

Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 26, 2013

X. Zhang, Ed.
Huawei Technologies
I. Minei, Ed.
Juniper Networks, Inc.
February 22, 2013

Applicability of PCEP Extensions for Stateful PCE
draft-zhang-pce-stateful-pce-app-03

Abstract

A Stateful PCE maintains information about LSP characteristics and resource usage within the network in order to provide traffic engineering calculations for its associated PCCs. This document describes general considerations for stateful PCEP and examines its applicability and benefits through a number of use cases. PCEP extensions required for stateful PCE usage are covered in separate documents.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 26, 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.	Overview of stateful PCE	4
4.	Deployment considerations	5
4.1.	Multi-PCE deployments	5
4.2.	LSP State Synchronization	5
4.3.	PCE Survivability	5
5.	Application scenarios	5
5.1.	Optimization of LSP placement	6
5.1.1.	Throughput Maximization and Bin Packing	6
5.1.2.	Deadlock	8
5.1.3.	Minimum Perturbation	9
5.1.4.	Predictability	10
5.2.	Stateful PCE in SDN	11
5.2.1.	Smart Bandwidth Adjustment	11
5.2.2.	Bandwidth Scheduling	12
5.3.	Recovery	12
5.3.1.	Protection	12
5.3.2.	Restoration	13
5.3.3.	SRLG Diversity	14
5.4.	Maintenance of Virtual Network Topology (VNT)	15
5.5.	LSP Re-optimization	15
5.6.	Resource defragmentation	16
5.7.	Future applications	16
5.7.1.	Impairment-Aware Routing and Wavelength Assignment (IA-RWA)	16
6.	Security Considerations	17
7.	Contributing authors	18
8.	Acknowledgements	19
9.	References	19
9.1.	Normative References	19
9.2.	Informative References	20
Appendix A.	Editorial notes and open issues	22
	Authors' Addresses	23

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. Extensions for support of GMPLS in PCEP are defined in [[I-D.ietf-pce-gmpls-pcep-extensions](#)].

As per [[RFC4655](#)], a PCE can be either stateful or stateless. Compared to a stateless PCE, a stateful PCE has access to not only the network states, but also to the set of active paths and their reserved resources in use in the network. In other words, the state in a stateful PCE is determined not only by the TED but also by the set of active LSPs and their corresponding reserved resources. Furthermore, a stateful PCE might also retain the information of LSPs under construction in order to reduce resource contention. Such augmented state allows the PCE to compute constrained paths while considering individual LSPs and their interaction.

This document describes how stateful PCE can solve various problems for MPLS-TE and GMPLS deployment use cases, and the benefits it brings to such deployments.

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, Revocation, Delegation Timeout Interval, LSP State Report, LSP Update Request, LSP State Database.

This document defines the following terms:

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.

3. Overview of stateful PCE

This section is included for the convenience of the reader, please refer to the specification documents for details of the operation.

[I-D.ietf-pce-stateful-pce] specifies a set of extensions to PCEP to enable stateful control of tunnels between and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect tunnel state synchronization between PCCs and PCEs, delegation of control over tunnels to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions.

[I-D.ietf-pce-stateful-pce] applies equally to MPLS-TE and GMPLS LSPs.

Several new functions were added in PCEP to support stateful PCEs and are described in [I-D.ietf-pce-stateful-pce]. 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.

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; the PCC may withdraw the delegation or the PCE may give up the delegation.

[I-D.sivabalan-pce-disco-stateful] defines the extensions needed to support autodiscovery of stateful PCEs when using the IGP for PCE discovery.

4. Deployment considerations

4.1. Multi-PCE deployments

Stateless and stateful PCEs can co-exist in the same network and be in charge of path computation of different types. To solve the problem of distinguishing between the two types of PCEs, either discovery or configuration can be used. The capability negotiation in [[I-D.ietf-pce-stateful-pce](#)] ensures correct operation when the PCE address is configured on the PCC.

4.2. LSP State Synchronization

A stateful PCE maintains two databases for path computation. The first database is the Traffic Engineering Database (TED) which includes the topology and resource state in the network. This information can be obtained by a stateful PCE using the same mechanisms as a stateless PCE (see [[RFC4655](#)]). The second database is the LSP state Database (LSP-DB), in which a PCE stores attributes of all active LSPs in the network, such as their path through the network, bandwidth/resource usage, switching types, and LSP constraints etc. The stateful PCE extensions support population of this database using information received from the network nodes via LSP Report messages. Population of the LSP database via other means is not precluded.

