::= <RP>
              <END-POINTS>
              [<LSP>]                <--- New Object
              [<LSPA>]
              [<BANDWIDTH>]
              [<metric-list>]
              [<RRO>[<BANDWIDTH>]]
              [<IRO>]
              [<LOAD-BALANCING>]
```

Where:

```
<metric-list>::=<METRIC>[<metric-list>]
```

4.4. MPLS-TE specific encoding for the PCRep Message for stateful PCE

A PCE MAY include the LSP object defined in [I-D.ietf-pce-stateful-pce] in the PCRep message if the stateful PCE capability has been negotiated on a PCEP session between the PCC and the PCE and the LSP object was included in the corresponding PCReq message from the PCC. The definition of the PCRep message (see [RFC5440], Section 6.5) is then extended as follows


```

<PCRep Message> ::= <Common Header>
                    <response-list>

```

Where:

```

<response-list> ::= <response> [<response-list>]

<response> ::= <RP>
               [<LSP>]                <--- New Object
               [<NO-PATH>]
               [<attribute-list>]
               [<path-list>]

<path-list> ::= <path> [<path-list>]

<path> ::= <ERO> <attribute-list>

```

Where:

```

<attribute-list> ::= [<LSPA>]
                    [<BANDWIDTH>]
                    [<metric-list>]
                    [<IRO>]

<metric-list> ::= <METRIC> [<metric-list>]

```

5. Object and TLV Formats

The PCEP objects defined in this document are compliant with the PCEP object format defined in [RFC5440]. The P flag and the I flag of the PCEP objects defined in this document MUST always be set to 0 on transmission and MUST be ignored on receipt since these flags are exclusively related to path computation requests.

5.1. LSP Identifiers TLVs

Whenever the value of an LSP identifier changes, a PCC MUST send out an LSP State Report, where the LSP Object carries the LSP Identifiers TLV that contains the new value. The LSP Identifiers TLV MUST also be included in the LSP object during state synchronization. There are two LSP Identifiers TLVs, one for IPv4 and one for IPv6.

The format of the IPV4-LSP-IDENTIFIERS TLV is shown in the following figure:

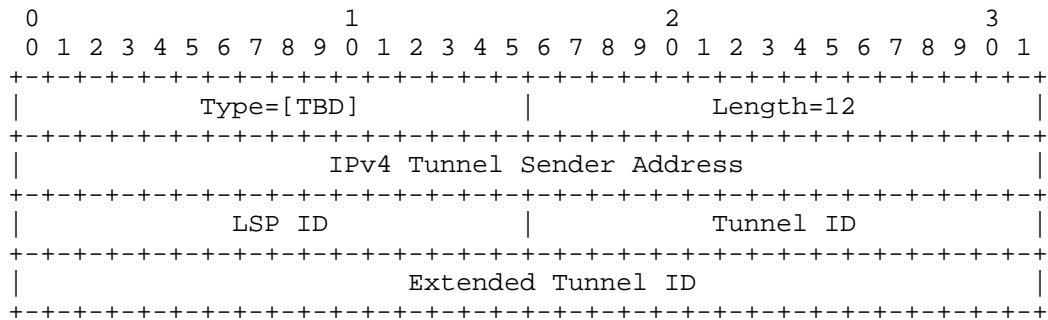


Figure 1: IPV4-LSP-IDENTIFIERS TLV format

The type of the TLV is [TBD] and it has a fixed length of 12 octets. The value contains the following fields:

IPv4 Tunnel Sender Address: contains the sender node's IPv4 address, as defined in [RFC3209], Section 4.6.2.1 for the LSP_TUNNEL_IPv4 Sender Template Object.

LSP ID: contains the 16-bit 'LSP ID' identifier defined in [RFC3209], Section 4.6.2.1 for the LSP_TUNNEL_IPv4 Sender Template Object.

Tunnel ID: contains the 16-bit 'Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.1 for the LSP_TUNNEL_IPv4 Session Object. Tunnel ID remains constant over the life time of a tunnel. However, when Global Path Protection or Global Default Restoration is used, both the primary and secondary LSPs have their own Tunnel IDs. A PCC will report a change in Tunnel ID when traffic switches over from primary LSP to secondary LSP (or vice versa).

Extended Tunnel ID: contains the 32-bit 'Extended Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.1 for the LSP_TUNNEL_IPv4 Session Object.

The format of the IPV6-LSP-IDENTIFIERS TLV is shown in 1 following figure:

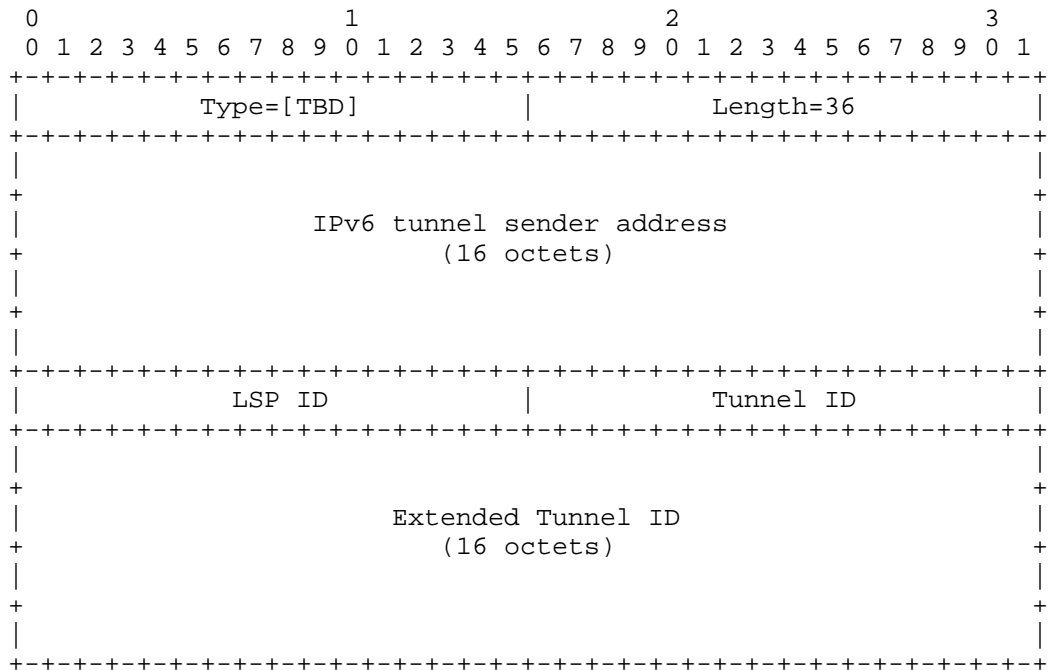


Figure 2: IPV6-LSP-IDENTIFIERS TLV format

The type of the TLV is [TBD] and it has a fixed length of 36 octets. The value contains the following fields:

IPv6 Tunnel Sender Address: contains the sender node's IPv6 address, as defined in [RFC3209], Section 4.6.2.2 for the LSP_TUNNEL_IPv6 Sender Template Object.

LSP ID: contains the 16-bit 'LSP ID' identifier defined in [RFC3209], Section 4.6.2.2 for the LSP_TUNNEL_IPv6 Sender Template Object.

Tunnel ID: contains the 16-bit 'Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.2 for the LSP_TUNNEL_IPv6 Session Object. Tunnel ID remains constant over the life time of a tunnel. However, when Global Path Protection or Global Default Restoration is used, both the primary and secondary LSPs have their own Tunnel IDs. A PCC will report a change in Tunnel ID when traffic switches over from primary LSP to secondary LSP (or vice versa).

Extended Tunnel ID: contains the 128-bit 'Extended Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.2 for the LSP_TUNNEL_IPv6 Session Object.

5.2. Tunnel ID TLV

The Tunnel ID TLV MAY be included in the LSPA object.

The format of the TUNNEL TLV is shown in the following figure:

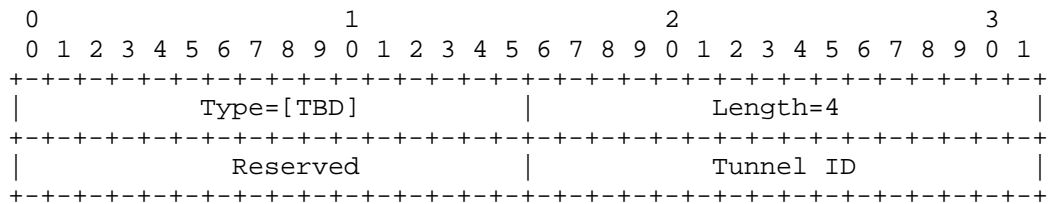


Figure 3: Tunnel-ID TLV format

The type of the TLV is [TBD] and it has a fixed length of 4 octets. The value contains a single field:

Tunnel ID: contains the 16-bit 'Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.1 for the LSP_TUNNEL_IPv4 Session Object. Tunnel ID remains constant over the life time of a tunnel. However, when Global Path Protection or Global Default Restoration is used, both the primary and secondary LSPs have their own Tunnel IDs.

5.3. LSP Update Error Code TLV

If an LSP Update Request failed, an LSP State Report MUST be sent to all connected stateful PCEs. LSP State Report MUST contain the LSP Update Error Code TLV, indicating the cause of the failure.

The format of the LSP-UPDATE-ERROR-CODE TLV is shown in the following figure:

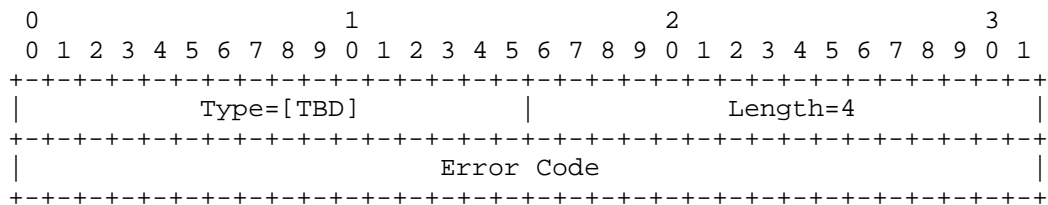


Figure 4: LSP-UPDATE-ERROR-CODE TLV format

The type of the TLV is [TBD] and it has a fixed length of 4 octets. The value contains the error code that indicates the cause of the LSP setup failure. Error codes will be defined in a later revision of this document.

6. IANA Considerations

This document requests IANA actions to allocate code points for the protocol elements defined in this document. Values shown here are suggested for use by IANA.

6.1. PCEP Objects

This document defines the following new PCEP Object-classes and Object-values:

Object-Class Value	Name	Reference
32	LSP Object-Type 1	This document

6.2. PCEP-Error Object

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

Error-Type	Meaning
6	Mandatory Object missing
Error-value=9:	ERO Object missing for a path in an LSP Update Request where TE-LSP setup is requested
Error-value=10:	BANDWIDTH Object missing for a path in an LSP Update Request where TE-LSP setup is requested
Error-value=11:	LSPA Object missing for a path in an LSP Update Request where TE-LSP setup is requested

6.3. PCEP TLV Type Indicators

This document defines the following new PCEP TLVs:

Value	Meaning	Reference
18	IPV4-LSP-IDENTIFIERS	This document
19	IPV6-LSP-IDENTIFIERS	This document
20	LSP-UPDATE-ERROR-CODE	This document
24	TUNNEL-ID	This document

7. Security Considerations

The security considerations listed in [I-D.ietf-pce-stateful-pce] apply to this document as well.

8. Acknowledgements

We would like to thank Adrian Farrel, Cyril Margaria and Ramon Casellas for their contributions to this document.

We would like to thank Shane Amante, Julien Meuric, Kohei Shiimoto, Paul Schultz and Raveendra Torvi for their comments and suggestions. Thanks also to Dhruv Dhoddy, Oscar Gonzales de Dios, Tomas Janciga, Stefan Kobza and Kexin Tang for helpful discussions.

9. References

9.1. 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-02 (work in progress),
October 2012.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [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.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.

- [RFC5088] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.

9.2. Informative References

- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC3346] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D., Christian, B., and W. Lai, "Applicability Statement for Traffic Engineering with MPLS", RFC 3346, August 2002.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash,

"Policy-Enabled Path Computation Framework", RFC 5394,
December 2008.

[RFC5557] Lee, Y., Le Roux, J.L., King, D., and E. Oki, "Path
Computation Element Communication Protocol (PCEP)
Requirements and Protocol Extensions in Support of Global
Concurrent Optimization", RFC 5557, July 2009.

Authors' Addresses

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: edc@google.com

Jan Medved
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: jmedved@cisco.com

Ina Minei
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: ina@juniper.net

Robert Varga
Pantheon Technologies SRO
Mlynske Nivy 56
Bratislava 821 05
Slovakia

Email: robert.varga@pantheon.sk

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: November 9, 2013

E. Crabbe
Google, Inc.
J. Medved
Cisco Systems, Inc.
I. Minei
Juniper Networks, Inc.
R. Varga
Pantheon Technologies SRO
May 8, 2013

Stateful PCE extensions for MPLS-TE LSPs
draft-crabbe-pce-stateful-pce-mpls-te-01

Abstract

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

[I-D.ietf-pce-stateful-pce] describes a set of extensions to PCEP to provide stateful control. This document describes the objects and TLVs to be used with these PCEP extensions to control Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP) via a stateful PCE.

Requirements Language

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

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 November 9, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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

1. Introduction	3
2. Terminology	3
3. MPLS-TE specific descriptors used in PCEP Messages	3
3.1. MPLS-TE specific descriptors for the PCRpt Message	3
3.2. MPLS-TE specific descriptors for the PCUpd Message	4
3.3. MPLS-TE specific encoding for the PCReq Message for stateful PCE	6
3.4. MPLS-TE specific encoding for the PCRep Message for stateful PCE	7
4. IANA Considerations	8
4.1. PCEP-Error Object	8
5. Security Considerations	9
6. Acknowledgements	9
7. References	9
7.1. Normative References	9
7.2. Informative References	10
Authors' Addresses	11

1. Introduction

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

[I-D.ietf-pce-stateful-pce] describes a set of extensions to PCEP to provide stateful control. This document describes the objects and TLVs to be used with these PCEP extensions to control Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP) via a stateful PCE.

2. 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] : Passive Stateful PCE, Active Stateful PCE, Delegation, Delegation Timeout Interval, LSP State Report, LSP Update Request, LSP Priority, LSP State Database, Revocation.

Within this document, when describing PCE-PCE communications, the requesting PCE fills the role of a PCC. This provides a saving in documentation without loss of function.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

3. MPLS-TE specific descriptors used in PCEP Messages

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable-length body made of a set of objects that can be either mandatory or optional. [I-D.ietf-pce-stateful-pce] describes the messages and objects needed in support of stateful PCE. The following sections contain MPLS-TE specific descriptors used in some of these messages.

3.1. MPLS-TE specific descriptors for the PCRpt Message

The format of the PCRpt message is defined in [I-D.ietf-pce-stateful-pce] as follows, and included here for easy reference:

```
<PCRpt Message> ::= <Common Header>
                     <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report>[<state-report-list>]
```

```
<state-report> ::= <LSP>
                  [<path-list>]
```

Where:

```
<path-list> ::= <path>[<path-list>]
```

For MPLS-TE LSPs, the path descriptor is defined as follows:

```
<path> ::= <ERO><attribute-list>
```

Where:

```
<attribute-list> ::= [<LSPA>]
                    [<BANDWIDTH>]
                    [<RRO>]
                    [<metric-list>]
```

```
<metric-list> ::= <METRIC>[<metric-list>]
```

The LSP State Report MAY contain a path descriptor for the primary path and one or more path descriptors for backup paths. A path descriptor MUST contain an ERO object as it was specified by a PCE or an operator. A path descriptor MUST contain the RRO object if a primary or secondary LSP is set up along the path in the network. A path descriptor MAY contain the LSPA, BANDWIDTH, and METRIC objects. The ERO, LSPA, BANDWIDTH, METRIC, and RRO objects are defined in [RFC5440].

3.2. MPLS-TE specific descriptors for the PCUpd Message

A Path Computation LSP Update Request message (also referred to as PCUpd message) is a PCEP message sent by a PCE to a PCC to update attributes of an LSP. A PCUpd message can carry more than one LSP Update Request. The Message-Type field of the PCEP common header for the PCUpd message is set to [TBD].

The format of the PCUpd message is defined in [I-D.ietf-pce-stateful-pce] and included here for easy reference:

```
<PCUpd Message> ::= <Common Header>
                        <update-request-list>
```

Where:

```
<update-request-list> ::= <update-request>[<update-request-list>]
```

```
<update-request> ::= <LSP>
                        [<path-list>]
```

Where:

```
<path-list> ::= <path>[<path-list>]
```

For MPLS-TE LSPs, the encoding of path descriptor is defined as follows:

```
<path> ::= <ERO><attribute-list>
```

Where:

```
<path> ::= <ERO><attribute-list>
```

Where:

```
<attribute-list> ::= [<LSPA>]
                        [<BANDWIDTH>]
                        [<metric-list>]
```

```
<metric-list> ::= <METRIC>[<metric-list>]
```

There is one mandatory object that MUST be included within each LSP Update Request in the PCUpd message: the LSP object (see [I-D.ietf-pce-stateful-pce]). If the LSP object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=[TBD] (LSP object missing).

The LSP Update Request MUST contain a path descriptor for the primary path, and MAY contain one or more path descriptors for backup paths. A path descriptor MUST contain an ERO object. A path descriptor MAY further contain the BANDWIDTH, IRO, and METRIC objects. The ERO, LSPA, BANDWIDTH, METRIC, and IRO objects are defined in [RFC5440].

Each LSP Update Request results in a separate LSP setup operation at a PCC. An LSP Update Request MUST contain all LSP parameters that a PCC wishes to set for the LSP. A PCC MAY set missing parameters from locally configured defaults. If the LSP specified the Update Request is already up, it will be re-signaled. The PCC will use make-before-break whenever possible in the re-signaling operation.

A PCC MUST respond with an LSP State Report to each LSP Update Request to indicate the resulting state of the LSP in the network. A PCC MAY respond with multiple LSP State Reports to report LSP setup progress of a single LSP.

If the rate of PCUpd messages sent to a PCC for the same target LSP exceeds the rate at which the PCC can signal LSPs into the network, the PCC MAY perform state compression and only re-signal the last modification in its queue.

Note that a PCC MUST process all LSP Update Requests - for example, an LSP Update Request is sent when a PCE returns delegation or puts an LSP into non-operational state. The protocol relies on TCP for message-level flow control.

Note also that it's up to the PCE to handle inter-LSP dependencies; for example, if ordering of LSP set-ups is required, the PCE has to wait for an LSP State Report for a previous LSP before triggering the LSP setup of a next LSP.

3.3. MPLS-TE specific encoding for the PCReq Message for stateful PCE

A PCC MAY include the LSP object defined in [I-D.ietf-pce-stateful-pce] in the PCReq message if the stateful PCE capability has been negotiated on a PCEP session between the PCC and a PCE. The definition of the PCReq message (see [RFC5440], Section 6.4) is then extended as follows:

```

<PCReq Message>::= <Common Header>
                   [<svec-list>]
                   <request-list>

```

Where:

```

<svec-list>::=<SVEC>[<svec-list>]
<request-list>::=<request>[<request-list>]

<request>::= <RP>
              <END-POINTS>
              [<LSP>]           <--- New Object
              [<LSPA>]
              [<BANDWIDTH>]
              [<metric-list>]
              [<RRO>[<BANDWIDTH>]]
              [<IRO>]
              [<LOAD-BALANCING>]

```

Where:

```

<metric-list>::=<METRIC>[<metric-list>]

```

3.4. MPLS-TE specific encoding for the PCRep Message for stateful PCE

A PCE MAY include the LSP object defined in [I-D.ietf-pce-stateful-pce] in the PCRep message if the stateful PCE capability has been negotiated on a PCEP session between the PCC and the PCE and the LSP object was included in the corresponding PCReq message from the PCC. The definition of the PCRep message (see [RFC5440], Section 6.5) is then extended as follows


```

<PCRep Message> ::= <Common Header>
                    <response-list>

```

Where:

```

<response-list> ::= <response> [<response-list>]

<response> ::= <RP>
               [<LSP>]                <--- New Object
               [<NO-PATH>]
               [<attribute-list>]
               [<path-list>]

<path-list> ::= <path> [<path-list>]

<path> ::= <ERO> <attribute-list>

```

Where:

```

<attribute-list> ::= [<LSPA>]
                    [<BANDWIDTH>]
                    [<metric-list>]
                    [<IRO>]

