Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 31, 2012

E. Crabbe
Google, Inc.
J. Medved
Cisco Systems, Inc.
R. Varga
Pantheon Technologies LLC
I. Minei
Juniper Networks, Inc.

February 28, 2012

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

Although PCEP explicitly makes no assumptions regarding the information available to the PCE, it also makes no provisions for synchronization or 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 this functionality, providing stateful control of Multiprotocol Label Switching (MPLS) Traffic Engineering Label Switched Paths (TE LSP) via PCEP.

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 will expire on August 31, 2012.

Copyright Notice

Table of Contents

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

This document 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.

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.

The following terms are defined in this document:

Passive Stateful PCE: uses LSP state information learned from PCCs to optimize path computations. It does not actively update LSP state. A PCC maintains synchronization with the PCE.

Active Stateful PCE: uses LSP state information learned from PCCs to optimize path computations. Additionally, it actively updates LSP parameters in those PCCs that delegated control over their LSPs to the PCE.

Delegation: An operation to grant a PCE temporary rights to modify a subset of LSPs parameters on one or more PCC's LSPs. LSPs are delegated from a PCC to a PCE.

Delegation Timeout Interval: when a PCEP session is terminated, a PCC waits for this time period before revoking LSP delegation to a PCE.

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

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

LSP Priority: a specific pair of MPLS setup and hold priority values.

LSP State Database: information about and attributes of all LSPs that are being reported to one or more PCEs via LSP State Reports.

Minimum Cut Set: the minimum set of links for a specific source destination pair which, when removed from the network, result in a specific source being completely isolated from specific destination. The summed capacity of these links is equivalent to the maximum capacity from the source to the destination by the max-flow min-cut theorem.

MPLS TE Global Default Restoration: once an LSP failure is detected by some downstream mode, the head-end LSP is notified by means of RSVP. Upon receiving the notification, the headend Label Switching Router (LSR) recomputes the path and signals the LSP along an alternate path. [NET-REC]

MPLS TE Global Path Protection: once an LSP failure is detected by some downstream mode, the head-end LSP is notified by means of RSVP. Upon receiving the notification, the headend LSR reroutes traffic using a pre-signaled backup (secondary) LSP. [NET-REC].

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 and Objectives

3.1. Motivation

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 it's

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, and in particular, the widely implemented but seldom deployed auto-bandwidth system.

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:

- Any reliable synchronization mechanism would result in significant control plane overhead

- Out-of-band ted synchronization would be complex and prone to race conditions

- Path calculations incorporating total network state would be highly complex

In general, stress on the MPLS TE control plane will be directly proportional to the size of the system being controlled and the 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 following use cases demonstrate a need for visibility into global inter-PCC LSP state in PCE path computations, and for a PCE control of sequence and timing in altering LSP path characteristics within and across PCEP sessions. Reference topologies for the use cases described later in this section are shown in Figures 1 and 2.

Unless otherwise cited, use cases assume that all LSPs listed exist at the same LSP priority.

```
          +-------+
          |   A   |
          |       |
          +-------+
                    \
                      +-------+                         +-------+
                      |   C   |-------------------------|   E   |
                      |       |                         |       |
                      +-------+        +-------+        +-------+
                    /          \       |   D   |      /
          +-------+              ------|       |------
          |   B   |                    +-------+
          |       |
          +-------+
```

Figure 1: Reference topology 1

```
               +-------+        +-------+        +-------+
               |   A   |        |   B   |        |   C   |
               |       |        |       |        |       |
               +---+---+        +---+---+        +---+---+
                   |                |                |
                   |                |                |
               +---+---+        +---+---+        +---+---+
               |   E   |        |   F   |        |   G   |
               |       +--------+       +--------+       |
               +-------+        +-------+        +-------+
```

Figure 2: Reference topology 2

3.1.2.1. Throughput Maximization and Bin Packing

Because LSP attribute changes in [RFC5440] are driven by PCReq messages under control of a PCC's local timers, the sequence of RSVP reservation arrivals occurring in the network will be randomized. This, coupled with a lack of global LSP state visibility on the part of a stateless PCE may result in suboptimal throughput in a given network topology.

Reference topology 2 in Figure 2 and Tables 1 and 2 show an example in which throughput is at 50% of optimal as a result of lack of visibility and synchronized control across PCC's. In this scenario, the decision must be made as to whether to route any portion of the E-G demand, as any demand routed for this source and destination will decrease system throughput.

Link	Metric	Capacity
A-E	1	10
B-F	1	10
C-G	1	10
E-F	1	10
F-G	1	10

Table 1: Link parameters for Throughput use case

Time	LSP	Src	Dst	Demand	Routable	Path
1	1	E	G	10	Yes	E-F-G
2	2	A	B	10	No	---
3	1	F	C	10	No	---

Table 2: Throughput use case demand time series

In many cases throughput maximization becomes a bin packing problem. While bin packing itself is an NP-hard problem, a number of common heuristics which run in polynomial time can provide significant improvements in throughput over random reservation event distribution, especially when traversing links which are members of the minimum cut set for a large subset of source destination pairs.

Tables 3 and 4 show a simple use case using Reference Topology 1 in

Figure 1, where LSP state visibility and control of reservation order across PCCs would result in significant improvement in total throughput.

Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	5
C-D	1	10
D-E	1	10

Table 3: Link parameters for Bin Packing use case

Time	LSP	Src	Dst	Demand	Routable	Path
1	1	A	E	5	Yes	A-C-D-E
2	2	B	E	10	No	---

Table 4: Bin Packing use case demand time series

3.1.2.2. Deadlock

Most existing RSVP-TE implementations will not tear down established LSPs in the event of the failure of the bandwidth increase procedure detailed in [RFC3209]. This behavior is directly implied to be correct in [RFC3209] and is often desirable from an operator's perspective, because either a) the destination prefixes are not reachable via any means other than MPLS or b) this would result in significant packet loss as demand is shifted to other LSPs in the overlay mesh.

In addition, there are currently few implementations offering ingress admission control at the LSP level. Again, having ingress admission control on a per LSP basis is not necessarily desirable from an operational perspective, as a) one must over-provision tunnels significantly in order to avoid deleterious effects resulting from stacked transport and flow control systems and b) there is currently no efficient commonly available northbound interface for dynamic configuration of per LSP ingress admission control (such an interface could easily be defined using the extensions present in this spec, but it beyond the scope of the current document).

Lack of ingress admission control coupled with the behavior in

[RFC3209] effectively results in mis-signaled LSPs during periods of contention for network capacity between LSPs in a given LSP priority. This in turn causes information loss in the TED with regard to actual network state, resulting in LSPs sharing common network interfaces with mis-signaled LSPs operating in a degraded state for significant periods of time, even when unused network capacity may potentially be available.

Reference Topology 1 in Figure 1 and Tables 5 and 6 show a use case that demonstrates this behavior. Two LSPs, LSP 1 and LSP 2 are signaled with demand 2 and routed along paths A-C-D-E and B-C-D-E respectively. At a later time, the demand of LSP 1 increases to 20. Under such a demand, the LSP cannot be resignaled. However, the existing LSP will not be torn down. In the absence of ingress policing, traffic on LSP 1 will cause degradation for traffic of LSP 2 (due to oversubscription on the links C-D and D-E), as well as information loss in the TED with regard to the actual network state.

The problem could be easily ameliorated by global visibility of LSP state coupled with PCC- external demand measurements and placement of two LSPs on disjoint links. Note that while the demand of 20 for LSP 1 could never be satisfied in the given topology, what could be achieved would be isolation from the ill-effects of the (unsatisfiable) increased demand.

Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	5
C-D	1	10
D-E	1	10

Table 5: Link parameters for the 'Deadlock' example

Time	LSP	Src	Dst	Demand	Routable	Path
1	1	A	E	2	Yes	A-C-D-E
2	2	B	E	2	Yes	B-C-D-E
3	1	A	E	20	No	---

Table 6: Deadlock LSP and demand time series

3.1.2.3. Minimum Perturbation

As a result of both the lack of visibility into global LSP state and the lack of control over event ordering across PCE sessions, unnecessary perturbations may be introduced into the network by a stateless PCE. Tables 7 and 8 show an example of an unnecessary network perturbation using Reference Topology 1 in Figure 1. In this case an unimportant (high LSP priority value) LSP (LSP1) is first set up along the shortest path. At time 2, which is assumed to be relatively close to time 1, a second more important (lower LSP-priority value) LSP is established, preempting LSP 1 and shifting it to the longer A-C-E path.

Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	10	10
C-D	1	10
D-E	1	10

Table 7: Link parameters for the 'Minimum-Perturbation' example

Time	LSP	Src	Dst	Demand	LSP Prio	Routable	Path
1	1	A	E	7	7	Yes	A-C-D-E
2	2	B	E	7	0	Yes	B-C-D-E
3	1	A	E	7	7	Yes	A-C-E

Table 8: Minimum-Perturbation LSP and demand time series

3.1.2.4. Predictability

Randomization of reservation events caused by lack of control over event ordering across PCE sessions results in poor predictability in LSP routing. An offline system applying a consistent optimization method will produce predictable results to within either the boundary of forecast error when reservations are over-provisioned by reasonable margins or to the variability of the signal and the forecast error when applying some hysteresis in order to minimize churn.

Reference Topology 1 and Tables 9, 10 and 11 show the impact of event ordering and predictability of LSP routing.

Link	Metric	Capacity
A-C	1	10
B-C	1	10
C-E	1	10
C-D	1	10
D-E	1	10

Table 9: Link parameters for the 'Predictability' example

Time	LSP	Src	Dst	Demand	Routable	Path
1	1	A	E	7	Yes	A-C-E
2	2	B	E	7	Yes	B-C-D-E

Table 10: Predictability LSP and demand time series 1

Time	LSP	Src	Dst	Demand	Routable	Path
1	2	B	E	7	Yes	B-C-E
2	1	A	E	7	Yes	A-C-D-E

Table 11: Predictability LSP and demand time series 2

3.1.2.5. Global Concurrent Optimization

Global Concurrent Optimization (GCO) defined in [RFC5557] is a network optimization mechanism that is able to simultaneously consider the entire topology of the network and the complete set of existing TE LSPs and their existing constraints, and look to optimize or reoptimize the entire network to satisfy all constraints for all TE LSPs. It allows for bulk path computations in order to avoid blocking problems and to achieve more optimal network-wide solutions.

Global control of LSP operation sequence in [RFC5557] is predicated on the use of what is effectively a stateful (or semi-stateful) NMS. The NMS can be either not local to the switch, in which case another northbound interface is required for LSP attribute changes, or local/ collocated, in which case there are significant issues with efficiency in resource usage. Stateful PCE adds a few features that:

- Roll the NMS visibility into the PCE and remove the requirement for an additional northbound interface

- Allow the PCE to determine when re-optimization is needed

- Allow the PCE to determine which LSPs should be re-optimized

- Allow a PCE to control the sequence of events across multiple PCCs, allowing for bulk (and truly global) optimization, LSP shuffling etc.

