

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 9, 2012

D. Dhody
Huawei Technologies India Pvt
Ltd
V. Manral
Hewlett-Packard Corp.
June 7, 2012

Extensions to the Path Computation Element Communication Protocol (PCEP)
to compute service aware Label Switched Path (LSP).
draft-dhody-pce-pcep-service-aware-03

Abstract

In certain networks like financial information network (stock/commodity trading) and enterprises using cloud based applications, Latency (delay), Latency-Variation (jitter) and Packet loss is becoming a key requirement for path computation along with other constraints and metrics. Latency, Latency-Variation and Packet Loss is associated with the Service Level Agreement (SLA) between customers and service providers.

[MPLS-DELAY-FWK] describes MPLS architecture to allow Latency (delay), Latency-Variation (jitter) and Packet loss as properties. [OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] describes mechanisms with which network performance information is distributed via OSPF and ISIS respectively. This document describes the extension to PCEP to carry Latency, Latency-Variation and Loss as constraints for end to end path computation.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

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

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

This Internet-Draft will expire on December 9, 2012.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Requirements Language	3
2. Terminology	3
3. Requirement for PCEP	4
4. Objects	4
4.1. Latency (Delay) Metric	5
4.1.1. Latency (Delay) Metric Value	6
4.2. Latency Variation (Jitter) Metric	6
4.2.1. Latency Variation (Jitter) Metric Value	6
4.3. Packet Loss Metric	7
4.3.1. Packet Loss Metric Value	7
4.4. Non-Understanding / Non-Support of Service Aware Path Computation	8
4.5. Mode of Operation	8
4.5.1. Examples	8
5. Protocol Consideration	9
5.1. Inter domain Consideration	9
5.1.1. Inter-AS Link	9
5.1.2. Inter-Layer Consideration	9
5.2. Reoptimization Consideration	10
5.3. P2MP	10
6. IANA Considerations	10
7. Security Considerations	10
8. Manageability Considerations	10
8.1. Control of Function and Policy	10
8.2. Information and Data Models	10
8.3. Liveness Detection and Monitoring	10
8.4. Verify Correct Operations	11
8.5. Requirements On Other Protocols	11
8.6. Impact On Network Operations	11
9. References	11
9.1. Normative References	11
9.2. Informative References	11

1. Introduction

Real time Network Performance is becoming a critical in the path computation in some networks. There exist mechanism described in [RFC6374] to measure latency, latency-Variation and packet loss after the LSP has been established, which is inefficient. It is important that latency, latency-variation and packet loss are considered during path selection process, even before the LSP is setup.

TED is populated with network performance information like link latency, latency variation and packet loss through [OSPF-TE-EXPRESS] or [ISIS-TE-EXPRESS]. Path Computation Client (PCC) can request Path Computation Element (PCE) to provide a path meeting end to end network performance criteria. This document extends Path Computation Element Communication Protocol (PCEP) [RFC5440] to handle network performance constraint.

PCE MAY use mechanism described in [MPLS-TE-EXPRESS-PATH] on how to use the link latency, latency variation and packet loss information for end to end path selection.

1.1. Requirements Language

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

2. Terminology

The following terminology is used in this document.

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

IS-IS: Intermediate System to Intermediate System.

OSPF: Open Shortest Path First.

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

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

TE: Traffic Engineering.

3. Requirement for PCEP

End-to-end service optimization based on latency, latency-variation and packet loss is a key requirement for service provider. Following key requirements associated with latency, latency-variation and loss is identified for PCEP:

1. Path Computation Element (PCE) supporting this draft MUST have the capability to compute end-to-end path with latency, latency-variation and packet loss constraints. It MUST also support the combination of network performance constraint (latency, latency-variation, loss...) with existing constraints (cost, hop-limit...)
2. Path Computation Client (PCC) supporting this draft MUST be able to request for network performance constraint in path request message as the key constraint to be optimized or to suggest boundary condition that should not be crossed.
3. PCEs are not required to support service aware path computation. Therefore, it MUST be possible for a PCE to reject a Path Computation Request message with a reason code that indicates no support for service-aware path computation.
4. PCEP supporting this draft SHOULD provide a means to return end to end network performance information of the computed path in the reply message.
5. PCEP supporting this draft SHOULD provide mechanism to compute multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) service aware paths.

It must be understood that such constraints are only meaningful if used consistently: for instance, if the delay of a computed path segment is exchanged between two PCEs residing in different domains, consistent ways of defining the delay must be used.

4. Objects

This section defines PCEP extensions (see [RFC5440]) so as to support network performance and service aware path computation.

[RFC5440] defines the optional METRIC object for several purposes. In a PCReq message, a PCC MAY insert one or more METRIC objects to indicate the metric that MUST be optimized or to indicate a bound on the path that MUST NOT be exceeded for the path to be considered as

acceptable by the PCC. In a PCRep message, the METRIC object MAY be inserted so as to provide the value for the computed path. It MAY also be inserted within a PCRep with the NO-PATH object to indicate that the metric constraint could not be satisfied.

As per [RFC5440] the format of the METRIC object body is as follows:

```

      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
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |                               |   Flags   |C|B|           T       |
      +-----+-----+-----+-----+-----+-----+-----+-----+
      |                               |   metric-value   |
      +-----+-----+-----+-----+-----+-----+-----+-----+

```

T (Type - 8 bits): Specifies the metric type.

Three values are currently defined:

- * T=1: IGP metric
- * T=2: TE metric
- * T=3: Hop Counts

Based on the section 3, PCEP is extended to define new METRIC types for network performance constraints.

4.1. Latency (Delay) Metric

The end to end Latency (Delay) for the path is represented by this metric.

- * T=13(IANA): Latency metric

PCC MAY use this latency metric In PCReq to request a path meeting the end to end latency requirement. In this case B bit MUST be set to suggest a bound (a maximum) for the path latency metric that must not be exceeded for the PCC to consider the computed path as acceptable. The path metric must be less than or equal to the value specified in the metric-value field.

PCC MAY also use this metric to ask PCE to optimize delay during path computation, in this case B flag will be cleared.

PCE MAY use this latency metric In PCRep along with NO-PATH object incase PCE cannot compute a path meeting this constraint. PCE MAY also use this metric to reply the computed end to end latency metric to PCC.

4.1.1.1. Latency (Delay) Metric Value

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] defines "Unidirectional Link Delay Sub-TLV" in a 24-bit field. [RFC5440] defines the METRIC object with 32-bit metric value.

The last 24-bits of the 32-bit metric value represents the end to end Latency (delay) quantified in units of microseconds and MUST be encoded as integer value. With the maximum value 16,777,215 representing 16.777215 sec.

```

  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
  +-----+-----+-----+-----+-----+-----+-----+-----+
  |      Resv      |      Latency (Delay) Metric      |
  +-----+-----+-----+-----+-----+-----+-----+-----+

```

4.2. Latency Variation (Jitter) Metric

The end to end Latency Variation (Jitter) for the path is represented by this metric.

* T=14(IANA): Latency Variation metric

PCC MAY use this latency variation metric In PCReq to request a path meeting the end to end latency variation requirement. In this case B bit MUST be set to suggest a bound (a maximum) for the path latency variation metric that must not be exceeded for the PCC to consider the computed path as acceptable. The path metric must be less than or equal to the value specified in the metric-value field.

PCC MAY also use this metric to ask PCE to optimize jitter during path computation, in this case B flag will be cleared.

PCE MAY use this latency variation metric In PCRep along with NO-PATH object incase PCE cannot compute a path meeting this constraint. PCE MAY also use this metric to reply the computed end to end latency variation metric to PCC.

4.2.1. Latency Variation (Jitter) Metric Value

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] defines "Unidirectional Delay Variation Sub-TLV" in a 24-bit field. [RFC5440] defines the METRIC object with 32-bit metric value.

The last 24-bits of the 32-bit metric value represents the end to end Latency variation (jitter) quantified in units of microseconds and MUST be encoded as integer value. With the maximum value 16,777,215

representing 16.777215 sec.

```

  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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      Resv      |      Latency variation (jitter) Metric      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

4.3. Packet Loss Metric

The end to end Packet Loss for the path is represented by this metric.

* T=15(IANA): Packet Loss metric

PCC MAY use this packet loss metric In PCReq to request a path meeting the end to end packet loss requirement. In this case B bit MUST be set to suggest a bound (a maximum) for the path packet loss metric that must not be exceeded for the PCC to consider the computed path as acceptable. The path metric must be less than or equal to the value specified in the metric-value field.

PCC MAY also use this metric to ask PCE to optimize packet loss during path computation, in this case B flag will be cleared.

PCE MAY use this packet loss metric In PCRep along with NO-PATH object incase PCE cannot compute a path meeting this constraint. PCE MAY also use this metric to reply the computed end to end packet loss metric to PCC.

4.3.1. Packet Loss Metric Value

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] defines "Unidirectional Link Loss Sub-TLV" in a 24-bit field. [RFC5440] defines the METRIC object with 32-bit metric value.

The last 24-bits of the 32-bit metric value represents the end to end packet loss quantified as a percentage of packets lost and MUST be encoded as integer. The basic unit is 0.000003%, with the maximum value 16,777,215 representing 50.331645% (16,777,215 * 0.000003%). This value is the highest packet loss percentage that can be expressed.

```

  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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      Resv      |      Packet loss Metric      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

4.4. Non-Understanding / Non-Support of Service Aware Path Computation

If the P bit is clear in the object header and PCE does not understand or does not support service aware path computation it SHOULD simply ignore this METRIC.

If the P Bit is set in the object header and PCE receives new METRIC type in path request and it understands the METRIC type, but the PCE is not capable of service aware path computation, the PCE MUST send a PCErr message with a PCEP-ERROR Object Error-Type = 4 (Not supported object) [RFC5440]. The path computation request MUST then be cancelled.

If the PCE does not understand the new METRIC type, then the PCE MUST send a PCErr message with a PCEP-ERROR Object Error-Type = 3 (Unknown object) [RFC5440].

4.5. Mode of Operation

As explained in [RFC5440], The METRIC object is optional and can be used for several purposes. In a PCReq message, a PCC MAY insert one or more METRIC objects:

- o To indicate the metric that MUST be optimized by the path computation algorithm (Latency, Latency-Variation or Loss)
- o To indicate a bound on the path METRIC (Latency, Latency-Variation or Loss) that MUST NOT be exceeded for the path to be considered as acceptable by the PCC.

In a PCRep message, the METRIC object MAY be inserted so as to provide the METRIC (Latency, Latency-Variation or Loss) for the computed path. It MAY also be inserted within a PCRep with the NO-PATH object to indicate that the metric constraint could not be satisfied.

The path computation algorithmic aspects used by the PCE to optimize a path with respect to a specific metric are outside the scope of this document.

All the rules of processing METRIC object as explained in [RFC5440] are applicable to the new metric types as well.

4.5.1. Examples

If a PCC sends a path computation request to a PCE where the metric to optimize is the latency and the packet loss must not exceed the value of M, two METRIC objects are inserted in the PCReq message:

- o First METRIC object with B=0, T=13 (TBA - IANA), C=1, metric-value=0x0000
- o Second METRIC object with B=1, T=15 (TBA - IANA), metric-value=M

If a path satisfying the set of constraints can be found by the PCE and there is no policy that prevents the return of the computed metric, the PCE inserts one METRIC object with B=0, T=13 (TBA - IANA), metric-value= computed end to end latency. Additionally, the PCE may insert a second METRIC object with B=1, T=15 (TBA - IANA), metric-value= computed end to end packet loss.

5. Protocol Consideration

There is no change in the message format of Path Request and Reply Message.

5.1. Inter domain Consideration

[RFC5441] describes the BRPC procedure to compute end to end optimized inter domain path by cooperating PCEs. The network performance constraints can be applied end to end in similar manner as IGP or TE cost.

All domains should have the same understanding of the METRIC (Latency-Variation etc) for end-to-end inter-domain path computation to make sense. Otherwise some form of Metric Normalization as described in [RFC5441] MAY need to be applied.

5.1.1. Inter-AS Link

The IGP in each neighbor domain can advertise its inter-domain TE link capabilities, this has been described in [RFC5316] (ISIS) and [RFC5392] (OSPF). The network performance link properties are described in [OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS], the same properties must be advertised using the mechanism described in [RFC5392] (OSPF) and [RFC5316] (ISIS).

5.1.2. Inter-Layer Consideration

PCEP supporting this draft SHOULD provide mechanism to support different Metric requirements for different Layers. This is important as the network performance metric would be different for Packet and Optical (TDM, LSC etc) Layers. In order to allow different Metric-Value to be applied within different network layers, multiple METRIC objects of the same type MAY be present. In such a case, the first METRIC object specifies an metric for the higher-layer network, and subsequent METRIC objects specify objection

functions of the subsequent lower-layer networks.

5.2. Reoptimization Consideration

PCC can monitor the setup LSPs and incase of degradation of network performance constraints, it MAY ask PCE for reoptimization as per [RFC5440].

5.3. P2MP

The scope of the network performance constraints as listed in this document needs to be described in terms of P2MP TE LSPs.

This section needs further discussions.

6. IANA Considerations

IANA has defined a registry for new METRIC type.

Type	Meaning
13(TBD)	Latency (delay) metric
14(TBD)	Latency Variation (jitter) metric
15(TBD)	Packet Loss metric

7. Security Considerations

This document defines three new METRIC Types which does not add any new security concerns to PCEP protocol.

8. Manageability Considerations

8.1. Control of Function and Policy

The only configurable item is the support of the new service-aware METRICS on a PCE which MAY be controlled by a policy module. If the new METRIC is not supported/allowed on a PCE, it MUST send a PCerr message as specified in Section 4.4.

8.2. Information and Data Models

[PCEP-MIB] describes the PCEP MIB, there are no new MIB Objects for this document.

8.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

8.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440].

8.5. Requirements On Other Protocols

PCE requires the TED to be populated with network performance information like link latency, latency variation and packet loss. This mechanism is described in [OSPF-TE-EXPRESS] or [ISIS-TE-EXPRESS].

8.6. Impact On Network Operations

Mechanisms defined in this document do not have any impact on network operations in addition to those already listed in [RFC5440].

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

9.2. Informative References

- [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.
- [RFC5441] Vasseur, JP., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.
- [RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS

- Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, December 2008.
- [RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, January 2009.
- [RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, September 2011.
- [MPLS-DELAY-FWK] Fu, X., Manral, V., McDysan, D., Malis, A., Giacalone, S., Betts, M., Wang, Q., and J. Drake, "Traffic Engineering architecture for services aware MPLS [draft-fuxh-mpls-delay-loss-te-framework]", Apr 2012.
- [OSPF-TE-EXPRESS] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions [draft-ietf-ospf-te-metric-extensions]", May 2012.
- [ISIS-TE-EXPRESS] Previdi, S., Giacalone, S., Ward, D., Drake, J., Atlas, A., and C. Filsfils, "IS-IS Traffic Engineering (TE) Metric Extensions [draft-previdi-isis-te-metric-extensions]", Mar 2012.
- [MPLS-TE-EXPRESS-PATH] Atlas, A., Drake, J., Ward, D., Giacalone, S., Previdi, S., and C. Filsfils, "Performance-based Path Selection for Explicitly Routed LSPs [draft-atlas-mpls-te-express-path]", Oct 2011.
- [PCEP-MIB] Kiran Koushik, A S., Stephan, E., Zhao, Q., and D. King, "PCE communication protocol(PCEP) Management Information Base [draft-ietf-pce-pcep-mib]", July 2010.