<metric-list> ::= <METRIC> [<metric-list>]

```

4. IANA Considerations

This document requests IANA actions to allocate code points for the protocol elements defined in this document. Values shown here are suggested for use by IANA.

4.1. PCEP-Error Object

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

Error-Type	Meaning
6	Mandatory Object missing
Error-value=9:	ERO Object missing for a path in an LSP Update Request where TE-LSP setup is requested
Error-value=10:	BANDWIDTH Object missing for a path in an LSP Update Request where TE-LSP setup is requested

Error-value=11: LSPA Object missing for a path in an LSP Update Request where TE-LSP setup is requested

5. Security Considerations

The security considerations listed in [I-D.ietf-pce-stateful-pce] apply to this document as well.

6. Acknowledgements

We would like to thank Adrian Farrel, Cyril Margaria and Ramon Casellas for their contributions to this document.

We would like to thank Shane Amante, Julien Meuric, Kohei Shiimoto, Paul Schultz and Raveendra Torvi for their comments and suggestions. Thanks also to Dhruv Dhoddy, Oscar Gonzales de Dios, Tomas Janciga, Stefan Kobza and Kexin Tang for helpful discussions.

7. References

7.1. 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-03 (work in progress), March 2013.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [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.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC5088] Le Roux, J.L., Vasseur, J.P., Ikejiri, Y., and R. Zhang,

"OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.

- [RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.

7.2. Informative References

- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC3346] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D., Christian, B., and W. Lai, "Applicability Statement for Traffic Engineering with MPLS", RFC 3346, August 2002.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394,

December 2008.

[RFC5557] Lee, Y., Le Roux, JL., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, July 2009.

Authors' Addresses

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: edc@google.com

Jan Medved
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: jmedved@cisco.com

Ina Minei
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: ina@juniper.net

Robert Varga
Pantheon Technologies SRO
Mlynske Nivy 56
Bratislava 821 05
Slovakia

Email: robert.varga@pantheon.sk

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 15, 2013

E. Crabbe
Google, Inc.
J. Medved
Cisco Systems, Inc.
I. Minei
R. Torvi
Juniper Networks, Inc.
October 12, 2012

PCEP Extensions for MPLS-TE LSP protection with stateful PCE
draft-crabbe-pce-stateful-pce-protection-00

Abstract

Stateful PCE [I-D.ietf-pce-stateful-pce] can apply global concurrent optimizations to optimize LSP placement. In a deployment where a PCE is used to compute all the paths, it may be beneficial for the protection paths to also be computed by the PCE. This document defines extensions needed for the setup and management of MPLS-TE protection paths by the PCE.

Requirements Language

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

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 April 15, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the

document authors. All rights reserved.

This document is subject to BCP 78 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

1. Introduction	3
2. Terminology	3
3. Architectural Overview	3
3.1. Path Protection Overview	3
3.2. Local Protection Overview	4
4. Extensions for the LSPA object	5
4.1. The Standby flag in the LSPA object	5
4.2. The Weight TLV	6
4.3. The Bypass TLV	6
4.4. The LOCALLY-PROTECTED-LSPS TLV	7
5. IANA considerations	9
5.1. PCEP-Error Object	9
5.2. PCEP TLV Type Indicators	9
6. Security Considerations	9
7. Acknowledgements	9
8. References	10
8.1. Normative References	10
8.2. Informative References	11
Authors' Addresses	11

1. Introduction

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

Stateful pce [I-D.ietf-pce-stateful-pce] specifies a set of extensions to PCEP to enable stateful control of paths such as MPLS TE LSPs between and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP state synchronization between PCCs and PCEs, delegation of control of LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions and focuses on a model where LSPs are configured on the PCC and control over them is delegated to the PCE.

Stateful PCE can apply global concurrent optimizations to optimize LSP placement. In a deployment where a PCE is used to compute all the paths, it may be beneficial for the protection paths to also be controlled through the PCE. This document defines extensions needed for the setup and management of protection paths by the PCE.

Benefits of controlling the protection paths include: better control over traffic after a failure and more deterministic path computation (paths not affected by overload after a failure).

2. 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, Delegation, Delegation Timeout Interval, LSP State Report, LSP Update Request.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

3. Architectural Overview

3.1. Path Protection Overview

Path protection refers to switching to a new path on failure. Several cases exist:

- (1) MPLS-TE Global Default Restoration - protection paths are computed dynamically by the LSR after the failure. This can be supported without any PCEP protocol changes by specifying a secondary path with an ERO of just the end points of the LSP. Once reestablished, the path is communicated to the PCE via the LSP State Report message.
- (2) MPLS-TE Global Path Protection - protection paths are fully specified ahead of the failure. The base Stateful PCE specification [I-D.ietf-pce-stateful-pce] supports sending multiple fully-specified paths in the PCUpd requests. There are 2 further sub-cases:
 - (a) Protection paths are pre-signaled ahead of the failure (standby paths).
 - (b) Protection paths are set up after the failure.

The protection path setup regimen (standby or not) is specified in the path using a new per-path flag in the LSPA object, the S (standby) flag (see section Section 4.1). Paths for which the S flag is set MUST have a name associated with them, specified using the SYMBOLIC-PATH-NAME TLV in the LSPA object.

Because multiple secondary standby paths are possible, there is also a need for the PCE to be able to specify the relative priorities between the paths (which one to take if there are 3 available). This is done through a weight assigned to each path. See details in Section 4.2.

Reversion from protection paths to the primary path when possible will be controlled by the PCE, by sending a new LSP Update Request. If the primary can be successfully signaled and the secondary does not have the S flag set, then the secondary MUST be torn down. Thus, there is no need to signal the desire for revertive behavior.

3.2. Local Protection Overview

Local protection refers to the ability to locally route around failure of an LSP. Two types of local protection are possible:

- (1) 1:1 protection - the protection path protects a single LSP.
- (2) 1:N protection - the protection path protects multiple LSPs traversing the protected resource.

It is assumed that the PCE knows what resources require protection through mechanisms outside the scope of this document. In a PCE-

controlled deployment, support of 1:1 protection has limited applicability, and can be achieved as a degenerate case of 1:N protection. For this reason, local protection will be discussed only for the 1:N case.

Local protection requires the setup of a bypass at the PLR. This bypass can be locally initiated and delegated, or PCE-initiated. In either case, the PLR must maintain a PCEP session to the PCE. A bypass identifier (the name of the bypass) is required for disambiguation as multiple bypasses are possible at the PLR. Mapping of LSPs to bypass is done through a new TLV, the LOCALLY-PROTECTED-LSPS TLV in the LSP Update message from PCE to PLR. See section Section 4.4. When an LSP requiring protection is set up through the PLR, the PLR checks if it has a mapping to a bypass and only provides protection if such a mapping exists. The status of bypasses and what LSPs are protected by them is communicated to the PCE via LSP Status Report messages.

4. Extensions for the LSPA object

4.1. The Standby flag in the LSPA object

The LSPA object is defined in [RFC5440] and replicated below for easy reference. This document defines a new flag, the S flag in the flags field of the LSPA object.

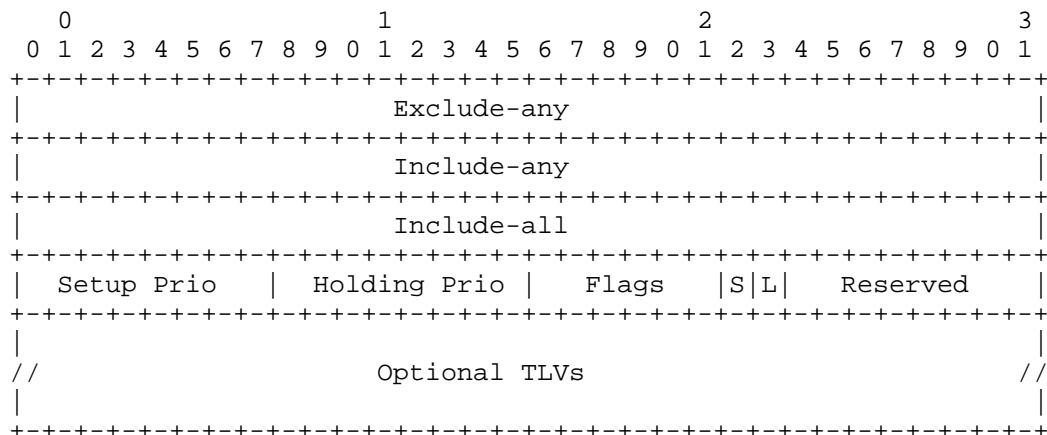


Figure 1: STATEFUL-PCE-CAPABILITY TLV format

The L flag is defined in [RFC5440].

If set to 1, the S Flag indicates this is a standby path.

If the S flag is set, the LSPA object MUST also carry the SYMBOLIC-PATH-NAME TLV as one of the optional TLVs. Failure to include the mandatory SYMBOLIC-PATH-NAME TLV when the S flag is set MUST trigger PCErr of type 6 (Mandatory Object missing) and value TBD (SYMBOLIC-PATH-NAME TLV missing for standby LSP).

4.2. The Weight TLV

This TLV will be discussed in a future version of tihs document.

4.3. The Bypass TLV

The facility backup method creates a bypass tunnel to protect a potential failure point. The bypass tunnel protects a set of LSPs with similar backup constraints [RFC4090].

A PCC can delegate a bypass tunnel to PCE control or a PCE can provision the bypass tunnel via a PCC. The procedures for bypass instantiation rely on the extensions defined in [I-D.crabbe-pce-pce-initiated-lsp] and will be detailed in a future version of this document.

The Bypass TLV carries information about the bypass tunnel. It is included in the LSPA Object in LSP State Report and LSP Update Request messages.

The format of the Bypass TLV is shown in the following figure:

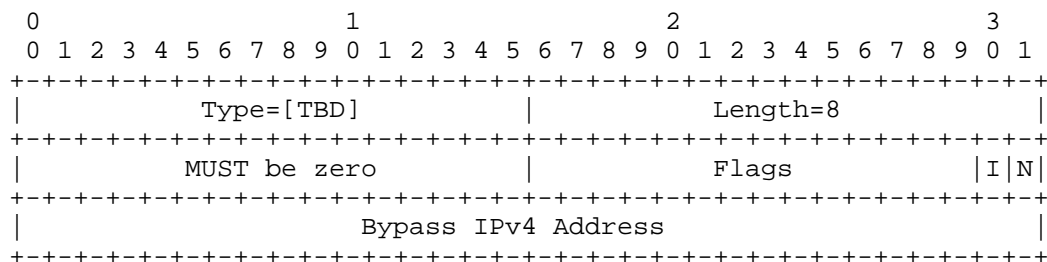


Figure 2: Bypass TLV format

The type of the TLV is [TBD] and it has a fixed length of 8 octets. The value contains the following fields:

Flags

N (Node Protection - 1 bit): The N Flag indicates whether the Bypass is used for node-protection. If the N flag is set to 1, the Bypass is used for node-protection. If the N flag is 0, the Bypass is used for link-protection.

I (Local Protection In Use - 1 bit): The I Flag indicates that local repair mechanism is in use.

Bypass IPv4 address: For link protection, the Bypass IPv4 Address is the nexthop address of the protected link in the paths of the protected LSPs. For node protection, the Bypass IPv4 Address is the node addresses of the protected node.

If the Bypass TLV is included, then the LSPA object MUST also carry the SYMBOLIC-PATH-NAME TLV as one of the optional TLVs. Failure to include the mandatory SYMBOLIC-PATH-NAME TLV MUST trigger PCErr of type 6 (Mandatory Object missing) and value TBD (SYMBOLIC-PATH-NAME TLV missing for bypass LSP)

4.4. The LOCALLY-PROTECTED-LSPS TLV

The LOCALLY-PROTECTED-LSPS TLV in the LSPA Object contains a list of LSPs protected by the bypass tunnel.

The format of the Bypass TLV is shown in the following figure:

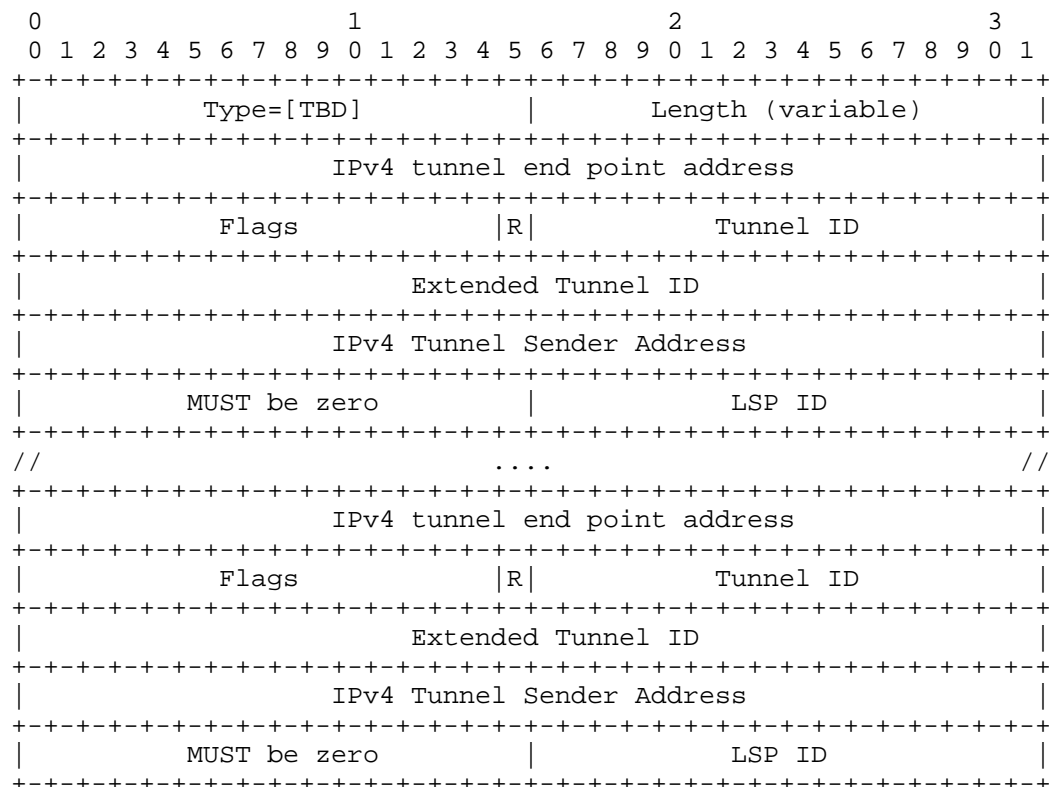


Figure 3: Locally protected LSPs TLV format

The type of the TLV is [TBD] and it is of variable length. The value contains one or more LSP descriptors including the following fields filled per [RFC3209].

IPv4 Tunnel end point address: [RFC3209]

Flags

R(Remove - 1 bit): The R Flag indicates that the LSP has been removed from the list of LSPs protected by the bypass tunnel.

Tunnel ID: [RFC3209]

Extended Tunnel ID: [RFC3209]

IPv4 Tunnel Sender address: [RFC3209]

LSP ID: [RFC3209]

5. IANA considerations

5.1. PCEP-Error Object

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:	SYMBOLIC-PATH-NAME TLV missing for a path where the S-bit is set in the LSPA object.
Error-value=TBD:	SYMBOLIC-PATH-NAME TLV missing for a bypass path.

5.2. PCEP TLV Type Indicators

This document defines the following new PCEP TLVs:

Value	Meaning	Reference
???	Bypass	This document
???	weight	This document
???	LOCALLY-PROTECTED-LSPS	This document

6. Security Considerations

The same security considerations apply at the PLR as those describe for the head end in [I-D.crabbe-pce-pce-initiated-lsp].

7. Acknowledgements

We would like to thank Ambrose Kwong for his contributions to this document.

8. References

8.1. Normative References

- [I-D.crabbe-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-crabbe-pce-pce-initiated-lsp-00 (work in progress), October 2012.
- [I-D.ietf-pce-stateful-pce]
Crabbe, E., Medved, J., Varga, R., and I. Minei, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-01 (work in progress), July 2012.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.
- [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.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC5088] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Vasseur, JP., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, April 2009.

8.2. Informative References

- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, September 1999.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.
- [RFC3346] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D., Christian, B., and W. Lai, "Applicability Statement for Traffic Engineering with MPLS", RFC 3346, August 2002.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, December 2008.
- [RFC5557] Lee, Y., Le Roux, J.L., King, D., and E. Oki, "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, July 2009.

Authors' Addresses

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: edc@google.com

Jan Medved
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: jmedved@cisco.com

Ina Minei
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: ina@juniper.net

Raveendra Torvi
Juniper Networks, Inc.
1194 N. Mathilda Ave.
Sunnyvale, CA 94089
US

Email: rtorvi@juniper.net

PCE Working Group
Internet-Draft
Intended status: Experimental
Expires: April 14, 2013

D. Dhody
U. Pallé
V. Kondreddy
Huawei Technologies India Pvt
Ltd
October 11, 2012

Supporting explicit inclusion or exclusion of abstract nodes for a
subset of P2MP destinations in Path Computation Element Communication
Protocol (PCEP).
draft-dhody-pce-pcep-p2mp-per-destination-03

Abstract

The ability to determine paths of point-to-multipoint (P2MP) Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs) is one the key requirements for Path Computation Element (PCE). [RFC6006] and [PCE-P2MP-PROCEDURES] describes these mechanisms for intra and inter domain path computation via PCE.

This document describes the need for explicitly specifying abstract nodes for inclusion or exclusion for a subset of destinations during the Point to Multipoint (P2MP) path computation via PCE. Further an extension to the PCE communication Protocol (PCEP) for P2MP is suggested to support this.

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 April 14, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the

document authors. All rights reserved.

This document is subject to BCP 78 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

1. Introduction	3
1.1. Requirements Language	3
2. Terminology	3
3. Motivation	4
3.1. Domain Sequence Tree in Inter Domain P2MP Path Computation	4
3.2. Explicit inclusion or exclusion of abstract nodes	6
4. Detailed Description	7
4.1. Objective	7
4.2. Request Message Format	7
4.3. Backward Compatibility	8
5. IANA Considerations	8
6. Security Considerations	9
7. Manageability Considerations	9
7.1. Control of Function and Policy	9
7.2. Information and Data Models	9
7.3. Liveness Detection and Monitoring	9
7.4. Verify Correct Operations	9
7.5. Requirements On Other Protocols	9
7.6. Impact On Network Operations	10
8. Acknowledgments	10
9. References	10
9.1. Normative References	10
9.2. Informative References	10

1. Introduction

The Path Computation Element (PCE) architecture is defined in [RFC4655]. [RFC5862] lay out the requirements for Path Computation Client (PCC) and Path Computation Element (PCE) to support Point-to-Multipoint (P2MP) path computation. [RFC6006] describe an extension to PCEP to compute optimal constrained intra-domain (G)MPLS P2MP TE LSPs. [PCE-P2MP-PROCEDURES] describes the mechanism for inter-domain P2MP path computation.

[PCE-P2MP-PROCEDURES] describes the core-tree procedure for computing inter-domain P2MP tree. It assumes that, due to deployment and commercial limitations, the sequence of domains for a path (the path domain tree) will be known in advance. For a group of destination which belong to a destination domain, the domain-sequence needs to be encoded separately as described in [DOMAIN-SEQ]. The mechanism of explicitly specifying abstract nodes for inclusion or exclusion for a subset of destinations can be used for this purpose, where abstract nodes are domains.

[RFC6006] describe a PCE-based path computation procedure to compute optimal constrained (G)MPLS P2MP TE LSPs. It describes mechanism to specify branch nodes that can or cannot be used via Branch Node Capability (BNC) object (which only supports IPv4 and IPv6 prefix sub-objects). This document adds the capability to explicitly specify any abstract nodes for inclusion or exclusion for a subset of destinations.

1.1. Requirements Language

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

2. Terminology

The following terminology is used in this document.

IRO: Include Route Object.

PCC: Path Computation Client: any client application requesting a path computation to be performed by a Path Computation Element.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

P2MP: Point-to-Multipoint

P2P: Point-to-Point

RRO: Record Route Object

RSVP: Resource Reservation Protocol

TE LSP: Traffic Engineering Label Switched Path.

XRO: Exclude Route Object.

3. Motivation

3.1. Domain Sequence Tree in Inter Domain P2MP Path Computation

[PCE-P2MP-PROCEDURES] describes the core-tree procedure for inter-domain path computation. The procedure assumes that the sequence of domains for a path (the path domain tree) will be known in advance due to deployment and commercial limitations (e.g., inter-AS peering agreements).

In the Figure 1 below, D1 is the root domain; D5 and D6 are the destination domains. The ingress is A in domain D1; egresses are X, Y in Domain D6 and Z in Domain D5.



Figure 1: Domain Topology Example

In the Figure 2 below, the P2MP tree spans 5 domains. Destination in D6 (X & Y) would use the domain-sequence: D1-D3-D4-D6; and destination in D5 (Z) would use the domain-sequence: D1-D3-D4-D5.

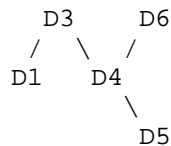


Figure 2: Domain Sequence Tree

Since destinations in different destination domain will have different domain sequence within the domain tree, it requires following encoding-

- o Destination X and Y: D1-D3-D4-D6
- o Destination Z : D1-D3-D4-D5

An extension in P2MP Path Computation request is needed to support this. (Refer Section 4.2)

The abstract nodes MAY include (but not limited to) domain subobjects AS number and IGP Area as described in [DOMAIN-SEQ].

3.2. Explicit inclusion or exclusion of abstract nodes

[RFC6006] describes four possible types of leaves in a P2MP request encoded in P2MP END-POINTS object.

- o New leaves to add
- o Old leaves to remove
- o Old leaves whose path can be modified/reoptimized
- o Old leaves whose path must be left unchanged

Currently [RFC6006] only allows a list of nodes that can be used as branch nodes or a list of nodes that cannot be used as branch nodes by using the Branch Node Capability (BNC) Object, which applies to all leaves (old and new) in the P2MP tree.

For an existing P2MP tree which may already have a branch node

through which most of the leaves are connected, but when adding a set of new leaves, administrator may want to exclude that branch node (as it may soon be overloaded) and would like to balance the final P2MP tree. This could be done explicitly by excluding a particular branch node or including a different branch node only for a set of new leaves encoded in the one P2MP END-POINTS object.

Administrator at the source can exert stronger control by providing explicit inclusion or exclusion of any abstract nodes (not limited to branch nodes) for a group (subset) of destinations.

4. Detailed Description

4.1. Objective

[RFC6006] defines Request Message Format and Objects, along with <end-point-rro-pair-list>. This section introduce the use of <IRO> and <XRO> which are added to the <end-point-rro-pair-list>.

To allow abstract nodes to be explicitly included or excluded for a subset of destinations (encoded in one <END-POINTS> object), changes are made as shown below.

The abstract node (encoded as subobject in <IRO> and <XRO>) MAY be an absolute hop, IP-Prefix, Autonomous system or IGP Area. The subobjects are described in [RFC3209], [RFC3477], [RFC4874] and [DOMAIN-SEQ].

Note that one P2MP Path request can have multiple <END-POINTS> objects and each P2MP <END-POINTS> object may have multiple destinations, the <IRO> and <XRO> is applied for all destinations in one such P2MP <END-POINTS> object.

4.2. Request Message Format

The format of PCReq message is modified as follows:

```

    <PCReq Message>::= <Common Header>
                        <request>
  where:
    <request>::= <RP>
                <end-point-iro-xro-rro-pair-list>
                [<OF>]
                [<LSPA>]
                [<BANDWIDTH>]
                [<metric-list>]
                [<IRO>]
                [<LOAD-BALANCING>]

  where:
    <end-point-iro-xro-rro-pair-list>::=
        <END-POINTS>
        [<IRO>]
        [<XRO>]
        [<RRO-List>][<BANDWIDTH>]
        [<end-point-iro-xro-rro-pair-list>]