4.3. PCE Survivability

For a stateful PCE, an important issue is to get the LSP state information resynchronized after a restart. [[I-D.ietf-pce-stateful-pce](#)] includes support of a synchronization function, allowing the PCC to synchronize its LSP state with the PCE. This can be applied equally to an LER client or another PCE, allowing for support of multiple ways of re-acquiring the LSP database on a restart. For example, the state can be retrieved from the network nodes, or from another stateful PCE. Because synchronization may also be skipped, if a PCE implementation has the means to retrieve its database in a different way (for example from a backup copy stored locally), the state can be restored without further overhead in the network.

5. Application scenarios

In the following sections, several use cases are described, showcasing scenarios that benefit from the deployment of a stateful PCE.

5.1. Optimization of LSP placement

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.

Although the use cases are for MPLS-TE deployments, they are equally applicable to GMPLS. Unless otherwise cited, use cases assume that all LSPs listed exist at the same LSP priority.

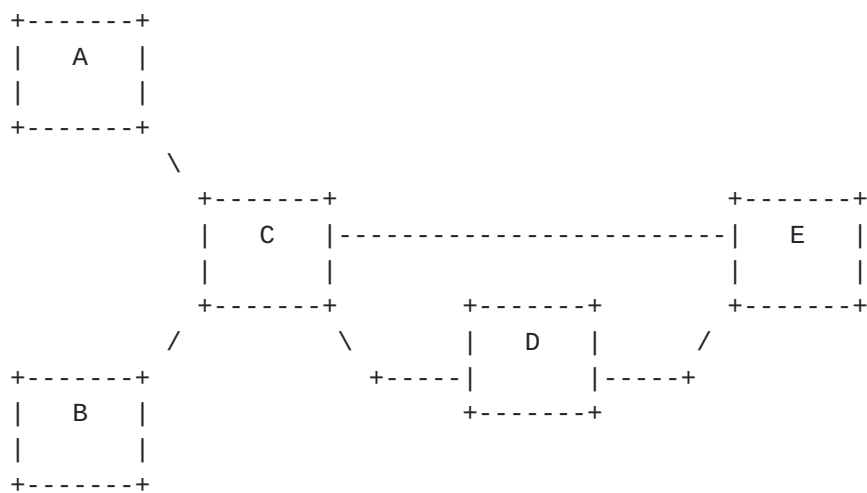


Figure 1: Reference topology 1

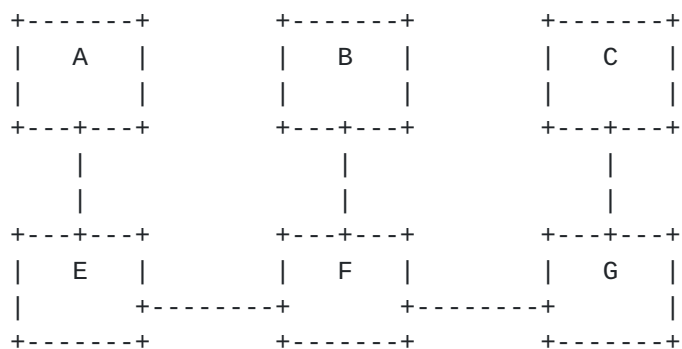


Figure 2: Reference topology 2

5.1.1. 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

5.1.1.2. 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 is beyond the scope of the current document).

Lack of ingress admission control coupled with the behavior in [RFC3209] effectively results in mis-signalized 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

5.1.3. 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

5.1.4. 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

5.2. Stateful PCE in SDN

SDN promises to incorporate more intelligence into the network by using smart centralized controllers. The use cases below show the integration between a stateful PCE and such a controller. Note that although from an implementation point of view, the SDN controller and stateful PCE could be combined, in the discussion below they are separate to show how stateful PCE enables the control-loop feedback central to SDN.