Authors' Addresses

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

EMail: dhruv.dhody@huawei.com

Vishwas Manral
Hewlett-Packard Corp.
191111 Pruneridge Ave.
Cupertino, CA 95014
USA

EMail: vishwas.manral@hp.com

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 24, 2013

D. Dhody
Huawei Technologies India Pvt
Ltd
V. Manral
Hewlett-Packard Corp.
Z. Ali
G. Swallow
Cisco Systems
K. Kumaki
KDDI Corporation
February 25, 2013

Extensions to the Path Computation Element Communication Protocol (PCEP)
to compute service aware Label Switched Path (LSP).
draft-dhody-pce-pcep-service-aware-05

Abstract

In certain networks like financial information network (stock/commodity trading) and enterprises using cloud based applications, Latency (delay), Latency-Variation (jitter) and Packet loss is becoming a key requirement for path computation along with other constraints and metrics. Latency, Latency-Variation and Packet Loss is associated with the Service Level Agreement (SLA) between customers and service providers.

[MPLS-DELAY-FWK] describes MPLS architecture to allow Latency (delay), Latency-Variation (jitter) and Packet loss as properties. [OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] describes mechanisms with which network performance information is distributed via OSPF and ISIS respectively. This document describes the extension to PCEP to carry Latency, Latency-Variation and Loss as constraints for end to end path computation.

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 June 4, 2013.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	4
1.1. Requirements Language	4
2. Terminology	4
3. PCEP Requirements	5
4. PCEP extensions	5
4.1. Latency (Delay) Metric	6
4.1.1. Latency (Delay) Metric Value	6
4.2. Latency Variation (Jitter) Metric	7
4.2.1. Latency Variation (Jitter) Metric Value	7
4.3. Packet Loss Metric	8
4.3.1. Packet Loss Metric Value	9
4.4. Non-Understanding / Non-Support of Service Aware Path Computation	9
4.5. Mode of Operation	9
4.5.1. Examples	10
5. Relationship with Objective function	11
6. Protocol Consideration	11
6.1. Inter domain Consideration	11
6.1.1. Inter-AS Link	12
6.1.2. Inter-Layer Consideration	12
6.2. Reoptimization Consideration	12
6.3. Point-to-Multipoint (P2MP)	12
6.3.1. P2MP Latency Metric	12
6.3.2. P2MP Latency Variation Metric	13
7. IANA Considerations	13
8. Security Considerations	13
9. Manageability Considerations	14
9.1. Control of Function and Policy	14
9.2. Information and Data Models	14
9.3. Liveness Detection and Monitoring	14
9.4. Verify Correct Operations	14
9.5. Requirements On Other Protocols	14
9.6. Impact On Network Operations	14
10. Acknowledgments	14
11. References	15
11.1. Normative References	15
11.2. Informative References	15
Appendix A. Contributor Addresses	16

1. Introduction

Real time Network Performance is becoming a critical in the path computation in some networks. There exist mechanism described in [RFC6374] to measure latency, latency-Variation and packet loss after the LSP has been established, which is inefficient. It is important that latency, latency-variation and packet loss are considered during path selection process, even before the LSP is setup.

TED is populated with network performance information like link latency, latency variation and packet loss through [OSPF-TE-EXPRESS] or [ISIS-TE-EXPRESS]. Path Computation Client (PCC) can request Path Computation Element (PCE) to provide a path meeting end to end network performance criteria. This document extends Path Computation Element Communication Protocol (PCEP) [RFC5440] to handle network performance constraint.

PCE MAY use mechanism described in [MPLS-TE-EXPRESS-PATH] on how to use the link latency, latency variation and packet loss information for end to end path selection.

1.1. Requirements Language

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

2. Terminology

The following terminology is used in this document.

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

IS-IS: Intermediate System to Intermediate System.

OSPF: Open Shortest Path First.

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

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

TE: Traffic Engineering.

3. PCEP Requirements

End-to-end service optimization based on latency, latency-variation and packet loss is a key requirement for service provider. Following key requirements associated with latency, latency-variation and loss are identified for PCEP:

1. Path Computation Element (PCE) supporting this draft MUST have the capability to compute end-to-end path with latency, latency-variation and packet loss constraints. It MUST also support the combination of network performance constraint (latency, latency-variation, loss...) with existing constraints (cost, hop-limit...)
2. Path Computation Client (PCC) MUST be able to request for network performance constraint in path request message as the key constraint to be optimized or to suggest boundary condition that should not be crossed.
3. PCEs are not required to support service aware path computation. Therefore, it MUST be possible for a PCE to reject a Path Computation Request message with a reason code that indicates no support for service-aware path computation.
4. PCEP SHOULD provide a means to return end to end network performance information of the computed path in the reply message.
5. PCEP SHOULD provide mechanism to compute multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) service aware paths.

It is assumed that such constraints are only meaningful if used consistently: for instance, if the delay of a computed path segment is exchanged between two PCEs residing in different domains, consistent ways of defining the delay must be used.

4. PCEP extensions

This section defines PCEP extensions (see [RFC5440]) for requirements outlined in Section 3. The proposed solution is used to support network performance and service aware path computation.

This document defines the following optional types for the METRIC object defined in [RFC5440].

For explanation of these metrics, the following terminology is used

and expanded along the way.

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

4.1. Latency (Delay) Metric

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

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

* T=13(IANA): Latency metric

PCC MAY use this latency metric In PCReq to request a path meeting the end to end latency requirement. In this case B bit MUST be set to suggest a bound (a maximum) for the path latency metric that must not be exceeded for the PCC to consider the computed path as acceptable. The path metric must be less than or equal to the value specified in the metric-value field.

PCC MAY also use this metric to ask PCE to optimize delay during path computation, in this case B flag will be cleared.

PCE MAY use this latency metric In PCRep along with NO-PATH object incase PCE cannot compute a path meeting this constraint. PCE MAY also use this metric to reply the computed end to end latency metric to PCC.

4.1.1. Latency (Delay) Metric Value

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] defines "Unidirectional Link Delay Sub-TLV" in a 24-bit field. [RFC5440] defines the METRIC object with 32-bit metric value. Consequently, encoding for Latency (Delay) Metric Value is defined as follows:

```

  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
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  | Reserved           |                               Latency (Delay) Metric |
  +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

```

Reserved (8 bits): Reserved field. This field MUST be set to zero on

transmission and MUST be ignored on receipt.

Latency (Delay) Metric (24 bits): Represents the end to end Latency (delay) quantified in units of microseconds and MUST be encoded as integer value. With the maximum value 16,777,215 representing 16.777215 sec.

4.2. Latency Variation (Jitter) Metric

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

- A Latency variation of link L is denoted DV(L).
- A P2P latency variation metric for the Path P = function {DV(L_pi), (i=1...K)}.

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

* T=14(IANA): Latency Variation metric

PCC MAY use this latency variation metric In PCReq to request a path meeting the end to end latency variation requirement. In this case B bit MUST be set to suggest a bound (a maximum) for the path latency variation metric that must not be exceeded for the PCC to consider the computed path as acceptable. The path metric must be less than or equal to the value specified in the metric-value field.

PCC MAY also use this metric to ask PCE to optimize jitter during path computation, in this case B flag will be cleared.

PCE MAY use this latency variation metric In PCRep along with NO-PATH object incase PCE cannot compute a path meeting this constraint. PCE MAY also use this metric to reply the computed end to end latency variation metric to PCC.

4.2.1. Latency Variation (Jitter) Metric Value

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] defines "Unidirectional Delay Variation Sub-TLV" in a 24-bit field. [RFC5440] defines the METRIC object with 32-bit metric value. Consequently, encoding for Latency Variation (Jitter) Metric Value is defined as follows:

```

  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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Reserved   |   Latency variation (jitter) Metric   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Reserved (8 bits): Reserved field. This field MUST be set to zero on transmission and MUST be ignored on receipt.

Latency variation (jitter) Metric (24 bits): Represents the end to end Latency variation (jitter) quantified in units of microseconds and MUST be encoded as integer value. With the maximum value 16,777,215 representing 16.777215 sec.

4.3. Packet Loss Metric

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] defines "Unidirectional Link Loss". Packet Loss Metric metric type of METRIC object in PCEP encodes a function of the link's unidirectional loss metric of all links along a P2P Path. Specifically, extending on the above mentioned terminology:

The end to end Packet Loss for the path is represented by this metric.

- A Packet loss of link L is denoted PL(L).

- A P2P packet loss metric for the Path P = function {PL(L_{pi}), (i=1...K)}.

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

* T=15(IANA): Packet Loss metric

PCC MAY use this packet loss metric In PCReq to request a path meeting the end to end packet loss requirement. In this case B bit MUST be set to suggest a bound (a maximum) for the path packet loss metric that must not be exceeded for the PCC to consider the computed path as acceptable. The path metric must be less than or equal to the value specified in the metric-value field.

PCC MAY also use this metric to ask PCE to optimize packet loss during path computation, in this case B flag will be cleared.

PCE MAY use this packet loss metric In PCRep along with NO-PATH object incase PCE cannot compute a path meeting this constraint. PCE

MAY also use this metric to reply the computed end to end packet loss metric to PCC.

4.3.1. Packet Loss Metric Value

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] defines "Unidirectional Link Loss Sub-TLV" in a 24-bit field. [RFC5440] defines the METRIC object with 32-bit metric value. Consequently, encoding for Packet Loss Metric Value is defined as follows:

```

  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
  +-----+-----+-----+-----+-----+-----+-----+-----+
  |  Reserved      |                               Packet loss Metric |
  +-----+-----+-----+-----+-----+-----+-----+-----+

```

Reserved (8 bits): Reserved field. This field MUST be set to zero on transmission and MUST be ignored on receipt.

Packet loss Metric (24 bits): Represents the end to end packet loss quantified as a percentage of packets lost and MUST be encoded as integer. The basic unit is 0.000003%, with the maximum value 16,777,215 representing 50.331645% ($16,777,215 * 0.000003\%$). This value is the highest packet loss percentage that can be expressed.

4.4. Non-Understanding / Non-Support of Service Aware Path Computation

If the P bit is clear in the object header and PCE does not understand or does not support service aware path computation it SHOULD simply ignore this METRIC.

If the P Bit is set in the object header and PCE receives new METRIC type in path request and it understands the METRIC type, but the PCE is not capable of service aware path computation, the PCE MUST send a PCErr message with a PCEP-ERROR Object Error-Type = 4 (Not supported object) [RFC5440]. The path computation request MUST then be cancelled.

If the PCE does not understand the new METRIC type, then the PCE MUST send a PCErr message with a PCEP-ERROR Object Error-Type = 3 (Unknown object) [RFC5440].

4.5. Mode of Operation

As explained in [RFC5440], The METRIC object is optional and can be used for several purposes. In a PCReq message, a PCC MAY insert one or more METRIC objects:

- o To indicate the metric that MUST be optimized by the path computation algorithm (Latency, Latency-Variation or Loss)
- o To indicate a bound on the path METRIC (Latency, Latency-Variation or Loss) that MUST NOT be exceeded for the path to be considered as acceptable by the PCC.

In a PCRep message, the METRIC object MAY be inserted so as to provide the METRIC (Latency, Latency-Variation or Loss) for the computed path. It MAY also be inserted within a PCRep with the NO-PATH object to indicate that the metric constraint could not be satisfied.

The path computation algorithmic aspects used by the PCE to optimize a path with respect to a specific metric are outside the scope of this document.

All the rules of processing METRIC object as explained in [RFC5440] are applicable to the new metric types as well.

In a PCReq message, a PCC MAY insert more than one METRIC object to be optimized, in such a case PCE should find the path that is optimal when both the metrics are considered together.

4.5.1. Examples

Example 1: If a PCC sends a path computation request to a PCE where two metric to optimize are the latency and the packet loss, two METRIC objects are inserted in the PCReq message:

- o First METRIC object with B=0, T=13 (TBA - IANA), C=1, metric-value=0x0000
- o Second METRIC object with B=0, T=15 (TBA - IANA), C=1, metric-value=0x0000

PCE in such a case should try to optimize both the metrics and find a path with the minimum latency and packet loss, if a path can be found by the PCE and there is no policy that prevents the return of the computed metric, the PCE inserts two METRIC object with B=0, T=13 (TBA - IANA), metric-value= computed end to end latency and second METRIC object with B=1, T=15 (TBA - IANA), metric-value= computed end to end packet loss.

Example 2: If a PCC sends a path computation request to a PCE where the metric to optimize is the latency and the packet loss must not exceed the value of M, two METRIC objects are inserted in the PCReq message:

- o First METRIC object with B=0, T=13 (TBA - IANA), C=1, metric-value=0x0000
- o Second METRIC object with B=1, T=15 (TBA - IANA), metric-value=M

If a path satisfying the set of constraints can be found by the PCE and there is no policy that prevents the return of the computed metric, the PCE inserts one METRIC object with B=0, T=13 (TBA - IANA), metric-value= computed end to end latency. Additionally, the PCE may insert a second METRIC object with B=1, T=15 (TBA - IANA), metric-value= computed end to end packet loss.

5. Relationship with Objective function

[RFC5541] defines mechanism to specify an optimization criteria, referred to as objective functions. The new metric types specified in this document can continue to use the existing Objective function.

Minimum Cost Path (MCP) is one such objective function.

- o A network comprises a set of N links $\{L_i, (i=1\dots N)\}$.
- o A path P is a list of K links $\{L_{pi}, (i=1\dots K)\}$.
- o Metric of link L is denoted M(L). This can be any metric, including the ones defined in this document.
- o The cost of a path P is denoted C(P), where $C(P) = \text{sum} \{M(L_{pi}), (i=1\dots K)\}$.

Name: Minimum Cost Path (MCP)

Description: Find a path P such that C(P) is minimized.

The new metric types for example latency (delay) can continue to use the above objective function to find the minimum cost path where cost is latency (delay). At the same time new objective functions can be defined in future to optimize these new metric types.

6. Protocol Consideration

There is no change in the message format of Path Request and Reply Message.

6.1. Inter domain Consideration

[RFC5441] describes the BRPC procedure to compute end to end optimized inter domain path by cooperating PCEs. The network

performance constraints can be applied end to end in similar manner as IGP or TE cost.

All domains should have the same understanding of the METRIC (Latency-Variation etc) for end-to-end inter-domain path computation to make sense. Otherwise some form of Metric Normalization as described in [RFC5441] MAY need to be applied.

6.1.1. Inter-AS Link

The IGP in each neighbor domain can advertise its inter-domain TE link capabilities, this has been described in [RFC5316] (ISIS) and [RFC5392] (OSPF). The network performance link properties are described in [OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS], the same properties must be advertised using the mechanism described in [RFC5392] (OSPF) and [RFC5316] (ISIS).

6.1.2. Inter-Layer Consideration

PCEP supporting this draft SHOULD provide mechanism to support different Metric requirements for different Layers. This is important as the network performance metric would be different for Packet and Optical (TDM, LSC etc) Layers. In order to allow different Metric-Value to be applied within different network layers, multiple METRIC objects of the same type MAY be present. In such a case, the first METRIC object specifies a metric for the higher-layer network, and subsequent METRIC objects specify objection functions of the subsequent lower-layer networks.

6.2. Reoptimization Consideration

PCC can monitor the setup LSPs and in case of degradation of network performance constraints, it MAY ask PCE for reoptimization as per [RFC5440].

6.3. Point-to-Multipoint (P2MP)

This document defines the following optional types for the METRIC object defined in [RFC5440] for P2MP TE LSPs. Additional metric types for P2MP TE LSPs are to be added in a future revision

6.3.1. P2MP Latency Metric

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

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

Value for P2MP latency metric is to be assigned by IANA

6.3.2. P2MP Latency Variation Metric

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

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

Value for P2MP latency variation metric is to be assigned by IANA

7. IANA Considerations

IANA has defined a registry for new METRIC type.

Type	Meaning
13(TBD)	Latency (delay) metric
14(TBD)	Latency Variation (jitter) metric
15(TBD)	Packet Loss metric
16(TBD)	P2MP latency metric
17(TBD)	P2MP latency variation metric

8. Security Considerations

This document defines three new METRIC Types which does not add any new security concerns to PCEP protocol.

9. Manageability Considerations

9.1. Control of Function and Policy

The only configurable item is the support of the new service-aware METRICS on a PCE which MAY be controlled by a policy module. If the new METRIC is not supported/allowed on a PCE, it MUST send a PCErr message as specified in Section 4.4.

9.2. Information and Data Models

[PCEP-MIB] describes the PCEP MIB, there are no new MIB Objects for this document.

9.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

9.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440].

9.5. Requirements On Other Protocols

PCE requires the TED to be populated with network performance information like link latency, latency variation and packet loss. This mechanism is described in [OSPF-TE-EXPRESS] or [ISIS-TE-EXPRESS].

9.6. Impact On Network Operations

Mechanisms defined in this document do not have any impact on network operations in addition to those already listed in [RFC5440].

10. Acknowledgments

We would like to thank Young Lee, Venugopal Reddy, Reerja Paul, Sandeep Kumar Boina, Suresh babu, Quintin Zhao and Chen Huaimo for their useful comments and suggestions.

11. References

11.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

11.2. Informative References

- [RFC5441] Vasseur, JP., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.
- [RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, December 2008.
- [RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, January 2009.
- [RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, June 2009.
- [RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, September 2011.
- [MPLS-DELAY-FWK] Fu, X., Manral, V., McDysan, D., Malis, A., Giacalone, S., Betts, M., Wang, Q., and J. Drake, "Traffic Engineering architecture for services aware MPLS [draft-fuxh-mpls-delay-loss-te-framework]", Oct 2012.
- [OSPF-TE-EXPRESS] Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions

[draft-ietf-ospf-te-metric-extensions]",
May 2012.

- [ISIS-TE-EXPRESS] Previdi, S., Giacalone, S., Ward, D., Drake, J., Atlas, A., and C. Filsfils, "IS-IS Traffic Engineering (TE) Metric Extensions [draft-previdi-isis-te-metric-extensions]", Oct 2012.
- [MPLS-TE-EXPRESS-PATH] Atlas, A., Drake, J., Ward, D., Giacalone, S., Previdi, S., and C. Filsfils, "Performance-based Path Selection for Explicitly Routed LSPs [draft-atlas-mpls-te-express-path]", June 2012.
- [PCEP-MIB] Kiran Koushik, A S., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "PCE communication protocol(PCEP) Management Information Base [draft-ietf-pce-pcep-mib]", July 2012.

Appendix A. Contributor Addresses

Clarence Filsfils
Cisco Systems
EMail: cfilsfil@cisco.com

Siva Sivabalan
Cisco Systems
EMail: msiva@cisco.com

Stefano Previdi
Cisco Systems
EMail: sprevidi@cisco.com

Udayasree Palle
Huawei Technologies India Pvt Ltd
Leela Palace
Bangalore, Karnataka 560008
INDIA
EMail: udayasree.palle@huawei.com

Authors' Addresses

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

EMail: dhruv.dhody@huawei.com

Vishwas Manral
Hewlett-Packard Corp.
191111 Pruneridge Ave.
Cupertino, CA 95014
USA

EMail: vishwas.manral@hp.com

Zafar Ali
Cisco Systems

EMail: zali@cisco.com

George Swallow
Cisco Systems

EMail: swallow@cisco.com

Kenji Kumaki
KDDI Corporation

EMail: ke-kumaki@kddi.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 22, 2014

T. Otani
K. Ogaki
KDDI
D. Caviglia
Ericsson
F. Zhang
Huawei Technologies
C. Margaria
Coriant R&D GmbH
July 21, 2013

Requirements for GMPLS applications of PCE
draft-ietf-pce-gmpls-aps-req-09.txt

Abstract

The initial effort of the PCE (Path computation element) WG was mainly focused on MPLS. As a next step, this draft describes functional requirements for GMPLS application of PCE.

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 January 22, 2014.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. GMPLS applications of PCE	3
2.1. Path computation in GMPLS network	3
2.2. Unnumbered Interface	5
2.3. Asymmetric Bandwidth Path Computation	5
3. Requirements for GMPLS application of PCE	5
3.1. Requirements on Path Computation Request	5
3.2. Requirements on Path Computation Reply	6
3.3. GMPLS PCE Management	8
4. Security Considerations	8
5. IANA Considerations	8
6. Acknowledgement	8
7. References	8
7.1. Normative References	8
7.2. Informative References	10
Authors' Addresses	11

1. Introduction

The initial effort of the PCE (Path computation element) WG was mainly focused on solving the path computation problem within a domain or over different domains in MPLS networks. As the same case with MPLS, service providers (SPs) have also come up with requirements for path computation in GMPLS-controlled networks [RFC3945] such as wavelength, TDM-based or Ethernet-based networks as well.

[RFC4655] and [RFC4657] discuss the framework and requirements for PCE on both packet MPLS networks and GMPLS-controlled networks. This document complements these RFCs by providing some considerations of GMPLS applications in the intra-domain and inter-domain networking environments and indicating a set of requirements for the extended definition of PCE-related protocols.

Note that the requirements for inter-layer and inter-area traffic engineering described in [RFC6457] and [RFC4927] are outside of the scope of this document.

Constraint-based shortest path first (CSPF) computation within a domain or over domains for signaling GMPLS Label Switched Paths (LSPs) is usually more stringent than that of MPLS TE LSPs [RFC4216],

because the additional constraints, e.g., interface switching capability, link encoding, link protection capability, SRLG (Shared risk link group) [RFC4202] and so forth need to be considered to establish GMPLS LSPs. GMPLS signaling protocol [RFC3473] is designed taking into account bi-directionality, switching type, encoding type and protection attributes of the TE links spanned by the path, as well as LSP encoding and switching type of the end points, appropriately.

This document provides requirements for GMPLS applications of PCE in support of GMPLS path computation, included are requirements for both intra-domain and inter-domain environments.

2. GMPLS applications of PCE

2.1. Path computation in GMPLS network

Figure 1 depicts a model GMPLS network, consisting of an ingress link, a transit link as well as an egress link. We will use this model to investigate consistent guidelines for GMPLS path computation. Each link at each interface has its own switching capability, encoding type and bandwidth.

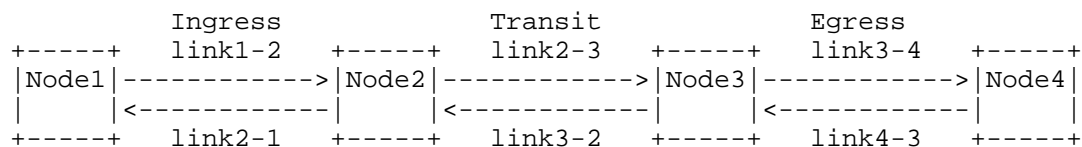


Figure 1: Path computation in GMPLS networks

For the simplicity in consideration, the below basic assumptions are made when the LSP is created.

- (1) Switching capabilities of outgoing links from the ingress and egress nodes (link1-2 and link4-3 in Figure 1) are consistent with each other.
- (2) Switching capabilities of all transit links including incoming links to the ingress and egress nodes (link2-1 and link3-4) are consistent with switching type of a LSP to be created.
- (3) Encoding-types of all transit links are consistent with encoding type of a LSP to be created.

GMPLS-controlled networks (e.g., GMPLS-based TDM networks) are usually responsible for transmitting data for the client layer.

These GMPLS-controlled networks can provide different types of connections for customer services based on different service bandwidth requests.

The applications and the corresponding additional requirements for applying PCE to, for example, GMPLS-based TDM networks, are described in Figure 2. In order to simplify the description, this document just discusses the scenario in SDH networks as an example. The scenarios in SONET or OTN are similar to this scenario.

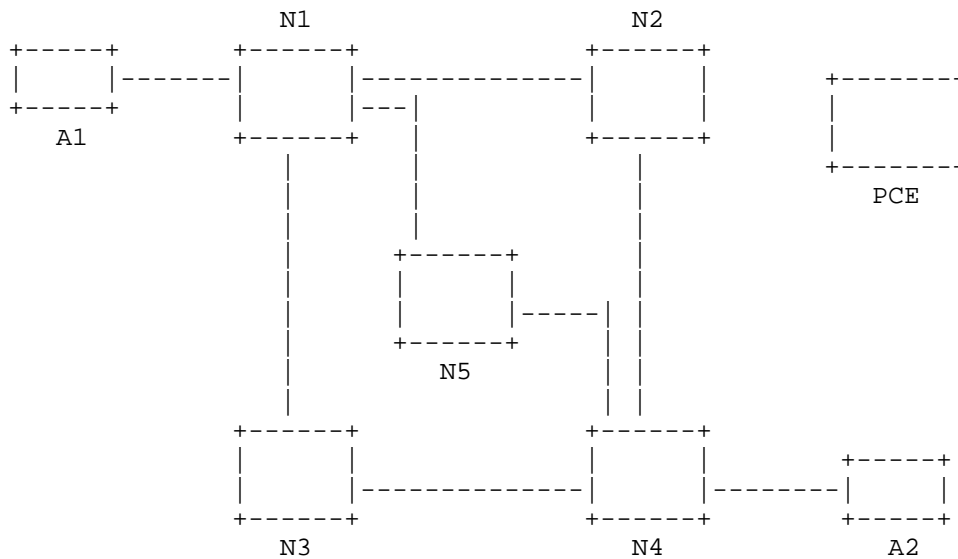


Figure 2: A simple TDM (SDH) network

Figure 2 shows a simple TDM (SDH) network topology, where N1, N2, N3, N4 and N5 are all SDH switches. Assume that one Ethernet service with 100M bandwidth is required from A1 to A2 over this network. The client Ethernet service could be provided by a VC4 container from N1 to N4, and it could also be provided by three concatenated VC3 containers (Contiguous or Virtual concatenation) from N1 to N4.

In this scenario, when the ingress node (e.g., N1) receives a client service transmitting request, the type of containers (one VC4 or three concatenated VC3) could be determined by PCC (Path computation client) (e.g., N1 or NMS), but could also be determined by PCE automatically based on policy [RFC5394]. If it is determined by PCC, PCC should be capable of specifying the ingress node and egress node, signal type, the type of the concatenation and the number of the concatenation in a PCReq (Path computation request) message. PCE

should consider those parameters during path computation. The route information (co-route or separated-route) should be specified in a PCRep (Path computation reply) message if path computation is performed successfully.

As described above, PCC should be capable of specifying TE attributes defined in the next section and PCE should compute a path accordingly.

Where a GMPLS network is consisting of inter-domain (e.g., inter-AS or inter-area) GMPLS-controlled networks, requirements on the path computation follows [RFC5376] and [RFC4726].

2.2. Unnumbered Interface

GMPLS supports unnumbered interface ID that is defined in [RFC3477], which means that the endpoints of the path may be unnumbered. It should also be possible to request a path consisting of the mixture of numbered links and unnumbered links, or a P2MP (Point-to-multipoint) path with different types of endpoints. Therefore, the PCC should be capable of indicating the unnumbered interface ID of the endpoints in the PCReq message.

2.3. Asymmetric Bandwidth Path Computation

As per [RFC6387], GMPLS signaling can be used for setting up an asymmetric bandwidth bidirectional LSP. If a PCE is responsible for the path computation, the PCE should be capable of computing a path for the bidirectional LSP with asymmetric bandwidth. It means that the PCC should be able to indicate the asymmetric bandwidth requirements in forward and reverse directions in the PCReq message.

3. Requirements for GMPLS application of PCE

3.1. Requirements on Path Computation Request

As for path computation in GMPLS-controlled networks as discussed in section 2, the PCE should appropriately consider the GMPLS TE attributes listed below once a PCC or another PCE requests a path computation. The path calculation request message from the PCC or the PCE must contain the information specifying appropriate attributes. According to [RFC5440], [PCE-WSO-REQ] and to RSVP procedures like explicit label control(ELC), the additional attributes introduced are as follows:

(1) Switching capability/type: as defined in [RFC3471], [RFC4203] and, all current and future values.

- (2) Encoding type: as defined in [RFC3471], [RFC4203] and, all current and future values.
- (3) Signal Type: as defined in [RFC4606] and, all current and future values.
- (4) Concatenation Type: In SDH/SONET and OTN, two kinds of concatenation modes are defined: contiguous concatenation which requires co-route for each member signal and requires all the interfaces along the path to support this capability, and virtual concatenation which allows diverse routes for the member signals and only requires the ingress and egress interfaces to support this capability. Note that for the virtual concatenation, it also may specify co-routed or separated-routed. See [RFC4606] and [RFC4328] about concatenation information.