    <RRO-List>::=<RRO>[<BANDWIDTH>][<RRO-List>]
    <metric-list>::=<METRIC>[<metric-list>]

```

From [RFC6006] usage of <end-point-rro-pair-list> is changed to <end-point-iro-xro-rro-pair-list> in this document.

[RFC6006] describes Branch Node Capability (BNC) Object which is different from the use of <IRO> and <XRO> to specify inclusion/exclusion of abstract nodes for a subset of destinations as described here.

4.3. Backward Compatibility

A legacy implementation that does not support explicit inclusion or exclusion of abstract nodes for a subset of P2MP destinations will act according to the procedures set out in [RFC5440], that is it will find the P2MP Path Request message out of order with respect to the format specified in [RFC6006].

5. IANA Considerations

There are no new IANA allocation in this document.

6. Security Considerations

PCEP security mechanisms as described in [RFC5440], [RFC6006] and [PCE-P2MP-PROCEDURES] are applicable for this document.

The new explicit inclusion or exclusion of abstract nodes for a subset of P2MP destination defined in this document allow finer and more specific control of the path computed by a PCE. Such control increases the risk if a PCEP message is intercepted, modified, or spoofed because it allows the attacker to exert control over the path that the PCE will compute or to make the path computation impossible. Therefore, the security techniques described in [RFC5440], [RFC6006] and [PCE-P2MP-PROCEDURES] are considered more important.

Note, however, that the route exclusion mechanisms also provide the operator with the ability to route around vulnerable parts of the network and may be used to increase overall network security.

7. Manageability Considerations

7.1. Control of Function and Policy

Mechanisms defined in this document do not add any new control function/policy requirements in addition to those already listed in [RFC6006].

7.2. Information and Data Models

Mechanisms defined in this document do not imply any new MIB requirements in addition to those already listed in [PCE-P2MP-MIB].

7.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC6006].

7.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC6006].

7.5. Requirements On Other Protocols

Mechanisms defined in this document do not imply any requirements on other protocols in addition to those already listed in [RFC6006].

7.6. Impact On Network Operations

Mechanisms defined in this document do not have any impact on network operations in addition to those already listed in [RFC6006].

8. Acknowledgments

We would like to thank Pradeep Shastry, Suresh babu, Quintin Zhao, Daniel King and Chen Huaimo for their useful comments and suggestions.

9. References

9.1. Normative References

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

9.2. 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.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC4874] Lee, CY., Farrel, A., and S. De Cnodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4874, April 2007.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5862] Yasukawa, S. and A. Farrel, "Path Computation Clients (PCC) - Path Computation Element (PCE) Requirements for Point-to-Multipoint MPLS-TE", RFC 5862, June 2010.

- [RFC6006] Zhao, Q., King, D., Verhaeghe, F., Takeda, T., Ali, Z., and J. Meuric, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 6006, September 2010.
- [PCE-P2MP-PROCEDURES] Zhao, Q., Dhody, D., Ali, Z., Saad,, T., Sivabalan,, S., and R. Casellas, "PCE-based Computation Procedure To Compute Shortest Constrained P2MP Inter-domain Traffic Engineering Label Switched Paths (draft-ietf-pce-pcep-inter-domain-p2mp-procedures-02)", May 2012.
- [PCE-P2MP-MIB] Zhao, Q., Dhody, D., Palle, U., and D. King, "Management Information Base for the PCE Communications Protocol (PCEP) When Requesting Point-to-Multipoint Services (draft-zhao-pce-pcep-p2mp-mib-05)", August 2012.
- [DOMAIN-SEQ] Dhody, D., Palle, U., and R. Casellas, "Standard Representation Of Domain Sequence (draft-ietf-pce-pcep-domain-sequence-01)", July 2012.

Authors' Addresses

Dhruv Dhody
Huawei Technologies India Pvt Ltd
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruv.dhody@huawei.com

Udayasree Palle
Huawei Technologies India Pvt Ltd
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: udayasree.palle@huawei.com

Venugopal Reddy Kondreddy
Huawei Technologies India Pvt Ltd
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: venugopalreddyk@huawei.com

PCE Working Group
Internet-Draft
Intended status: Experimental
Expires: March 31, 2018

D. Dhody
R. Palleti
U. Palle
V. Kondreddy
Huawei Technologies
September 27, 2017

Supporting Explicit Inclusion or Exclusion of Abstract Nodes for a
Subset of P2MP Destinations in Path Computation Element Communication
Protocol (PCEP).

draft-dhody-pce-pcep-p2mp-per-destination-12

Abstract

The ability to determine paths of point-to-multipoint (P2MP) Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs) is one the key requirements for Path Computation Element (PCE). The PCEP has been extended for intra and inter domain path computation via PCE(s) for P2MP TE LSP.

This document describes the motivation and PCEP extension for explicitly specifying abstract nodes for inclusion or exclusion for a subset of destinations during P2MP path computation via PCE(s).

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 <https://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 March 31, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://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

1. Introduction	2
1.1. Requirements Language	3
2. Terminology	3
3. Motivation	4
3.1. Domain Sequence Tree in Inter Domain P2MP Path Computation	4
3.2. Explicit inclusion or exclusion of abstract nodes	6
4. Detailed Description	6
4.1. Objective	7
4.2. Request Message Format	7
4.3. Report Message Format	9
4.4. Backward Compatibility	10
5. IANA Considerations	10
6. Security Considerations	10
7. Manageability Considerations	10
7.1. Control of Function and Policy	10
7.2. Information and Data Models	11
7.3. Liveness Detection and Monitoring	11
7.4. Verify Correct Operations	11
7.5. Requirements On Other Protocols	11
7.6. Impact On Network Operations	11
8. Acknowledgments	11
9. References	11
9.1. Normative References	11
9.2. Informative References	12
Authors' Addresses	13

1. Introduction

The PCE architecture is defined in [RFC4655]. [RFC5862] lay out the requirements for PCEP to support P2MP path computation. [I-D.ietf-pce-rfc6006bis] describe an extension to PCEP to compute optimal constrained intra-domain (G)MPLS P2MP TE LSPs. [RFC7334] describes the mechanism for inter-domain P2MP path computation.

Further [I-D.ietf-pce-rfc6006bis] describes mechanism to specify a list of nodes that can be used as branch nodes or a list of nodes that cannot be used as branch nodes via Branch Node Capability (BNC) object. The BNC object is used to specify which nodes have the capability to act as a branch nodes or which nodes lack the capability. It supports IPv4 and IPv6 prefix sub-objects only.

This document explains the need to add the capability to explicitly specify any abstract nodes (not just nodes with branch node capability) for inclusion or exclusion for a subset of destinations.

[RFC7334] describes the core-tree procedure to compute inter-domain P2MP tree. It assumes that, due to deployment and commercial limitations, the sequence of domains for a path (the path domain tree) will be known in advance. For a group of destination which belong to a particular destination domain, the domain-sequence needs to be encoded separately as described in [RFC7897]. The mechanism, as described in this document, of explicitly specifying abstract nodes for inclusion or exclusion for a subset of destinations can be used for this purpose, where abstract nodes are domains.

Stateful PCEs are shown to be helpful in many application scenarios, in both MPLS and GMPLS networks, as illustrated in [RFC8051]. These scenarios apply equally to P2P and P2MP TE LSPs. [RFC8231] provides the fundamental extensions needed for stateful PCE to support general functionality for P2P TE LSP. [I-D.ietf-pce-pce-initiated-lsp] provides the an extensions needed for stateful PCE-initiated P2P TE LSP. Complementarily, [I-D.ietf-pce-stateful-pce-p2mp] focuses on the extensions that are necessary in order for the deployment of stateful PCEs to support P2MP TE LSPs.

1.1. Requirements Language

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.

2. Terminology

The following terminology is used in this document.

IRO: Include Route Object.

PCC: Path Computation Client: any client application requesting a path computation to be performed by a Path Computation Element.

PCE: Path Computation Element. An entity (component, application, or network node) that is capable of computing a network path or route based on a network graph and applying computational constraints.

PCEP: Path Computation Element Protocol.

P2MP: Point-to-Multipoint

P2P: Point-to-Point

RRO: Record Route Object

RSVP: Resource Reservation Protocol

TE LSP: Traffic Engineering Label Switched Path.

XRO: Exclude Route Object.

3. Motivation

3.1. Domain Sequence Tree in Inter Domain P2MP Path Computation

[RFC7334] describes the core-tree procedure for inter-domain path computation. The procedure assumes that the sequence of domains for a path (the path domain tree) will be known in advance due to deployment and commercial limitations (e.g., inter-AS peering agreements).

In the Figure 1 below, D1 is the root domain; D4, D5 and D6 are the destination domains. The ingress is Ro in domain D1; egresses are M, N in Domain D4; R, S in Domain D5; and U, V in Domain D6.

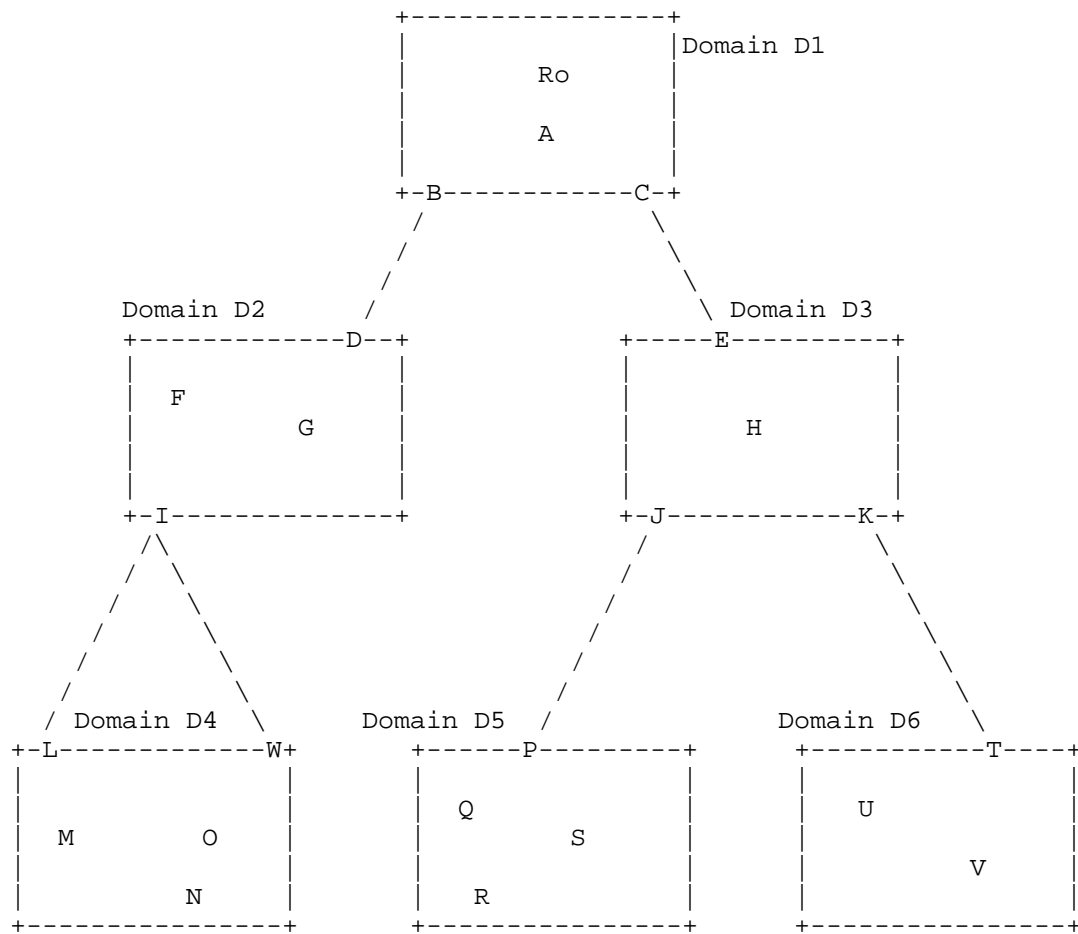


Figure 1: Domain Topology Example

The domain tree can be represented as a series of domain sequences:

Domain D1, Domain D3, Domain D6

Domain D1, Domain D3, Domain D5

Domain D1, Domain D2, Domain D4

Since destinations in different destination domain will have different domain sequence within the domain tree, it requires following encoding that binds destinations to a particular domain sequence.

- o Destination M and N: D1-D2-D4
- o Destination R and S: D1-D3-D5
- o Destination U and V: D1-D3-D6

An extension in P2MP Path Computation request is needed to support this. (Refer Section 4.2)

The abstract nodes MAY include (but not limited to) domain subobjects - AS number and IGP Area as described in [RFC7897].

3.2. Explicit inclusion or exclusion of abstract nodes

[I-D.ietf-pce-rfc6006bis] describes four possible types of leaves in a P2MP request encoded in P2MP END-POINTS object.

- o New leaves to add
- o Old leaves to remove
- o Old leaves whose path can be modified/reoptimized
- o Old leaves whose path must be left unchanged

[I-D.ietf-pce-rfc6006bis] only allows to encode a list of nodes that have (or have not) the branch node capability by using the Branch Node Capability (BNC) Object. This object apply to all destinations (old and new) in the P2MP tree.

For an existing P2MP tree with an overloaded branch node, when adding a set of new leaves, administrator may want to exclude that particular branch node to balance the final P2MP tree. This cannot be achieved via the BNC object but by explicitly excluding a particular node or including a different node, for the P2MP END-POINTS object for new leaves only.

Administrator at the Ingress can exert stronger control by providing explicit inclusion or exclusion of any abstract nodes (not limited to specifying nodes with branch node capability) for a group (subset) of destinations and not all destinations.

4. Detailed Description

4.1. Objective

[I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp] defines Request Message Format and Objects, along with <end-point-rro-pair-list>. This section introduce the use of <IRO> and <XRO> which are added to the <end-point-rro-pair-list>.

To allow abstract nodes to be explicitly included or excluded for a subset of destinations (encoded in one <END-POINTS> object), changes are made as shown below.

The abstract node (encoded as subobject in <IRO> and <XRO>) MAY be an absolute hop, IP-Prefix, AS or IGP Area. The subobjects are described in [RFC3209], [RFC3477], [RFC4874] and [RFC7897].

Note that one P2MP Path request can have multiple <END-POINTS> objects and each P2MP <END-POINTS> object may have multiple destinations, the <pce-list>, <IRO> and <XRO> is applied for all destinations in one such P2MP <END-POINTS> object.

4.2. Request Message Format

The format of PCReq message, with [I-D.ietf-pce-stateful-pce-p2mp] as base, is modified as follows:

```

    <PCReq Message> ::= <Common Header>
                        [<svec-list>]
                        <request-list>

where:

<svec-list> ::= <SVEC>
               [<OF>]
               [<metric-list>]
               [<svec-list>]

<request-list> ::= <request> [<request-list>]

<request> ::= <RP>
              <end-point-pce-iro-xro-rro-pair-list>
              [<LSP>]
              [<OF>]
              [<LSPA>]
              [<BANDWIDTH>]
              [<metric-list>]
              [<IRO> | <BNC>]
              [<LOAD-BALANCING>]

<end-point-pce-iro-xro-rro-pair-list> ::=
              <END-POINTS>
              [<IRO>]
              [<XRO>]
              [<RRO-List>] [<BANDWIDTH>]
              [<end-point-pce-iro-xro-rro-pair-list>]

<RRO-List> ::= (<RRO> | <SRRO>) [<RRO-List>]
<metric-list> ::= <METRIC> [<metric-list>]

```

From [I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp], usage of <end-point-rro-pair-list> is changed to <end-point-pce-iro-xro-rro-pair-list> in this document.

[I-D.ietf-pce-rfc6006bis] describes Branch Node Capability (BNC) Object which is different from the use of <IRO> and <XRO> to specify inclusion/exclusion of abstract nodes for a subset of destinations as described here.

4.3. Report Message Format

[I-D.ietf-pce-stateful-pce-p2mp] defines a report message format and objects. This document extends the message to allow explicit inclusion and exclusion of abstract nodes for a group of destinations.

```
<PCRpt Message> ::= <Common Header>
                        <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report>
                        [<state-report-list>]
```

```
<state-report> ::= [<SRP>]
                   <LSP>
                   <end-point-intended-path-pair-list>
                   [<actual_attribute_list>]
                   <end-point-actual-path-pair-list>]
                   <intended-attribute-list>
```

Where:

```
<end-point-intended-path-pair-list> ::=
    [<END-POINTS>]
    [<S2LS>]
    [<IRO>]
    [<XRO>]
    <intended_path>
    [<end-point-intended-path-pair-list>]
```

```
<end-point-actual-path-pair-list> ::=
    [<END-POINTS>]
    [<S2LS>]
    <actual_path>
    [<end-point-actual-path-pair-list>]
```