5.2.1. Smart Bandwidth Adjustment

The bandwidth requirement of LSPs often change over time, requiring resizing the LSP. Currently router software performs this function by monitoring the actual bandwidth usage, triggering a recomputation and ressignaling when a threshold is reached. A central controller can use additional information (such as historical trending data, information from specific applications or policy information) in

order to make the determination of when and along which path an LSP should be resized. The controller can rely on a stateful PCE to perform the central function.

5.2.2. Bandwidth Scheduling

Bandwidth Scheduling allows network operators to reserve resources in advance upon request from the customers to transmit large bulk of data with specified starting time and duration, such as in support of scheduled data transmission between data centers.

Traditionally, this can be supported by NMS operation through path pre-establishment and activation on the agreed starting time. However, this does not provide efficient network usage since the established paths exclude the possibility of being used by other services even when they are not used for undertaking any service. It can also be accomplished through GMPLS protocol extensions by carrying the related request information (e.g., starting time and duration) across the network. Nevertheless, this method inevitably increases the complexity of signaling and routing process.

A stateful PCE can support this application with better efficiency since it can alleviate the burden of processing on network elements as well as enable the flexibility of resources usage by only excluding the time slot(s) reserved for bandwidth scheduling requests. The details of organizing bandwidth scheduling related information as well as its impact on LSP-DB is subject to network providers policy and administrative consideration and thus outside of the scope of this document.

5.3. Recovery

5.3.1. Protection

For protection purposes, a PCC may send a request to a PCE for computing a set of paths for a given LSP. Alternatively, the PCC can send multiple requests to the PCE, asking for working and backup LSPs separately. Either way, the resources bound to backup paths can be shared by different LSPs to improve the overall network efficiency, such as m:n protection or pre-configured shared mesh recovery techniques as specified in [\[RFC4427\]](#). If resource sharing is supported for LSP protection, the information relating to existing LSPs is required to avoid allocation of shared protection resources to two LSPs that might fail together and cause protection contention issues. Stateful PCEs can easily accommodate this need using the information stored in its LSP-DB.

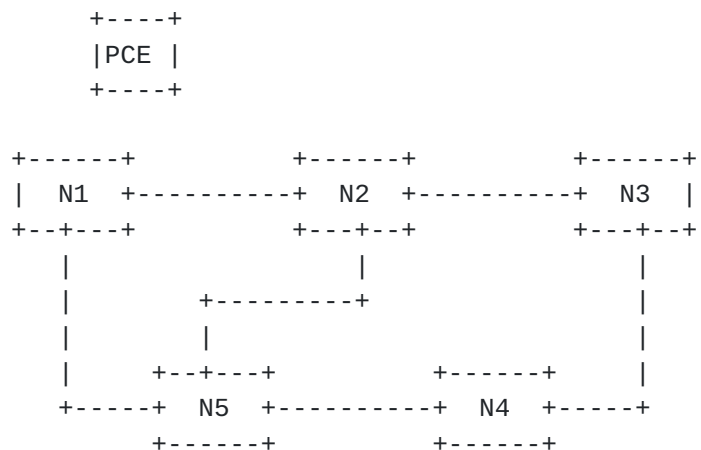


Figure 3: Reference topology 3

For example, in the network depicted in 3, suppose there exists LSP1 (N1->N5) with backup route following N1->N2->N5. A request arrives asking for a working and backup path pair to be computed for a request from N2 to N5. If the PCE decides N2->N1->N5 to be the best working route, then the backup path should not use the same protection resource with LSP1 since the new LSP shares part of its resource with LSP1 (i.e., these two LSPs are in the same shared risk group). Alternatively, there is no such constraint if N2->N3->N4->N5 is chosen to be the right candidate for undertaking the request.

5.3.2. Restoration

In case of a link failure, such as fiber cut, multiple LSPs may fail at the same time. Thus, the source nodes of the affected LSPs will be informed of the failure by the nodes detecting the failure. These source nodes will send requests to a PCE for rerouting. In order to reuse the resource taken by an existing LSP, the source node can send a PCReq message including the XRO object with F bit set, together with RRO object, as specified in [\[RFC5521\]](#).