- (5) Concatenation Number: Indicates the number of signals that are requested to be contiguously or virtually concatenated. Also see [RFC4606] and [RFC4328].
- (6) Technology-specific label(s) such as defined in [RFC4606], [RFC6060], [RFC6002] or [RFC6205].
- (7) e2e Path protection type: as defined in [RFC4872], e.g., 1+1 protection, 1:1 protection, (pre-planned) rerouting, etc.
- (8) Administrative group: as defined in [RFC3630]
- (9) Link Protection type: as defined in [RFC4203]
- (10) Support for unnumbered interfaces: as defined in [RFC3477]
- (11) Support for asymmetric bandwidth request: as defined in [RFC6387]
- (12) Support for explicit label control during the path computation.
- (13) Support of label restrictions in the requests/responses, similarly to RSVP-TE ERO (Explicit route object) and XRO (Exclude route object) as defined in [RFC3473] and [RFC4874].

3.2. Requirements on Path Computation Reply

As described above, a PCE should compute the path that satisfies the constraints which are specified in the PCReq message. Then the PCE should send a PCRep message including the computation result to the PCC. For Path Computation Reply message (PCRep) in GMPLS networks, there are some additional requirements. The PCEP (PCE communication protocol) PCRep message must be extended to meet the following requirements.

(1) Path computation with concatenation

In the case of path computation involving concatenation, when a PCE receives the PCReq message specifying the concatenation constraints described in section 3.1, the PCE should compute a path accordingly.

For path computation involving contiguous concatenation, a single route is required and all the interfaces along the route should support contiguous concatenation capability. Therefore, the PCE should compute a path based on the contiguous concatenation capability of each interface and only one ERO which should carry the route information for the response.

For path computation involving virtual concatenation, only the ingress/egress interfaces need to support virtual concatenation capability and there may be diverse routes for the different member signals. Therefore, multiple EROs may be needed for the response. Each ERO may represent the route of one or multiple member signals. In the case where one ERO represents several member signals among the total member signals, the number of member signals along the route of the ERO must be specified.

(2) Label constraint

In the case that a PCC does not specify the exact label(s) when requesting a label-restricted path and the PCE is capable of performing the route computation and label assignment computation procedure, the PCE needs to be able to specify the label of the path in a PCRep message.

Wavelength restriction is a typical case of label restriction. More generally in GMPLS-controlled networks label switching and selection constraints may apply and a PCC may request a PCE to take label constraint into account and return an ERO containing the label or set of label that fulfil the PCC request.

(3) Roles of the routes

When a PCC specifies the protection type of an LSP, the PCE should compute the working route and the corresponding protection route(s).

Therefore, the PCRep should allow to distinguish the working (nominal) and the protection routes. According to these routes, RSVP-TE procedure appropriately creates both the working and the protection LSPs for example with ASSOCIATION object [RFC6689].

3.3. GMPLS PCE Management

This document does not change any of the management or operational details for networks that utilise PCE. Please refer to [RFC4655] for an overview of this scenery. However, this document proposes the introduction of several PCEP objects and data for the better integration of PCE with GMPLS networks. Those protocol elements will need to be visible in any management tools that apply to the PCE, PCC, and PCEP. That includes, but is not limited to, adding appropriate objects to existing PCE MIB modules that are used for modelling and monitoring PCEP deployments [PCEP-MIB]. Ideas for what objects are needed may be guided by the relevant GMPLS extensions in GMPLS-TE-STD-MIB [RFC4802]."

4. Security Considerations

PCEP extensions to support GMPLS should be considered under the same security as current PCE work and this extension will not change the underlying security issues. Sec. 10 of [RFC5440] describes the list of security considerations in PCEP. At the time [RFC5440] was published, TCP Authentication Option (TCP-AO) had not been fully specified for securing the TCP connections that underlie PCEP sessions. TCP-AO [RFC5925] has now been published and PCEP implementations should fully support TCP-AO according to [RFC6952].

5. IANA Considerations

This document has no actions for IANA.

6. Acknowledgement

The author would like to express the thanks to Ramon Casellas, Julien Meuric, Adrian Farrel, Yaron Sheffer and Shuichi Okamoto for their comments.

7. References

7.1. Normative References

- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.

- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, October 2004.
- [RFC4202] Kompella, K. and Y. Rekhter, "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4202, October 2005.
- [RFC4203] Kompella, K. and Y. Rekhter, "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, October 2005.
- [RFC4328] Papadimitriou, D., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, January 2006.
- [RFC4606] Mannie, E. and D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", RFC 4606, August 2006.
- [RFC4802] Nadeau, T. and A. Farrel, "Generalized Multiprotocol Label Switching (GMPLS) Traffic Engineering Management Information Base", RFC 4802, February 2007.
- [RFC4872] Lang, J., Rekhter, Y., and D. Papadimitriou, "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.
- [RFC4927] Le Roux, J., "Path Computation Element Communication Protocol (PCECP) Specific Requirements for Inter-Area MPLS and GMPLS Traffic Engineering", RFC 4927, June 2007.
- [RFC5376] Bitar, N., Zhang, R., and K. Kumaki, "Inter-AS Requirements for the Path Computation Element Communication Protocol (PCECP)", RFC 5376, November 2008.

- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC6002] Berger, L. and D. Fedyk, "Generalized MPLS (GMPLS) Data Channel Switching Capable (DCSC) and Channel Set Label Extensions", RFC 6002, October 2010.
- [RFC6060] Fedyk, D., Shah, H., Bitar, N., and A. Takacs, "Generalized Multiprotocol Label Switching (GMPLS) Control of Ethernet Provider Backbone Traffic Engineering (PBB-TE)", RFC 6060, March 2011.
- [RFC6205] Otani, T. and D. Li, "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC 6205, March 2011.
- [RFC6387] Takacs, A., Berger, L., Caviglia, D., Fedyk, D., and J. Meuric, "GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths (LSPs)", RFC 6387, September 2011.
- [RFC6689] Berger, L., "Usage of the RSVP ASSOCIATION Object", RFC 6689, July 2012.

7.2. Informative References

- [PCE-WSO-REQ]
 - Lee, Y., Bernstein, G., Martensson, J., Takeda, T., Tsuritani, T., and O. de Dios, "PCEP Requirements for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-routing-wavelength-09 (work in progress), June 2013.
- [PCEP-MIB]
 - Koushik, A., Emile, S., Zhao, Q., King, D., and J. Hardwick, "PCE communication protocol (PCEP) Management Information Base", draft-ietf-pce-pcep-mib-05 (work in progress), July 2013.
- [RFC4216] Zhang, R. and J. Vasseur, "MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements", RFC 4216, November 2005.
- [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.

- [RFC4726] Farrel, A., Vasseur, J., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.
- [RFC4874] Lee, CY., Farrel, A., and S. De Cnodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4874, April 2007.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, December 2008.
- [RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.
- [RFC6457] Takeda, T. and A. Farrel, "PCC-PCE Communication and PCE Discovery Requirements for Inter-Layer Traffic Engineering", RFC 6457, December 2011.
- [RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, May 2013.

Authors' Addresses

Tomohiro Otani
KDDI Corporation
2-3-2 Nishi-shinjuku
Shinjuku-ku, Tokyo
Japan

Phone: +81-(3) 3347-6006
Email: tm-otani@kddi.com

Kenichi Ogaki
KDDI Corporation
3-10-10 Iidabashi
Chiyoda-ku, Tokyo
Japan

Phone: +81-(3) 6678-0284
Email: ke-oogaki@kddi.com

Diego Caviglia
Ericsson
16153 Genova Cornigliano
Italy

Phone: +390106003736
Email: diego.caviglia@ericsson.com

Fatai Zhang
Huawei Technologies Co., Ltd.
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District, Shenzhen 518129
P.R.China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Cyril Margaria
Coriant R&D GmbH
St Martin Strasse 76
Munich, 81541
Germany

Phone: +49 89 5159 16934
Email: cyril.margaria@coriant.com

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: June 14, 2020

C. Margaria, Ed.
Juniper
O. Gonzalez de Dios, Ed.
Telefonica Investigacion y Desarrollo
F. Zhang, Ed.
Huawei Technologies
December 12, 2019

PCEP extensions for GMPLS
draft-ietf-pce-gmpls-pcep-extensions-16

Abstract

A Path Computation Element (PCE) provides path computation functions for Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks. Additional requirements for GMPLS are identified in RFC7025.

This memo provides extensions to the Path Computation Element communication Protocol (PCEP) for the support of the GMPLS control plane to address those requirements.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

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

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

This Internet-Draft will expire on June 14, 2020.

Copyright Notice

Copyright (c) 2019 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Terminology	3
1.2. PCEP Requirements for GMPLS	5
1.3. Requirements Applicability	5
1.3.1. Requirements on Path Computation Request	6
1.3.2. Requirements on Path Computation Response	7
1.4. Existing Support for GMPLS in Base PCEP Objects and its Limitations	7
2. PCEP Objects and Extensions	10
2.1. GMPLS Capability Advertisement	10
2.1.1. GMPLS Computation TLV in the Existing PCE Discovery Protocol	10
2.1.2. OPEN Object Extension GMPLS-CAPABILITY TLV	10
2.2. RP Object Extension	11
2.3. BANDWIDTH Object Extensions	12
2.4. LOAD-BALANCING Object Extensions	14
2.5. END-POINTS Object Extensions	16
2.5.1. Generalized Endpoint Object Type	17
2.5.2. END-POINTS TLV Extensions	20
2.6. IRO Extension	24
2.7. XRO Extension	24
2.8. LSPA Extensions	26
2.9. NO-PATH Object Extension	26
2.9.1. Extensions to NO-PATH-VECTOR TLV	27
3. Additional Error-Types and Error-Values Defined	27
4. Manageability Considerations	29
4.1. Control of Function through Configuration and Policy	29
4.2. Information and Data Models	29
4.3. Liveness Detection and Monitoring	29
4.4. Verifying Correct Operation	30
4.5. Requirements on Other Protocols and Functional Components	30
4.6. Impact on Network Operation	30
5. IANA Considerations	30
5.1. PCEP Objects	30
5.2. Endpoint type field in Generalized END-POINTS Object	31
5.3. New PCEP TLVs	32
5.4. RP Object Flag Field	32
5.5. New PCEP Error Codes	32
5.6. New NO-PATH-VECTOR TLV Fields	33

5.7. New Subobject for the Include Route Object	34
5.8. New Subobject for the Exclude Route Object	34
5.9. New GMPLS-CAPABILITY TLV Flag Field	35
6. Security Considerations	35
7. Contributing Authors	36
8. Acknowledgments	38
9. References	38
9.1. Normative References	38
9.2. Informative References	42
Appendix A. LOAD-BALANCING Usage for SDH Virtual Concatenation .	43
Authors' Addresses	43

1. Introduction

Although [RFC4655] defines the PCE architecture and framework for both MPLS and GMPLS networks, most preexisting PCEP RFCs [RFC5440], [RFC5521], [RFC5541], [RFC5520] are focused on MPLS networks, and do not cover the wide range of GMPLS networks. This document complements these RFCs by addressing the extensions required for GMPLS applications and routing requests, for example for Optical Transport Network (OTN) and Wavelength Switched Optical Network (WSN) networks.

The functional requirements to be addressed by the PCEP extensions to support these applications are fully described in [RFC7025] and [RFC7449].

1.1. Terminology

This document uses terminologies from the PCE architecture document [RFC4655], the PCEP documents including [RFC5440], [RFC5521], [RFC5541], [RFC5520], [RFC7025] and [RFC7449], and the GMPLS documents such as [RFC3471], [RFC3473] and so on. Note that it is expected the reader is familiar with these documents. The following abbreviations are used in this document

ODU ODU Optical Channel Data Unit [G.709-v3]

OTN Optical Transport Network [G.709-v3]

L2SC Layer-2 Switch Capable [RFC3471]

TDM Time-Division Multiplex Capable [RFC3471]

LSC Lambda Switch Capable [RFC3471]

SONET Synchronous Optical Networking

SDH Synchronous Digital Hierarchy

PCC Path Computation Client

RSVP-TE Resource Reservation Protocol - Traffic Engineering

LSP Label Switched Path

TE-LSP Traffic Engineering LSP

IRO Include Route Object

ERO Explicit Route Object

XRO eXclude Route Object

RRO Record Route Object

LSPA LSP Attribute

SRLG Shared Risk Link Group

NVC Number of Virtual Components [RFC4328] [RFC4606]

NCC Number of Contiguous Components [RFC4328] [RFC4606]

MT Multiplier [RFC4328] [RFC4606]

RCC Requested Contiguous Concatenation [RFC4606]

PCReq Path Computation Request [RFC5440]

PCRep Path Computation Reply [RFC5440]

MEF Metro Ethernet Forum

SSON Spectrum-Switched Optical Network

P2MP Point to Multi-Point

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

1.2. PCEP Requirements for GMPLS

The document [RFC7025] describes the set of PCEP requirements to support GMPLS TE-LSPs. This document assumes a significant familiarity with [RFC7025] and existing PCEP extensions. As a short overview, those requirements can be broken down into the following categories.

- o Which data flow is switched by the LSP: a combination of Switching type (for instance L2SC or TDM), LSP Encoding type (e.g., Ethernet, SONET/SDH) and sometimes the Signal Type (e.g., in case of TDM/LSC switching capability).
- o Data flow specific traffic parameters, which are technology specific. For instance, in SDH/SONET and [G.709-v3] OTN networks the Concatenation Type and the Concatenation Number have an influence on the switched data and on which link it can be supported
- o Support for asymmetric bandwidth requests.
- o Support for unnumbered interface identifiers, as defined in [RFC3477]
- o Label information and technology specific label(s) such as wavelength labels as defined in [RFC6205]. A PCC should also be able to specify a label restriction similar to the one supported by RSVP-TE in [RFC3473].
- o Ability to indicate the requested granularity for the path ERO: node, link or label. This is to allow the use of the explicit label control feature of RSVP-TE.

The requirements of [RFC7025] apply to several objects conveyed by PCEP, this is described in Section 1.3. Some of the requirements of [RFC7025] are already supported in existing documents, as described in Section 1.4.

This document describes a set of PCEP extensions, including new object types, TLVs, encodings, error codes and procedures, in order to fulfill the aforementioned requirements not covered in existing RFCs.

1.3. Requirements Applicability

This section follows the organization of [RFC7025] Section 3 and indicates, for each requirement, the affected piece of information carried by PCEP and its scope.

1.3.1. Requirements on Path Computation Request

- (1) Switching capability/type: as described in [RFC3471] this piece of information is used with the Encoding Type and Signal Type to fully describe the switching technology and data carried by the TE-LSP. This is applicable to the TE-LSP itself and also to the TE-LSP endpoint (Carried in the END-POINTS object for MPLS networks in [RFC5440]) when considering multiple network layers. Inter-layer path computation requirements are addressed in [RFC8282] which addressing the TE-LSP itself, but the TE-LSP endpoints are not addressed.
- (2) Encoding type: see (1).
- (3) Signal type: see (1).
- (4) Concatenation type: this parameter and the Concatenation Number (5) are specific to some TDM (SDH and ODU) switching technology. They MUST be described together and are used to derive the requested resource allocation for the TE-LSP. It is scoped to the TE-LSP and is related to the [RFC5440] BANDWIDTH object in MPLS networks. See [RFC4606] and [RFC4328] about concatenation information.
- (5) Concatenation number: see (4).
- (6) Technology-specific label(s): as described in [RFC3471] the GMPLS Labels are specific to each switching technology. They can be specified on each link and also on the TE-LSP endpoints, in WSON networks for instance, as described in [RFC6163]. The label restriction can apply to endpoints and on each hop, the related PCEP objects are END-POINTS, IRO, XRO and RRO.
- (7) End-to-End (E2E) path protection type: as defined in [RFC4872], this is applicable to the TE-LSP. In MPLS networks the related PCEP object is LSPA (carrying local protection information).
- (8) Administrative group: as defined in [RFC3630], this information is already carried in the LSPA object.
- (9) Link protection type: as defined in [RFC4872], this is applicable to the TE-LSP and is carried in association with the E2E path protection type.
- (10) Support for unnumbered interfaces: as defined in [RFC3477]. Its scope and related objects are the same as labels

- (11) Support for asymmetric bandwidth requests: as defined [RFC6387], the scope is similar to (4)
- (12) Support for explicit label control during the path computation. This affects the TE-LSP and amount of information returned in the ERO.
- (13) Support of label restrictions in the requests/responses: This is described in (6).

1.3.2. Requirements on Path Computation Response

- (1) Path computation with concatenation: This is related to Path Computation request requirement (4). In addition there is a specific type of concatenation called virtual concatenation that allows different routes to be used between the endpoints. It is similar to the semantic and scope of the LOAD-BALANCING in MPLS networks.
- (2) Label constraint: The PCE should be able to include Labels in the path returned to the PCC, the related object is the ERO object.
- (3) Roles of the routes: as defined in [RFC4872], this is applicable to the TE-LSP and is carried in association with the E2E path protection type.

1.4. Existing Support for GMPLS in Base PCEP Objects and its Limitations

The support provided by specifications in [RFC8282] and [RFC5440] for the requirements listed in [RFC7025] is summarized in Table 1 and Table 2. In some cases the support may not be complete, as noted, and additional support need to be provided in this specification.

Req.	Name	Support
1	Switching capability/type	SWITCH-LAYER (RFC8282)
2	Encoding type	SWITCH-LAYER (RFC8282)
3	Signal type	SWITCH-LAYER (RFC8282)
4	Concatenation type	No
5	Concatenation number	No
6	Technology-specific label	(Partial) ERO (RFC5440)
7	End-to-End (E2E) path protection type	No
8	Administrative group	LSPA (RFC5440)
9	Link protection type	No
10	Support for unnumbered interfaces	(Partial) ERO (RFC5440)