```
<intended_path> ::= (<ERO>|<SERO>)
                    [<intended_path>]
```

```
<actual_path> ::= (<RRO>|<SRRO>)
                  [<actual_path>]
```

<intended_path> is represented by the ERO, SERO object. The <actual_attribute_list> consists of the actual computed and signaled values of the <BANDWIDTH> and <metric-lists> objects defined in [RFC5440]. <actual_path> is represented by the RRO, SERO object.

The <end-point-intended-path-pair-list> is extended to add the IRO and XRO object for a group of destinations in the END-POINTS object.

4.4. Backward Compatibility

A legacy implementation that does not support explicit inclusion or exclusion of abstract nodes for a subset of P2MP destinations will act according to the procedures set out in [RFC5440], that is it will find the P2MP Path Request message out of order with respect to the format specified in [I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp].

5. IANA Considerations

There are no new IANA allocation in this document.

6. Security Considerations

PCEP security mechanisms as described in [RFC5440], [I-D.ietf-pce-rfc6006bis], [RFC7334] and [I-D.ietf-pce-stateful-pce-p2mp] are applicable for this document.

The new explicit inclusion or exclusion of abstract nodes for a subset of P2MP destination defined in this document allow finer and more specific control of the path computed by a PCE. Such control increases the risk if a PCEP message is intercepted, modified, or spoofed because it allows the attacker to exert control over the path that the PCE will compute or to make the path computation impossible. Therefore, the security techniques described in [RFC5440], [I-D.ietf-pce-rfc6006bis], [RFC7334] and [I-D.ietf-pce-stateful-pce-p2mp] are considered more important.

Note, however, that the route exclusion mechanisms also provide the operator with the ability to route around vulnerable parts of the network and may be used to increase overall network security.

7. Manageability Considerations

7.1. Control of Function and Policy

Mechanisms defined in this document do not add any new control function/policy requirements in addition to those already listed in [I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp].

7.2. Information and Data Models

Mechanisms defined in this document do not imply any new MIB requirements.

7.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp].

7.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp].

7.5. Requirements On Other Protocols

Mechanisms defined in this document do not imply any requirements on other protocols in addition to those already listed in [I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp].

7.6. Impact On Network Operations

Mechanisms defined in this document do not have any impact on network operations in addition to those already listed in [I-D.ietf-pce-rfc6006bis] and [I-D.ietf-pce-stateful-pce-p2mp].

8. Acknowledgments

We would like to thank Pradeep Shastry, Suresh babu, Quintin Zhao, Daniel King and Chen Huaimo for their useful comments and suggestions.

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<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>>.
- [I-D.ietf-pce-rfc6006bis]
Zhao, Q., Dhody, D., Palleti, R., and D. King, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", draft-ietf-pce-rfc6006bis-04 (work in progress), September 2017.
- [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-10 (work in progress), June 2017.
- [I-D.ietf-pce-stateful-pce-p2mp]
Palle, U., Dhody, D., Tanaka, Y., and V. Beeram, "Path Computation Element (PCE) Protocol Extensions for Stateful PCE usage for Point-to-Multipoint Traffic Engineering Label Switched Paths", draft-ietf-pce-stateful-pce-p2mp-04 (work in progress), July 2017.

9.2. 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>>.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, DOI 10.17487/RFC3477, January 2003, <<https://www.rfc-editor.org/info/rfc3477>>.

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC4874] Lee, CY., Farrel, A., and S. De Cnodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4874, DOI 10.17487/RFC4874, April 2007, <<https://www.rfc-editor.org/info/rfc4874>>.
- [RFC5862] Yasukawa, S. and A. Farrel, "Path Computation Clients (PCC) - Path Computation Element (PCE) Requirements for Point-to-Multipoint MPLS-TE", RFC 5862, DOI 10.17487/RFC5862, June 2010, <<https://www.rfc-editor.org/info/rfc5862>>.
- [RFC7334] Zhao, Q., Dhody, D., King, D., Ali, Z., and R. Casellas, "PCE-Based Computation Procedure to Compute Shortest Constrained Point-to-Multipoint (P2MP) Inter-Domain Traffic Engineering Label Switched Paths", RFC 7334, DOI 10.17487/RFC7334, August 2014, <<https://www.rfc-editor.org/info/rfc7334>>.
- [RFC7897] Dhody, D., Palle, U., and R. Casellas, "Domain Subobjects for the Path Computation Element Communication Protocol (PCEP)", RFC 7897, DOI 10.17487/RFC7897, June 2016, <<https://www.rfc-editor.org/info/rfc7897>>.
- [RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a Stateful Path Computation Element (PCE)", RFC 8051, DOI 10.17487/RFC8051, January 2017, <<https://www.rfc-editor.org/info/rfc8051>>.

Authors' Addresses

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

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

EMail: udayasreereddy@gmail.com

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

EMail: venugopalreddyk@huawei.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 21, 2017

E. Crabbe
Oracle
I. Minei
Google, Inc.
J. Medved
Cisco Systems, Inc.
R. Varga
Pantheon Technologies SRO
June 19, 2017

PCEP Extensions for Stateful PCE
draft-ietf-pce-stateful-pce-21

Abstract

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

Although PCEP explicitly makes no assumptions regarding the information available to the PCE, it also makes no provisions for PCE control of timing and sequence of path computations within and across PCEP sessions. This document describes a set of extensions to PCEP to enable stateful control of MPLS-TE and GMPLS LSPs via PCEP.

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 December 21, 2017.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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

1. Introduction	3
1.1. Requirements Language	4
2. Terminology	4
3. Motivation and Objectives for Stateful PCE	5
3.1. Motivation	5
3.1.1. Background	5
3.1.2. Why a Stateful PCE?	6
3.1.3. Protocol vs. Configuration	7
3.2. Objectives	7
4. New Functions to Support Stateful PCEs	8
5. Overview of Protocol Extensions	9
5.1. LSP State Ownership	9
5.2. New Messages	9
5.3. Error Reporting	10
5.4. Capability Advertisement	10
5.5. IGP Extensions for Stateful PCE Capabilities Advertisement	11
5.6. State Synchronization	12
5.7. LSP Delegation	15
5.7.1. Delegating an LSP	15
5.7.2. Revoking a Delegation	16
5.7.3. Returning a Delegation	18
5.7.4. Redundant Stateful PCEs	18
5.7.5. Redefinition on PCE Failure	19
5.8. LSP Operations	19
5.8.1. Passive Stateful PCE Path Computation Request/Response	19
5.8.2. Switching from Passive Stateful to Active Stateful .	21
5.8.3. Active Stateful PCE LSP Update	22
5.9. LSP Protection	23
5.10. PCEP Sessions	23
6. PCEP Messages	23
6.1. The PCRpt Message	24
6.2. The PCUpd Message	26
6.3. The PCErr Message	28
6.4. The PCReq Message	29

6.5.	The PCRep Message	30
7.	Object Formats	30
7.1.	OPEN Object	30
7.1.1.	Stateful PCE Capability TLV	30
7.2.	SRP Object	31
7.3.	LSP Object	33
7.3.1.	LSP-IDENTIFIERS TLVs	35
7.3.2.	Symbolic Path Name TLV	38
7.3.3.	LSP Error Code TLV	39
7.3.4.	RSVP Error Spec TLV	40
8.	IANA Considerations	41
8.1.	PCE Capabilities in IGP Advertisements	41
8.2.	PCEP Messages	41
8.3.	PCEP Objects	42
8.4.	LSP Object	42
8.5.	PCEP-Error Object	43
8.6.	Notification Object	43
8.7.	PCEP TLV Type Indicators	44
8.8.	STATEFUL-PCE-CAPABILITY TLV	44
8.9.	LSP-ERROR-CODE TLV	45
9.	Manageability Considerations	45
9.1.	Control Function and Policy	45
9.2.	Information and Data Models	46
9.3.	Liveness Detection and Monitoring	47
9.4.	Verifying Correct Operation	47
9.5.	Requirements on Other Protocols and Functional Components	47
9.6.	Impact on Network Operation	47
10.	Security Considerations	48
10.1.	Vulnerability	48
10.2.	LSP State Snooping	48
10.3.	Malicious PCE	49
10.4.	Malicious PCC	49
11.	Contributing Authors	49
12.	Acknowledgements	50
13.	References	50
13.1.	Normative References	50
13.2.	Informative References	51
	Authors' Addresses	53

1. Introduction

[RFC5440] describes the Path Computation Element Communication Protocol (PCEP). PCEP defines the communication between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between PCEs, enabling computation of Multiprotocol Label Switching (MPLS) for Traffic Engineering Label Switched Path (TE LSP) characteristics. Extensions for support of Generalized MPLS (GMPLS) in PCEP are defined in [I-D.ietf-pce-gmpls-pcep-extensions]

This document specifies a set of extensions to PCEP to enable stateful control of LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect Label Switched Path (LSP) state synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions.

Extensions to permit the PCE to drive creation of an LSP are defined in [I-D.ietf-pce-pce-initiated-lsp], which specifies PCE-initiated LSP creation.

1.1. Requirements Language

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

2. Terminology

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

This document uses the following terms defined in [RFC4655]: TED.

This document uses the following terms defined in [RFC3031]: LSP.

This document uses the following terms defined in [RFC8051]: Stateful PCE, Passive Stateful PCE, Active Stateful PCE, Delegation, LSP State Database.

The following terms are defined in this document:

Revocation: an operation performed by a PCC on a previously delegated LSP. Revocation revokes the rights granted to the PCE in the delegation operation.

Redelegation Timeout Interval: the period of time a PCC waits for, when a PCEP session is terminated, before revoking LSP delegation to a PCE and attempting to redelegate LSPs associated with the terminated PCEP session to an alternate PCE. The Redelegation Timeout Interval is a PCC-local value that can be either operator-configured or dynamically computed by the PCC based on local policy.

State Timeout Interval: the period of time a PCC waits for, when a PCEP session is terminated, before flushing LSP state associated with that PCEP session and reverting to operator-defined default parameters or behaviors. The State Timeout Interval is a PCC-

local value that can be either operator-configured or dynamically computed by the PCC based on local policy.

LSP State Report: an operation to send LSP state (Operational / Admin Status, LSP attributes configured at the PCC and set by a PCE, etc.) from a PCC to a PCE.

LSP Update Request: an operation where an Active Stateful PCE requests a PCC to update one or more attributes of an LSP and to re-signal the LSP with updated attributes.

SRP-ID-number: a number used to correlate errors and LSP State Reports to LSP Update Requests. It is carried in the SRP (Stateful PCE Request Parameters) Object described in Section 7.2.

Within this document, PCEP communications are described through PCC-PCE relationship. The PCE architecture also supports the PCE-PCE communication, by having the requesting PCE fill the role of a PCC, as usual.

The message formats in this document are specified using Routing Backus-Naur Format (RBNF) encoding as specified in [RFC5511].

3. Motivation and Objectives for Stateful PCE

3.1. Motivation

[RFC8051] presents several use cases, demonstrating scenarios that benefit from the deployment of a stateful PCE. The scenarios apply equally to MPLS-TE and GMPLS deployments.

3.1.1. Background

Traffic engineering has been a goal of the MPLS architecture since its inception ([RFC3031], [RFC2702], [RFC3346]). In the traffic engineering system provided by [RFC3630], [RFC5305], and [RFC3209] information about network resources utilization is only available as total reserved capacity by traffic class on a per interface basis; individual LSP state is available only locally on each LER for its own LSPs. In most cases, this makes good sense, as distribution and retention of total LSP state for all LERs within in the network would be prohibitively costly.

Unfortunately, this visibility in terms of global LSP state may result in a number of issues for some demand patterns, particularly within a common setup and hold priority. This issue affects online traffic engineering systems.

A sufficiently over-provisioned system will by definition have no issues routing its demand on the shortest path. However, lowering the degree to which network over-provisioning is required in order to run a healthy, functioning network is a clear and explicit promise of MPLS architecture. In particular, it has been a goal of MPLS to provide mechanisms to alleviate congestion scenarios in which "traffic streams are inefficiently mapped onto available resources; causing subsets of network resources to become over-utilized while others remain underutilized" ([RFC2702]).

3.1.2. Why a Stateful PCE?

[RFC4655] defines a stateful PCE to be one in which the PCE maintains "strict synchronization between the PCE and not only the network states (in term of topology and resource information), but also the set of computed paths and reserved resources in use in the network." [RFC4655] also expressed a number of concerns with regard to a stateful PCE, specifically:

- o Any reliable synchronization mechanism would result in significant control plane overhead
- o Out-of-band TED synchronization would be complex and prone to race conditions
- o Path calculations incorporating total network state would be highly complex

In general, stress on the control plane will be directly proportional to the size of the system being controlled and the tightness of the control loop, and indirectly proportional to the amount of over-provisioning in terms of both network capacity and reservation overhead.

Despite these concerns in terms of implementation complexity and scalability, several TE algorithms exist today that have been demonstrated to be extremely effective in large TE systems, providing both rapid convergence and significant benefits in terms of optimality of resource usage [MXMN-TE]. All of these systems share at least two common characteristics: the requirement for both global visibility of a flow (or in this case, a TE LSP) state and for ordered control of path reservations across devices within the system being controlled. While some approaches have been suggested in order to remove the requirements for ordered control (See [MPLS-PC]), these approaches are highly dependent on traffic distribution, and do not allow for multiple simultaneous LSP priorities representing diffserv classes.

The use cases described in [RFC8051] demonstrate a need for visibility into global inter-PCC LSP state in PCE path computations, and for PCE control of sequence and timing in altering LSP path characteristics within and across PCEP sessions.

3.1.3. Protocol vs. Configuration

Note that existing configuration tools and protocols can be used to set LSP state, such as a Command Line Interface (CLI) tool. However, this solution has several shortcomings:

- o Scale & Performance: configuration operations often have transactional semantics which are typically heavyweight and often require processing of additional configuration portions beyond the state being directly acted upon, with corresponding cost in CPU cycles, negatively impacting both PCC stability LSP update rate capacity.
- o Security: when a PCC opens a configuration channel allowing a PCE to send configuration, a malicious PCE may take advantage of this ability to take over the PCC. In contrast, the PCEP extensions described in this document only allow a PCE control over a very limited set of LSP attributes.
- o Interoperability: each vendor has a proprietary information model for configuring LSP state, which limits interoperability of a stateful PCE with PCCs from different vendors. The PCEP extensions described in this document allow for a common information model for LSP state for all vendors.
- o Efficient State Synchronization: configuration channels may be heavyweight and unidirectional, therefore efficient state synchronization between a PCC and a PCE may be a problem.

3.2. Objectives

The objectives for the protocol extensions to support stateful PCE described in this document are as follows:

- o Allow a single PCC to interact with a mix of stateless and stateful PCEs simultaneously using the same protocol, i.e. PCEP.
- o Support efficient LSP state synchronization between the PCC and one or more active or passive stateful PCEs.
- o Allow a PCC to delegate control of its LSPs to an active stateful PCE such that a given LSP is under the control of a single PCE at any given time.

- * A PCC may revoke this delegation at any time during the lifetime of the LSP. If LSP delegation is revoked while the PCEP session is up, the PCC MUST notify the PCE about the revocation.
- * A PCE may return an LSP delegation at any point during the lifetime of the PCEP session. If LSP delegation is returned by the PCE while the PCEP session is up, the PCE MUST notify the PCC about the returned delegation.
- o Allow a PCE to control computation timing and update timing across all LSPs that have been delegated to it.
- o Enable uninterrupted operation of PCC's LSPs in the event of a PCE failure or while control of LSPs is being transferred between PCEs.

4. New Functions to Support Stateful PCEs

Several new functions are required in PCEP to support stateful PCEs. A function can be initiated either from a PCC towards a PCE (C-E) or from a PCE towards a PCC (E-C). The new functions are:

Capability advertisement (E-C,C-E): both the PCC and the PCE must announce during PCEP session establishment that they support PCEP Stateful PCE extensions defined in this document.

LSP state synchronization (C-E): after the session between the PCC and a stateful PCE is initialized, the PCE must learn the state of a PCC's LSPs before it can perform path computations or update LSP attributes in a PCC.

LSP Update Request (E-C): a PCE requests modification of attributes on a PCC's LSP.

LSP State Report (C-E): a PCC sends an LSP state report to a PCE whenever the state of an LSP changes.

LSP control delegation (C-E,E-C): a PCC grants to a PCE the right to update LSP attributes on one or more LSPs; the PCE becomes the authoritative source of the LSP's attributes as long as the delegation is in effect (See Section 5.7); the PCC may withdraw the delegation or the PCE may give up the delegation at any time.

Similarly to [RFC5440], no assumption is made about the discovery method used by a PCC to discover a set of PCEs (e.g., via static configuration or dynamic discovery) and on the algorithm used to select a PCE.

5. Overview of Protocol Extensions

5.1. LSP State Ownership

In PCEP (defined in [RFC5440]), LSP state and operation are under the control of a PCC (a PCC may be an LSR or a management station). Attributes received from a PCE are subject to PCC's local policy. The PCEP extensions described in this document do not change this behavior.

An active stateful PCE may have control of a PCC's LSPs that were delegated to it, but the LSP state ownership is retained by the PCC. In particular, in addition to specifying values for LSP's attributes, an active stateful PCE also decides when to make LSP modifications.

Retaining LSP state ownership on the PCC allows for:

- o a PCC to interact with both stateless and stateful PCEs at the same time
- o a stateful PCE to only modify a small subset of LSP parameters, i.e. to set only a small subset of the overall LSP state; other parameters may be set by the operator, for example through command line interface (CLI) commands
- o a PCC to revert delegated LSP to an operator-defined default or to delegate the LSPs to a different PCE, if the PCC get disconnected from a PCE with currently delegated LSPs

5.2. New Messages

In this document, we define the following new PCEP messages:

Path Computation State Report (PCRpt): a PCEP message sent by a PCC to a PCE to report the status of one or more LSPs. Each LSP State Report in a PCRpt message MAY contain the actual LSP's path, bandwidth, operational and administrative status, etc. An LSP Status Report carried on a PCRpt message is also used in delegation or revocation of control of an LSP to/from a PCE. The PCRpt message is described in Section 6.1.

Path Computation Update Request (PCUpd): a PCEP message sent by a PCE to a PCC to update LSP parameters, on one or more LSPs. Each LSP Update Request on a PCUpd message MUST contain all LSP parameters that a PCE wishes to be set for a given LSP. An LSP Update Request carried on a PCUpd message is also used to return LSP delegations if at any point PCE no longer desires control of an LSP. The PCUpd message is described in Section 6.2.

The new functions defined in Section 4 are mapped onto the new messages as shown in the following table.

Function	Message
Capability Advertisement (E-C,C-E)	Open
State Synchronization (C-E)	PCRpt
LSP State Report (C-E)	PCRpt
LSP Control Delegation (C-E,E-C)	PCRpt, PCUpd
LSP Update Request (E-C)	PCUpd

Table 1: New Function to Message Mapping

5.3. Error Reporting

Error reporting is done using the procedures defined in [RFC5440], and reusing the applicable error types and error values of [RFC5440] wherever appropriate. The current document defines new error values for several error types to cover failures specific to stateful PCE.

5.4. Capability Advertisement

During PCEP Initialization Phase, PCEP Speakers (PCE or PCC) advertise their support of stateful PCEP extensions. A PCEP Speaker includes the "Stateful PCE Capability" TLV, described in Section 7.1.1, in the OPEN Object to advertise its support for PCEP stateful extensions. The Stateful Capability TLV includes the 'LSP Update' Flag that indicates whether the PCEP Speaker supports LSP parameter updates.

The presence of the Stateful PCE Capability TLV in PCC's OPEN Object indicates that the PCC is willing to send LSP State Reports whenever LSP parameters or operational status changes.

The presence of the Stateful PCE Capability TLV in PCE's OPEN message indicates that the PCE is interested in receiving LSP State Reports whenever LSP parameters or operational status changes.

The PCEP extensions for stateful PCEs MUST NOT be used if one or both PCEP Speakers have not included the Stateful PCE Capability TLV in their respective OPEN message. If the PCEP Speaker on the PCC supports the extensions of this draft but did not advertise this capability, then upon receipt of PCUpd message from the PCE, it MUST generate a PCErr with error-type 19 (Invalid Operation), error-value 2 (Attempted LSP Update Request if the stateful PCE capability was not advertised)(see Section 8.5) and it SHOULD terminate the PCEP

session. If the PCEP Speaker on the PCE supports the extensions of this draft but did not advertise this capability, then upon receipt of a PCRpt message from the PCC, it MUST generate a PCErr with error-type 19 (Invalid Operation), error-value 5 (Attempted LSP State Report if stateful PCE capability was not advertised) (see Section 8.5) and it SHOULD terminate the PCEP session.

LSP delegation and LSP update operations defined in this document may only be used if both PCEP Speakers set the LSP-UPDATE-CAPABILITY Flag in the "Stateful Capability" TLV to 'Updates Allowed (U Flag = 1)'. If this is not the case and LSP delegation or LSP update operations are attempted, then a PCErr with error-type 19 (Invalid Operation) and error-value 1 (Attempted LSP Update Request for a non-delegated LSP) (see Section 8.5) MUST be generated. Note that, even if one of the PCEP speakers does not set the LSP-UPDATE-CAPABILITY flag in its "Stateful Capability" TLV, a PCE can still operate as a passive stateful PCE by accepting LSP State Reports from the PCC in order to build and maintain an up to date view of the state of the PCC's LSPs.

5.5. IGP Extensions for Stateful PCE Capabilities Advertisement

When PCCs are LSRs participating in the IGP (OSPF or IS-IS), and PCEs are either LSRs or servers also participating in the IGP, an effective mechanism for PCE discovery within an IGP routing domain consists of utilizing IGP advertisements. Extensions for the advertisement of PCE Discovery Information are defined for OSPF and for IS-IS in [RFC5088] and [RFC5089] respectively.

The PCE-CAP-FLAGS sub-TLV, defined in [RFC5089], is an optional sub-TLV used to advertise PCE capabilities. It MAY be present within the PCED sub-TLV carried by OSPF or IS-IS. [RFC5088] and [RFC5089] provide the description and processing rules for this sub-TLV when carried within OSPF and IS-IS, respectively.

The format of the PCE-CAP-FLAGS sub-TLV is included below for easy reference:

Type: 5

Length: Multiple of 4.

Value: This contains an array of units of 32 bit flags with the most significant bit as 0. Each bit represents one PCE capability.

PCE capability bits are defined in [RFC5088]. This document defines new capability bits for the stateful PCE as follows:

Bit	Capability
11	Active Stateful PCE capability
12	Passive Stateful PCE capability

Note that while active and passive stateful PCE capabilities may be advertised during discovery, PCEP Speakers that wish to use stateful PCEP MUST negotiate stateful PCEP capabilities during PCEP session setup, as specified in the current document. A PCC MAY initiate stateful PCEP capability negotiation at PCEP session setup even if it did not receive any IGP PCE capability advertisements.

5.6. State Synchronization

The purpose of State Synchronization is to provide a checkpoint-in-time state replica of a PCC's LSP state in a PCE. State Synchronization is performed immediately after the Initialization phase ([RFC5440]).

During State Synchronization, a PCC first takes a snapshot of the state of its LSPs state, then sends the snapshot to a PCE in a sequence of LSP State Reports. Each LSP State Report sent during State Synchronization has the SYNC Flag in the LSP Object set to 1. The set of LSPs for which state is synchronized with a PCE is determined by the PCC's local configuration (see more details in Section 9.1) and MAY also be determined by stateful PCEP capabilities defined in other documents, such as [I-D.ietf-pce-stateful-sync-optimizations].

The end of synchronization marker is a PCRpt message with the SYNC Flag set to 0 for an LSP Object with PLSP-ID equal to the reserved value 0 (see Section 7.3). In this case, the LSP Object SHOULD NOT include the SYMBOLIC-PATH-NAME TLV and SHOULD include the LSP-IDENTIFIERS TLV with the special value of all zeroes. The PCRpt message MUST include an empty ERO as its intended path and SHOULD NOT include the optional RRO object for its actual path. If the PCC has no state to synchronize, it SHOULD only send the end of synchronization marker.

A PCE SHOULD NOT send PCUpd messages to a PCC before State Synchronization is complete. A PCC SHOULD NOT send PCReq messages to a PCE before State Synchronization is complete. This is to allow the PCE to get the best possible view of the network before it starts computing new paths.

Either the PCE or the PCC MAY terminate the session using the PCEP session termination procedures during the synchronization phase. If the session is terminated, the PCE MUST clean up state it received from this PCC. The session reestablishment MUST be re-attempted per

the procedures defined in [RFC5440], including use of a back-off timer.

If the PCC encounters a problem which prevents it from completing the LSP state synchronization, it MUST send a PCErr message with error-type 20 (LSP State Synchronization Error) and error-value 5 (indicating an internal PCC error) to the PCE and terminate the session.

The PCE does not send positive acknowledgements for properly received synchronization messages. It MUST respond with a PCErr message with error-type 20 (LSP State Synchronization Error) and error-value 1 (indicating an error in processing the PCRpt) (see Section 8.5) if it encounters a problem with the LSP State Report it received from the PCC and it MUST terminate the session.

A PCE implementing a limit on the resources a single PCC can occupy, MUST send a PCNtf message with Notification Type 4 (Stateful PCE resource limit exceeded) and Notification Value 1 (Entering resource limit exceeded state) in response to the PCRpt message triggering this condition in the synchronization phase and MUST terminate the session.

The successful State Synchronization sequence is shown in Figure 1.

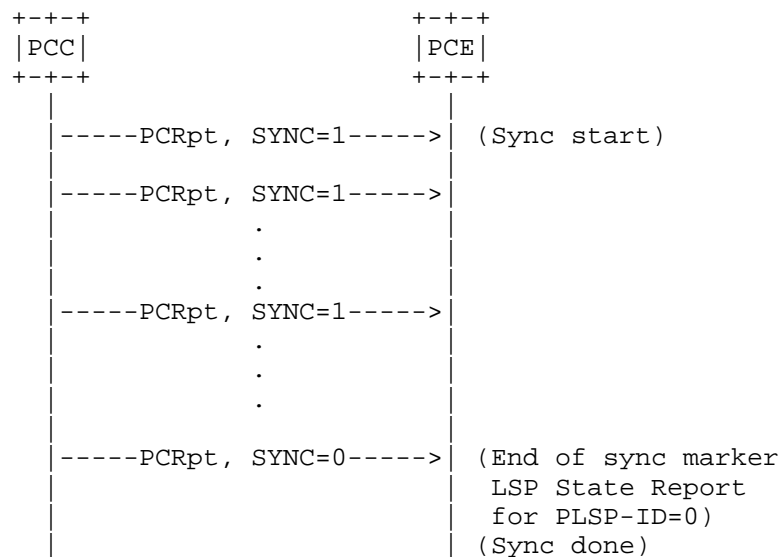


Figure 1: Successful state synchronization

The sequence where the PCE fails during the State Synchronization phase is shown in Figure 2.

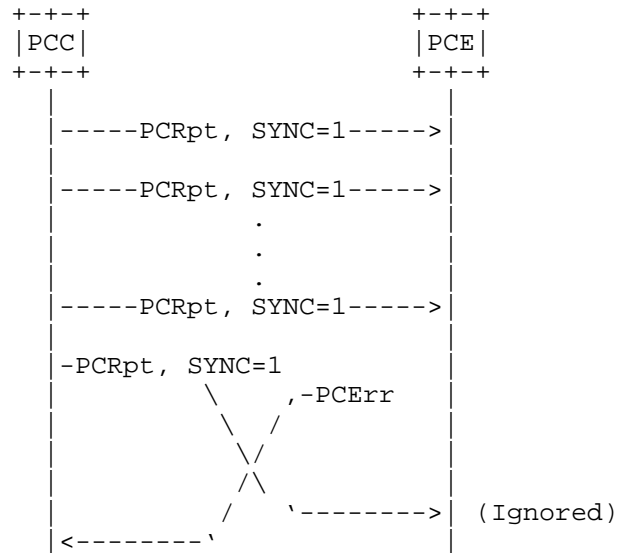


Figure 2: Failed state synchronization (PCE failure)

The sequence where the PCC fails during the State Synchronization phase is shown in Figure 3.

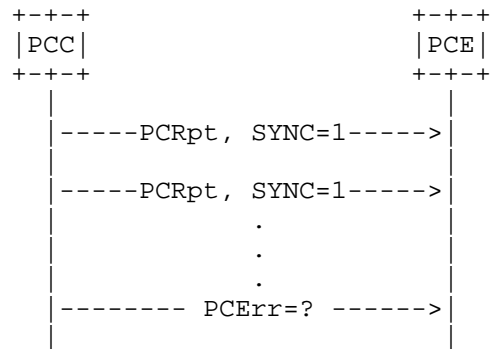


Figure 3: Failed state synchronization (PCC failure)

Optimizations to the synchronization procedures and alternate mechanisms of providing the synchronization function are outside the scope of this document and are discussed elsewhere (see [I-D.ietf-pce-stateful-sync-optimizations]).

5.7. LSP Delegation

If during Capability advertisement both the PCE and the PCC have indicated that they support LSP Update, then the PCC may choose to grant the PCE a temporary right to update (a subset of) LSP attributes on one or more LSPs. This is called "LSP Delegation", and it MAY be performed at any time after the Initialization phase, including during the State Synchronization phase.

A PCE MAY return an LSP delegation at any time if it no longer wishes to update the LSP's state. A PCC MAY revoke an LSP delegation at any time. Delegation, Revocation, and Return are done individually for each LSP.

In the event of a delegation being rejected or returned by a PCE, the PCC SHOULD react based on local policy. It can, for example, either retry delegating to the same PCE using an exponentially increasing timer or delegate to an alternate PCE.

5.7.1. Delegating an LSP

A PCC delegates an LSP to a PCE by setting the Delegate flag in LSP State Report to 1. If the PCE does not accept the LSP Delegation, it MUST immediately respond with an empty LSP Update Request which has the Delegate flag set to 0. If the PCE accepts the LSP Delegation, it MUST set the Delegate flag to 1 when it sends an LSP Update Request for the delegated LSP (note that this may occur at a later time). The PCE MAY also immediately acknowledge a delegation by sending an empty LSP Update Request which has the Delegate flag set to 1.

The delegation sequence is shown in Figure 4.

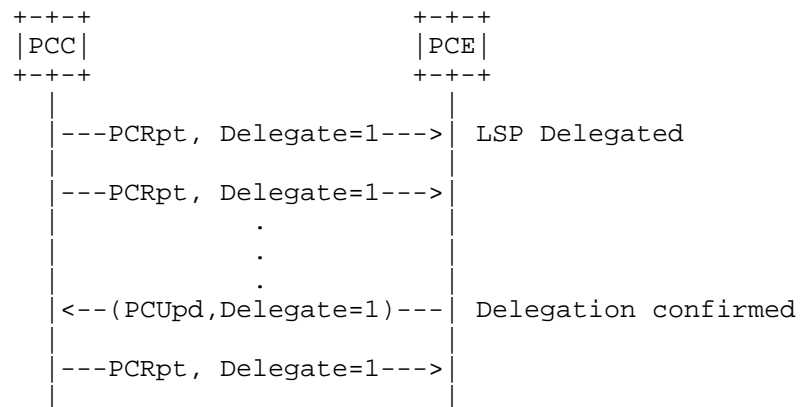


Figure 4: Delegating an LSP

Note that for an LSP to remain delegated to a PCE, the PCC MUST set the Delegate flag to 1 on each LSP State Report sent to the PCE.

5.7.2. Revoking a Delegation

5.7.2.1. Explicit Revocation

When a PCC decides that a PCE is no longer permitted to modify an LSP, it revokes that LSP's delegation to the PCE. A PCC may revoke an LSP delegation at any time during the LSP's life time. A PCC revoking an LSP delegation MAY immediately remove the updated parameters provided by the PCE and revert to the operator-defined parameters, but to avoid traffic loss, it SHOULD do so in a make-before-break fashion. If the PCC has received but not yet acted on PCUpd messages from the PCE for the LSP whose delegation is being revoked, then it SHOULD ignore these PCUpd messages when processing the message queue. All effects of all messages for which processing started before the revocation took place MUST be allowed to complete and the result MUST be given the same treatment as any LSP that had been previously delegated to the PCE (e.g. the state MAY immediately revert to the operator-defined parameters).

If a PCEP session with the PCE to which the LSP is delegated exists in the UP state during the revocation, the PCC MUST notify that PCE by sending an LSP State Report with the Delegate flag set to 0, as shown in Figure 5.

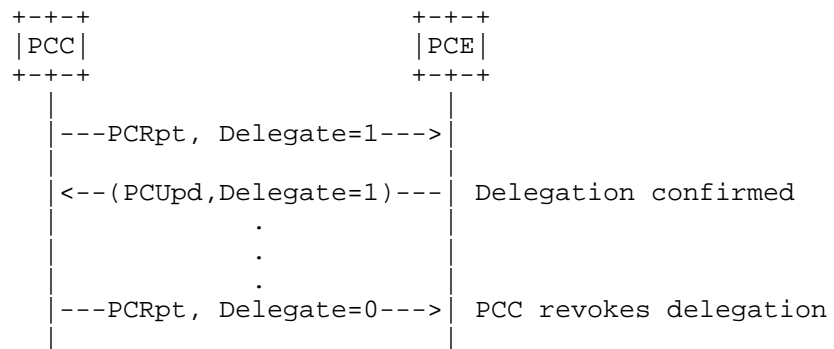


Figure 5: Revoking a Delegation

After an LSP delegation has been revoked, a PCE can no longer update LSP's parameters; an attempt to update parameters of a non-delegated LSP will result in the PCC sending a PCErr message with error-type 19 (Invalid Operation), error-value 1 (attempted LSP Update Request for a non-delegated LSP) (see Section 8.5).

5.7.2.2. Revocation on Redelegating Timeout

When a PCC's PCEP session with a PCE terminates unexpectedly, the PCC MUST wait the time interval specified in Redelegating Timeout Interval before revoking LSP delegations to that PCE and attempting to redelegate LSPs to an alternate PCE. If a PCEP session with the original PCE can be reestablished before the Redelegating Timeout Interval timer expires, LSP delegations to the PCE remain intact.

Likewise, when a PCC's PCEP session with a PCE terminates unexpectedly, and the PCC does not succeed in redelegating its LSPs, the PCC MUST wait for the State Timeout Interval before flushing any LSP state associated with that PCE. Note that the State Timeout Interval timer may expire before the PCC has redelegated the LSPs to another PCE, for example if a PCC is not connected to any active stateful PCE or if no connected active stateful PCE accepts the delegation. In this case, the PCC MUST flush any LSP state set by the PCE upon expiration of the State Timeout Interval and revert to operator-defined default parameters or behaviors. This operation SHOULD be done in a make-before-break fashion.

The State Timeout Interval MUST be greater than or equal to the Redelegating Timeout Interval and MAY be set to infinity (meaning that until the PCC specifically takes action to change the parameters set by the PCE, they will remain intact).

5.7.3. Returning a Delegation

In order to keep a delegation, a PCE MUST set the Delegate flag to 1 on each LSP Update Request sent to the PCC. A PCE that no longer wishes to update an LSP's parameters SHOULD return the LSP delegation back to the PCC by sending an empty LSP Update Request which has the Delegate flag set to 0. If a PCC receives an LSP Update Request with the Delegate flag set to 0 (whether the LSP Update Request is empty or not), it MUST treat this as a delegation return.

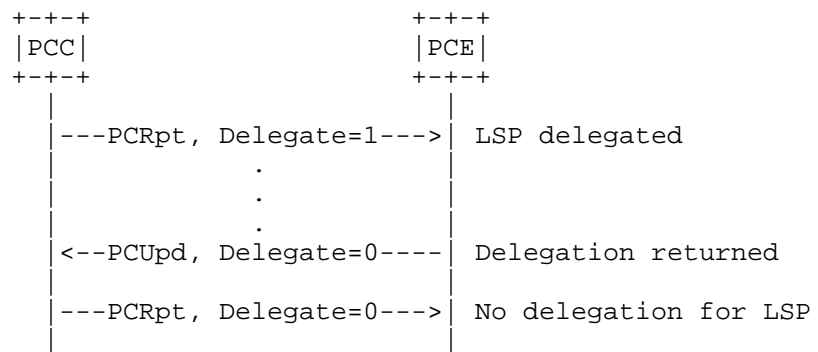


Figure 6: Returning a Delegation

If a PCC cannot delegate an LSP to a PCE (for example, if a PCC is not connected to any active stateful PCE or if no connected active stateful PCE accepts the delegation), the LSP delegation on the PCC will time out within a configurable Redelegating Timeout Interval and the PCC MUST flush any LSP state set by a PCE at the expiration of the State Timeout Interval and revert to operator-defined default parameters or behaviors.

5.7.4. Redundant Stateful PCEs

In a redundant configuration where one PCE is backing up another PCE, the backup PCE may have only a subset of the LSPs in the network delegated to it. The backup PCE does not update any LSPs that are not delegated to it. In order to allow the backup to operate in a hot-standby mode and avoid the need for state synchronization in case the primary fails, the backup receives all LSP State Reports from a PCC. When the primary PCE for a given LSP set fails, after expiry of the Redelegating Timeout Interval, the PCC SHOULD delegate to the redundant PCE all LSPs that had been previously delegated to the failed PCE. Assuming that the State Timeout Interval had been configured to be greater than the Redelegating Timeout Interval (as MANDATORY), and assuming that the primary and redundant PCEs take

similar decisions, this delegation change will not cause any changes to the LSP parameters.

5.7.5. Redelegation on PCE Failure

On failure, the goal is to: 1) avoid any traffic loss on the LSPs that were updated by the PCE that crashed 2) minimize the churn in the network in terms of ownership of the LSPs, 3) not leave any "orphan" (undelegated) LSPs and 4) be able to control when the state that was set by the PCE can be changed or purged. The values chosen for the Redelegation Timeout and State Timeout values affect the ability to accomplish these goals.

This section summarizes the behaviour with regards to LSP delegation and LSP state on a PCE failure.

If the PCE crashes but recovers within the Redelegation Timeout, both the delegation state and the LSP state are kept intact.

If the PCE crashes but does not recover within the Redelegation Timeout, the delegation state is returned to the PCC. If the PCC can redelegate the LSPs to another PCE, and that PCE accepts the delegations, there will be no change in LSP state. If the PCC cannot redelegate the LSPs to another PCE, then upon expiration of the State Timeout Interval, the state set by the PCE is removed and the LSP reverts to operator-defined parameters, which may cause a change in the LSP state. Note that an operator may choose to use an infinite State Timeout Interval if he wishes to maintain the PCE state indefinitely. Note also that flushing the state should be implemented using make-before-break to avoid traffic loss.

If there is a standby PCE, the Redelegation Timeout may be set to 0 through policy on the PCC, causing the LSPs to be redelegated immediately to the PCC, which can delegate them immediately to the standby PCE. Assuming that the PCC can redelegate the LSP to the standby PCE within the State Timeout Interval, and assuming the standby PCE takes similar decisions as the failed PCE, the LSP state will be kept intact.

5.8. LSP Operations

5.8.1. Passive Stateful PCE Path Computation Request/Response

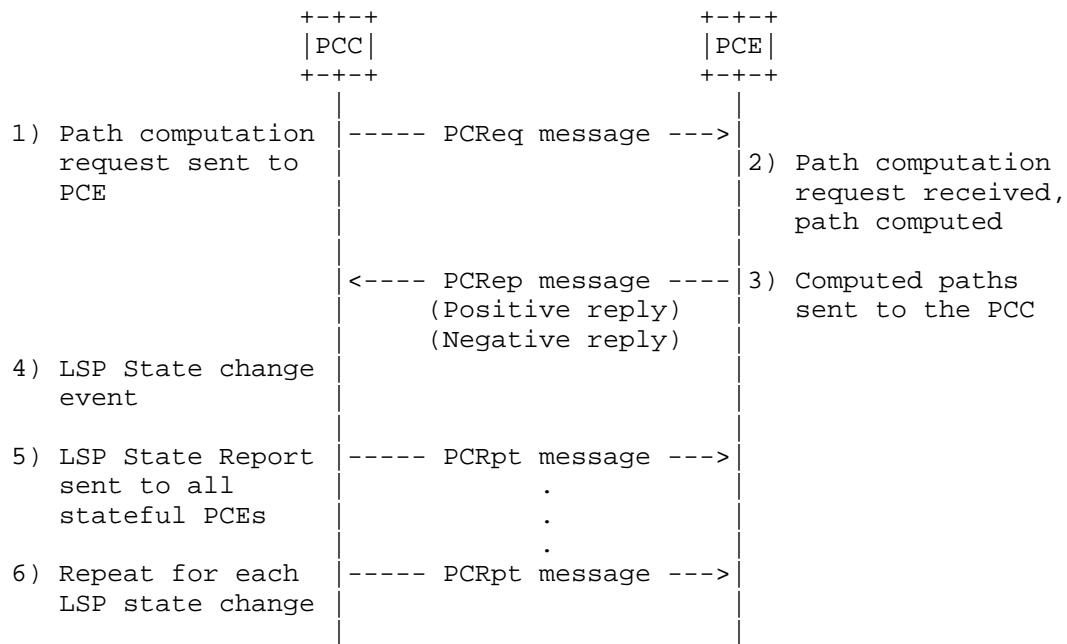


Figure 7: Passive Stateful PCE Path Computation Request/Response

Once a PCC has successfully established a PCEP session with a passive stateful PCE and the PCC's LSP state is synchronized with the PCE (i.e. the PCE knows about all PCC's existing LSPs), if an event is triggered that requires the computation of a set of paths, the PCC sends a path computation request to the PCE ([RFC5440], Section 4.2.3). The PCReq message MAY contain the LSP Object to identify the LSP for which the path computation is requested.

Upon receiving a path computation request from a PCC, the PCE triggers a path computation and returns either a positive or a negative reply to the PCC ([RFC5440], Section 4.2.4).

Upon receiving a positive path computation reply, the PCC receives a set of computed paths and starts to setup the LSPs. For each LSP, it MAY send an LSP State Report carried on a PCRpt message to the PCE, indicating that the LSP's status is "Going-up".

Once an LSP is up or active, the PCC MUST send an LSP State Report carried on a PCRpt message to the PCE, indicating that the LSP's status is 'Up' or 'Active' respectively. If the LSP could not be set up, the PCC MUST send an LSP State Report indicating that the LSP is "Down" and stating the cause of the failure. Note that due to timing constraints, the LSP status may change from 'Going-up' to 'Up' (or

'Down') before the PCC has had a chance to send an LSP State Report indicating that the status is 'Going-up'. In such cases, the PCC MAY choose to only send the PCRpt indicating the latest status ('Active', 'Up' or 'Down').

Upon receiving a negative reply from a PCE, a PCC MAY resend a modified request or take any other appropriate action. For each requested LSP, it SHOULD also send an LSP State Report carried on a PCRpt message to the PCE, indicating that the LSP's status is 'Down'.

There is no direct correlation between PCRep and PCRpt messages. For a given LSP, multiple LSP State Reports will follow a single PCRep message, as a PCC notifies a PCE of the LSP's state changes.

A PCC MUST send each LSP State Report to each stateful PCE that is connected to the PCC.

Note that a single PCRpt message MAY contain multiple LSP State Reports.

The passive stateful model for stateful PCEs is described in [RFC4655], Section 6.8.

5.8.2. Switching from Passive Stateful to Active Stateful

This section deals with the scenario of an LSP transitioning from a passive stateful to an active stateful mode of operation. When the LSP has no working path, prior to delegating the LSP, the PCC MUST first use the procedure defined in Section 5.8.1 to request the initial path from the PCE. This is required because the action of delegating the LSP to a PCE using a PCRpt message is not an explicit request to the PCE to compute a path for the LSP. The only explicit way for a PCC to request a path from PCE is to send a PCReq message. The PCRpt message MUST NOT be used by the PCC to attempt to request a path from the PCE.

When the LSP is delegated after its setup, it may be useful for the PCC to communicate to the PCE the locally configured intended configuration parameters, so that the PCE may reuse them in its computations. Such parameters MAY be acquired through an out of band channel, or MAY be communicated in the PCRpt message delegating the LSPs, by including them as part of the intended-attribute-list as explained in Section 6.1. An implementation MAY allow policies on the PCC to determine the configuration parameters to be sent to the PCE.

5.8.3. Active Stateful PCE LSP Update

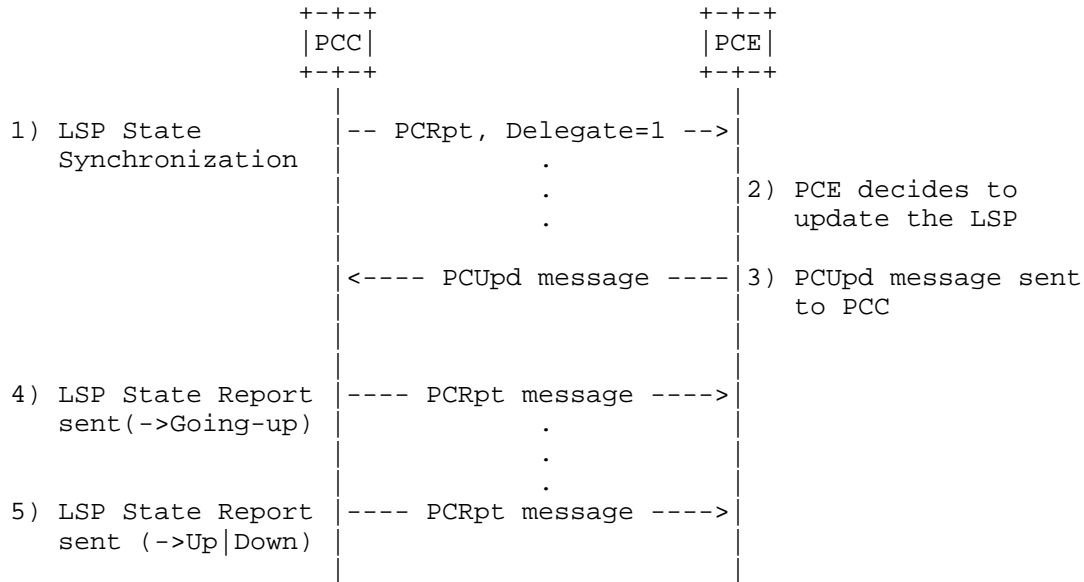


Figure 8: Active Stateful PCE

Once a PCC has successfully established a PCEP session with an active stateful PCE, the PCC's LSP state is synchronized with the PCE (i.e. the PCE knows about all PCC's existing LSPs). After LSPs have been delegated to the PCE, the PCE can modify LSP parameters of delegated LSPs.

To update an LSP, a PCE MUST send the PCC an LSP Update Request using a PCUpd message. The LSP Update Request contains a variety of objects that specify the set of constraints and attributes for the LSP's path. Each LSP Update Request MUST have a unique identifier, the SRP-ID-number, carried in the SRP (Stateful PCE Request Parameters) Object described in Section 7.2. The SRP-ID-number is used to correlate errors and state reports to LSP Update Requests. A single PCUpd message MAY contain multiple LSP Update Requests.

Upon receiving a PCUpd message the PCC starts to setup LSPs specified in LSP Update Requests carried in the message. For each LSP, it MAY send an LSP State Report carried on a PCRpt message to the PCE, indicating that the LSP's status is 'Going-up'. If the PCC decides that the LSP parameters proposed in the PCUpd message are unacceptable, it MUST report this error by including the LSP-ERROR-CODE TLV (Section 7.3.3) with LSP error-value="Unacceptable parameters" in the LSP object in the PCRpt message to the PCE. Based

on local policy, it MAY react further to this error by revoking the delegation. If the PCC receives a PCUpd message for an LSP object identified with a PLSP-ID that does not exist on the PCC, it MUST generate a PCErr with error-type 19 (Invalid Operation), error-value 3, (Attempted LSP Update Request for an LSP identified by an unknown PSP-ID) (see Section 8.5).

Once an LSP is up, the PCC MUST send an LSP State Report (PCRpt message) to the PCE, indicating that the LSP's status is 'Up'. If the LSP could not be set up, the PCC MUST send an LSP State Report indicating that the LSP is 'Down' and stating the cause of the failure. A PCC MAY compress LSP State Reports to only reflect the most up to date state, as discussed in the previous section.

A PCC MUST send each LSP State Report to each stateful PCE that is connected to the PCC.

PCErr and PCRpt messages triggered as a result of a PCUpd message MUST include the SRP-ID-number from the PCUpd. This provides correlation of requests and errors and acknowledgement of state processing. The PCC MAY compress state when processing PCUpd. In this case, receipt of a higher SRP-ID-number implicitly acknowledges processing all the updates with lower SRP-ID-number for the specific LSP (as per Section 7.2).

A PCC MUST NOT send to any PCE a Path Computation Request for a delegated LSP. Should the PCC decide it wants to issue a Path Computation Request on a delegated LSP, it MUST perform Delegation Revocation procedure first.

5.9. LSP Protection

LSP protection and interaction with stateful PCE, as well as the extensions necessary to implement this functionality will be discussed in a separate document.

5.10. PCEP Sessions

A permanent PCEP session MUST be established between a stateful PCE and the PCC. In the case of session failure, session reestablishment MUST be re-attempted per the procedures defined in [RFC5440].

6. PCEP Messages

As defined in [RFC5440], a PCEP message consists of a common header followed by a variable-length body made of a set of objects. For each PCEP message type, a set of rules is defined that specify the set of objects that the message can carry.

6.1. The PCRpt Message

A Path Computation LSP State Report message (also referred to as PCRpt message) is a PCEP message sent by a PCC to a PCE to report the current state of an LSP. A PCRpt message can carry more than one LSP State Reports. A PCC can send an LSP State Report either in response to an LSP Update Request from a PCE, or asynchronously when the state of an LSP changes. The Message-Type field of the PCEP common header for the PCRpt message is 10.

The format of the PCRpt message is as follows:

```
<PCRpt Message> ::= <Common Header>
                        <state-report-list>
```

Where:

```
<state-report-list> ::= <state-report>[<state-report-list>]
```

```
<state-report> ::= [<SRP>]
                    <LSP>
                    <path>
```

Where:

```
<path> ::= <intended-path>
           [<actual-attribute-list><actual-path>]
           <intended-attribute-list>
```

```
<actual-attribute-list> ::= [<BANDWIDTH>]
                           [<metric-list>]
```

Where:

```
<intended-path> is represented by the ERO object defined in
section 7.9 of [RFC5440].
<actual-attribute-list> consists of the actual computed and
signaled values of the <BANDWIDTH> and <metric-lists> objects
defined in [RFC5440].
<actual-path> is represented by the RRO object defined in
section 7.10 of [RFC5440].
<intended-attribute-list> is the attribute-list defined in
section 6.5 of [RFC5440] and extended by PCEP extensions.
```

The SRP object (see Section 7.2) is OPTIONAL. If the PCRpt message is not in response to a PCUpd message, the SRP object MAY be omitted. When the PCC does not include the SRP object, the PCE MUST treat this as an SRP object with an SRP-ID-number equal to the reserved value 0x00000000. The reserved value 0x00000000 indicates that the state reported is not as a result of processing a PCUpd message.

If the PCRpt message is in response to a PCUpd message, the SRP object MUST be included and the value of the SRP-ID-number in the SRP Object MUST be the same as that sent in the PCUpd message that triggered the state that is reported. If the PCC compressed several PCUpd messages for the same LSP by only processing the one with the highest number, then it should use the SRP-ID-number of that request. No state compression is allowed for state reporting, e.g. PCRpt messages MUST NOT be pruned from the PCC's egress queue even if subsequent operations on the same LSP have been completed before the PCRpt message has been sent to the TCP stack. The PCC MUST explicitly report state changes (including removal) for paths it manages.

The LSP object (see Section 7.3) is REQUIRED, and it MUST be included in each LSP State Report on the PCRpt message. If the LSP object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value 8 (LSP object missing).

If the LSP transitioned to non-operational state, the PCC SHOULD include the LSP-ERROR-TLV (Section 7.3.3) with the relevant LSP Error Code to report the error to the PCE.

The intended path, represented by the ERO object, is REQUIRED. If the ERO object is missing, the receiving PCE MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value 9 (ERO object missing). The ERO may be empty if the PCE does not have a path for a delegated LSP.

The actual path, represented by the RRO object, SHOULD be included in PCRpt by the PCC when the path is up or active, but MAY be omitted if the path is down due to a signaling error or another failure.

The intended-attribute-list maps to the attribute-list in Section 6.5 of [RFC5440] and is used to convey the requested parameters of the LSP path. This is needed in order to support the switch from passive to active stateful PCE as described in Section 5.8.2. When included as part of the intended-attribute-list, the meaning of the BANDWIDTH object is the requested bandwidth as intended by the operator. In this case, the BANDWIDTH Object-Type of 1 SHOULD be used. Similarly, to indicate a limiting constraint, the METRIC object SHOULD be included as part of the intended-attribute-list with the B flag set and with a specific metric value. To indicate the optimization metric, the METRIC object SHOULD be included as part of the intended-attribute-list with the B flag unset and the metric value set to zero. Note that the intended-attribute-list is optional and thus may be omitted. In this case, the PCE MAY use the values in the actual-attribute-list as the requested parameters for the path.

The actual-attribute-list consists of the actual computed and signaled values of the BANDWIDTH and METRIC objects defined in [RFC5440]. When included as part of the actual-attribute-list, Object-Type 2 ([RFC5440]) SHOULD be used for the BANDWIDTH object and the C flag SHOULD be set in the METRIC object ([RFC5440]).

Note that the ordering of intended-path, actual-attribute-list, actual-path and intended-attribute-list is chosen to retain compatibility with implementations of an earlier version of this standard.

A PCE may choose to implement a limit on the resources a single PCC can occupy. If a PCRpt is received that causes the PCE to exceed this limit, the PCE MUST notify the PCC using a PCNtf message with Notification Type 4 (Stateful PCE resource limit exceeded) and Notification Value 1 (Entering resource limit exceeded state) and MUST terminate the session.

6.2. The PCUpd Message

A Path Computation LSP Update Request message (also referred to as PCUpd message) is a PCEP message sent by a PCE to a PCC to update attributes of an LSP. A PCUpd message can carry more than one LSP Update Request. The Message-Type field of the PCEP common header for the PCUpd message is 11.

The format of a PCUpd message is as follows:

```
<PCUpd Message> ::= <Common Header>
                        <update-request-list>
```

Where:

```
<update-request-list> ::= <update-request>[<update-request-list>]
```

```
<update-request> ::= <SRP>
                        <LSP>
                        <path>
```

Where:

```
<path> ::= <intended-path><intended-attribute-list>
```

Where:

```
<intended-path> is represented by the ERO object defined in
section 7.9 of [RFC5440].
<intended-attribute-list> is the attribute-list defined in [RFC5440]
and extended by PCEP extensions.
```

There are three mandatory objects that MUST be included within each LSP Update Request in the PCUpd message: the SRP Object (see

Section 7.2), the LSP object (see Section 7.3) and the ERO object (as defined in [RFC5440], which represents the intended path. If the SRP object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=10 (SRP object missing). If the LSP object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=8 (LSP object missing). If the ERO object is missing, the receiving PCC MUST send a PCErr message with Error-type=6 (Mandatory Object missing) and Error-value=9 (ERO object missing).

The ERO in the PCUpd may be empty if the PCE cannot find a valid path for a delegated LSP. One typical situation resulting in this empty ERO carried in the PCUpd message is that a PCE can no longer find a strict SRLG-disjoint path for a delegated LSP after a link failure. The PCC SHOULD implement a local policy to decide the appropriate action to be taken: either tear down the LSP, or revoke the delegation and use a locally computed path, or keep the existing LSP.

A PCC only acts on an LSP Update Request if permitted by the local policy configured by the network manager. Each LSP Update Request that the PCC acts on results in an LSP setup operation. An LSP Update Request MUST contain all LSP parameters that a PCE wishes to be set for the LSP. A PCC MAY set missing parameters from locally configured defaults. If the LSP specified in the Update Request is already up, it will be re-signaled.

The PCC SHOULD minimize the traffic interruption, and MAY use the make-before-break procedures described in [RFC3209] in order to achieve this goal. If the make-before-break procedures are used, two paths will briefly co-exist. The PCC MUST send separate PCRpt messages for each, identified by the LSP-IDENTIFIERS TLV. When the old path is torn down after the head end switches over the traffic, this event MUST be reported by sending a PCRpt message with the LSP-IDENTIFIERS-TLV of the old path and the R bit set. The SRP-ID-number that the PCC associates with this PCRpt MUST be 0x00000000. Thus, a make-before-break operation will typically result in at least two PCRpt messages, one for the new path and one for the removal of the old path (more messages may be possible if intermediate states are reported).

If the path setup fails due to an RSVP signaling error, the error is reported to the PCE. The PCC will not attempt to resignal the path until it is prompted again by the PCE with a subsequent PCUpd message.

A PCC MUST respond with an LSP State Report to each LSP Update Request it processed to indicate the resulting state of the LSP in

the network (even if this processing did not result in changing the state of the LSP). The SRP-ID-number included in the PCRpt MUST match that in the PCUpd. A PCC MAY respond with multiple LSP State Reports to report LSP setup progress of a single LSP. In that case, the SRP-ID-number MUST be included for the first message, for subsequent messages the reserved value 0x00000000 SHOULD be used.

Note that a PCC MUST process all LSP Update Requests - for example, an LSP Update Request is sent when a PCE returns delegation or puts an LSP into non-operational state. The protocol relies on TCP for message-level flow control.

If the rate of PCUpd messages sent to a PCC for the same target LSP exceeds the rate at which the PCC can signal LSPs into the network, the PCC MAY perform state compression on its ingress queue. The compression algorithm is based on the fact that each PCUpd request contains the complete LSP state the PCE wishes to be set and works as follows: when the PCC starts processing a PCUpd message at the head of its ingress queue, it may search the queue forward for more recent PCUpd messages pertaining that particular LSP, prune all but the latest one from the queue and process only the last one as that request contains the most up-to-date desired state for the LSP. The PCC MUST NOT send PCRpt nor PCErr messages for requests which were pruned from the queue in this way. This compression step may be performed only while the LSP is not being signaled, e.g. if two PCUpd arrive for the same LSP in quick succession and the PCC started the signaling of the changes relevant to the first PCUpd, then it MUST wait until the signaling finishes (and report the new state via a PCRpt) before attempting to apply the changes indicated in the second PCUpd.

Note also that it is up to the PCE to handle inter-LSP dependencies; for example, if ordering of LSP set-ups is required, the PCE has to wait for an LSP State Report for a previous LSP before starting the update of the next LSP.

If the PCUpd cannot be satisfied (for example due to unsupported object or TLV), the PCC MUST respond with a PCErr message indicating the failure (see Section 7.3.3).

6.3. The PCErr Message

If the stateful PCE capability has been advertised on the PCEP session, the PCErr message MAY include the SRP object. If the error reported is the result of an LSP update request, then the SRP-ID-number MUST be the one from the PCUpd that triggered the error. If the error is unsolicited, the SRP object MAY be omitted. This is

equivalent to including an SRP object with SRP-ID-number equal to the reserved value 0x00000000.

The format of a PCErr message from [RFC5440] is extended as follows:

```

<PCErr Message> ::= <Common Header>
                    ( <error-obj-list> [<Open>] ) | <error>
                    [<error-list>]

<error-obj-list> ::= <PCEP-ERROR> [<error-obj-list>]

<error> ::= [<request-id-list> | <stateful-request-id-list>]
           <error-obj-list>

<request-id-list> ::= <RP> [<request-id-list>]

<stateful-request-id-list> ::= <SRP> [<stateful-request-id-list>]

<error-list> ::= <error> [<error-list>]

```

6.4. The PCReq Message

A PCC MAY include the LSP object in the PCReq message (see Section 7.3) if the stateful PCE capability has been negotiated on a PCEP session between the PCC and a PCE.

The definition of the PCReq message from [RFC5440] is extended to optionally include the LSP object after the END-POINTS object. The encoding from [RFC5440] will become:

```

<PCReq Message> ::= <Common Header>
                    [<svec-list>]
                    <request-list>

```

Where:

```

<svec-list> ::= <SVEC> [<svec-list>]
<request-list> ::= <request> [<request-list>]

<request> ::= <RP>
              <END-POINTS>
              [<LSP>]
              [<LSPA>]
              [<BANDWIDTH>]
              [<metric-list>]
              [<RRO> [<BANDWIDTH>]]
              [<IRO>]
              [<LOAD-BALANCING>]

```

6.5. The PCRep Message

A PCE MAY include the LSP object in the PCRep message (see (Section 7.3) if the stateful PCE capability has been negotiated on a PCEP session between the PCC and the PCE and the LSP object was included in the corresponding PCReq message from the PCC.

The definition of the PCRep message from [RFC5440] is extended to optionally include the LSP object after the RP object. The encoding from [RFC5440] will become:

```
<PCRep Message> ::= <Common Header>
                        <response-list>
```

Where:

```
<response-list> ::= <response> [<response-list>]

<response> ::= <RP>
                [<LSP>]
                [<NO-PATH>]
                [<attribute-list>]
                [<path-list>]
```

7. Object Formats

The PCEP objects defined in this document are compliant with the PCEP object format defined in [RFC5440]. The P flag and the I flag of the PCEP objects defined in the current document MUST be set to 0 on transmission and SHOULD be ignored on receipt since the P and I flags are exclusively related to path computation requests.

7.1. OPEN Object

This document defines one new optional TLV for use in the OPEN Object.

7.1.1. Stateful PCE Capability TLV

The STATEFUL-PCE-CAPABILITY TLV is an optional TLV for use in the OPEN Object for stateful PCE capability advertisement. Its format is shown in the following figure:



Figure 9: STATEFUL-PCE-CAPABILITY TLV format

The type (16 bits) of the TLV is 16. The length field is 16 bit-long and has a fixed value of 4.

The value comprises a single field - Flags (32 bits):

U (LSP-UPDATE-CAPABILITY - 1 bit): if set to 1 by a PCC, the U Flag indicates that the PCC allows modification of LSP parameters; if set to 1 by a PCE, the U Flag indicates that the PCE is capable of updating LSP parameters. The LSP-UPDATE-CAPABILITY Flag must be advertised by both a PCC and a PCE for PCUpd messages to be allowed on a PCEP session.

Unassigned bits are considered reserved. They MUST be set to 0 on transmission and MUST be ignored on receipt.

A PCEP speaker operating in passive stateful PCE mode advertises the stateful PCE capability with the U flag set to 0. A PCEP speaker operating in active stateful PCE mode advertises the stateful PCE capability with the U Flag set to 1.

Advertisement of the stateful PCE capability implies support of LSPs that are signaled via RSVP, as well as the objects, TLVs and procedures defined in this document.

7.2. SRP Object

The SRP (Stateful PCE Request Parameters) object MUST be carried within PCUpd messages and MAY be carried within PCRpt and PCErr messages. The SRP object is used to correlate between update requests sent by the PCE and the error reports and state reports sent by the PCC.

SRP Object-Class is 33.

SRP Object-Type is 1.

The format of the SRP object body is shown in Figure 10:

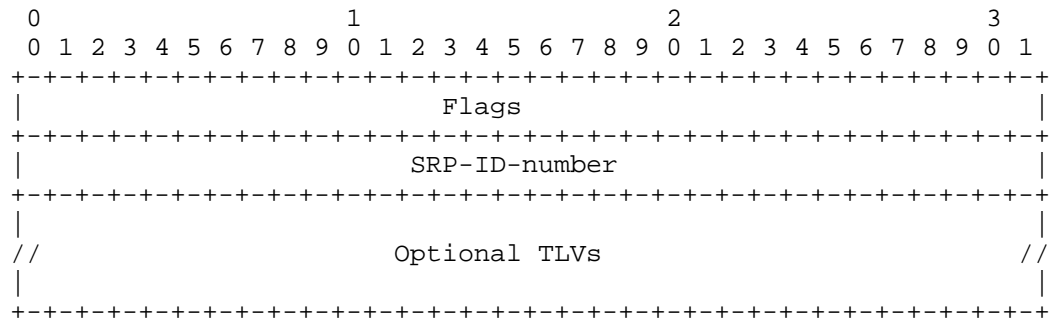


Figure 10: The SRP Object format

The SRP object body has a variable length and may contain additional TLVs.

Flags (32 bits): None defined yet.

SRP-ID-number (32 bits): The SRP-ID-number value in the scope of the current PCEP session uniquely identify the operation that the PCE has requested the PCC to perform on a given LSP. The SRP-ID-number is incremented each time a new request is sent to the PCC, and may wrap around.

The values 0x00000000 and 0xFFFFFFFF are reserved.

Optional TLVs MAY be included within the SRP object body. The specification of such TLVs is outside the scope of this document.

Every request to update an LSP receives a new SRP-ID-number. This number is unique per PCEP session and is incremented each time an operation is requested from the PCE. Thus, for a given LSP there may be more than one SRP-ID-number unacknowledged at a given time. The value of the SRP-ID-number is echoed back by the PCC in PCErr and PCRpt messages to allow for correlation between requests made by the PCE and errors or state reports generated by the PCC. If the error or report were not as a result of a PCE operation (for example in the case of a link down event), the reserved value of 0x00000000 is used for the SRP-ID-number. The absence of the SRP object is equivalent to an SRP object with the reserved value of 0x00000000. An SRP-ID-number is considered unacknowledged and cannot be reused until a PCErr or PCRpt arrives with an SRP-ID-number equal or higher for the same LSP. In case of SRP-ID-number wrapping the last SRP-ID-number before the wrapping MUST be explicitly acknowledged, to avoid a situation where SRP-ID-numbers remain unacknowledged after the wrap.

This means that the PCC may need to issue two PCUpd messages on detecting a wrap.

7.3. LSP Object

The LSP object MUST be present within PCRpt and PCUpd messages. The LSP object MAY be carried within PCReq and PCRep messages if the stateful PCE capability has been negotiated on the session. The LSP object contains a set of fields used to specify the target LSP, the operation to be performed on the LSP, and LSP Delegation. It also contains a flag indicating to a PCE that the LSP state synchronization is in progress. This document focuses on LSPs that are signaled with RSVP, many of the TLVs used with the LSP object mirror RSVP state.

LSP Object-Class is 32.

LSP Object-Type is 1.

The format of the LSP object body is shown in Figure 11:

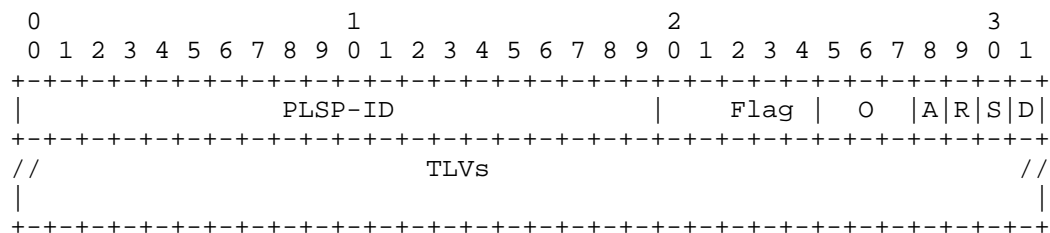


Figure 11: The LSP Object format

PLSP-ID (20 bits): A PCEP-specific identifier for the LSP. A PCC creates a unique PLSP-ID for each LSP that is constant for the lifetime of a PCEP session. The PCC will advertise the same PLSP-ID on all PCEP sessions it maintains at a given times. The mapping of the Symbolic Path Name to PLSP-ID is communicated to the PCE by sending a PCRpt message containing the SYMBOLIC-PATH-NAME TLV. All subsequent PCEP messages then address the LSP by the PLSP-ID. The values of 0 and 0xFFFFF are reserved. Note that the PLSP-ID is a value that is constant for the lifetime of the PCEP session, during which time for an RSVP-signaled LSP there might be a different RSVP identifiers (LSP-id, tunnel-id) allocated to it.

Flags (12 bits), starting from the least significant bit:

D (Delegate - 1 bit): On a PCRpt message, the D Flag set to 1 indicates that the PCC is delegating the LSP to the PCE. On a

PCUpd message, the D flag set to 1 indicates that the PCE is confirming the LSP Delegation. To keep an LSP delegated to the PCE, the PCC must set the D flag to 1 on each PCRpt message for the duration of the delegation - the first PCRpt with the D flag set to 0 revokes the delegation. To keep the delegation, the PCE must set the D flag to 1 on each PCUpd message for the duration of the delegation - the first PCUpd with the D flag set to 0 returns the delegation.

S (SYNC - 1 bit): The S Flag MUST be set to 1 on each PCRpt sent from a PCC during State Synchronization. The S Flag MUST be set to 0 in other messages sent from the PCC. When sending a PCUpd message, the PCE MUST set the S Flag to 0.

R(Remove - 1 bit): On PCRpt messages the R Flag indicates that the LSP has been removed from the PCC and the PCE SHOULD remove all state from its database. Upon receiving an LSP State Report with the R Flag set to 1 for an RSVP-signaled LSP, the PCE SHOULD remove all state for the path identified by the LSP-IDENTIFIERS TLV from its database. When the all-zeros LSP-IDENTIFIERS TLV is used, the PCE SHOULD remove all state for the PLSP-ID from its database. When sending a PCUpd message, the PCE MUST set the R Flag to 0.

A(Administrative - 1 bit): On PCRpt messages, the A Flag indicates the PCC's target operational status for this LSP. On PCUpd messages, the A Flag indicates the LSP status that the PCE desires for this LSP. In both cases, a value of '1' means that the desired operational state is active, and a value of '0' means that the desired operational state is inactive. A PCC ignores the A flag on a PCUpd message unless the operator's policy allows the PCE to control the corresponding LSP's administrative state.

O(Operational - 3 bits): On PCRpt messages, the O Field represents the operational status of the LSP.

The following values are defined:

0 - DOWN: not active.

1 - UP: signalled.

2 - ACTIVE: up and carrying traffic.

3 - GOING-DOWN: LSP is being torn down, resources are being released.

4 - GOING-UP: LSP is being signalled.

5-7 - Reserved: these values are reserved for future use.

Unassigned bits are considered reserved. They MUST be set to 0 on transmission and MUST be ignored on receipt. When sending a PCUpd message, the PCE MUST set the O Field to 0.

TLVs that may be included in the LSP Object are described in the following sections. Other optional TLVs, that are not defined in this document, MAY also be included within the LSP Object body.

7.3.1. LSP-IDENTIFIERS TLVs

The LSP-IDENTIFIERS TLV MUST be included in the LSP object in PCRpt messages for RSVP-signaled LSPs. If the TLV is missing, the PCE will generate an error with error-type 6 (mandatory object missing) and error-value 11 (LSP-IDENTIFIERS TLV missing) and close the session. The LSP-IDENTIFIERS TLV MAY be included in the LSP object in PCUpd messages for RSVP-signaled LSPs. The special value of all zeros for this TLV is used to refer to all paths pertaining to a particular PLSP-ID. There are two LSP-IDENTIFIERS TLVs, one for IPv4 and one for IPv6.

It is the responsibility of the PCC to send to the PCE the identifiers for each RSVP incarnation of the tunnel. For example, in a make-before-break scenario, the PCC MUST send a separate PCRpt for the old and for the reoptimized paths, and explicitly report removal of any of these paths using the R bit in the LSP object.

The format of the IPV4-LSP-IDENTIFIERS TLV is shown in the following figure:

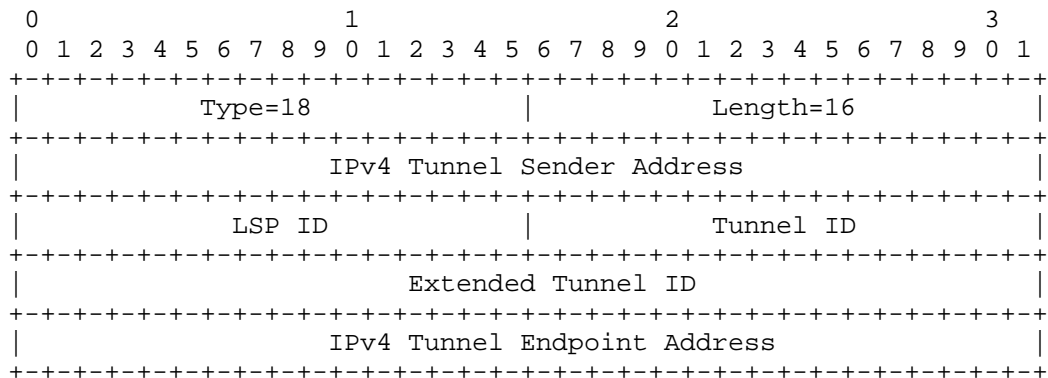


Figure 12: IPV4-LSP-IDENTIFIERS TLV format

The type (16 bits) of the TLV is 18. The length field is 16 bit-long and has a fixed value of 16. The value contains the following fields:

IPv4 Tunnel Sender Address: contains the sender node's IPv4 address, as defined in [RFC3209], Section 4.6.2.1 for the LSP_TUNNEL_IPv4 Sender Template Object.

LSP ID: contains the 16-bit 'LSP ID' identifier defined in [RFC3209], Section 4.6.2.1 for the LSP_TUNNEL_IPv4 Sender Template Object. A value of 0 MUST be used if the LSP is not yet signaled.

Tunnel ID: contains the 16-bit 'Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.1 for the LSP_TUNNEL_IPv4 Session Object.

Extended Tunnel ID: contains the 32-bit 'Extended Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.1 for the LSP_TUNNEL_IPv4 Session Object.

IPv4 Tunnel Endpoint Address: contains the egress node's IPv4 address, as defined in [RFC3209], Section 4.6.1.1 for the LSP_TUNNEL_IPv4 Sender Template Object.

The format of the IPV6-LSP-IDENTIFIERS TLV is shown in the following figure:

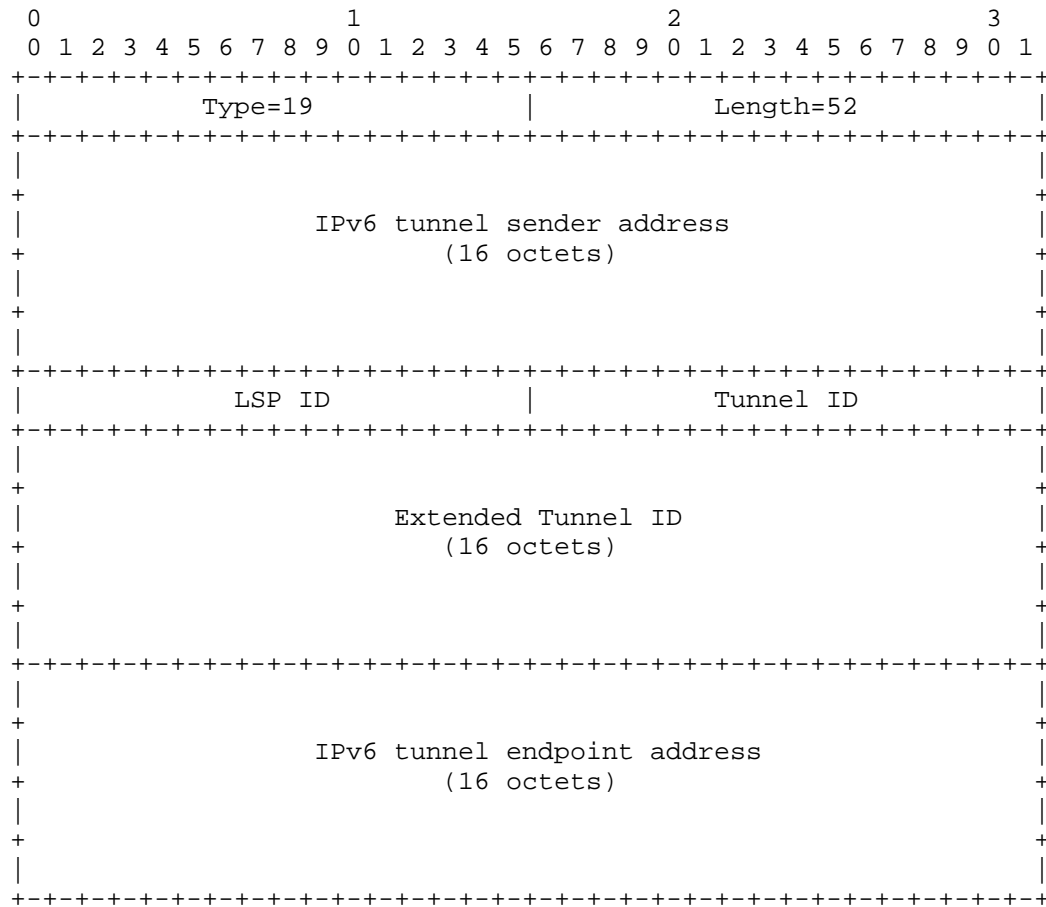


Figure 13: IPV6-LSP-IDENTIFIERS TLV format

The type (16 bits) of the TLV is 19. The length field is 16 bit-long and has a fixed value of 52. The value contains the following fields:

IPv6 Tunnel Sender Address: contains the sender node's IPv6 address, as defined in [RFC3209], Section 4.6.2.2 for the LSP_TUNNEL_IPv6 Sender Template Object.

LSP ID: contains the 16-bit 'LSP ID' identifier defined in [RFC3209], Section 4.6.2.2 for the LSP_TUNNEL_IPv6 Sender Template Object. A value of 0 MUST be used if the LSP is not yet signaled.

Tunnel ID: contains the 16-bit 'Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.2 for the LSP_TUNNEL_IPv6 Session Object.

Extended Tunnel ID: contains the 128-bit 'Extended Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.2 for the LSP_TUNNEL_IPv6 Session Object.

IPv6 Tunnel Endpoint Address: contains the egress node's IPv6 address, as defined in [RFC3209], Section 4.6.1.2 for the LSP_TUNNEL_IPv6 Session Object.

The Tunnel ID remains constant over the life time of a tunnel.

7.3.2. Symbolic Path Name TLV

Each LSP MUST have a symbolic path name that is unique in the PCC. The symbolic path name is a human-readable string that identifies an LSP in the network. The symbolic path name MUST remain constant throughout an LSP's lifetime, which may span across multiple consecutive PCEP sessions and/or PCC restarts. The symbolic path name MAY be specified by an operator in a PCC's configuration. If the operator does not specify a unique symbolic name for an LSP, then the PCC MUST auto-generate one.

The PCE uses the symbolic path name as a stable identifier for the LSP. If the PCEP session restarts, or the PCC restarts, or the PCC re-delegates the LSP to a different PCE, the symbolic path name for the LSP remains constant and can be used to correlate across the PCEP session instances.

The other protocol identifiers for the LSP cannot reliably be used to identify the LSP across multiple PCEP sessions, for the following reasons.

- o The PLSP-ID is unique only within the scope of a single PCEP session.
- o The LSP-IDENTIFIERS TLV is only guaranteed to be present for LSPs that are signalled with RSVP-TE, and may change during the lifetime of the LSP.

The SYMBOLIC-PATH-NAME TLV MUST be included in the LSP object in the LSP State Report (PCRpt) message when during a given PCEP session an LSP is first reported to a PCE. A PCC sends to a PCE the first LSP State Report either during State Synchronization, or when a new LSP is configured at the PCC.

The initial PCRpt creates a binding between the symbolic path name and the PLSP-ID for the LSP which lasts for the duration of the PCEP session. The PCC MAY omit the symbolic path name from subsequent LSP

State Reports for that LSP on that PCEP session, and just use the PLSP-ID.

The format of the SYMBOLIC-PATH-NAME TLV is shown in the following figure:

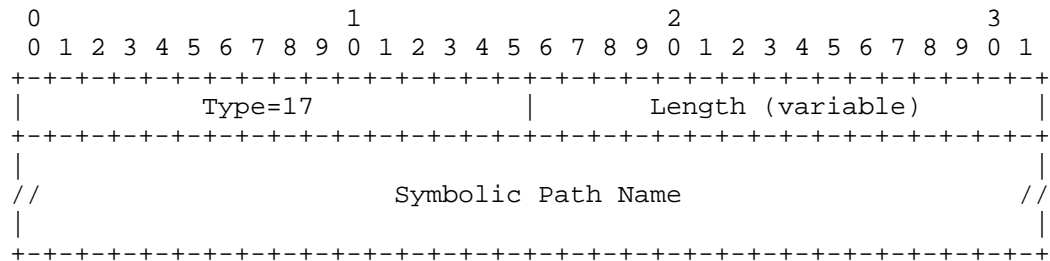


Figure 14: SYMBOLIC-PATH-NAME TLV format