3.1.3. Protocol vs. Configuration

Note that existing configuration tools and protocols can be used to set LSP state. However, this solution has several shortcomings:

- Scale & Performance: configuration operations 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.

- Scale & Performance: configuration operations often have transactional semantics which are typically heavyweight and require additional CPU cycles, negatively impacting PCC update rate capacity.

- Security: opening up a configuration channel to a PCE would allow a malicious PCE to take over a PCC. The PCEP extensions described in this document only allow a PCE control over a very limited set of LSP attributes.

- Interoperability: each vendor has a proprietary information model for configuring LSP state, which prevents interoperability of a 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.

- Efficient State Synchronization: configuration channels may be heavyweight and unidirectional, therefore efficient state synchronization between a PCE 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:

- Allow a single PCC to interact with a mix of stateless and stateful PCEs simultaneously using the same PCEP.

- Support efficient LSP state synchronization between the PCC and one or more active or passive stateful PCEs.

- Allow a PCC to delegate control of its LSPs to an active stateful PCE such that a single LSP is under the control a single PCE at any given time. A PCC may revoke this delegation at any point during the lifetime of the PCEP session. A PCE may return this delegation at any point during the lifetime of the PCEP session.

- Allow a PCE to control computation timing and update timing across all LSPs that have been delegated to it.

- Allow a PCE to specify protection / restoration settings for all LSPs that have been delegated to it.

- Enable uninterrupted operation of PCC's LSPs in the event PCE failure or while control of LSPs is being transferred between PCEs.

4. New Functions to Support Stateful PCEs

Several new functions will be 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 negotiation (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 5.5); the PCC may withdraw the delegation or the PCE may give up the delegation

In addition to new PCEP functions, stateful capabilities discovery

will be required in OSPF ([RFC5088]) and IS-IS ([RFC5089]). Stateful capabilities discovery is not in scope of this document.

5. Architectural Overview of Protocol Extensions

5.1. LSP State Ownership

In the PCEP protocol (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 protocol extensions described in this document do not change this behavior.

An active stateful PCE may have control of a PCC's LSPs be 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:

- a PCC to interact with both stateless and stateful PCEs at the same time

- 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 through CLI commands

- 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 Status Report in a PCRpt message can 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 PCRep 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 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 Negotiation (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
ISIS stateful capability advertisement	ISIS PCE-CAP-FLAGS sub-TLV
OSPF stateful capability advertisement	OSPF RI LSA, PCE TLV, PCE-CAP-FLAGS sub-TLV

Table 12: New Function to Message Mapping

5.3. Capability Negotiation

During PCEP Initialization Phase, PCEP Speakers (PCE pr PCC) negotiate the use 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 protocol extensions for stateful PCEs MAY only be used if both sides have included the Stateful PCE Capability TLV in their respective OPEN messages, otherwise a PCErr with code "Stateful PCE capability not negotiated" (see Section 8.4) will be generated and the PCEP session will be terminated.

LSP delegation and LSP update operations defined in this document MAY

only be used if both PCEP Speakers set the LSP-UPDATE Flag in the "Stateful Capability" TLV to 'Updates Allowed (U Flag = 1)', otherwise a PCErr with code "Delegation not negotiated" (see Section 8.4) will be generated. Note that even if the update capability has not been negotiated, a PCE can still receive LSP Status Reports from a PCC and build and maintain an up to date view of the state of the PCC's LSPs.

5.4. 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 negotiated stateful PCEP capabilities and PCC's local configuration (see more details in Section 9.1).

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.

If the PCC encounters a problem which prevents it from completing the state transfer, it MUST send a PCErr message to the PCE and terminate the session using the PCEP session termination procedure.

The PCE does not send positive acknowledgements for properly received synchronization messages. It MUST respond with a PCErr message indicating "PCRpt error" (see) if it encounters a problem with the LSP State Report it received from the PCC. Either the PCE or the PCC MAY terminate the session if the PCE encounters a problem during the synchronization.

The successful State Synchronization sequence is shown in Figure 3.

```
                     +-+-+                    +-+-+
                     |PCC|                    |PCE|
                     +-+-+                    +-+-+
                       |                        |
                       |-----PCRpt, SYNC=1----->| (Sync start)
                       |                        |
                       |-----PCRpt, SYNC=1----->|
                       |            .           |
                       |            .           |
                       |            .           |
                       |-----PCRpt, SYNC=1----->| (Sync done)
                       |            .           |
                       |            .           |
                       |            .           |
                       |                        |
                       |-----PCRpt, SYNC=0----->| (Regular
                       |                        |  LSP State Report)
```

Figure 3: Successful state synchronization

The sequence where the PCE fails during the State Synchronization phase is shown in Figure 4.

```
                     +-+-+                    +-+-+
                     |PCC|                    |PCE|
                     +-+-+                    +-+-+
                       |                        |
                       |-----PCRpt, SYNC=1----->|
                       |                        |
                       |-----PCRpt, SYNC=1----->|
                       |            .           |
                       |            .           |
                       |            .           |
                       |-----PCRpt, SYNC=1----->|
                       |                        |
                       |-PCRpt, SYNC=1          |
                       |         \    ,-PCErr=?-|
                       |          \  /          |
                       |           \/           |
                       |           /\           |
                       |          /   `-------->| (Ignored)
                       |<--------`              |
```