11	Support for asymmetric bandwidth requests	No
12	Support for explicit label control during the path computation	No
13	Support of label restrictions in the requests/responses	No

Table 1: RFC7025 Section 3.1 requirements support

Req.	Name	Support
1	Path computation with concatenation	No
2	Label constraint	No
3	Roles of the routes	No

Table 2: RFC7025 Section 3.2 requirements support

As described in Section 1.3 PCEP as of [RFC5440], [RFC5521] and [RFC8282], supports the following objects, included in requests and responses, related to the described requirements.

From [RFC5440]:

- o END-POINTS: related to requirements (1, 2, 3, 6, 10 and 13). The object only supports numbered endpoints. The context specifies whether they are node identifiers or numbered interfaces.
- o BANDWIDTH: related to requirements (4, 5 and 11). The data rate is encoded in the bandwidth object (as IEEE 32 bit float). [RFC5440] does not include the ability to convey an encoding proper to all GMPLS-controlled networks.

- o ERO: related to requirements (6, 10, 12 and 13). The ERO content is defined in RSVP in [RFC3209][RFC3473][RFC3477][RFC7570] and supports all the requirements already.
- o LSPA: related to requirements (7, 8 and 9). The requirement 8 (setup and holding priorities) is already supported.

From [RFC5521]:

- o XRO:
 - * This object allows excluding (strict or not) resources and is related to requirements (6, 10 and 13). It also includes the requested diversity (node, link or SRLG).
 - * When the F bit is set, the request indicates that the existing path has failed and the resources present in the RRO can be reused.

From [RFC8282]:

- o SWITCH-LAYER: addresses requirements (1, 2 and 3) for the TE-LSP and indicates which layer(s) should be considered. The object can be used to represent the RSVP-TE generalized label request. It does not address the endpoints case of requirements (1, 2 and 3).
- o REQ-ADAP-CAP: indicates the adaptation capabilities requested, can also be used for the endpoints in case of mono-layer computation

The gaps in functional coverage of the base PCEP objects are:

The BANDWIDTH and LOAD-BALANCING objects do not describe the details of the traffic request (requirements 4 and 5, for example NVC, multiplier) in the context of GMPLS networks, for instance TDM or OTN networks.

The END-POINTS object does not allow specifying an unnumbered interface, nor potential label restrictions on the interface (requirements 6, 10 and 13). Those parameters are of interest in case of switching constraints.

The Include/eXclude Route Objects (IRO/XRO) do not allow the inclusion/exclusion of labels (requirements 6, 10 and 13).

Base attributes do not allow expressing the requested link protection level and/or the end-to-end protection attributes.

The PCEP extensions defined later in this document to cover the gaps are:

Two new object types are defined for the BANDWIDTH object (Generalized bandwidth, Generalized bandwidth of existing TE-LSP for which a reoptimization is requested).

A new object type is defined for the LOAD-BALANCING object (Generalized Load Balancing).

A new object type is defined for the END-POINTS object (Generalized Endpoint).

A new TLV is added to the Open message for capability negotiation.

A new TLV is added to the LSPA object.

The Label TLV is now allowed in the IRO and XRO objects.

In order to indicate the used routing granularity in the response, a new flag in the RP object is added.

2. PCEP Objects and Extensions

This section describes the necessary PCEP objects and extensions. The PCReq and PCRep messages are defined in [RFC5440]. This document does not change the existing grammars.

2.1. GMPLS Capability Advertisement

2.1.1. GMPLS Computation TLV in the Existing PCE Discovery Protocol

IGP-based PCE Discovery (PCED) is defined in [RFC5088] and [RFC5089] for the OSPF and IS-IS protocols. Those documents have defined bit 0 in PCE-CAP-FLAGS Sub-TLV of the PCED TLV as "Path computation with GMPLS link constraints". This capability is optional and can be used to detect GMPLS-capable PCEs. PCEs that set the bit to indicate support of GMPLS path computation MUST follow the procedures in Section 2.1.2 to further qualify the level of support during PCEP session establishment.

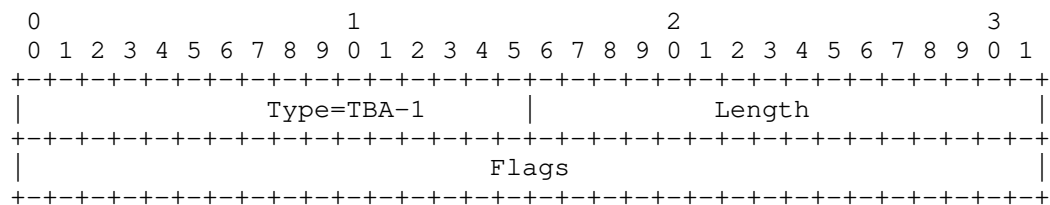
2.1.2. OPEN Object Extension GMPLS-CAPABILITY TLV

In addition to the IGP advertisement, a PCEP speaker MUST be able to discover the other peer GMPLS capabilities during the Open message exchange. This capability is also useful to avoid misconfigurations. This document defines a GMPLS-CAPABILITY TLV for use in the OPEN object to negotiate the GMPLS capability. The inclusion of this TLV

in the Open message indicates that the PCEP speaker support the PCEP extensions defined in the document. A PCEP speaker that is able to support the GMPLS extensions defined in this document MUST include the GMPLS-CAPABILITY TLV on the Open message. If one of the PCEP peers does not include the GMPLS-CAPABILITY TLV in the Open message, the peers MUST NOT make use of the objects and TLVs defined in this document.

If the PCEP speaker supports the extensions of this specification but did not advertise the GMPLS-CAPABILITY capability, upon receipt of a message from the PCE including an extension defined in this document, it MUST generate a PCEP Error (PCErr) with Error-Type=10 (Reception of an invalid object) and Error-value=TBA-42 (Missing GMPLS-CAPABILITY TLV), and it SHOULD terminate the PCEP session.

IANA has allocated value TBA-1 from the "PCEP TLV Type Indicators" sub-registry, as documented in Section 5.3 ("New PCEP TLVs"). The description is "GMPLS-CAPABILITY". Its format is shown in the following figure.



No Flags are defined in this document, they are reserved for future use.

2.2. RP Object Extension

Explicit label control (ELC) is a procedure supported by RSVP-TE, where the outgoing labels are encoded in the ERO. As a consequence, the PCE can provide such labels directly in the path ERO. Depending on policies or switching layer, it can be necessary for the PCC to use explicit label control or explicit link ids, thus it needs to indicate in the PCReq which granularity it is expecting in the ERO. This corresponds to requirement 12 of [RFC7025]. The possible granularities can be node, link or label. The granularities are inter-dependent, in the sense that link granularity implies the presence of node information in the ERO; similarly, a label granularity implies that the ERO contains node, link and label information.

A new 2-bit routing granularity (RG) flag (Bits TBA-13) is defined in the RP object. The values are defined as follows

0: reserved
1: node
2: link
3: label

Table 3: RG flag

The flag in the RP object indicates the requested route granularity. The PCE SHOULD follow this granularity and MAY return a NO-PATH if the requested granularity cannot be provided. The PCE MAY return any granularity on the route based on its policy. The PCC can decide if the ERO is acceptable based on its content.

If a PCE honored the requested routing granularity for a request, it MUST indicate the selected routing granularity in the RP object included in the response. Otherwise, the PCE MUST use the reserved RG to leave the check of the ERO to the PCC. The RG flag is backward-compatible with [RFC5440]: the value sent by an implementation (PCC or PCE) not supporting it will indicate a reserved value.

2.3. BANDWIDTH Object Extensions

From [RFC5440] the object carrying the requested size for the TE-LSP is the BANDWIDTH object. The object types 1 and 2 defined in [RFC5440] do not describe enough information to describe the TE-LSP bandwidth in GMPLS networks. The BANDWIDTH object encoding has to be extended to allow the object to express the bandwidth as described in [RFC7025]. RSVP-TE extensions for GMPLS provide a set of encodings allowing such representation in an unambiguous way, this is encoded in the RSVP-TE TSpec and FlowSpec objects. This document extends the BANDWIDTH object with new object types reusing the RSVP-TE encoding.

The following possibilities are supported by the extended encoding:

- o Asymmetric bandwidth (different bandwidth in forward and reverse direction), as described in [RFC6387]
- o GMPLS (SDH/SONET, G.709, ATM, MEF, etc.) parameters.

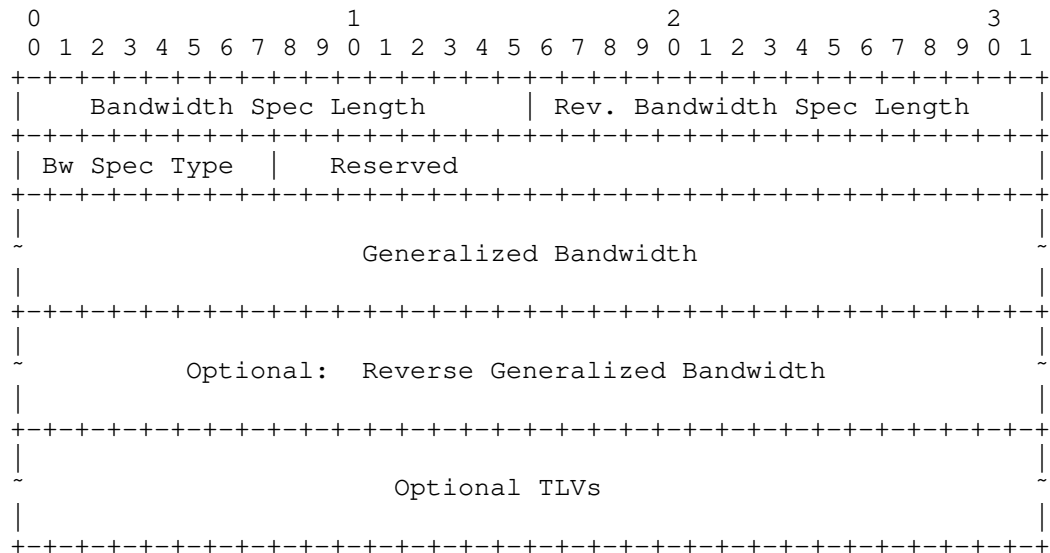
This corresponds to requirements 3, 4, 5 and 11 of [RFC7025] Section 3.1.

This document defines two Object Types for the BANDWIDTH object:

TBA-2 Generalized bandwidth

TBA-3 Generalized bandwidth of an existing TE-LSP for which a reoptimization is requested

The definitions below apply for Object Type TBA-2 and TBA-3. The body is as follows:



The BANDWIDTH object type TBA-2 and TBA-3 have a variable length. The 16-bit Bandwidth Spec Length field indicates the length of the Generalized Bandwidth field. The Bandwidth Spec Length MUST be strictly greater than 0. The 16-bit Reverse Bandwidth Spec Length field indicates the length of the Reverse Generalized Bandwidth field. The Reverse Bandwidth Spec Length MAY be equal to 0.

The Bw Spec Type field determines which type of bandwidth is represented by the object.

The Bw Spec Type corresponds to the RSVP-TE SENDER_TSPEC (Object Class 12) C-Types

The encoding of the fields Generalized Bandwidth and Reverse Generalized Bandwidth is the same as the Traffic Parameters carried in RSVP-TE, it can be found in the following references. It is to be noted that the RSVP-TE traffic specification MAY also include TLVs (e.g., [RFC6003] different from the PCEP TLVs).

Bw Spec	Type Name	Reference
2	Intserv	[RFC2210]
4	SONET/SDH	[RFC4606]
5	G.709	[RFC4328]
6	Ethernet	[RFC6003]
7	OTN-TDM	[RFC7139]
8	SSON	[RFC7792]

Table 4: Generalized Bandwidth and Reverse Generalized Bandwidth field encoding

When a PCC requests a bi-directional path with symmetric bandwidth, it SHOULD only specify the Generalized Bandwidth field, and set the Reverse Bandwidth Spec Length to 0. When a PCC needs to request a bi-directional path with asymmetric bandwidth, it SHOULD specify the different bandwidth in the forward and reverse directions with a Generalized Bandwidth and Reverse Generalized Bandwidth fields.

The procedure described in [RFC5440] for the PCRep is unchanged: a PCE MAY include the BANDWIDTH objects in the response to indicate the BANDWIDTH of the path.

As specified in [RFC5440] in the case of the reoptimization of a TE-LSP, the bandwidth of the existing TE-LSP MUST also be included in addition to the requested bandwidth if and only if the two values differ. The Object Type TBA-3 MAY be used instead of the previously specified object type 2 to indicate the existing TE-LSP bandwidth originally specified with object type TBA-2. A PCC that requested a path with a BANDWIDTH object of object type 1 MUST use object type 2 to represent the existing TE-LSP BANDWIDTH.

OPTIONAL TLVs MAY be included within the object body to specify more specific bandwidth requirements. No TLVs for the Object Type TBA-2 and TBA-3 are defined by this document.

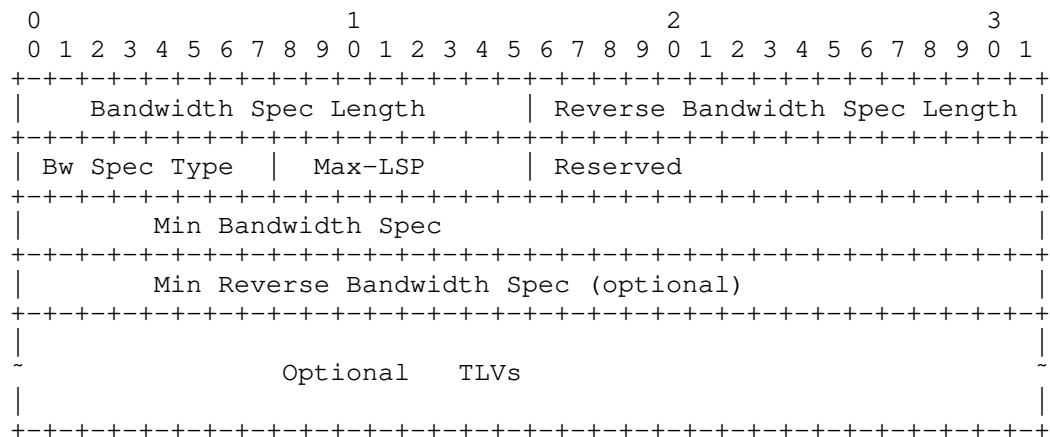
2.4. LOAD-BALANCING Object Extensions

The LOAD-BALANCING object [RFC5440] is used to request a set of at most Max-LSP TE-LSP having in total the bandwidth specified in BANDWIDTH, with each TE-LSP having at least a specified minimum bandwidth. The LOAD-BALANCING follows the bandwidth encoding of the BANDWIDTH object, and thus the existing definition from [RFC5440] does not describe enough details for the bandwidth specification expected by GMPLS.

Similarly to the BANDWIDTH object, a new object type is defined to allow a PCC to represent the bandwidth types supported by GMPLS networks.

This document defines the Generalized Load Balancing object type TBA-4 for the LOAD-BALANCING object. The Generalized Load Balancing object type has a variable length.

The format of the Generalized Load Balancing object type is as follows:



Bandwidth Spec Length (16 bits): the total length of the Min Bandwidth Spec field. The length MUST be strictly greater than 0.

Reverse Bandwidth Spec Length (16 bits): the total length of the Min Reverse Bandwidth Spec field. It MAY be equal to 0.

Bw Spec Type (8 bits): the bandwidth specification type, it corresponds to the RSVP-TE SENDER_TSPEC (Object Class 12) C-Types.

Max-LSP (8 bits): maximum number of TE-LSPs in the set.

Min Bandwidth Spec (variable): specifies the minimum bandwidth specification of each element of the TE-LSP set.

Min Reverse Bandwidth Spec (variable): specifies the minimum reverse bandwidth specification of each element of the TE-LSP set.

The encoding of the fields Min Bandwidth Spec and Min Reverse Bandwidth Spec is the same as in RSVP-TE SENDER_TSPEC object, it can be found in Table 4 from Section 2.3 from this document.

When a PCC requests a bi-directional path with symmetric bandwidth while specifying load balancing constraints it SHOULD specify the Min Bandwidth Spec field, and set the Reverse Bandwidth Spec Length to 0. When a PCC needs to request a bi-directional path with asymmetric bandwidth while specifying load balancing constraints, it MUST specify the different bandwidth in forward and reverse directions through a Min Bandwidth Spec and Min Reverse Bandwidth Spec fields.

OPTIONAL TLVs MAY be included within the object body to specify more specific bandwidth requirements. No TLVs for the Generalized Load Balancing object type are defined by this document.

The semantic of the LOAD-BALANCING object is not changed. If a PCC requests the computation of a set of TE-LSPs with at most N TE-LSPs so that it can carry generalized bandwidth X, each TE-LSP must at least transport bandwidth B, it inserts a BANDWIDTH object specifying X as the required bandwidth and a LOAD-BALANCING object with the Max-LSP and Min Bandwidth Spec fields set to N and B, respectively. When the BANDWIDTH and Min Bandwidth Spec can be summarized as scalars, the sum of all TE-LSPs bandwidth in the set is greater than X. The mapping of X over N path with (at least) bandwidth B is technology and possibly node specific. Each standard definition of the transport technology is defining those mappings and are not repeated in this document. A simplified example for SDH is described in Appendix A

In all other cases, including for technologies based on statistical multiplexing (e.g., InterServ, Ethernet), the exact bandwidth management (e.g., Ethernet's Excessive Rate) is left to the PCE's policies, according to the operator's configuration. If required, further documents may introduce a new mechanism to finely express complex load balancing policies within PCEP.

The BANDWIDTH and LOAD-BALANCING Bw Spec Type can be different depending on the endpoint nodes architecture. When the PCE is not able to handle those two Bw Spec Type, it MUST return a NO-PATH with the bit "LOAD-BALANCING could not be performed with the bandwidth constraints" set in the NO-PATH-VECTOR TLV.

2.5. END-POINTS Object Extensions

The END-POINTS object is used in a PCEP request message to specify the source and the destination of the path for which a path computation is requested. From [RFC5440], the source IP address and the destination IP address are used to identify those. A new Object Type is defined to address the following possibilities:

- o Different source and destination endpoint types.

- o Label restrictions on the endpoint.
- o Specification of unnumbered endpoints type as seen in GMPLS networks.

The Object encoding is described in the following sections.

In path computation within a GMPLS context the endpoints can:

- o Be unnumbered as described in [RFC3477].
- o Have labels associated to them, specifying a set of constraints on the allocation of labels.
- o Have different switching capabilities

The IPv4 and IPv6 endpoints are used to represent the source and destination IP addresses. The scope of the IP address (Node or numbered Link) is not explicitly stated. It is also possible to request a Path between a numbered link and an unnumbered link, or a P2MP path between different type of endpoints.

This document defines the Generalized Endpoint object type TBA-5 for the END-POINTS object. This new type also supports the specification of constraints on the endpoint label to be used. The PCE might know the interface restrictions but this is not a requirement. This corresponds to requirements 6 and 10 of [RFC7025].

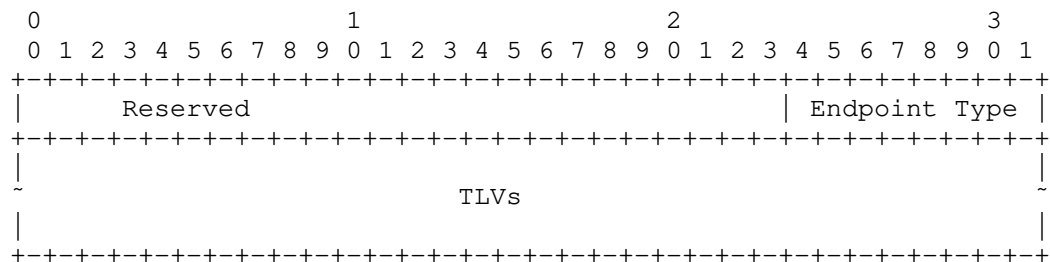
2.5.1. Generalized Endpoint Object Type

The Generalized Endpoint object type format consists of a body and a list of TLVs scoped to this object. The TLVs give the details of the endpoints and are described in Section 2.5.2. For each Endpoint Type, a different grammar is defined. The TLVs defined to describe an endpoint are:

1. IPv4 address endpoint.
2. IPv6 address endpoint.
3. Unnumbered endpoint.
4. Label request.
5. Label set.

The Label set TLV is used to restrict or suggest the label allocation in the PCE. This TLV expresses the set of restrictions which may

apply to signaling. Label restriction support can be an explicit or a suggested value (Label set describing one label, with the L bit respectively cleared or set), mandatory range restrictions (Label set with L bit cleared) and optional range restriction (Label set with L bit set). Endpoints label restriction may not be part of the RRO or IRO. They can be included when following [RFC4003] in signaling for egress endpoint, but ingress endpoint properties can be local to the PCC and not signaled. To support this case the label set allows indication which label are used in case of reoptimization. The label range restrictions are valid in GMPLS-controlled networks, either by PCC policy or depending on the switching technology used, for instance on given Ethernet or ODU equipment having limited hardware capabilities restricting the label range. Label set restriction also applies to WSON networks where the optical senders and receivers are limited in their frequency tunability ranges, consequently restricting the possible label ranges on the interface in GMPLS. The END-POINTS Object with Generalized Endpoint object type is encoded as follow:



Reserved bits SHOULD be set to 0 when a message is sent and ignored when the message is received.

The Endpoint Type is defined as follow:

Value	Type	Meaning
0	Point-to-Point	
1	Point-to-Multipoint	New leaves to add
2		Old leaves to remove
3		Old leaves whose path can be modified/reoptimized
4		Old leaves whose path has to be left unchanged
5-244	Reserved	
245-255	Experimental range	

Table 5: Generalized Endpoint endpoint types

The Endpoint Type is used to cover both point-to-point and different point-to-multipoint endpoints. A PCE may accept only Endpoint Type 0: Endpoint Types 1-4 apply if the PCE implementation supports P2MP path calculation. A PCE not supporting a given Endpoint Type SHOULD respond with a PCERR with Error-Type=4 (Not supported object), Error-value=TBA-15 (Unsupported endpoint type in END-POINTS Generalized Endpoint object type). As per [RFC5440], a PCE unable to process Generalized Endpoints may respond with Error-Type=3 (Unknown Object), Error-value=2 (Unrecognized object Type) or Error-Type=4 (Not supported object), Error-value=2 (Not supported object Type). The TLVs present in the request object body MUST follow the following [RFC5511] grammar:

```
<generalized-endpoint-tlvs> ::=
  <p2p-endpoints> | <p2mp-endpoints>

<p2p-endpoints> ::=
  <endpoint> [<endpoint-restriction-list>]
  <endpoint> [<endpoint-restriction-list>]

<p2mp-endpoints> ::=
  <endpoint> [<endpoint-restriction-list>]
  <endpoint> [<endpoint-restriction-list>]
  [<endpoint> [<endpoint-restriction-list>]]...
```

For endpoint type Point-to-Point, 2 endpoint TLVs MUST be present in the message. The first endpoint is the source and the second is the destination.

For endpoint type Point-to-Multipoint, several END-POINT objects MAY be present in the message and the exact meaning depending on the endpoint type defined for the object. The first endpoint TLV is the root and other endpoints TLVs are the leaves. The root endpoint MUST be the same for all END-POINTS objects for that P2MP tree request. If the root endpoint is not the same for all END-POINTS, a PCERR with Error-Type=17 (P2MP END-POINTS Error), Error-value=4 (The PCE cannot satisfy the request due to inconsistent END-POINTS) MUST be returned. The procedure defined in [RFC8306] Section 3.10 also apply to the Generalized Endpoint with Point-to-Multipoint endpoint types.

An endpoint is defined as follows:

```

<endpoint>::=<IPV4-ADDRESS>|<IPV6-ADDRESS>|<UNNUMBERED-ENDPOINT>
<endpoint-restriction-list> ::=                                <endpoint-restriction>
                                [<endpoint-restriction-list>]

<endpoint-restriction> ::=
                                [<LABEL-REQUEST>][<label-restriction-list>]