```
Type (16 bits): The type is 17.
```

Length (16 bits): indicates the total length of the TLV in octets and MUST be greater than 0. The TLV MUST be zero-padded so that the TLV is 4-octet aligned.

Symbolic Path Name (variable): symbolic name for the LSP, unique in the PCC. It SHOULD be a string of printable ASCII characters, without a NULL terminator.

7.3.3. LSP Error Code TLV

The LSP Error code TLV is an optional TLV for use in the LSP object to convey error information. When an LSP Update Request fails, an LSP State Report **MUST** be sent to report the current state of the LSP, and **SHOULD** contain the LSP-ERROR-CODE TLV indicating the reason for the failure. Similarly, when a PCrpt is sent as a result of an LSP transitioning to non-operational state, the LSP-ERROR-CODE TLV **SHOULD** be included to indicate the reason for the transition.

The format of the LSP-ERROR-CODE TLV is shown in the following figure:

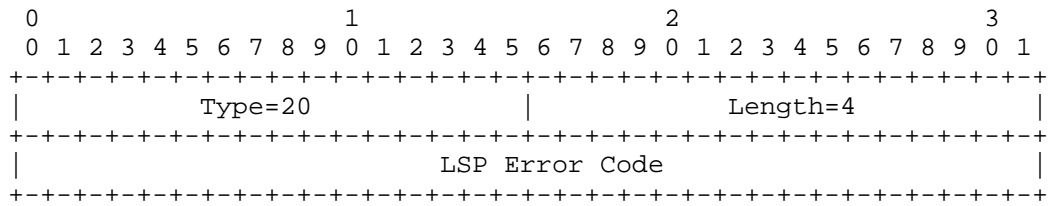


Figure 15: LSP-ERROR-CODE TLV format

The type (16 bits) of the TLV is 20. The length field is 16 bit-long and has a fixed value of 4. The value contains an error code that indicates the cause of the failure.

The following LSP Error Codes are currently defined:

Value	Meaning
1	Unknown reason
2	Limit reached for PCE-controlled LSPs
3	Too many pending LSP update requests
4	Unacceptable parameters
5	Internal error
6	LSP administratively brought down
7	LSP preempted
8	RSVP signaling error

7.3.4. RSVP Error Spec TLV

The RSVP-ERROR-SPEC TLV is an optional TLV for use in the LSP object to carry RSVP error information. It includes the RSVP ERROR_SPEC or USER_ERROR_SPEC Object ([RFC2205] and [RFC5284]) which were returned to the PCC from a downstream node. If the set up of an LSP fails at a downstream node which returned an ERROR_SPEC to the PCC, the PCC SHOULD include in the PCRpt for this LSP the LSP-ERROR-CODE TLV with LSP Error Code = "RSVP signaling error" and the RSVP-ERROR-SPEC TLV with the relevant RSVP ERROR_SPEC or USER_ERROR_SPEC Object.

The format of the RSVP-ERROR-SPEC TLV is shown in the following figure:

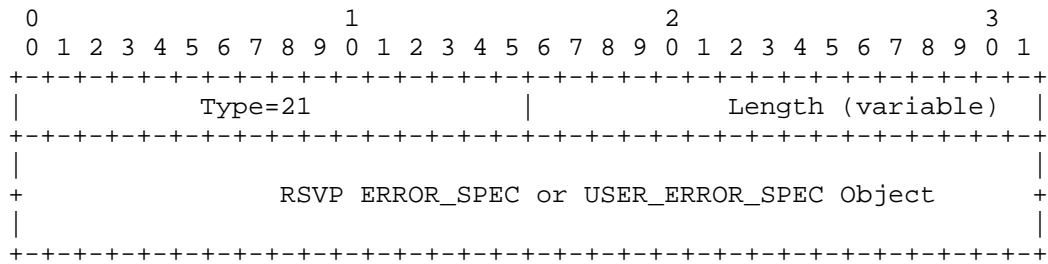


Figure 16: RSVP-ERROR-SPEC TLV format

Type (16 bits): The type is 21.

Length (16 bits): indicates the total length of the TLV in octets. The TLV MUST be zero-padded so that the TLV is 4-octet aligned.

Value (variable): contains the RSVP_ERROR_SPEC or USER_ERROR_SPEC Object: as specified in [RFC2205] and [RFC5284], including the object header.

8. IANA Considerations

This document requests IANA actions to allocate code points for the protocol elements defined in this document.

8.1. PCE Capabilities in IGP Advertisements

IANA is requested to confirm the early allocation of the following bits in the OSPF Parameters "PCE Capability Flags" registry, and to update the reference in the registry to point to this document, when it is an RFC:

Bit	Meaning	Reference
11	Active Stateful PCE capability	This document
12	Passive Stateful PCE capability	This document

8.2. PCEP Messages

IANA is requested to confirm the early allocation of the following message types within the "PCEP Messages" sub-registry of the PCEP Numbers registry, and to update the reference in the registry to point to this document, when it is an RFC:

Value	Meaning	Reference
10	Report	This document
11	Update	This document

8.3. PCEP Objects

IANA is requested to confirm the early allocation of the following object-class values and object types within the "PCEP Objects" sub-registry of the PCEP Numbers registry, and to update the reference in the registry to point to this document, when it is an RFC:

Object-Class Value	Name	Reference
32	LSP Object-Type 1	This document
33	SRP Object-Type 1	This document

8.4. LSP Object

This document requests that a new sub-registry, named "LSP Object Flag Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field of the LSP object. New values are to be assigned by Standards Action [RFC5226]. Each bit should be tracked with the following qualities:

- o Bit number (counting from bit 0 as the most significant bit)
- o Capability description
- o Defining RFC

The following values are defined in this document:

Bit	Description	Reference
0-4	Reserved	This document
5-7	Operational (3 bits)	This document
8	Administrative	This document
9	Remove	This document
10	SYNC	This document
11	Delegate	This document

8.5. PCEP-Error Object

IANA is requested to confirm the early allocation of the following Error Types and Error Values within the "PCEP-ERROR Object Error Types and Values" sub-registry of the PCEP Numbers registry, and to update the reference in the registry to point to this document, when it is an RFC:

Error-Type	Meaning
6	Mandatory Object missing
	Error-value=8: LSP Object missing
	Error-value=9: ERO Object missing
	Error-value=10: SRP Object missing
	Error-value=11: LSP-IDENTIFIERS TLV missing
19	Invalid Operation
	Error-value=1: Attempted LSP Update Request for a non-delegated LSP. The PCEP-ERROR Object is followed by the LSP Object that identifies the LSP.
	Error-value=2: Attempted LSP Update Request if the stateful PCE capability was not advertised.
	Error-value=3: Attempted LSP Update Request for an LSP identified by an unknown PLSP-ID.
	Error-value=5: Attempted LSP State Report if stateful PCE capability was not advertised.
20	LSP State synchronization error.
	Error-value=1: A PCE indicates to a PCC that it can not process (an otherwise valid) LSP State Report. The PCEP-ERROR Object is followed by the LSP Object that identifies the LSP.
	Error-value=5: A PCC indicates to a PCE that it can not complete the state synchronization,

8.6. Notification Object

IANA is requested to confirm the early allocation of the following Notification Types and Notification Values within the "Notification Object" sub-registry of the PCEP Numbers registry, and to update the reference in the registry to point to this document, when it is an RFC:

Notification-Type	Meaning
4	Stateful PCE resource limit exceeded

Notification-value=1:	Entering resource limit exceeded state
-----------------------	--

Note to IANA: the early allocation included an additional Notification value 2 for "Exiting resource limit exceeded state". This Notification value is no longer required.

8.7. PCEP TLV Type Indicators

IANA is requested to confirm the early allocation of the following TLV Type Indicator values within the "PCEP TLV Type Indicators" sub-registry of the PCEP Numbers registry, and to update the reference in the registry to point to this document, when it is an RFC:

Value	Meaning	Reference
16	STATEFUL-PCE-CAPABILITY	This document
17	SYMBOLIC-PATH-NAME	This document
18	IPV4-LSP-IDENTIFIERS	This document
19	IPV6-LSP-IDENTIFIERS	This document
20	LSP-ERROR-CODE	This document
21	RSVP-ERROR-SPEC	This document

8.8. STATEFUL-PCE-CAPABILITY TLV

This document requests that a new sub-registry, named "STATEFUL-PCE-CAPABILITY TLV Flag Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field in the STATEFUL-PCE-CAPABILITY TLV of the PCEP OPEN object (class = 1). New values are to be assigned by Standards Action [RFC5226]. Each bit should be tracked with the following qualities:

- o Bit number (counting from bit 0 as the most significant bit)
- o Capability description
- o Defining RFC

The following values are defined in this document:

Bit	Description	Reference
31	LSP-UPDATE-CAPABILITY	This document

8.9. LSP-ERROR-CODE TLV

This document requests that a new sub-registry, named "LSP-ERROR-CODE TLV Error Code Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the LSP Error code field of the LSP-ERROR-CODE TLV. This field specifies the reason for failure to update the LSP.

New values are to be assigned by Standards Action [RFC5226]. Each value should be tracked with the following qualities: value, description and defining RFC. The following values are defined in this document:

Value	Meaning
1	Unknown reason
2	Limit reached for PCE-controlled LSPs
3	Too many pending LSP update requests
4	Unacceptable parameters
5	Internal error
6	LSP administratively brought down
7	LSP preempted
8	RSVP signaling error

9. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440] apply to PCEP extensions defined in this document. In addition, requirements and considerations listed in this section apply.

9.1. Control Function and Policy

In addition to configuring specific PCEP session parameters, as specified in [RFC5440], Section 8.1, a PCE or PCC implementation MUST allow configuring the stateful PCEP capability and the LSP Update capability. A PCC implementation SHOULD allow the operator to specify multiple candidate PCEs for and a delegation preference for each candidate PCE. A PCC SHOULD allow the operator to specify an LSP delegation policy where LSPs are delegated to the most-preferred online PCE. A PCC MAY allow the operator to specify different LSP delegation policies.

A PCC implementation which allows concurrent connections to multiple PCEs SHOULD allow the operator to group the PCEs by administrative domains and it MUST NOT advertise LSP existence and state to a PCE if the LSP is delegated to a PCE in a different group.

A PCC implementation SHOULD allow the operator to specify whether the PCC will advertise LSP existence and state for LSPs that are not

controlled by any PCE (for example, LSPs that are statically configured at the PCC).

A PCC implementation SHOULD allow the operator to specify both the Redelelegation Timeout Interval and the State Timeout Interval. The default value of the Redelelegation Timeout Interval SHOULD be set to 30 seconds. An operator MAY also configure a policy that will dynamically adjust the Redelelegation Timeout Interval, for example setting it to zero when the PCC has an established session to a backup PCE. The default value for the State Timeout Interval SHOULD be set to 60 seconds.

After the expiration of the State Timeout Interval, the LSP reverts to operator-defined default parameters. A PCC implementation MUST allow the operator to specify the default LSP parameters. To achieve a behavior where the LSP retains the parameters set by the PCE until such time that the PCC makes a change to them, a State Timeout Interval of infinity SHOULD be used. Any changes to LSP parameters SHOULD be done in make-before-break fashion.

LSP Delegation is controlled by operator-defined policies on a PCC. LSPs are delegated individually - different LSPs may be delegated to different PCEs. An LSP is delegated to at most one PCE at any given point in time. A PCC implementation SHOULD support the delegation policy, when all PCC's LSPs are delegated to a single PCE at any given time. Conversely, the policy revoking the delegation for all PCC's LSPs SHOULD also be supported.

A PCC implementation SHOULD allow the operator to specify delegation priority for PCEs. This effectively defines the primary PCE and one or more backup PCEs to which primary PCE's LSPs can be delegated when the primary PCE fails.

Policies defined for stateful PCEs and PCCs should eventually fit in the Policy-Enabled Path Computation Framework defined in [RFC5394], and the framework should be extended to support Stateful PCEs.

9.2. Information and Data Models

The PCEP YANG module [I-D.ietf-pce-pcep-yang] should include

- o advertised stateful capabilities and synchronization status per PCEP session
- o the delegation status of each configured LSP.

The PCEP MIB [RFC7420] could also be updated to include this information.

9.3. Liveness Detection and Monitoring

PCEP extensions defined in this document do not require any new mechanisms beyond those already defined in [RFC5440], Section 8.3.

9.4. Verifying Correct Operation

Mechanisms defined in [RFC5440], Section 8.4 also apply to PCEP extensions defined in this document. In addition to monitoring parameters defined in [RFC5440], a stateful PCC-side PCEP implementation SHOULD provide the following parameters:

- o Total number of LSP updates
- o Number of successful LSP updates
- o Number of dropped LSP updates
- o Number of LSP updates where LSP setup failed

A PCC implementation SHOULD provide a command to show for each LSP whether it is delegated, and if so, to which PCE.

A PCC implementation SHOULD allow the operator to manually revoke LSP delegation.

9.5. Requirements on Other Protocols and Functional Components

PCEP extensions defined in this document do not put new requirements on other protocols.

9.6. Impact on Network Operation

Mechanisms defined in [RFC5440], Section 8.6 also apply to PCEP extensions defined in this document.

Additionally, a PCEP implementation SHOULD allow a limit to be placed on the number of LSPs delegated to the PCE and on the rate of PCUpd and PCRpt messages sent by a PCEP speaker and processed from a peer. It SHOULD also allow sending a notification when a rate threshold is reached.

A PCC implementation SHOULD allow a limit to be placed on the rate of LSP Updates to the same LSP to avoid signaling overload discussed in Section 10.3.

10. Security Considerations

10.1. Vulnerability

This document defines extensions to PCEP to enable stateful PCEs. The nature of these extensions and the delegation of path control to PCEs results in more information being available for a hypothetical adversary and a number of additional attack surfaces which must be protected.

The security provisions described in [RFC5440] remain applicable to these extensions. However, because the protocol modifications outlined in this document allow the PCE to control path computation timing and sequence, the PCE defense mechanisms described in [RFC5440] section 7.2 are also now applicable to PCC security.

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, using Transport Layer Security (TLS) [I-D.ietf-pce-pceps], as per the recommendations and best current practices in [RFC7525].

The following sections identify specific security concerns that may result from the PCEP extensions outlined in this document along with recommended mechanisms to protect PCEP infrastructure against related attacks.

10.2. LSP State Snooping

The stateful nature of this extension explicitly requires LSP status updates to be sent from PCC to PCE. While this gives the PCE the ability to provide more optimal computations to the PCC, it also provides an adversary with the opportunity to eavesdrop on decisions made by network systems external to PCE. This is especially true if the PCC delegates LSPs to multiple PCEs simultaneously.

Adversaries may gain access to this information by eavesdropping on unsecured PCEP sessions, and might then use this information in various ways to target or optimize attacks on network infrastructure. For example by flexibly countering anti-DDoS measures being taken to protect the network, or by determining choke points in the network where the greatest harm might be caused.

PCC implementations which allow concurrent connections to multiple PCEs SHOULD allow the operator to group the PCEs by administrative domains and they MUST NOT advertise LSP existence and state to a PCE if the LSP is delegated to a PCE in a different group.

10.3. Malicious PCE

The LSP delegation mechanism described in this document allows a PCC to grant effective control of an LSP to the PCE for the duration of a PCEP session. While this enables PCE control of the timing and sequence of path computations within and across PCEP sessions, it also introduces a new attack vector: an attacker may flood the PCC with PCUpd messages at a rate which exceeds either the PCC's ability to process them or the network's ability to signal the changes, either by spoofing messages or by compromising the PCE itself.

A PCC is free to revoke an LSP delegation at any time without needing any justification. A defending PCC can do this by enqueueing the appropriate PCRpt message. As soon as that message is enqueued in the session, the PCC is free to drop any incoming PCUpd messages without additional processing.

10.4. Malicious PCC

A stateful session also results in an increased attack surface by placing a requirement for the PCE to keep an LSP state replica for each PCC. It is RECOMMENDED that PCE implementations provide a limit on resources a single PCC can occupy. A PCE implementing such a limit MUST send a PCNtf message with notification-type 4 (Stateful PCE resource limit exceeded) and notification-value 1 (Entering resource limit exceeded state) upon receiving an LSP state report causing it to exceed this threshold.

Delegation of LSPs can create further strain on PCE resources and a PCE implementation MAY preemptively give back delegations if it finds itself lacking the resources needed to effectively manage the delegation. Since the delegation state is ultimately controlled by the PCC, PCE implementations SHOULD provide throttling mechanisms to prevent strain created by flaps of either a PCEP session or an LSP delegation.

11. Contributing Authors

Xian Zhang
Huawei Technology
F3-5-B R&D Center
Huawei Industrial Base, Bantian, Longgang District
Shenzhen, Guangdong 518129
P.R.China
EMail: zhang.xian@huawei.com

Dhruv Dhody
Huawei Technology

Leela Palace
Bangalore, Karnataka 560008
INDIA
EMail: dhruv.dhody@huawei.com

Siva Sivabalan
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario K2K 3E8
Canada
EMail: msiva@cisco.com

12. Acknowledgements

We would like to thank Adrian Farrel, Cyril Margaria and Ramon Casellas for their contributions to this document.

We would like to thank Shane Amante, Julien Meuric, Kohei Shiimoto, Paul Schultz and Raveendra Torvi for their comments and suggestions. Thanks also to Jon Hardwick, Oscar Gonzales de Dios, Tomas Janciga, Stefan Kobza, Kexin Tang, Matej Spanik, Jon Parker, Marek Zavodsky, Ambrose Kwong, Ashwin Sampath, Calvin Ying, Mustapha Aissaoui, Stephane Litkowski and Olivier Dugeon for helpful comments and discussions.

13. References

13.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, DOI 10.17487/RFC2205, September 1997, <<http://www.rfc-editor.org/info/rfc2205>>.
- [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, <<http://www.rfc-editor.org/info/rfc3209>>.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, DOI 10.17487/RFC5088, January 2008, <<http://www.rfc-editor.org/info/rfc5088>>.

- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, DOI 10.17487/RFC5089, January 2008, <<http://www.rfc-editor.org/info/rfc5089>>.
- [RFC5284] Swallow, G. and A. Farrel, "User-Defined Errors for RSVP", RFC 5284, DOI 10.17487/RFC5284, August 2008, <<http://www.rfc-editor.org/info/rfc5284>>.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<http://www.rfc-editor.org/info/rfc5440>>.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, DOI 10.17487/RFC5511, April 2009, <<http://www.rfc-editor.org/info/rfc5511>>.
- [RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a Stateful Path Computation Element (PCE)", RFC 8051, DOI 10.17487/RFC8051, January 2017, <<http://www.rfc-editor.org/info/rfc8051>>.

13.2. Informative References

- [I-D.ietf-pce-gmpls-pcep-extensions]
Margarita, C., Dios, O., and F. Zhang, "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions-11 (work in progress), October 2015.
- [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-09 (work in progress), March 2017.
- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and j. jeffrant@gmail.com, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-02 (work in progress), March 2017.
- [I-D.ietf-pce-pceps]
Lopez, D., Dios, O., Wu, Q., and D. Dhody, "Secure Transport for PCEP", draft-ietf-pce-pceps-14 (work in progress), May 2017.

- [I-D.ietf-pce-stateful-sync-optimizations]
Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X.,
and D. Dhody, "Optimizations of Label Switched Path State
Synchronization Procedures for a Stateful PCE", draft-
ietf-pce-stateful-sync-optimizations-10 (work in
progress), March 2017.
- [MPLS-PC] Chaieb, I., Le Roux, J.L., and B. Cousin, "Improved MPLS-TE
LSP Path Computation using Preemption", Global
Information Infrastructure Symposium, July 2007.
- [MXMN-TE] Danna, E., Mandal, S., and A. Singh, "Practical linear
programming algorithm for balancing the max-min fairness
and throughput objectives in traffic engineering",
INFOCOM, 2012 Proceedings IEEE Page(s): 846-854, 2012.
- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J.
McManus, "Requirements for Traffic Engineering Over MPLS",
RFC 2702, DOI 10.17487/RFC2702, September 1999,
<<http://www.rfc-editor.org/info/rfc2702>>.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
Label Switching Architecture", RFC 3031,
DOI 10.17487/RFC3031, January 2001,
<<http://www.rfc-editor.org/info/rfc3031>>.
- [RFC3346] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D.,
Christian, B., and W. Lai, "Applicability Statement for
Traffic Engineering with MPLS", RFC 3346,
DOI 10.17487/RFC3346, August 2002,
<<http://www.rfc-editor.org/info/rfc3346>>.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering
(TE) Extensions to OSPF Version 2", RFC 3630,
DOI 10.17487/RFC3630, September 2003,
<<http://www.rfc-editor.org/info/rfc3630>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
Element (PCE)-Based Architecture", RFC 4655,
DOI 10.17487/RFC4655, August 2006,
<<http://www.rfc-editor.org/info/rfc4655>>.
- [RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation
Element (PCE) Communication Protocol Generic
Requirements", RFC 4657, DOI 10.17487/RFC4657, September
2006, <<http://www.rfc-editor.org/info/rfc4657>>.

- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, DOI 10.17487/RFC5226, May 2008, <<http://www.rfc-editor.org/info/rfc5226>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <<http://www.rfc-editor.org/info/rfc5305>>.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, DOI 10.17487/RFC5394, December 2008, <<http://www.rfc-editor.org/info/rfc5394>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<http://www.rfc-editor.org/info/rfc7420>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<http://www.rfc-editor.org/info/rfc7525>>.

Authors' Addresses

Edward Crabbe
Oracle
1501 4th Ave, suite 1800
Seattle, WA 98101
US

Email: edward.crabbe@oracle.com

Ina Minei
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: inaminei@google.com

Jan Medved
Cisco Systems, Inc.
170 West Tasman Dr.
San Jose, CA 95134
US

Email: jmedved@cisco.com

Robert Varga
Pantheon Technologies SRO
Mlynske Nivy 56
Bratislava 821 05
Slovakia

Email: robert.varga@pantheon.tech

Network Working Group
Internet Draft
Intended Status: Standards Track
Expires: October 29, 2012

K. Kumaki, Ed.
KDDI Corporation
T. Murai
Furukawa Network Solutions Corp.
D. Dhody
Huawei Technology
P. Jiang
KDDI Corporation
April 29, 2012

PCEP extensions for a BGP/MPLS IP-VPN
draft-kumaki-murai-pce-pcep-extension-l3vpn-09.txt

Abstract

IP Virtual Private Networks (IP-VPNs) allow Service Providers to offer customers connectivity between sites across an IP Backbone. These VPNs can be supported using BGP/MPLS and the connections can be created by using MPLS Traffic Engineered (TE) Label Switched Paths (LSPs). The paths of these LSPs must be computed to provide the connectivity between customer sites. Path selection may be dependent on a variety of factors including traffic engineering constraints and bandwidth requirements.