If a stateless PCE is exploited, it might respond to the rerouting requests separately if they arrive at different times. Thus, it might result in sub-optimal resource usage. Even worse, it might unnecessarily block some of the rerouting requests due to insufficient resources for later-arrived rerouting messages. If a stateful PCE is used to fulfill this task, it can re-compute the affected LSPs concurrently while reusing part of the existing LSPs resources when it is informed of the failed link identifier provided by the first request. This is made possible since the stateful PCE can check what other LSPs are affected by the failed link and their route information by inspecting its LSP-DB. As a result, a better

performance, such as better resource usage, minimal probability of blocking upcoming new rerouting requests sent as a result of the link failure, can be achieved.

In order to further reduce the amount of LSP rerouting messages flow in the network, the notification can be performed at the node(s) which detect the link failure. For example, suppose there are two LSPs in the network as shown in Figure 3: (i) LSP1: N1->N5->N4->N3; (ii) LSP2: N2->N5->N4. They traverse the failed link between N5-N4. When N4 detects the failure, it can send a notification message to a stateful PCE. Note that the stateful PCE stores the path information of the LSPs that are affected by the link failure, so it does not need to acquire this information from N4. Moreover, it can make use of the bandwidth resources occupied by the affected LSPs when performing path recalculation. After N4 receives the new paths from the PCE, it notifies the ingress nodes of the LSPs, i.e., N1 and N2, and specifies the new paths which should be used as the rerouting paths. To support this, it would require extensions to existing signaling protocol.

Alternatively, if the target is to avoid resource contention within the time-window of high LSP requests, a stateful PCE can retain the under-construction LSP resource usage information for a given time and exclude it from being used for forthcoming LSPs request. In this way, it can ensure that the resource will not be double-booked and thus the issue of resource contention and computation crank-backs can be resolved.

5.3.3. SRLG Diversity

An alternative way to achieve efficient recovery is to maintain SRLG disjointness between LSPs. This can be achieved at provisioning time, if the routes of all the LSPs are requested together, using a synchronized computation of the different LSPs with SRLG disjointness constraint. If the LSPs need to be provisioned at different times (more general, the routes are requested at different times, e.g. in the case of a restoration), the PCC can specify, as constraints to the path computation a set of Shared Risk Link Groups (SRLGs) using the Explicit route Object [[RFC5521](#)]. However, for the latter to be effective, it is needed that the entity that requests the route to the PCE maintains updated SRLG information of all the LSPs to which it must maintain the disjointness.

Using a stateful PCE allows the maintenance of the updated SRLG information of the established LSPs in a centralized manner. Having such information in the PCE facilitates the PCC to specify, as constraint to the path computation, the SRLG disjointness of a set of already established LSPs by only providing the LSP identifiers.

5.4. Maintenance of Virtual Network Topology (VNT)

In Multi-Layer Networks (MLN), a Virtual Network Topology (VNT) [[RFC5212](#)] consists of a set of one or more TE LSPs in the lower layer which provides TE links to the upper layer. In [[RFC5623](#)], the PCE-based architecture is proposed to support path computation in MLN networks in order to achieve inter-layer TE.

The establishment/teardown of a TE link in VNT needs to take into consideration the state of existing LSPs and/or new LSP request(s) in the higher layer. As specified in [[RFC5623](#)], a VNT manager (VNTM) is in charge of the topology in the upper layer by connections in the lower layer. Hence, when a stateless PCE is requested to compute a new TE link, it will need interaction with VNTM for detailed TE link information. To be more specific, without detailed LSP information, this process would be inefficient or even infeasible for stateless PCE(s), unless with cooperation with VNTM. On the other hand, a stateful PCE seems more suitable to make the decision of when and how to modify the VNT either to accommodate new LSP requests or to re-optimize resource use across layers irrespective of PCE models. As described in [Section 2.2](#), path computation for a VNT change can be performed by the PCE if a single PCE model is adopted. On the other hand, if a per-layer PCE model is more appropriate, coordination between PCEs is required.

5.5. LSP Re-optimization

In order to make efficient usage of network resource, re-optimization of one or more LSPs dynamically through online planning is desirable. In case of a stateless PCE, in order to optimize network resource usage dynamically through online planning, PCC (e.g., NMS) should send a request to PCE together with detailed path/bandwidth information of the LSPs that need to be concurrently optimized. This would require a PCC (e.g., NMS) to determine when and which LSPs should be optimized. Given all of the existing LSP state information kept at a stateful PCE, it allows automation of this process without the PCC (e.g., NMS) to supply give the re-optimization commands and the existing LSP state information. Moreover, since a stateful PCE can maintain the information regarding to all LSPs that are currently under signaling, it makes the optimization procedures be performed more intelligently and effectively.