Figure 4: Failed state synchronization (PCE failure)

The sequence where the PCC fails during the State Synchronization phase is shown in Figure 5.

```
                     +-+-+                    +-+-+
                     |PCC|                    |PCE|
                     +-+-+                    +-+-+
                       |                        |
                       |-----PCRpt, SYNC=1----->|
                       |                        |
                       |-----PCRpt, SYNC=1----->|
                       |            .           |
                       |            .           |
                       |            .           |
                       |-------- PCErr=? ------>|
                       |                        |
```

Figure 5: Failed state synchronization (PCC failure)

5.4.1. State Synchronization Avoidance

State Synchronization MAY be skipped following a PCEP session restart if the state of both PCEP peers did not change during the period prior to session re-initialization. To be able to make this determination, state must be exchanged and maintained by both PCE and PCC during normal operation. This is accomplished by keeping track of the changes to the LSP State Database. When State Synchronization avoidance is enabled on a PCEP session, a PCC includes the LSP-DB-VERSION TLV as an optional TLV in the LSP Object on each LSP State Report. The LSP-DB-VERSION TLV contains a PCC's LSP State Database version, which is incremented each time a change is made to the PCC's local LSP State Database. The LSP State Database version is an unsigned 64-bit value that MUST be incremented by 1 for each successive change in the LSP state database. The LSP State Database version MUST start at 1 and may wrap around. Values 0 and 0xFFFFFFFFFFFFFFFF are reserved.

State Synchronization Avoidance is negotiated on a PCEP session during session startup.

If both PCEP speakers set the INCLUDE-DB-VERSION Flag in the OPEN object's STATEFUL-PCE-CAPABILITY TLV to 1, the PCC will include the LSP-DB-VERSION TLV in each LSP Object. The TLV will contain the PCC's latest LSP State Database version.

If a PCE's LSP State Database survived the restart of a PCEP session, the PCE will include the LSP-DB-VERSION TLV in its OPEN object, and the TLV will contain the last LSP State Database version received on an LSP State Update from the PCC in a previous PCEP session. If a PCC's LSP State Database survived the restart, the PCC will include the LSP-DB-VERSION TLV in its OPEN object and the TLV will contain the last LSP State Database version sent on an LSP State Update from

the PCC in the previous PCEP session. If a PCEP Speaker's LSP State Database did not survive the restart of a PCEP session, the PCEP Speaker MUST NOT include the LSP-DB-VERSION TLV in the OPEN Object.

If both PCEP Speakers include the LSP-DB-VERSION TLV in the OPEN Object and the TLV values match, the PCC MAY skip State Synchronization. Otherwise, the PCE MUST purge any remaining LSP state and expect the PCC to perform State Synchronization; if the PCC attempts to skip State Synchronization (i.e. the SYNC Flag = 0 on the first LSP State Report from the PCC), the PCE MUST send back a PCError with Error-type 20 Error-value 2 'LSP Database version mismatch', and close the PCEP session.

Note that a PCE MAY force State Synchronization by not including the LSP-DB-VERSION TLV in its OPEN object.

Figure 6 shows an example sequence where State Synchronization is skipped.

```
                  +-+-+                    +-+-+
                  |PCC|                    |PCE|
                  +-+-+                    +-+-+
                    |                        |
                    |-PCOpen-,               |
                    |  DBv=42 \    ,--PCOpen-|
                    |   IDB=1  \  /   DBv=42 |
                    |           \/     IDB=1 |
                    |           /\           |
                    |          /   `-------->| (OK to skip sync)
        (Skip sync) |<--------`              |
                    |            .           |
                    |            .           |
                    |            .           |
                    |                        |
                    |--PCRpt,DBv=43,SYNC=0-->| (Regular
                    |                        |  LSP State Update)
                    |--PCRpt,DBv=44,SYNC=0-->| (Regular
                    |                        |  LSP State Update)
                    |--PCRpt,DBv=45,SYNC=0-->|
                    |                        |
```

Figure 6: State Synchronization skipped

Figure 7 shows an example sequence where State Synchronization is performed due to LSP State Database version mismatch during the PCEP session setup. Note that the same State Synchronization sequence would happen if either the PCE or the PCE would not include the LSP-DB-VERSION TLV in their respective Open messages.

```
                     +-+-+                    +-+-+
                     |PCC|                    |PCE|
                     +-+-+                    +-+-+
                       |                        |
                       |-PCOpen-,               |
                       |  DBv=46 \    ,--PCOpen-|
                       |   IDB=1  \  /   DBv=42 |
                       |           \/     IDB=1 |
                       |           /\           |
                       |          /   `-------->| (Expect sync)
             (Do sync) |<--------`              | (Purge LSP State)
                       |                        |
                       |--PCRpt,DBv=46,SYNC=1-->| (Sync start)
                       |            .           |
                       |            .           |
                       |            .           |
                       |--PCRpt,DBv=46,SYNC=1-->| (Sync done)
                       |            .           |
                       |            .           |
                       |            .           |
                       |--PCRpt,DBv=47,SYNC=0-->| (Regular
                       |                        |  LSP State Update)
                       |--PCRpt,DBv=48,SYNC=0-->| (Regular
                       |                        |  LSP State Update)
                       |--PCRpt,DBv=49,SYNC=0-->|
                       |                        |
```