It is highly desirable for VPN customers to be able to dynamically establish their MPLS TE LSPs for interconnectivity between BGP/MPLS IP-VPN sites. The Path Computation Element (PCE) can determine the optimal paths of TE LSPs within an MPLS network. This document defines the PCEP extensions for the dynamic creation of MPLS TE LSPs between BGP/MPLS IP-VPN sites.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

This Internet-Draft will expire on October 29, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 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.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction.....	3
2. Problem Statement.....	4
3. Protocol Extensions and Procedures.....	5
3.1 Type Definition.....	5
3.2 PCE Capabilities.....	6
3.2.1 PCReq message Processing at Ingress PE (PCE).....	6
3.2.2 PCReq message Processing at Egress PE (PCE).....	7
3.2.3 PCRep message Processing at Egress PE (PCE).....	7
3.2.4 PCRep message Processing at Ingress PE (PCE).....	7
4. Security Considerations.....	8
5. IANA Considerations.....	8
6. References.....	8
6.1 Normative References.....	8
6.2 Informative References.....	8
7. Acknowledgments.....	9
8. Authors' Addresses.....	9

1. Introduction

[RFC4364] describes how a Service Provider could use an IP backbone to provide IP Virtual Private Networks (VPNs) to its customers. Each VPN contains at least one Customer Edge (CE) router which is attached to a Provider (PE) router. It is possible for CE routers to be attached to multiple PE routers. The Border Gateway Protocol (BGP) [RFC4271] can be used to exchange VPN route information between the PE routers.

[RFC4655] describes the motivation and architecture for the Path Computation Element (PCE). The PCE can be used to compute the routes for MPLS and GMPLS Traffic Engineering (TE) Label Switched Paths (LSPs). A path computation request is issued by a Path Computation Client (PCC) to the PCE. The communication protocol between PCCs and PCEs is called the Path Computation Element Communication Protocol (PCEP) and is defined in [RFC5440].

This document examines why it would be advantageous to use the PCE architecture to establish connectivity between IP/MPLS VPNs and defines extensions to PCEP to support that function.

The requirements for establishing connectivity between CE and PE sites are defined in [RFC5824]. The requirements placed on PCEP for the use of a PCE in a variety of VPN environments are set out in [PCE-VPN-REQ]. This work only examines a small subset of the requirements from [PCE-VPN-REQ] because it limits the use of PCEs to specific objectives in BGP/MPLS IP-VPNs.

In order to establish a customer MPLS TE LSP over a BGP/MPLS IP-VPN, a PCE needs to know the VPNv4/VPNv6 tail-end addresses (source and destination) and the addresses for the PEs that provide access to them. Additionally the PCE needs to calculate the route of an end-to-end customer MPLS TE LSP. [RFC5441] describes the Backward Recursive Path Computation (BRPC) technique that could be used to compute the route of a customer MPLS TE LSP.

In order to discover PCEs participating in a BGP/MPLS IP-VPNs, [PCE-BGP-VPN] is proposed, but this information could be configured or discovered through another means. Note that it is assumed that PCE functions are normally included in PE routers they could also be placed in other nodes that are fully accessible to all PCCs in the VPN.

This document defines new object types in the PCEP END-POINTS object to calculate the end-to-end customer MPLS TE LSP used for BGP/IP-VPNs, and describes a procedure for the PCEP message processing. The new object types are defined in Section 3.1 and the specific procedure is described in Section 3.2.

1. VPN1 and VPN2 are completely different customers.
2. C0 and C4 are head-end routers.
3. C3 and C7 are tail-end routers.
4. The same address (e.g., 192.0.2.1) is assigned to both C3 and C7.

PCE1 (PE1) is able to distinguish to which VPN the received requests apply from the interface over which the requests were received. PCE1 can forward the request to PCE2 having been configured with or discovered ([PCE-BGP-VPN]) the existence and address of PCE2. However, based on the PCEP specification defined in [RFC5440] and the fact that the two messages come from the same cooperating PCE (PCE1), PCE2 cannot determine to which VPN the computation requests apply. Therefore, PCE2 cannot calculate the requested paths.

In order to distinguish between the VPN1 PCReq messages and the VPN2 PCReq messages, a VPN identifier is required in PCReq messages. This identifier can be duplicated into the PCRep messages to achieve symmetry and allow cross-checking.

This document defines a new object type for the END-POINTS object to be used to facilitate VPN identification.

3. Protocol Extensions and Procedures

The new END-POINTS Object-Types for the PCEP request allow the PCE to distinguish the VRF instance that is associated with the incoming PCEP message.

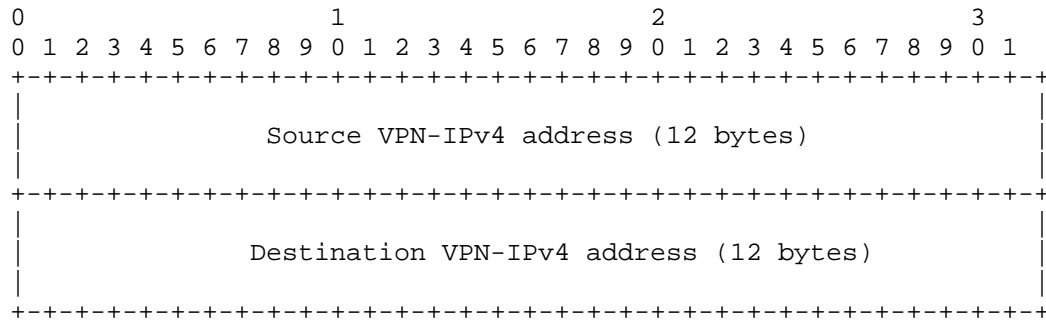
3.1 Type Definition

The END-POINTS Object is defined in [RFC5440].

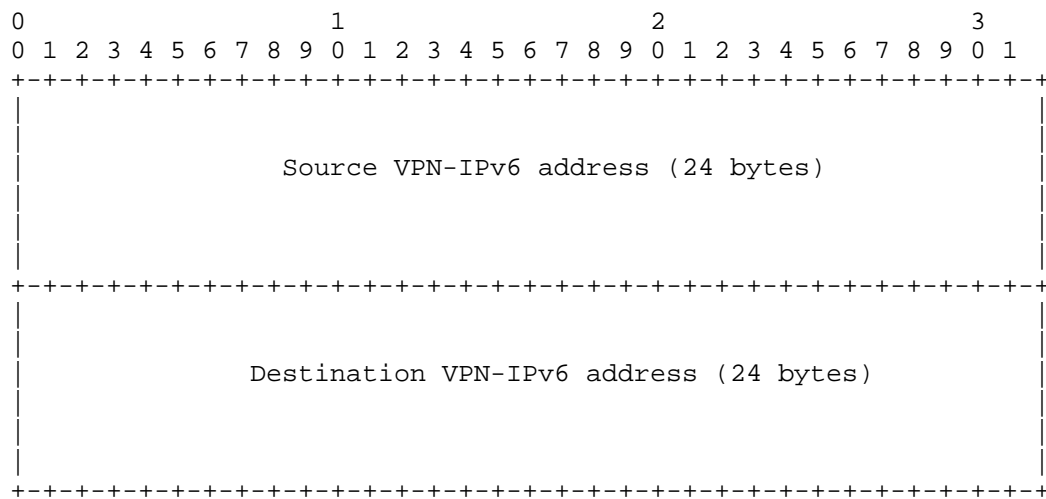
Two new Object-Types are defined to carry VPN-IPv4 addresses and VPN-IPv6 addresses.

END-POINTS Object-Type is to be assigned by IANA (recommended values: 3 for VPN-IPv4 and 4 for VPN-IPv6)

The format of the END-POINTS object body for VPN-IPv4 is as follows:



The format of the END-POINTS object body for VPN-IPv6 (Object-Type=4) is as follows:



3.2 PCE Capabilities

It is assumed that the BGP/MPLS IP-VPN ingress and egress PE routers have PCE capabilities. External PCE architectures will require further study and will be discussed in future revisions of this document.

3.2.1 PCReq Message Processing at Ingress PE (PCE)

When an ingress PE (PCE) receives a PCReq message from a PCC/PCE, it can distinguish the VRF instance that is associated with an incoming interface:

1. The ingress PE processes the destination IPv4/IPv6 address in the END-POINTS object as the destination VPN-IPv4/VPN-IPv6 address for the VRF instance.

2. The destination VPN-IPv4/VPN-IPv6 address is looked up in the context of VRF instance, and the BGP next-hop for this destination is identified.
3. The destination VPN-IPv4/VPN-IPv6 address is then added to END-POINTS object consisting of the original destination IPv4/IPv6 address in END-POINTS object followed by the 8 octet Route Distinguisher (RD).

Note that the RD is specified by the BGP next-hop for the destination VPN-IPv4/VPN-IPv6 address. The source VPN-IPv4/VPN-IPv6 address in the new END-POINTS object consists of the original IPv4/IPv6 address in END-POINTS object and the RD. Also the RD is used by this ingress PE to advertise customer's prefix including the source VPN-IPv4/VPN-IPv6 address into the VRF instance.

4. If necessary, the ingress PE will then send the PCReq message to next PCE (the egress PE for BGP/MPLS IP-VPNs).
5. Finally, the ingress PE should replace the incoming END-POINTS object from the PCC/PCE into the new END-POINTS object.

3.2.2 PCReq Message Processing at Egress PE (PCE)

When an egress PE (PCE) receives a PCReq message from an ingress PE(PCE), it is able to distinguish the VRF instance from the destination VPN-IPv4/VPN-IPv6 address in the new END-POINTS object. The egress PE will send a PCReq message to next PCE (PE) if needed.

The egress PE will then remove the RD from the source and the destination VPN-IPv4/VPN-IPv6 addresses in the new END-POINTS object received from the ingress PE. Finally, the egress PE should store the new END-POINTS object for a PCReq message in a VRF instance.

3.2.3 PCRep message Processing at Egress PE (PCE)

When an egress PE (PCE) receives a PCRep message for a PCReq message from a previous PCE (i.e. CE), it will look up the new END-POINTS object associated with the PCReq message for the PCRep message. The egress PE performs a path computation. Note that the path computation procedure itself is out of scope in this document. Afterwards, the egress PE adds the new END-POINTS object in a PCRep message and sends it to an ingress PE.

3.2.4 PCRep Message Processing at Ingress PE (PCE)

When an ingress PE (PCE) receives a PCRep message for a PCReq message from an egress PE (PCE), it can distinguish a VRF instance from the source VPN-IPv4/VPN-IPv6 address in the new END-POINTS object.

Therefore, it is now possible to generate a PCRep message to send to an appropriate PCC/PCE.

4. Security Considerations

This document defines PCEP extensions for BGP/MPLS IP-VPNs. The security of the PCE extensions relies on the security of PCEP [RFC5440]. It is important that implementations conform to security features defined in [RFC5440].

5. IANA Considerations

IANA maintains a registry of PCEP parameters. As described in Section 3.1 (Type Definition), two Object-Types have been defined. IANA is requested to make the following allocations from the "PCEP Objects" sub-registry.

Object-Class Value	Name	Object-Type	Reference
4	END-POINTS	1: IPv4 addresses	[RFC5440]
		2: IPv6 addresses	[RFC5440]
		3: VPN-IPv4 addresses	[This.I-D]
		4: VPN-IPv6 addresses	[This.I-D]
		5-15: Unassigned	

The values 3 and 4 are suggested.

6. References

6.1 Normative References

- [RFC4271] Rekhter, Y. and Li, T., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.
- [RFC5440] Vasseur, J.-P., et al., "Path Computation Element(PCE) communication Protocol (PCEP) - Version 1", RFC5440, March 2009.

6.2 Informative References

- [RFC4364] Rosen, E., Rekhter, Y., "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.
- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "Path Computation Element (PCE) Architecture", RFC 4655, August 2006.
- [RFC5824] Kumaki, K., Zhang, R. and Kamite, Y., "Requirements for supporting Customer RSVP and RSVP-TE over a BGP/MPLS IP-VPN", RFC 5824, April 2010.
- [PCE-VPN-REQ] Yasukawa, S. and Farrel, A. "PCC-PCE Communication Requirements for VPNs", draft-ietf-pce-vpn-req, March 2011.
- [RFC5441] Vasseur, J.-P., et al., "A Backward Recursive PCE-based Computation (BRPC) Procedure To Compute Shortest Constrained Inter-domain Traffic Engineering Label Switched Paths", RFC5441, April 2009.
- [PCE-BGP-VPN] Kumaki, K. and Murai, "BGP protocol extensions for Path Computation Element (PCE) Discovery in a BGP/MPLS IP-VPN ", draft-kumaki-pce-bgp-disco-attribute, September 2010.

7. Acknowledgments

The author would like to express thanks to Makoto Nakamura for his helpful and useful comments and feedback.

8. Authors' Addresses

Kenji Kumaki
KDDI Corporation
Garden Air Tower
Iidabashi, Chiyoda-ku,
Tokyo 102-8460, JAPAN
Email: ke-kumaki@kddi.com

Tomoki Murai
FURUKAWA NETWORK SOLUTION CORP.
5-1-9, HIGASHI-YAWATA, HIRATSUKA
Kanagawa 254-0016, JAPAN
Email: murai@fnsc.co.jp

Dhruv Dhody
Huawei Technology
Leela Palace
Bangalore, Karnataka, 560008, INDIA
Email: dhruv.dhody@huawei.com

Peng Jiang
KDDI Corporation
Garden Air Tower
Iidabashi, Chiyoda-ku,
Tokyo 102-8460, JAPAN
Email: pe-jiang@kddi.com

Tomohiro Yamagata
KDDI Corporation
Garden Air Tower
Iidabashi, Chiyoda-ku,
Tokyo 102-8460, JAPAN
Email: to-yamagata@kddi.com

Chikara Sasaki
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Fujimino
Saitama 356-8502, JAPAN
Email: ch-sasaki@kddilabs.jp

Network Working Group
Internet Draft
Intended Status: Experimental
Expires: March 28, 2016

K. Kumaki, Ed.
KDDI Corporation
T. Murai
Furukawa Network Solutions Corp.
T. Miyasaka
KDDI Corporation
September 25, 2015

PCEP extensions for a BGP/MPLS IP-VPN
draft-kumaki-murai-pce-pcep-extension-l3vpn-17.txt

Abstract

IP Virtual Private Networks (IP-VPNs) allow Service Providers to offer customers connectivity between sites across an IP Backbone. These VPNs can be supported using BGP/MPLS and the connections can be created by using MPLS Traffic Engineered (TE) Label Switched Paths (LSPs). The paths of these LSPs must be computed to provide the connectivity between customer sites. Path selection may be dependent on a variety of factors including traffic engineering constraints and bandwidth requirements.

It is highly desirable for VPN customers to be able to dynamically establish their MPLS TE LSPs for interconnectivity between BGP/MPLS IP-VPN sites. The Path Computation Element (PCE) can determine the optimal paths of TE LSPs within an MPLS network. This document defines how to extend the PCEP to support the dynamic creation of MPLS TE LSPs between BGP/MPLS IP-VPN sites.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

This Internet-Draft will expire on March 28, 2015.

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 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.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction.....	3
2. Problem Statement.....	4
3. Protocol Extensions and Procedures.....	5
3.1 Type Definition.....	5
3.2 PCE Capabilities.....	6
3.2.1 PCReq message Processing at Ingress PE (PCE).....	6
3.2.2 PCReq message Processing at Egress PE (PCE).....	7
3.2.3 PCRep message Processing at Egress PE (PCE).....	7
3.2.4 PCRep message Processing at Ingress PE (PCE).....	7
4. Security Considerations.....	8
5. IANA Considerations.....	8
6. References.....	8
6.1 Normative References.....	8
6.2 Informative References.....	8
7. Acknowledgments.....	9
8. Authors' Addresses.....	9

1. Introduction

[RFC4364] describes how a Service Provider could use an IP backbone to provide IP Virtual Private Networks (VPNs) to its customers. Each VPN contains at least one Customer Edge (CE) router which is attached to a Provider (PE) router. It is possible for CE routers to be attached to multiple PE routers. The Border Gateway Protocol (BGP) [RFC4271] can be used to exchange VPN route information between the PE routers.

[RFC4655] describes the motivation and architecture for the Path Computation Element (PCE). The PCE can be used to compute the routes for MPLS and GMPLS Traffic Engineering (TE) Label Switched Paths (LSPs). A path computation request is issued by a Path Computation Client (PCC) to the PCE. The communication protocol between PCCs and PCEs is called the Path Computation Element Communication Protocol (PCEP) and is defined in [RFC5440].

This document describes experimental mechanisms that use the PCE architecture to establish connectivity between IP/MPLS VPNs. And defines extensions to PCEP to support this experimental mechanisms.

The requirements for establishing connectivity between CE and PE sites are defined in [RFC5824]. The extensions to support RSVP-TE between customer sites when a single PE supports multiple VPNs are experimented in [RFC6882].

In order to establish a customer MPLS TE LSP over a BGP/MPLS IP-VPN, a PCE needs to know the VPNv4/VPNv6 tail-end addresses (source and destination) and the addresses for the PEs that provide access to them. Additionally the PCE needs to calculate the route of an end-to-end customer MPLS TE LSP. [RFC5441] describes the Backward Recursive Path Computation (BRPC) technique that could be used to compute the route of a customer MPLS TE LSP.

It is assumed in this experiment that PCE functions are included in PE routers that are fully accessible to all PCCs in the VPN.

This experiment defines new object types in the PCEP END-POINTS object to calculate the end-to-end customer MPLS TE LSP used for BGP/IP-VPNs, and verifies a procedure for the PCEP message processing. The new object types are defined in Section 3.1 and the specific procedure is described in Section 3.2.

2. Problem Statement

PCEs in the context of BGP/MPLS IP-VPNs are shown in figure 1. Here, we make the following set of assumptions.

1. VPN1 and VPN2 are completely different customers.
2. C0 and C4 are head-end routers.
3. C3 and C7 are tail-end routers.
4. The same address (e.g., 192.0.2.1) is assigned to both C3 and C7.

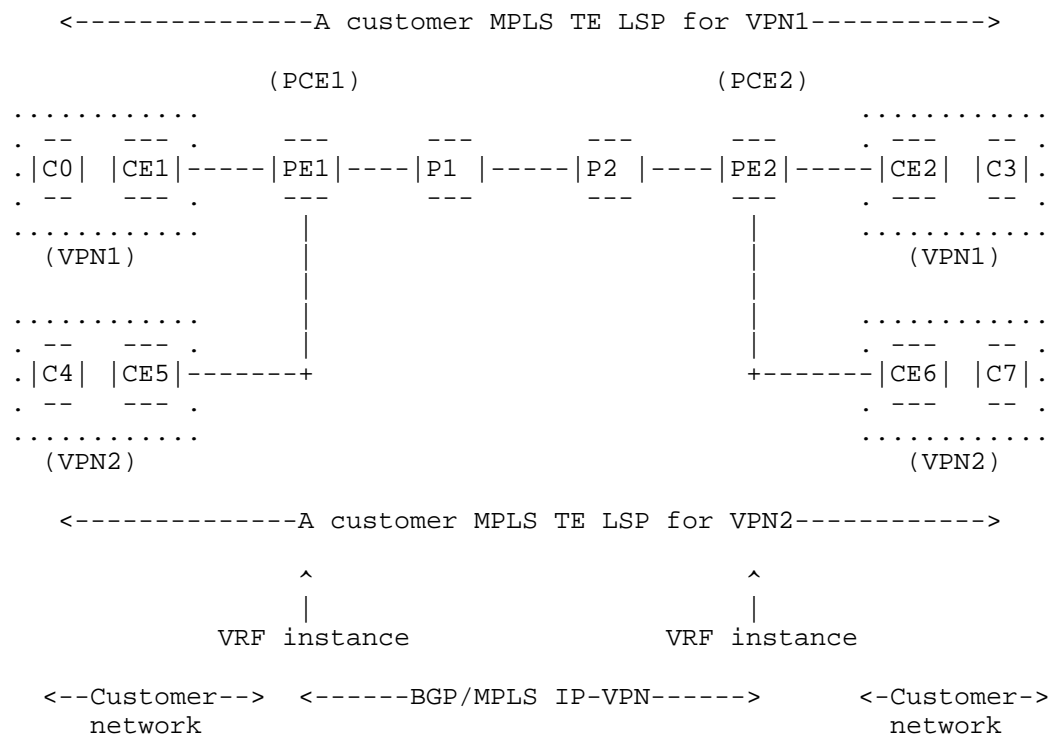


Figure 1 PCEs in the context of BGP/MPLS IP-VPNs

Consider that customers in VPN1 and VPN2 would like to establish customer MPLS TE LSPs between their sites (i.e., between CE0 and CE3, and between CE4 and CE7). The following PCEP requests will occur:

1. C0 send a PCReq message to PCE1 to compute a path for a customer MPLS TE LSPs between C0 and C3.
2. C4 send a PCReq message to PCE1 to compute a path for a customer MPLS TE LSPs between C4 and C7.

PCE1 (PE1) is able to distinguish to which VPN the received requests apply from the interface over which the requests were received. PCE1 can forward the request to PCE2 having been configured with or discovered the existence and address of PCE2. However, based on the PCEP specification defined in [RFC5440] and the fact that the two messages come from the same cooperating PCE (PCE1), PCE2 cannot determine to which VPN the computation requests apply. Therefore, PCE2 cannot calculate the requested paths.

In order to distinguish between the VPN1 PCReq messages and the VPN2 PCReq messages, a VPN identifier is required in PCReq messages. This identifier can be duplicated into the PCRep messages to achieve symmetry and allow cross-checking.

This experiment defines a new object type for the END-POINTS object to be used to facilitate VPN identification.

3. Protocol Extensions and Procedures

The new END-POINTS Object-Types for the PCEP request allow the PCE to distinguish the VRF instance that is associated with the incoming PCEP message.

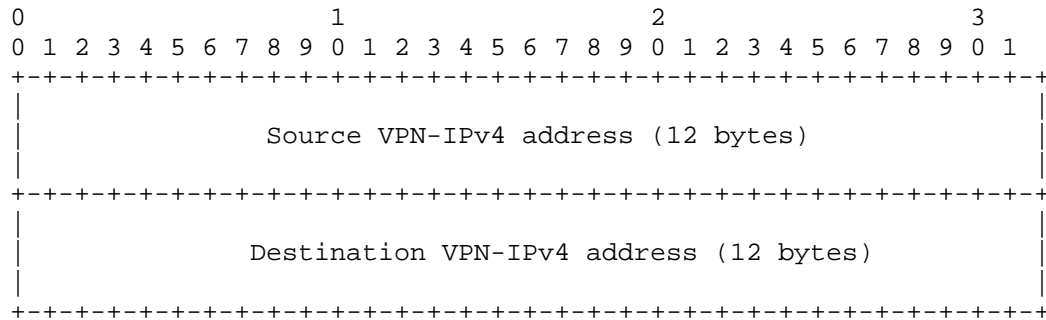
3.1 Type Definition

The END-POINTS Object is defined in [RFC5440].

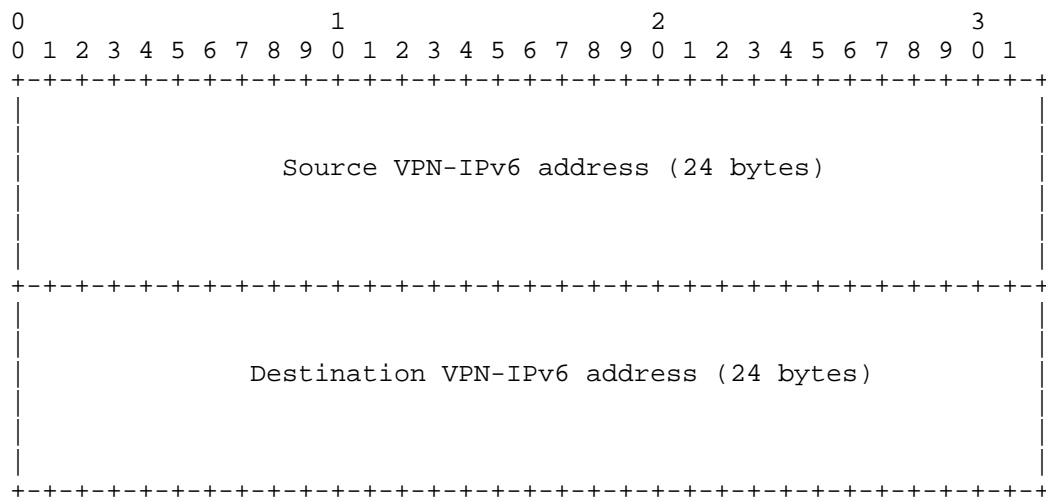
Two new Object-Types are defined to carry VPN-IPv4 addresses and VPN-IPv6 addresses.

END-POINTS Object-Type is to be assigned by IANA (recommended values: 3 for VPN-IPv4 and 4 for VPN-IPv6)

The format of the END-POINTS object body for VPN-IPv4 is as follows:



The format of the END-POINTS object body for VPN-IPv6 (Object-Type=4) is as follows:



3.2 PCE Capabilities

It is assumed that the BGP/MPLS IP-VPN ingress and egress PE routers have PCE capabilities. External PCE architectures will require further study and will be discussed in future revisions of this document.

3.2.1 PCReq Message Processing at Ingress PE (PCE)

When an ingress PE (PCE) receives a PCReq message from a PCC/PCE, it can distinguish the VRF instance that is associated with an incoming interface:

1. The ingress PE processes the destination IPv4/IPv6 address in the END-POINTS object as the destination VPN-IPv4/VPN-IPv6 address for the VRF instance.

2. The destination VPN-IPv4/VPN-IPv6 address is looked up in the context of VRF instance, and the BGP next-hop for this destination is identified.
3. The destination VPN-IPv4/VPN-IPv6 address is then added to END-POINTS object consisting of the original destination IPv4/IPv6 address in END-POINTS object followed by the 8 octet Route Distinguisher (RD).

Note that the RD is specified by the BGP next-hop for the destination VPN-IPv4/VPN-IPv6 address. The source VPN-IPv4/VPN-IPv6 address in the new END-POINTS object consists of the original IPv4/IPv6 address in END-POINTS object and the RD. Also the RD is used by this ingress PE to advertise customer's prefix including the source VPN-IPv4/VPN-IPv6 address into the VRF instance.

4. If necessary, the ingress PE will then send the PCReq message to next PCE (the egress PE for BGP/MPLS IP-VPNs).
5. Finally, the ingress PE should replace the incoming END-POINTS object from the PCC/PCE into the new END-POINTS object.

3.2.2 PCReq Message Processing at Egress PE (PCE)

When an egress PE (PCE) receives a PCReq message from an ingress PE(PCE), it is able to distinguish the VRF instance from the destination VPN-IPv4/VPN-IPv6 address in the new END-POINTS object. The egress PE will send a PCReq message to next PCE (PE) if needed.

The egress PE will then remove the RD from the source and the destination VPN-IPv4/VPN-IPv6 addresses in the new END-POINTS object received from the ingress PE. Finally, the egress PE should store the new END-POINTS object for a PCReq message in a VRF instance.

3.2.3 PCRep message Processing at Egress PE (PCE)

When an egress PE (PCE) receives a PCRep message for a PCReq message from a previous PCE (i.e. CE), it will look up the new END-POINTS object associated with the PCReq message for the PCRep message. The egress PE performs a path computation. Note that the path computation procedure itself is out of scope in this document. Afterwards, the egress PE adds the new END-POINTS object in a PCRep message and sends it to an ingress PE.

3.2.4 PCRep Message Processing at Ingress PE (PCE)

When an ingress PE (PCE) receives a PCRep message for a PCReq message from an egress PE (PCE), it can distinguish a VRF instance from the source VPN-IPv4/VPN-IPv6 address in the new END-POINTS object.

Therefore, it is now possible to generate a PCRep message to send to an appropriate PCC/PCE.

4. Security Considerations

This document defines PCEP extensions for BGP/MPLS IP-VPNs. The security of the PCE extensions relies on the security of PCEP [RFC5440]. It is important that implementations conform to security features defined in [RFC5440].

5. IANA Considerations

IANA maintains a registry of PCEP parameters. As described in Section 3.1 (Type Definition), two Object-Types have been defined. IANA is requested to make the following allocations from the "PCEP Objects" sub-registry.

Object-Class Value	Name	Object-Type	Reference
4	END-POINTS	1: IPv4 addresses	[RFC5440]
		2: IPv6 addresses	[RFC5440]
		3: VPN-IPv4 addresses	[This.I-D]
		4: VPN-IPv6 addresses	[This.I-D]
		5-15: Unassigned	

The values 3 and 4 are suggested.

6. References

6.1 Normative References

- [RFC4271] Rekhter, Y. and Li, T., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.
- [RFC5440] Vasseur, J.-P., et al., "Path Computation Element(PCE) communication Protocol (PCEP) - Version 1", RFC5440, March 2009.

6.2 Informative References

- [RFC4364] Rosen, E., Rekhter, Y., "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.
- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "Path Computation Element (PCE) Architecture", RFC 4655, August 2006.
- [RFC5824] Kumaki, K., Zhang, R., and Kamite, Y., "Requirements for supporting Customer RSVP and RSVP-TE over a BGP/MPLS IP-VPN", RFC 5824, April 2010.
- [RFC6882] Kumaki, K., Murai, T., Matsushima, S., and Jiang, P., "Support for Resource Reservation Protocol Traffic Engineering (RSVP-TE) in Layer 3 Virtual Private Networks (L3VPNs)", RFC 6882, March 2013.
- [RFC5441] Vasseur, J.-P., et al., "A Backward Recursive PCE-based Computation (BRPC) Procedure To Compute Shortest Constrained Inter-domain Traffic Engineering Label Switched Paths", RFC5441, April 2009.

7. Acknowledgments

The author would like to express thanks to Makoto Nakamura for his helpful and useful comments and feedback.

8. Authors' Addresses

Kenji Kumaki
KDDI Corporation
1-20-1 Nishishinjuku
Shinjuku-ku, Tokyo 160-0023
Japan
Email: ke-kumaki@kddi.com

Tomoki Murai
FURUKAWA NETWORK SOLUTION CORP.
5-1-9, HIGASHI-YAWATA, HIRATSUKA
Kanagawa 254-0016, JAPAN
Email: murai@fnsc.co.jp

Takuya Miyasaka
KDDI Corporation
1-20-1 Nishishinjuku
Shinjuku-ku, Tokyo 160-0023
Japan
Email: ta-miyasaka@kddi.com

Chikara Sasaki
KDDI Corporation
1-20-1 Nishishinjuku
Shinjuku-ku, Tokyo 160-0023
Japan
Email: ch-sasaki@kddi.com

Peng Jiang
KDDI Corporation
1-20-1 Nishishinjuku
Shinjuku-ku, Tokyo 160-0023
Japan
Email: pe-jiang@kddi.com