A special case of LSP re-optimization is Global Concurrent Optimization (GCO) [[RFC5557](#)]. 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:

- o Roll the NMS visibility into the PCE and remove the requirement for an additional northbound interface
- o Allow the PCE to determine when re-optimization is needed, with which level (GCO or a more incremental optimization)
- o Allow the PCE to determine which LSPs should be re-optimized
- o Allow a PCE to control the sequence of events across multiple PCCs, allowing for bulk (and truly global) optimization, LSP shuffling etc.

5.6. Resource defragmentation

In networks with link bundles, if LSPs are dynamically allocated and released over time, the resource becomes fragmented. The overall available resource on a (bundle) link might be sufficient for a new LSP request. But if the available resource is not continuous, the request would be rejected. In order to perform the defragmentation procedure, stateful PCEs can be used, since existing TE LSPs information is required to accurately assess spectrum resources on the LSPs, and perform de-fragmentation while ensuring a minimal disruption of the network, e.g., based on active LSP priorities .

A case of particular interest to GMPLS-based transport networks is the frequency defragmentation in flexible grid. In Flexible grid networks [[I-D.ogrcetal-ccamp-flexi-grid-fwk](#)], LSPs with different slot widths (such as 12.5G, 25G etc.) can co-exist so as to accommodate the services with different bandwidth requests. Therefore, even if the overall spectrum can meet the service request, it may not be usable if they are not contiguous. Thus, with the help of existing LSP state information, stateful PCE can make the resource grouped together to be usable. Moreover, stateful PCE can proactively choose routes for upcoming path requests to reduce the chance of spectrum defragmentation.

5.7. Future applications

5.7.1. Impairment-Aware Routing and Wavelength Assignment (IA-RWA)

In WSON networks [[RFC6163](#)], a wavelength-switched LSP traverses one or more fiber links. The bit rates of the client signals carried by the wavelength LSPs may be the same or different. Hence, a fiber link may transmit a number of wavelength LSPs with equal or mixed bit rate signals. For example, a fiber link may multiplex the

wavelengths with only 10G signals, mixed 10G and 40G signals, or mixed 40G and 100G signals.

IA-RWA in WSONs refers to the RWA process (i.e., lightpath computation) that takes into account the optical layer/transmission imperfections by considering as additional (i.e., physical layer) constraints. To be more specific, linear and non-linear effects associated with the optical network elements should be incorporated into the route and wavelength assignment procedure. For example, the physical imperfection can result in the interference of two adjacent lightpaths. Thus, a guard band should be reserved between them to alleviate these effects. The width of the guard band between two adjacent wavelengths depends on their characteristics, such as modulation formats and bit rates. Two adjacent wavelengths with different characteristics (e.g., different bit rates) may need a wider guard band and with same characteristics may need a narrower guard band. For example, 50GHz spacing may be acceptable for two adjacent wavelengths with 40G signals. But for two adjacent wavelengths with different bit rates (e.g., 10G and 40G), a larger spacing such as 300GHz spacing may be needed. Hence, the characteristics (states) of the existing wavelength LSPs should be considered for a new RWA request in WSON.

In summary, when stateful PCEs are used to perform the IA-RWA procedure, they need to know the characteristics of the existing wavelength LSPs. The impairment information relating to existing and to-be-established LSPs can be obtained by nodes in WSON networks via external configuration or other means such as monitoring or estimation based on a vendor-specific impair model. However, WSON related routing protocols, i.e., [[I-D.ietf-ccamp-wson-signal-compatibility-ospf](#)] and [[I-D.ietf-ccamp-gmpls-general-constraints-ospf-te](#)], only advertise limited information (i.e., availability) of the existing wavelengths, without defining the supported client bit rates. It will incur substantial amount of control plane overhead if routing protocols are extended to support dissemination of the new information relevant for the IA-RWA process. In this scenario, stateful PCE(s) would be a more appropriate mechanism to solve this problem. Stateful PCE(s) can exploit impairment information of LSPs stored in LSP-DB to provide accurate RWA calculation.