Figure 7: State Synchronization performed

Figure 8 shows an example sequence where State Synchronization is skipped, but because one or both PCEP Speakers set the INCLUDE-DB-VERSION Flag to 0, the PCC does not send LSP-DB-VERSION TLVs to the PCE. If the current PCEP session restarts, the PCEP Speakers will have to perform State Synchronization, since the PCE will not know the PCC's latest LSP State Database version.

```
                     +-+-+                    +-+-+
                     |PCC|                    |PCE|
                     +-+-+                    +-+-+
                       |                        |
                       |-PCOpen-,               |
                       |  DBv=42 \    ,--PCOpen-|
                       |   IDB=0  \  /   DBv=42 |
                       |           \/     IDB=0 |
                       |           /\           |
                       |          /   `-------->| (OK to skip sync)
           (Skip sync) |<--------`              |
                       |            .           |
                       |            .           |
                       |            .           |
                       |------PCRpt,SYNC=0----->| (Regular
                       |                        |  LSP State Update)
                       |------PCRpt,SYNC=0----->| (Regular
                       |                        |  LSP State Update)
                       |------PCRpt,SYNC=0----->|
                       |                        |
```

Figure 8: State Synchronization skipped, no LSP-DB-VERSION TLVs sent from PCC

5.5. LSP Delegation

If during Capability negotiation 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.

LSP Delegation is controlled by operator-defined policies on a PCC. LSPs are delegated individually - different LSPs may be delegated to different PCEs, and an LSP may be delegated to one or more PCEs. The delegation policy, when all PCC's LSPs are delegated to a single PCE at any given time, SHOULD be supported by all delegation-capable PCCs. Conversely, the policy revoking the delegation for all PCC's LSPs SHOULD also be supported

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

5.5.1. Delegating an LSP

A PCC delegates an LSP to a PCE by setting the Delegate flag in LSP State Report to 1. A PCE confirms the delegation when it sends the first LSP Update Request for the delegated LSP to the PCC by setting the Delegate flag to 1. Note that a PCE does not immediately confirm to the PCC the acceptance of LSP Delegation; Delegation acceptance is confirmed when the PCC wishes to update the LSP via the LSP Update Request. If a 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.

The delegation sequence is shown in Figure 9.

```
                  +-+-+                    +-+-+
                  |PCC|                    |PCE|
                  +-+-+                    +-+-+
                    |                        |
                    |---PCRpt, Delegate=1--->| LSP Delegated
                    |                        |
                    |---PCRpt, Delegate=1--->|
                    |            .           |
                    |            .           |
                    |            .           |
                    |<--(PCUpd,Delegate=1)---| Delegation confirmed
                    |                        |
                    |---PCRpt, Delegate=1--->|
                    |                        |
```

Figure 9: 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 Status Report sent to the PCE.

5.5.2. Revoking a Delegation

A PCC revokes an LSP delegation by sending an LSP State Report with the Delegate flag set to 0. A PCC MAY revoke an LSP delegation at any time during the PCEP session life time. When a PCC's PCEP session with the PCE terminates, the PCC SHALL wait a time interval specified in 'Delegation Timeout Interval' and then revoke all LSP delegations to the PCE .

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 indicating "LSP is not delegated" (see Section 8.4).

The revocation sequence is shown in Figure 10.

```
                    +-+-+                    +-+-+
                    |PCC|                    |PCE|
                    +-+-+                    +-+-+
                      |                        |
                      |---PCRpt, Delegate=1--->|
                      |                        |
                      |<--(PCUpd,Delegate=1)---| Delegation confirmed
                      |            .           |
                      |            .           |
                      |            .           |
                      |---PCRpt, Delegate=0--->| Delegation revoked
                      |                        |
```

Figure 10: Revoking a Delegation

If a PCC can not 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 Delegation Timeout Interval and the PCC SHALL flush any LSP state set by a PCE.

5.5.3. Returning a Delegation

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. Note that in order to keep a delegation, the PCE MUST set the Delegate flag to 1 on each LSP Update Request sent to the PCC.

```
                    +-+-+                    +-+-+
                    |PCC|                    |PCE|
                    +-+-+                    +-+-+
                      |                        |
                      |---PCRpt, Delegate=1--->| LSP delegated
                      |            .           |
                      |            .           |
                      |            .           |
                      |<--PCUpd, Delegate=0----| Delegation returned
                      |                        |
                      |---PCRpt, Delegate=0--->| No delegation for LSP
                      |                        |
```

Figure 11: Returning a Delegation

If a PCC can not 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 Delegation Timeout Interval and the PCC MUST flush any LSP state set by a PCE.

5.5.4. Redundant Stateful PCEs

Note that a PCE may not have any delegated LSPs: in a redundant configuration where one PCE is backing up another PCE, the backup PCE will not have any delegated LSPs. The backup PCE does not update any LSPs, but it receives all LSP State Reports from a PCC. When the primary PCE fails, a PCC will delegate to the secondary PCE all LSPs that had been previously delegated to the failed PCE.