<label-restriction-list> ::= <label-restriction>
                                [<label-restriction-list>]
<label-restriction> ::= <LABEL-SET>

```

The different TLVs are described in the following sections. A PCE MAY support any or all of IPV4-ADDRESS, IPV6-ADDRESS, and UNNUMBERED-ENDPOINT TLVs. When receiving a PCReq, a PCE unable to resolve the identifier in one of those TLVs MUST respond using a PCRep with NO-PATH and set the bit "Unknown destination" or "Unknown source" in the NO-PATH-VECTOR TLV. The response SHOULD include the END-POINTS object with only the unsupported TLV(s).

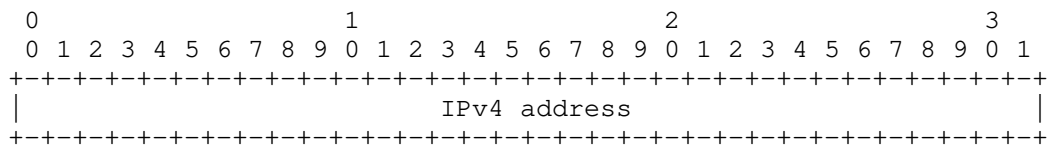
A PCE MAY support either or both of the LABEL-REQUEST and LABEL-SET TLVs. If a PCE finds a non-supported TLV in the END-POINTS the PCE MUST respond with a PCErr message with Error-Type=4 (Not supported object) and Error-value=TBA-15 (Unsupported TLV present in END-POINTS Generalized Endpoint object type) and the message SHOULD include the END-POINTS object in the response with only the endpoint and endpoint restriction TLV it did not understand. A PCE supporting those TLVs but not being able to fulfil the label restriction MUST send a response with a NO-PATH object which has the bit "No endpoint label resource" or "No endpoint label resource in range" set in the NO-PATH-VECTOR TLV. The response SHOULD include an END-POINTS object containing only the TLV(s) related to the constraints the PCE could not meet.

2.5.2. END-POINTS TLV Extensions

All endpoint TLVs have the standard PCEP TLV header as defined in [RFC5440] Section 7.1. For the Generalized Endpoint Object Type the TLVs MUST follow the ordering defined in Section 2.5.1.

2.5.2.1. IPV4-ADDRESS TLV

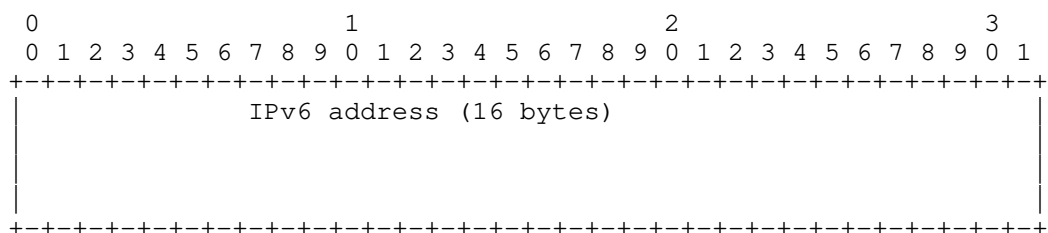
This TLV represents a numbered endpoint using IPv4 numbering, the format of the IPV4-ADDRESS TLV value (TLV-Type=TBA-6) is as follows:



This TLV MAY be ignored, in which case a PCRep with NO-PATH SHOULD be returned, as described in Section 2.5.1.

2.5.2.2. IPV6-ADDRESS TLV

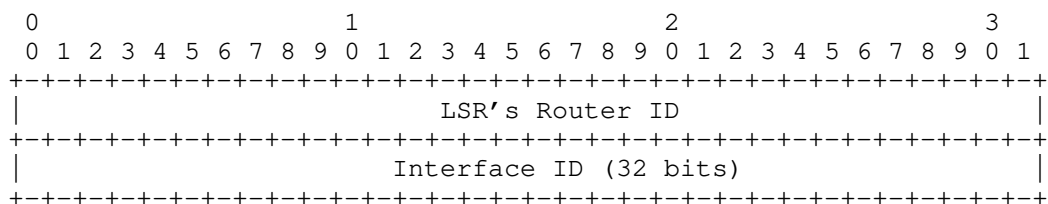
This TLV represents a numbered endpoint using IPV6 numbering, the format of the IPV6-ADDRESS TLV value (TLV-Type=TBA-7) is as follows:



This TLV MAY be ignored, in which case a PCRep with NO-PATH SHOULD be returned, as described in Section 2.5.1.

2.5.2.3. UNNUMBERED-ENDPOINT TLV

This TLV represents an unnumbered interface. This TLV has the same semantic as in [RFC3477]. The TLV value is encoded as follows (TLV-Type=TBA-8)



This TLV MAY be ignored, in which case a PCRep with NO-PATH SHOULD be returned, as described in Section 2.5.1.

2.5.2.4. LABEL-REQUEST TLV

The LABEL-REQUEST TLV indicates the switching capability and encoding type of the following label restriction list for the endpoint. The value format and encoding is the same as described in [RFC3471]

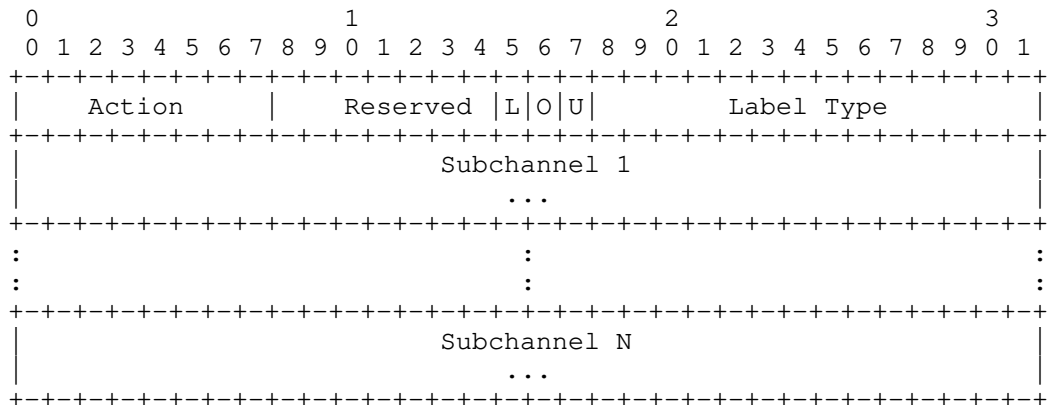
Section 3.1 Generalized label request. The LABEL-REQUEST TLV uses TLV-Type=TBA-9. The Encoding Type indicates the encoding type, e.g., SONET/SDH/GigE etc., of the LSP with which the data is associated. The Switching type indicates the type of switching that is being requested on the endpoint. G-PID identifies the payload. This TLV and the following one are defined to satisfy requirement 13 of [RFC7025] for the endpoint. It is not directly related to the TE-LSP label request, which is expressed by the SWITCH-LAYER object.

On the path calculation request only the GENERALIZED-BANDWIDTH and SWITCH-LAYER need to be coherent, the endpoint labels could be different (supporting a different LABEL-REQUEST). Hence the label restrictions include a Generalized label request in order to interpret the labels. This TLV MAY be ignored, in which case a PCRep with NO-PATH SHOULD be returned, as described in Section 2.5.1.

2.5.2.5. LABEL-SET TLV

Label or label range restrictions can be specified for the TE-LSP endpoints. Those are encoded using the LABEL-SET TLV. The label value need to be interpreted with a description on the Encoding and switching type. The REQ-ADAP-CAP object from [RFC8282] can be used in case of mono-layer request, however in case of multilayer it is possible to have more than one object, so it is better to have a dedicated TLV for the label and label request. These TLVs MAY be ignored, in which case a response with NO-PATH SHOULD be returned, as described in Section 2.5.1. TLVs are encoded as follows (following [RFC5440]):

- o LABEL-SET TLV, Type=TBA-10. The TLV Length is variable, Encoding follows [RFC3471] Section 3.5 "Label set" with the addition of a U bit, O bit and L bit. The L bit is used to represent a suggested set of labels, following the semantic of SUGGESTED_LABEL defined by [RFC3471].



A LABEL-SET TLV represents a set of possible labels that can be used on an interface. If the L bit is cleared, the label allocated on the first endpoint **MUST** be within the label set range. The action parameter in the Label set indicates the type of list provided. These parameters are described by [RFC3471] Section 3.5.1.

The U, O and L bits have the following meaning:

- U: Upstream direction: The U bit is set for upstream (revers) direction in case of bidirectional LSP.
- O: Old Label: set when the TLV represent the old (previously allocated) label in case of re-optimization. The R bit of the RP object **MUST** be set to 1. If the L bit is set, this bit **SHOULD** be set to 0 and ignored on receipt. When this bit is set, the Action field **MUST** be set to 0 (Inclusive List) and the Label Set **MUST** contain one subchannel.
- L: Loose Label: set when the TLV indicates to the PCE a set of preferred (ordered) labels to be used. The PCE **MAY** use those labels for label allocation.

Labels TLV bits

Several LABEL_SET TLVs **MAY** be present with the O bit cleared, LABEL_SET TLVs with L bit set can be combined with a LABEL_SET TLV with L bit cleared. There **MUST NOT** be more than two LABEL_SET TLVs present with the O bit set. If there are two LABEL_SET TLVs present, there **MUST NOT** be more than one with the U bit set, and there **MUST NOT** be more than one with the U bit cleared. For a given U bit value, if more than one LABEL_SET TLV with the O bit set is present, the first TLV **MUST** be processed and the following TLVs with the same U and O bit **MUST** be ignored.

A LABEL-SET TLV with the O and L bit set MUST trigger a PCErr message with Error-Type=10 (Reception of an invalid object) Error-value=TBA-25 (Wrong LABEL-SET TLV present with O and L bit set).

A LABEL-SET TLV with the O bit set and an Action Field not set to 0 (Inclusive list) or containing more than one subchannel MUST trigger a PCErr message with Error-Type=10 (Reception of an invalid object) Error-value=TBA-26 (Wrong LABEL-SET TLV present with O bit and wrong format).

If a LABEL-SET TLV is present with O bit set, the R bit of the RP object MUST be set, otherwise a PCErr message MUST be sent with Error-Type=10 (Reception of an invalid object) Error-value=TBA-24 (LABEL-SET TLV present with O bit set but without R bit set in RP).

2.6. IRO Extension

The IRO as defined in [RFC5440] is used to include specific objects in the path. RSVP-TE allows the inclusion of a label definition. In order to fulfill requirement 13 of [RFC7025] the IRO needs to support the new subobject type as defined in [RFC3473]:

Type	Sub-object
TBA-38	LABEL

The Label subobject MUST follow a subobject identifying a link, currently an IP address subobject (Type 1 or 2) or an interface ID (type 4) subobject. If an IP address subobject is used, then the given IP address MUST be associated with a link. More than one label subobject MAY follow each link subobject. The procedure associated with this subobject is as follows.

If the PCE is able to allocate labels (e.g., via explicit label control) the PCE MUST allocate one label from within the set of label values for the given link. If the PCE does not assign labels, then it sends a response with a NO-PATH object, containing a NO-PATH-VECTOR TLV with the bit 'No label resource in range' set.

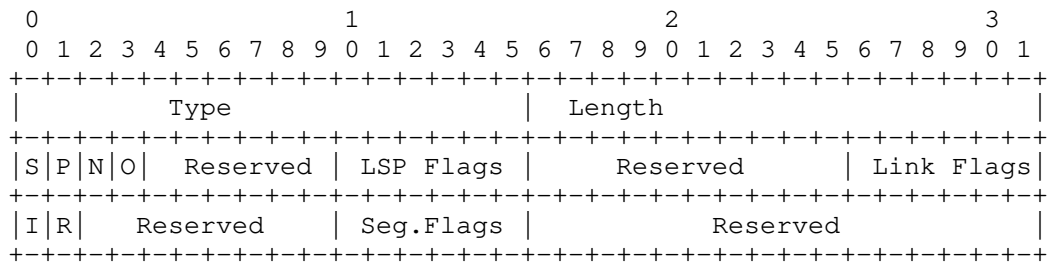
2.7. XRO Extension

The XRO as defined in [RFC5521] is used to exclude specific objects in the path. RSVP-TE allows the exclusion of certain labels ([RFC6001]). In order to fulfill requirement 13 of [RFC7025] Section 3.1, the PCEP's XRO needs to support a new subobject to enable label exclusion.

The encoding of the XRO Label subobject follows the encoding of the Label ERO subobject defined in [RFC3473] and XRO subobject defined in

2.8. LSPA Extensions

The LSPA carries the LSP attributes. In the end-to-end recovery context, this also includes the protection state information. A new TLV is defined to fulfil requirement 7 of [RFC7025] Section 3.1 and requirement 3 of [RFC7025] Section 3.2. This TLV contains the information of the PROTECTION object defined by [RFC4872] and can be used as a policy input. The LSPA object MAY carry a PROTECTION-ATTRIBUTE TLV defined as: Type TBA-12: PROTECTION-ATTRIBUTE



The content is as defined in [RFC4872] Section 14, [RFC4873] Section 6.1.

LSP (protection) Flags or Link flags field can be used by a PCE implementation for routing policy input. The other attributes are only meaningful for a stateful PCE.

This TLV is OPTIONAL and MAY be ignored by the PCE. If ignored by the PCE, it MUST NOT include the TLV in the LSPA of the response. When the TLV is used by the PCE, a LSPA object and the PROTECTION-ATTRIBUTE TLV MUST be included in the response. Fields that were not considered MUST be set to 0.

2.9. NO-PATH Object Extension

The NO-PATH object is used in PCRep messages in response to an unsuccessful path computation request (the PCE could not find a path satisfying the set of constraints). In this scenario, PCE MUST include a NO-PATH object in the PCRep message. The NO-PATH object MAY carry the NO-PATH-VECTOR TLV that specifies more information on the reasons that led to a negative reply. In case of GMPLS networks there could be some additional constraints that led to the failure such as protection mismatch, lack of resources, and so on. Several new flags have been defined in the 32-bit flag field of the NO-PATH-VECTOR TLV but no modifications have been made in the NO-PATH object.

2.9.1. Extensions to NO-PATH-VECTOR TLV

The modified NO-PATH-VECTOR TLV carrying the additional information is as follows:

Bit number TBA-32 - Protection Mismatch (1-bit). Specifies the mismatch of the protection type in the PROTECTION-ATTRIBUTE TLV in the request.

Bit number TBA-33 - No Resource (1-bit). Specifies that the resources are not currently sufficient to provide the path.

Bit number TBA-34 - Granularity not supported (1-bit). Specifies that the PCE is not able to provide a path with the requested granularity.

Bit number TBA-35 - No endpoint label resource (1-bit). Specifies that the PCE is not able to provide a path because of the endpoint label restriction.

Bit number TBA-36 - No endpoint label resource in range (1-bit). Specifies that the PCE is not able to provide a path because of the endpoint label set restriction.

Bit number TBA-37 - No label resource in range (1-bit). Specifies that the PCE is not able to provide a path because of the label set restriction.

3. Additional Error-Types and Error-Values Defined

A PCEP-ERROR object is used to report a PCEP error and is characterized by an Error-Type that specifies the type of error while Error-value that provides additional information about the error. An additional error type and several error values are defined to represent some of the errors related to the newly identified objects related to GMPLS networks. For each PCEP error, an Error-Type and an Error-value are defined. Error-Type 1 to 10 are already defined in [RFC5440]. Additional Error-values are defined for Error-Types 4 and 10. A new Error-Type is defined (value TBA-27).

The Error-Type TBA-27 (path computation failure) is used to reflect constraints not understood by the PCE, for instance when the PCE is not able to understand the generalized bandwidth. If the constraints are understood, but the PCE is unable to find with those constraints, the NO-PATH is to be used.

Error-Type Error-value

4	Not supported object
	value=TBA-14: Bandwidth Object type TBA-2 or TBA-3 not supported
	value=TBA-15: Unsupported endpoint type in END-POINTS Generalized Endpoint object type
	value=TBA-16: Unsupported TLV present in END-POINTS Generalized Endpoint object type
	value=TBA-17: Unsupported granularity in the RP object flags
10	Reception of an invalid object
	value=TBA-18: Bad Bandwidth Object type TBA-2 (Generalized bandwidth) or TBA-3 (Generalized bandwidth of existing TE-LSP for which a reoptimization is requested)
	value=TBA-20: Unsupported LSP Protection Flags in PROTECTION-ATTRIBUTE TLV
	value=TBA-21: Unsupported Secondary LSP Protection Flags in PROTECTION-ATTRIBUTE TLV
	value=TBA-22: Unsupported Link Protection Type in PROTECTION-ATTRIBUTE TLV
	value=TBA-24: LABEL-SET TLV present with 0 bit set but without R bit set in RP
	value=TBA-25: Wrong LABEL-SET TLV present with 0 and L bit set
	value=TBA-26: Wrong LABEL-SET with 0 bit set and wrong format
	value=TBA-42: Missing GMPLS-CAPABILITY TLV
TBA-27	Path computation failure
	value=0: Unassigned
	value=TBA-28: Unacceptable request message
	value=TBA-29: Generalized bandwidth value not supported
	value=TBA-30: Label Set constraint could not be met
	value=TBA-31: Label constraint could not be met

4. Manageability Considerations

This section follows the guidance of [RFC6123].

4.1. Control of Function through Configuration and Policy

This document makes no change to the basic operation of PCEP and so the requirements described in [RFC5440] Section 8.1. also apply to this document. In addition to those requirements a PCEP implementation may allow the configuration of the following parameters:

- Accepted RG in the RP object.

- Default RG to use (overriding the one present in the PCReq)

- Accepted BANDWIDTH object type TBA-2 and TBA-3 parameters in request, default mapping to use when not specified in the request

- Accepted LOAD-BALANCING object type TBA-4 parameters in request.

- Accepted endpoint type and allowed TLVs in object END-POINTS with object type Generalized Endpoint.

- Accepted range for label restrictions in label restriction in END-POINTS, or IRO or XRO objects

- PROTECTION-ATTRIBUTE TLV acceptance and suppression.

The configuration of the above parameters is applicable to the different sessions as described in [RFC5440] Section 8.1 (by default, per PCEP peer, etc.).

4.2. Information and Data Models

This document makes no change to the basic operation of PCEP and so the requirements described in [RFC5440] Section 8.2. also apply to this document. This document does not introduce any new ERO sub objects, so that the, ERO information model is already covered in [RFC4802].

4.3. Liveness Detection and Monitoring

This document makes no change to the basic operation of PCEP and so there are no changes to the requirements for liveness detection and monitoring set out in [RFC4657] and [RFC5440] Section 8.3.

4.4. Verifying Correct Operation

This document makes no change to the basic operations of PCEP and considerations described in [RFC5440] Section 8.4. New errors defined by this document should satisfy the requirement to log error events.

4.5. Requirements on Other Protocols and Functional Components

No new Requirements on Other Protocols and Functional Components are made by this document. This document does not require ERO object extensions. Any new ERO subobject defined in the TEAS or CCAMP working group can be adopted without modifying the operations defined in this document.

4.6. Impact on Network Operation

This document makes no change to the basic operations of PCEP and considerations described in [RFC5440] Section 8.6. In addition to the limit on the rate of messages sent by a PCEP speaker, a limit MAY be placed on the size of the PCEP messages.

5. IANA Considerations

IANA assigns values to the PCEP objects and TLVs. IANA is requested to make some allocations for the newly defined objects and TLVs defined in this document. Also, IANA is requested to manage the space of flags that are newly added in the TLVs.

5.1. PCEP Objects

As described in Section 2.3, Section 2.4 and Section 2.5.1 new Objects types are defined. IANA is requested to make the following Object-Type allocations from the "PCEP Objects" sub-registry.

Object 5
Class
Name BANDWIDTH