6. Security Considerations

This document does not introduce any new security considerations.

7. Contributing authors

The following people all contributed significantly to this document and are listed below in alphabetical order:

Ramon Casellas
CTTC - Centre Tecnologic de Telecomunicacions de Catalunya
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain
Email: ramon.casellas@cttc.es

Edward Crabbe
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US
Email: edc@google.com

Dhruv Dhody
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA
Email: dhruvd@huawei.com

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain
Phone: +34 913374013
Email: ogondio@tid.es

Young Lee
Huawei
1700 Alma Drive, Suite 100
Plano, TX 75075
US
Phone: +1 972 509 5599 x2240
Fax: +1 469 229 5397
Email: ylee@huawei.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

Fatai Zhang
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Xiaobing Zi
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China
Phone: +86-755-28973229
Email: zixiaobing@huawei.com

8. Acknowledgements

We would like to thank Cyril Margaria for the useful comments and discussions.

9. References

9.1. Normative References

- [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.

- [RFC4427] Mannie, E. and D. Papadimitriou, "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", [RFC 4427](#), March 2006.
- [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.
- [RFC5212] Shiomoto, K., Papadimitriou, D., Le Roux, JL., Vigoureux, M., and D. Brungard, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", [RFC 5212](#), July 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.
- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", [RFC 5521](#), April 2009.
- [RFC5623] Oki, E., Takeda, T., Le Roux, JL., and A. Farrel, "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", [RFC 5623](#), September 2009.
- [RFC6163] Lee, Y., Bernstein, G., and W. Imajuku, "Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSOs)", [RFC 6163](#), April 2011.

9.2. Informative References

- [I-D.crabbe-pce-stateful-pce-mpls-te]
Crabbe, E., Medved, J., Minei, I., and R. Varga, "Stateful PCE extensions for MPLS-TE LSPs",
[draft-crabbe-pce-stateful-pce-mpls-te-00](#) (work in

progress), October 2012.

[I-D.ietf-ccamp-gmpls-general-constraints-ospf-te]

Zhang, F., Lee, Y., Han, J., Bernstein, G., and Y. Xu,
"OSPF-TE Extensions for General Network Element
Constraints",
[draft-ietf-ccamp-gmpls-general-constraints-ospf-te-04](#)
(work in progress), July 2012.

[I-D.ietf-ccamp-wson-signal-compatibility-ospf]

Lee, Y. and G. Bernstein, "GMPLS OSPF Enhancement for
Signal and Network Element Compatibility for Wavelength
Switched Optical Networks",
[draft-ietf-ccamp-wson-signal-compatibility-ospf-11](#) (work
in progress), February 2013.

[I-D.ietf-pce-gmpls-pcep-extensions]

Margarita, C., Dios, O., and F. Zhang, "PCEP extensions for
GMPLS", [draft-ietf-pce-gmpls-pcep-extensions-07](#) (work in
progress), October 2012.

[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.

[I-D.ogrcetal-ccamp-flexi-grid-fwk]

Dios, O., Casellas, R., Zhang, F., Fu, X., Ceccarelli, D.,
and I. Hussain, "Framework for GMPLS based control of
Flexi-grid DWDM networks",
[draft-ogrcetal-ccamp-flexi-grid-fwk-01](#) (work in progress),
October 2012.

[I-D.sivabalan-pce-disco-stateful]

Sivabalan, S. and J. Medved, "IGP Extensions for Stateful
PCE Discovery", [draft-sivabalan-pce-disco-stateful-00](#)
(work in progress), January 2013.

[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", 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.

Appendix A. Editorial notes and open issues

This section will be removed prior to publication.

The following open issues remain:

Use cases from [draft-ietf-pce-stateful-pce](#) To avoid loss of information, the use cases will be removed from [\[I-D.ietf-pce-stateful-pce\]](#) only after this document becomes a working group document.

This document WILL NOT repeat terminology defined in other documents or attempt to place any additional requirements on stateful PCE.

Authors' Addresses

Xian Zhang (editor)
Huawei Technologies
F3-5-B R&D Center, Huawei Base Bantian, Longgang District
Shenzhen, Guangdong 518129
P.R.China

Email: zhang.xian@huawei.com

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

Email: ina@juniper.net