5.6. LSP Operations

5.6.1. Passive Stateful PCE Path Computation Request/Response

```
                     +-+-+                    +-+-+
                     |PCC|                    |PCE|
                     +-+-+                    +-+-+
                       |                        |
   1) Path computation |----- PCReq message --->|
      request sent to  |                        |2) Path computation
      PCE              |                        |   request received,
                       |                        |   path computed
                       |                        |
                       |<---- PCRep message ----|3) Computed paths
                       |     (Positive reply)   |   sent to the PCC
                       |     (Negative reply)   |
   4) LSP Status change|                        |
      event            |                        |
                       |                        |
   5) LSP Status Report|----- PCRpt message --->|
      sent to all      |            .           |
      stateful PCEs    |            .           |
                       |            .           |
   6) Repeat for each  |----- PCRpt message --->|
      LSP status change|                        |
                       |                        |
```

Figure 12: 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 sends an LSP State Report carried on a PCRpt message to the PCE, indicating that the LSP's status is 'Pending'.

Once an LSP is up, the PCC sends an LSP State Report carried on a PCRpt message to the PCE, indicating that the LSP's status is 'Up'. If the LSP could not be set up, the PCC sends 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 'Pending' to 'Up' (or 'Down') before the PCC has had a chance to send an LSP State Report indicating that the status is 'Pending'. In such cases, the PCC may choose to only send the PCRpt indicating the latest status ('Up' or 'Down').

Upon receiving a negative reply from a PCE, a PCC may decide to resend a modified request or take any other appropriate action. For each requested LSP, it also sends 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 PC Reply, as a PCC notifies a PCE of the LSP's state changes.

A PCC sends 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 PCE is the model for stateful PCEs is described in [RFC4655], Section 6.8.

5.6.2. Active Stateful PCE LSP Update

```
                     +-+-+                     +-+-+
                     |PCC|                     |PCE|
                     +-+-+                     +-+-+
                       |                         |
   1) LSP State        |-- PCRpt, Delegate=1 -->|
      Synchronization  |            .            |
      or add new LSP   |            .            |2) PCE decides to
                       |            .            |   update the LSP
                       |                         |
                       |<---- PCUpd message ----|3) PCUpd message sent
                       |                         |   to PCC
                       |                         |
                       |                         |
   4) LSP Status Report|---- PCRpt message ---->|
      sent(->Pending)  |            .            |
                       |            .            |
                       |            .            |
   5) LSP Status Report|---- PCRpt message ---->|
      sent (->Up|Down) |                         |
                       |                         |
```

Figure 13: 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) and LSPs have been delegated to the PCE, the PCE can modify LSP parameters of delegated LSPs.

A PCE sends an LSP Update Request carried on a PCUpd message to the PCC. The LSP Update Request contains a variety of objects that specify the set of constraints and attributes for the LSP's path. Additionally, the PCC may specify the urgency of such request by assigning a request priority. 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 sends an LSP State Report carried on a PCRpt message to the PCE, indicating that the LSP's status is 'Pending'.

Once an LSP is up, the PCC sends 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 sends an LSP State Report indicating that the LSP is 'Down' and stating the cause of the failure. A PCC may choose to compress LSP State Updates to only reflect the most up

to date state, as discussed in the previous section.

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

A PCC MUST NOT send to any PCE a Path Computation Request for a delegated LSP.

5.7. LSP Protection

With a stateless PCE or a passive stateful PCE, LSP protection and restoration settings may be operator-configured locally at a PCC. A PCE may be merely asked to compute the protected (primary) and backup (secondary) paths for the LSP.

An active stateful PCE controls the LSPs that are delegated to it, and must therefore be able to set via PCEP the desired protection / restoration mechanism for each delegated LSP. PCEP extensions for stateful PCEs SHOULD support, at a minimum, the following protection mechanisms:

- MPLS TE Global Default Restoration

- MPLS TE Global Path Protection

- FRR One-to-One Backup

- FRR Facility Backup - link protection, node protection, or both

5.8. Transport

A Permanent PCEP session MUST be established between a stateful PCEs and the PCC.

State cleanup after session termination, as well as session setup failures will be described in a later version of this document.

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 that can be either mandatory or optional. An object is said to be mandatory in a PCEP message when the object must be included for the message to be considered valid. For each PCEP message type, a set of rules is defined that specify the set of objects that the message can carry. An implementation MUST form the PCEP messages using the object ordering specified in this document.

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 set to [TBD].

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

Where:

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

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

Where:

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

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

The LSP object (see Section 7.2) is mandatory, 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=[TBD] (LSP object missing).

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].

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 PCRpt message is set to [TBD].

The format of a PCUpd message is as follows:

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

Where:

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

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

Where:

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

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

Where:

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

<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 Section 7.2). 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.

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 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.

7.1. OPEN Object

This document defines a new optional TLV for the OPEN Object to support stateful PCE capability negotiation.

7.1.1. Stateful PCE Capability TLV

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

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |               Type=[TBD]      |            Length=2           |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |              Flags        |S|U|
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 14: STATEFUL-PCE-CAPABILITY TLV format

The type of the TLV is [TBD] and it has a fixed length of 2 octets.

The value comprises a single field - Flags (16 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 wishes to update 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.

S (INCLUDE-DB-VERSION - 1 bit): if set to 1 by both PCEP Speakers, the PCC will include the LSP-DB-VERSION TLV in each LSP Object.

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

7.1.2. LSP State Database Version TLV

LSP-DB-VERSION is an optional TLV that MAY be included in the OPEN Object when a PCEP Speaker wishes to determine if State Synchronization can be skipped when a PCEP session is restarted. If sent from a PCE, the TLV contains the local LSP State Database version from the last valid LSP State Report received from a PCC. If sent from a PCC, the TLV contains the PCC's local LSP State Database version, which is incremented each time the LSP State Database is updated.

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

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Type=[TBD]          |            Length=8           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                      LSP State DB Version                     |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 15: LSP-DB-VERSION TLV format

The type of the TLV is [TBD] and it has a fixed length of 8 octets. The value contains a 64-bit unsigned integer.

7.2. 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 is also contains a flag to indicate to a PCE that the initial LSP state synchronization has been done.

LSP Object-Class is [TBD].

LSP Object-Type is 1.

The format of the LSP object body is shown in Figure 16:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |       Session-internal LSP-ID         |     Flags     |R|O|S|D|
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    //                        Optional TLVs                        //
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 16: The LSP Object format

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

Session-internal LSP-ID (20 bits): Per-PCEP session identifier for an LSP. In each PCEP session the PCC creates a unique LSP-ID for each LSP that will remain constant for the duration of the session. The

mapping of the LSP Symbolic Name to LSP-ID is communicated to the PCE by sending a PCRpt message containing the 'LSP Symbolic Name' TLV. All subsequent PCEP messages then address the LSP by its Session-internal LSP-ID.

Flags (12 bits):

D (Delegate - 1 bit): on a PCRpt message, the D Flag set to 1 indicates that the PCC is delegating the LSP 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 LSP State Report sent from a PCC during State Synchronization. The S Flag MUST be set to 0 otherwise.

O (Operational - 1 bit): On PCRpt messages the O Flag indicates the LSP status. Value of '1' means that the LSP is operational, i.e. it is either being signaled or it is active. Value of '0' means that the LSP is not operational, i.e it is de-routed and the PCC is not attempting to set it up. On PCUpd messages the flag indicates the desired status for the LSP. Value of '1' means that the desired LSP state is operational, value of '0' means that the target LSP should be non-operational. Setting the LSP status from the PCE SHALL NOT override the operator: if a pce-controlled LSP has been configured to be non-operational, setting the LSP's status to '1' from an PCE will not make it operational.

R (Remove - 1 bit): On PCRpt messages the R Flag indicates that the LSP has been removed from the PCC. Upon receiving an LSP State Update with the R Flag set to 1, the PCE SHOULD remove all state related to the LSP from its database.

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

TLVs that are currently defined for the LSP Object are described in the following sections.

7.2.1. The LSP Symbolic Name TLV

Each LSP MUST have a symbolic name that is unique in the PCC. The LSP Symbolic Name MUST remain constant throughout an LSP's lifetime,

which may span across multiple consecutive PCEP sessions and/or PCC restarts. The LSP Symbolic Name MAY be specified by an operator in a PCC's CLI configuration. If the operator does not specify a Symbolic Name for an LSP, the PCC MUST auto-generate one.

The LSP-SYMBOLIC-NAME TLV MUST be included in the LSP State Report 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. LSP State Report MAY be included in subsequent LSP State Reports for the LSP.

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

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type=[TBD]          |       Length (variable)       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
//                      Symbolic LSP Name                      //
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 17: LSP-SYMBOLIC-NAME TLV format

The type of the TLV is [TBD] and it has a variable length, which MUST be greater than 0.

7.2.2. 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:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Type=[TBD]          |           Length=12           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   IPv4 Tunnel Sender Address                  |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             LSP ID            |           Tunnel ID           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                       Extended Tunnel ID                      |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 18: IPV4-LSP-IDENTIFIERS TLV format

The type of the TLV is [TBD] and it has a fixed length of 8 octets. The value contains two 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 128-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 the following figure:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Type=[TBD]          |           Length=36           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +                                                               +
    |                  IPv6 tunnel sender address                   |
    +                          (16 octets)                          +
    |                                                               |
    +                                                               +
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |             LSP ID            |           Tunnel ID           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +                                                               +
    |                       Extended Tunnel ID                      |
    +                          (16 octets)                          +
    |                                                               |
    +                                                               +
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 19: IPV6-LSP-IDENTIFIERS TLV format

The type of the TLV is [TBD] and it has a fixed length of 20 octets. The value contains two 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 32-bit 'Extended Tunnel ID' identifier defined in [RFC3209], Section 4.6.1.2 for the LSP_TUNNEL_IPv6 Session Object.

7.2.3. 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:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Type=[TBD]          |            Length=4           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Error Code                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 20: 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.

7.2.4. RSVP ERROR_SPEC TLVs

If the set up of an LSP failed at a downstream node which returned an ERROR_SPEC to the PCC, the ERROR_SPEC MUST be included in the LSP State Report. Depending on whether RSVP signaling was performed over IPv4 or IPv6, the LSP Object will contain an IPV4-ERROR_SPEC TLV or an IPV6-ERROR_SPEC TLV.

The format of the IPV4-RSVP-ERROR-SPEC TLV is shown in the following figure:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Type=[TBD]          |           Length=8            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    +                IPv4 ERROR_SPEC object (rfc2205)               +
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 21: IPV4-RSVP-ERROR-SPEC TLV format

The type of the TLV is [TBD] and it has a fixed length of 8 octets. The value contains the RSVP IPv4 ERROR_SPEC object defined in [RFC2205]. Error codes allowed in the ERROR_SPEC object are defined in [RFC2205] and [RFC3209].

The format of the IPV6-RSVP-ERROR-SPEC TLV is shown in the following figure:

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           Type=[TBD]          |           Length=20           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    //               IPv6 ERROR_SPEC object (rfc2205)             //
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 22: IPV6-RSVP-ERROR-SPEC TLV format

The type of the TLV is [TBD] and it has a fixed length of 20 octets. The value contains the RSVP IPv6 ERROR_SPEC object defined in [RFC2205]. Error codes allowed in the ERROR_SPEC object are defined in [RFC2205] and [RFC3209].

7.2.5. LSP State Database Version TLV

The LSP-DB-VERSION TLV can be included as an optional TLV in the LSP object. The LSP-DB-VERSION TLV is discussed in Section 5.4.1 which covers state synchronization avoidance. The format of the TLV is described in Section 7.1.2, where the details of its use in the OPEN message are listed.

If State Synchronization Avoidance has been enabled on a PCEP session (as described in Section 5.4.1) , a PCC MUST include the LSP-DB-VERSION TLV in each LSP Object sent out on the session. If the TLV

is missing, the PCE will generate an error with error-type 6 (mandatory object missing) and Error Value 12 (LSP-DB-VERSION TLV missing) and close the session. If State Synchronization Avoidance has not been enabled on a PCEP session, the PCC SHOULD NOT include the LSP-DB-VERSION TLV in the LSP Object and the PCE SHOULD ignore it were it to receive one.

Since a PCE does not send LSP updates to a PCC, a PCC should never encounter this TLV. A PCC SHOULD ignore the LSP-DB-VERSION TLV, were it to receive one from a PCE.

7.2.6. Delegation Parameters TLVs

Multiple delegation parameters, such as sub-delegation permissions, authentication parameters, etc. need to be communicated from a PCC to a PCE during the delegation operation. Delegation parameters will be carried in multiple delegation parameter TLVs, which will be defined in future revisions of this document.

8. 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.

8.1. PCEP Messages

This document defines the following new PCEP messages:

Value	Meaning	Reference
10	Report	This document
11	Update	This document

8.2. 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

8.3. LSP Object

This document requests that a registry is created to manage the Flags field of the LSP object. New values are to be assigned by Standards Action [RFC5226]. Each bit should be tracked with the following qualities:

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

The following values are defined in this document:

Bit	Description	Reference
28	Remove	This document
29	Operational	This document
30	SYNC	This document
31	Delegate	This document

8.4. 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 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
              Error-value=12: LSP-DB-VERSION 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 active
                                 stateful PCE capability was not
                                 negotiated active PCE.
        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=2:  LSP Database version mismatch.
                 Error-value=3:  The LSP-DB-VERSION TLV Missing when
                                 State Synchronization Avoidance
                                 enabled.
```

8.5. PCEP TLV Type Indicators

This document defines the following new PCEP TLVs:

Value	Meaning	Reference
16	STATEFUL-PCE-CAPABILITY	This document
17	LSP-SYMBOLIC-NAME	This document
18	IPV4-LSP-IDENTIFIERS	This document
19	IPV6-LSP-IDENTIFIERS	This document
20	LSP-UPDATE-ERROR-CODE	This document
21	IPV4-RSVP-ERROR-SPEC	This document
22	IPV6-RSVP-ERROR-SPEC	This document
23	LSP-DB-VERSION	This document

8.6. STATEFUL-PCE-CAPABILITY TLV

This document requests that a registry is created to manage the Flags field in the STATEFUL-PCE-CAPABILITY TLV in the OPEN 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
30	INCLUDE-DB-VERSION	This document
31	LSP-UPDATE-CAPABILITY	This document

8.7. LSP-UPDATE-ERROR-CODE TLV

This document requests that a registry is created to manage the Error Codes in the LSP-UPDATE-ERROR-CODE TLV. New values are to be assigned by Standards Action [RFC5226].

The following Error Codes are defined in this document:

Value	Meaning	Reference
1	LSP Setup failed outside of the node. Must be followed by the RSVP-ERROR-SPEC TLV, which indicates the failure cause.	This document
2	LSP not operational	This document

9. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440] apply to PCEP protocol 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 the Delegation Timeout Interval. The default value of the Delegation Timeout Interval SHOULD be set to 30 seconds.

When an LSP can no longer be delegated to a PCE, after the expiration of the Delegation Timeout Interval, the LSP MAY either: 1) retain its current parameters or 2) revert to operator-defined default LSP parameters. This behavior SHOULD be configurable and in the case when (2) is supported, a PCC implementation MUST allow the operator to specify the default LSP parameters.

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

PCEP session configuration and information in the PCEP MIB module SHOULD be extended to include negotiated stateful capabilities, synchronization status, and delegation status (at the PCC list PCEs with delegated LSPs).

9.3. Liveness Detection and Monitoring

PCEP protocol 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 protocol 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

- Number of dropped LSP updates

- Number of LSP updates where LSP setup failed

A PCC implementation SHOULD provide a command to show to which PCEs LSPs are delegated.

A PCC implementation SHOULD allow the operator to manually revoke LSP delegation.

9.5. Requirements on Other Protocols and Functional Components

PCEP protocol 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 protocol extensions defined in this document.

Additionally, a PCEP implementation SHOULD allow a limit to be placed on the rate 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.

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 result in 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.

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. Acknowledgements

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

We would like to thank Shane Amante, Julien Meuric, Kohei Shiomoto, 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.

12. References

12.1. Normative References

[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.

12.2. Informative References

[MPLS-PC] Chaieb, I., Le Roux, JL., 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", pre-print, 2011.

[NET-REC] Vasseur, JP., Pickavet, M., and P. Demeester, "Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS", The Morgan Kaufmann Series in Networking, June 2004.

[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

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

Email: robert.varga@pantheon.sk

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

Email: ina@juniper.net