Object-Type TBA-2: Generalized bandwidth
TBA-3: Generalized bandwidth of an existing TE-LSP for
which a reoptimization is requested
Reference This document (Section 2.3)

Object 14
Class
Name LOAD-BALANCING
Object-Type TBA-4: Generalized Load Balancing

Reference This document (Section 2.4)

Object 4
Class
Name END-POINTS
Object-Type TBA-5: Generalized Endpoint
Reference This document (Section 2.5)

5.2. Endpoint type field in Generalized END-POINTS Object

IANA is requested to create a registry to manage the Endpoint Type field of the END-POINTS object, Object Type Generalized Endpoint and manage the code space.

New endpoint type in the Reserved range are assigned by Standards Action [RFC8126]. Each endpoint type should be tracked with the following attributes:

- o Endpoint type
- o Description
- o Defining RFC

New endpoint type in the Experimental range are for experimental use; these will not be registered with IANA and MUST NOT be mentioned by RFCs.

The following values have been defined by this document.
(Section 2.5.1, Table 5):

Value	Type	Meaning
0	Point-to-Point	
1	Point-to-Multipoint	New leaves to add
2		Old leaves to remove
3		Old leaves whose path can be modified/reoptimized
4		Old leaves whose path has to be left unchanged
5-244	Unassigned	
245-255	Experimental range	

5.3. New PCEP TLVs

IANA manages the PCEP TLV code point registry (see [RFC5440]). This is maintained as the "PCEP TLV Type Indicators" sub-registry of the "Path Computation Element Protocol (PCEP) Numbers" registry. IANA is requested to do the following allocation. Note: TBA-11 is not used

Value	Meaning	Reference
TBA-6	IPV4-ADDRESS	This document (Section 2.5.2.1)
TBA-7	IPV6-ADDRESS	This document (Section 2.5.2.2)
TBA-8	UNNUMBERED-ENDPOINT	This document (Section 2.5.2.3)
TBA-9	LABEL-REQUEST	This document (Section 2.5.2.4)
TBA-10	LABEL-SET	This document (Section 2.5.2.5)
TBA-12	PROTECTION-ATTRIBUTE	This document (Section 2.8)
TBA-1	GMPLS-CAPABILITY	This document (Section 2.1.2)

5.4. RP Object Flag Field

As described in Section 2.2 new flag are defined in the RP Object Flag IANA is requested to make the following Object-Type allocations from the "RP Object Flag Field" sub-registry.

Bit	Description	Reference
TBA-13	routing granularity (2 bits) (RG)	This document, Section 2.2

5.5. New PCEP Error Codes

As described in Section 3, new PCEP Error-Types and Error-values are defined. IANA is requested to make the following allocation in the "PCEP-ERROR Object Error Types and Values" registry.

Error	name	Reference
Type=4	Not supported object	[RFC5440]
Value=TBA-14:	Bandwidth Object type TBA-2 or TBA-3 not supported	This Document
Value=TBA-15:	Unsupported endpoint type in END-POINTS Generalized Endpoint object type	This Document
Value=TBA-16:	Unsupported TLV present in END-POINTS Generalized Endpoint object type	This Document
Value=TBA-17:	Unsupported granularity in the RP object flags	This Document
Type=10	Reception of an invalid object	[RFC5440]
Value=TBA-18:	Bad Bandwidth Object type TBA-2 (Generalized bandwidth) or TBA-3 (Generalized bandwidth of existing TE-LSP for which a reoptimization is requested)	This Document
Value=TBA-20:	Unsupported LSP Protection Flags in PROTECTION-ATTRIBUTE TLV	This Document
Value=TBA-21:	Unsupported Secondary LSP Protection Flags in PROTECTION-ATTRIBUTE TLV	This Document
Value=TBA-22:	Unsupported Link Protection Type in PROTECTION-ATTRIBUTE TLV	This Document
Value=TBA-24:	LABEL-SET TLV present with 0 bit set but without R bit set in RP	This Document
Value=TBA-25:	Wrong LABEL-SET TLV present with 0 and L bit set	This Document
Value=TBA-26:	Wrong LABEL-SET with 0 bit set and wrong format	This Document
Value=TBA-42:	Missing GMPLS-CAPABILITY TLV	This Document
Type=TBA-27	Path computation failure	This Document
Value=0	Unassigned	This Document
Value=TBA-28:	Unacceptable request message	This Document
Value=TBA-29:	Generalized bandwidth value not supported	This Document
Value=TBA-30:	Label Set constraint could not be met	This Document
Value=TBA-31:	Label constraint could not be met	This Document

5.6. New NO-PATH-VECTOR TLV Fields

As described in Section 2.9.1, new NO-PATH-VECTOR TLV Flag Fields have been defined. IANA is requested to do the following allocations in the "NO-PATH-VECTOR TLV Flag Field" sub-registry.

Bit number TBA-32 - Protection Mismatch (1-bit). Specifies the mismatch of the protection type of the PROTECTION-ATTRIBUTE TLV in the request.

Bit number TBA-33 - No Resource (1-bit). Specifies that the resources are not currently sufficient to provide the path.

Bit number TBA-34 - Granularity not supported (1-bit). Specifies that the PCE is not able to provide a path with the requested granularity.

Bit number TBA-35 - No endpoint label resource (1-bit). Specifies that the PCE is not able to provide a path because of the endpoint label restriction.

Bit number TBA-36 - No endpoint label resource in range (1-bit). Specifies that the PCE is not able to provide a path because of the endpoint label set restriction.

Bit number TBA-37 - No label resource in range (1-bit). Specifies that the PCE is not able to provide a path because of the label set restriction.

Bit number TBA-40 - LOAD-BALANCING could not be performed with the bandwidth constraints (1 bit). Specifies that the PCE is not able to provide a path because it could not map the BANDWIDTH into the parameters specified by the LOAD-BALANCING.

5.7. New Subobject for the Include Route Object

The "PCEP Parameters" registry contains a subregistry "IRO Subobjects" with an entry for the Include Route Object (IRO).

IANA is requested to add a further subobject that can be carried in the IRO as follows:

Subobject type	Reference
TBA-38 Label subobject	This Document

5.8. New Subobject for the Exclude Route Object

The "PCEP Parameters" registry contains a subregistry "XRO Subobjects" with an entry for the XRO object (Exclude Route Object).

IANA is requested to add a further subobject that can be carried in the XRO as follows:

Subobject type	Reference
TBA-39 Label subobject	This Document

5.9. New GMPLS-CAPABILITY TLV Flag Field

IANA is requested to create a sub-registry to manage the Flag field of the GMPLS-CAPABILITY TLV within the "Path Computation Element Protocol (PCEP) Numbers" registry.

New bit numbers are to be assigned by Standards Action [RFC8126]. 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 initial contents of the sub-registry are empty, with all bits marked unassigned

6. Security Considerations

GMPLS controls multiple technologies and types of network elements. The LSPs that are established using GMPLS, whose paths can be computed using the PCEP extensions to support GMPLS described in this document, can carry a high volume of traffic and can be a critical part of a network infrastructure. The PCE can then play a key role in the use of the resources and in determining the physical paths of the LSPs and thus it is important to ensure the identity of PCE and PCC, as well as the communication channel. In many deployments there will be a completely isolated network where an external attack is of very low probability. However, there are other deployment cases in which the PCC-PCE communication can be more exposed and there could be more security considerations. Three main situations in case of an attack in the GMPLS PCE context could happen:

- o PCE Identity theft: A legitimate PCC could request a path for a GMPLS LSP to a malicious PCE, which poses as a legitimate PCE. The answer can make that the LSP traverses some geographical place known to the attacker where confidentiality (sniffing), integrity (traffic modification) or availability (traffic drop) attacks could be performed by use of an attacker-controlled middlebox device. Also, the resulting LSP can omit constraints given in the requests (e.g., excluding certain fibers, avoiding some SRLGs) which could make that the LSP which will be later set-up can look perfectly fine, but will be in a risky situation. Also, the result can lead to the creation of an LSP that does not provide the desired quality and gives less resources than necessary.

- o PCC Identity theft: A malicious PCC, acting as a legitimate PCC, requesting LSP paths to a legitimate PCE can obtain a good knowledge of the physical topology of a critical infrastructure. It could get to know enough details to plan a later physical attack.
- o Message inspection: As in the previous case, knowledge of an infrastructure can be obtained by sniffing PCEP messages.

The security mechanisms can provide authentication and confidentiality for those scenarios where the PCC-PCE communication cannot be completely trusted. [RFC8253] provides origin verification, message integrity and replay protection, and ensures that a third party cannot decipher the contents of a message.

In order to protect against the malicious PCE case the PCC SHOULD have policies in place to accept or not the path provided by the PCE. Those policies can verify if the path follows the provided constraints. In addition, technology specific data plane mechanism can be used (following [RFC5920] Section 5.8) to verify the data plane connectivity and deviation from constraints.

The document [RFC8253] describes the usage of Transport Layer Security (TLS) to enhance PCEP security. The document describes the initiation of the TLS procedures, the TLS handshake mechanisms, the TLS methods for peer authentication, the applicable TLS ciphersuites for data exchange, and the handling of errors in the security checks. PCE and PCC SHOULD use [RFC8253] mechanism to protect against malicious PCC and PCE.

Finally, as mentioned by [RFC7025] the PCEP extensions to support GMPLS should be considered under the same security as current PCE work and this extension will not change the underlying security issues. However, given the critical nature of the network infrastructures under control by GMPLS, the security issues described above should be seriously considered when deploying a GMPLS-PCE based control plane for such networks. For more information on the security considerations on a GMPLS control plane, not only related to PCE/PCEP, [RFC5920] provides an overview of security vulnerabilities of a GMPLS control plane.

7. Contributing Authors

Elie Sfeir
Coriant
St Martin Strasse 76
Munich, 81541
Germany

Email: elie.sfeir@coriant.com

Franz Rambach
Nockherstrasse 2-4,
Munich 81541
Germany

Phone: +49 178 8855738
Email: franz.rambach@cgi.com

Francisco Javier Jimenez Chico
Telefonica Investigacion y Desarrollo
C/ Emilio Vargas 6
Madrid, 28043
Spain

Phone: +34 91 3379037
Email: fjjc@tid.es

Huawei Technologies

Suresh BR
Shenzhen
China
Email: sureshbr@huawei.com

Young Lee
1700 Alma Drive, Suite 100
Plano, TX 75075
USA

Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

SenthilKumar S
Shenzhen
China
Email: senthilkumars@huawei.com

Jun Sun
Shenzhen
China
Email: johnsun@huawei.com

CTTC - Centre Tecnologic de Telecomunicacions de Catalunya

Ramon Casellas
PMT Ed B4 Av. Carl Friedrich Gauss 7
08860 Castelldefels (Barcelona)
Spain
Phone: (34) 936452916
Email: ramon.casellas@cttc.es

8. Acknowledgments

The research of Ramon Casellas, Francisco Javier Jimenez Chico, Oscar Gonzalez de Dios, Cyril Margaria, and Franz Rambach leading to these results has received funding from the European Community's Seventh Framework Program FP7/2007-2013 under grant agreement no 247674 and no 317999.

The authors would like to thank Julien Meuric, Lyndon Ong, Giada Lander, Jonathan Hardwick, Diego Lopez, David Sinicrope, Vincent Roca, Dhruv Dhody, Adrian Farrel and Tianran Zhou for their review and useful comments to the document.

Thanks to Alisa Cooper, Benjamin Kaduk, Elwun-davies, Martin Vigoureux, Roman Danyliw, and Suresh Krishnan for the IESG comments

9. References

9.1. Normative References

- [G.709-v3] ITU-T, "Interfaces for the optical transport network, Recommendation G.709/Y.1331", June 2016, <<https://www.itu.int/rec/T-REC-G.709-201606-I/en>>.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC2210] Wroclawski, J., "The Use of RSVP with IETF Integrated Services", RFC 2210, DOI 10.17487/RFC2210, September 1997, <<https://www.rfc-editor.org/info/rfc2210>>.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<https://www.rfc-editor.org/info/rfc3209>>.

- [RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, DOI 10.17487/RFC3471, January 2003, <<https://www.rfc-editor.org/info/rfc3471>>.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <<https://www.rfc-editor.org/info/rfc3473>>.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, DOI 10.17487/RFC3477, January 2003, <<https://www.rfc-editor.org/info/rfc3477>>.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <<https://www.rfc-editor.org/info/rfc3630>>.
- [RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", RFC 4003, DOI 10.17487/RFC4003, February 2005, <<https://www.rfc-editor.org/info/rfc4003>>.
- [RFC4328] Papadimitriou, D., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Extensions for G.709 Optical Transport Networks Control", RFC 4328, DOI 10.17487/RFC4328, January 2006, <<https://www.rfc-editor.org/info/rfc4328>>.
- [RFC4606] Mannie, E. and D. Papadimitriou, "Generalized Multi-Protocol Label Switching (GMPLS) Extensions for Synchronous Optical Network (SONET) and Synchronous Digital Hierarchy (SDH) Control", RFC 4606, DOI 10.17487/RFC4606, August 2006, <<https://www.rfc-editor.org/info/rfc4606>>.
- [RFC4802] Nadeau, T., Ed. and A. Farrel, Ed., "Generalized Multiprotocol Label Switching (GMPLS) Traffic Engineering Management Information Base", RFC 4802, DOI 10.17487/RFC4802, February 2007, <<https://www.rfc-editor.org/info/rfc4802>>.

- [RFC4872] Lang, J., Ed., Rekhter, Y., Ed., and D. Papadimitriou, Ed., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, DOI 10.17487/RFC4872, May 2007, <<https://www.rfc-editor.org/info/rfc4872>>.
- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, DOI 10.17487/RFC4873, May 2007, <<https://www.rfc-editor.org/info/rfc4873>>.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, DOI 10.17487/RFC5088, January 2008, <<https://www.rfc-editor.org/info/rfc5088>>.
- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, DOI 10.17487/RFC5089, January 2008, <<https://www.rfc-editor.org/info/rfc5089>>.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<https://www.rfc-editor.org/info/rfc5440>>.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, DOI 10.17487/RFC5511, April 2009, <<https://www.rfc-editor.org/info/rfc5511>>.
- [RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, DOI 10.17487/RFC5520, April 2009, <<https://www.rfc-editor.org/info/rfc5520>>.
- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, DOI 10.17487/RFC5521, April 2009, <<https://www.rfc-editor.org/info/rfc5521>>.
- [RFC5541] Le Roux, JL., Vasseur, JP., and Y. Lee, "Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)", RFC 5541, DOI 10.17487/RFC5541, June 2009, <<https://www.rfc-editor.org/info/rfc5541>>.

- [RFC6001] Papadimitriou, D., Vigoureux, M., Shiomoto, K., Brungard, D., and JL. Le Roux, "Generalized MPLS (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)", RFC 6001, DOI 10.17487/RFC6001, October 2010, <<https://www.rfc-editor.org/info/rfc6001>>.
- [RFC6003] Papadimitriou, D., "Ethernet Traffic Parameters", RFC 6003, DOI 10.17487/RFC6003, October 2010, <<https://www.rfc-editor.org/info/rfc6003>>.
- [RFC6205] Otani, T., Ed. and D. Li, Ed., "Generalized Labels for Lambda-Switch-Capable (LSC) Label Switching Routers", RFC 6205, DOI 10.17487/RFC6205, March 2011, <<https://www.rfc-editor.org/info/rfc6205>>.
- [RFC6387] Takacs, A., Berger, L., Caviglia, D., Fedyk, D., and J. Meuric, "GMPLS Asymmetric Bandwidth Bidirectional Label Switched Paths (LSPs)", RFC 6387, DOI 10.17487/RFC6387, September 2011, <<https://www.rfc-editor.org/info/rfc6387>>.
- [RFC7139] Zhang, F., Ed., Zhang, G., Belotti, S., Ceccarelli, D., and K. Pithewan, "GMPLS Signaling Extensions for Control of Evolving G.709 Optical Transport Networks", RFC 7139, DOI 10.17487/RFC7139, March 2014, <<https://www.rfc-editor.org/info/rfc7139>>.
- [RFC7570] Margaria, C., Ed., Martinelli, G., Balls, S., and B. Wright, "Label Switched Path (LSP) Attribute in the Explicit Route Object (ERO)", RFC 7570, DOI 10.17487/RFC7570, July 2015, <<https://www.rfc-editor.org/info/rfc7570>>.
- [RFC7792] Zhang, F., Zhang, X., Farrel, A., Gonzalez de Dios, O., and D. Ceccarelli, "RSVP-TE Signaling Extensions in Support of Flexi-Grid Dense Wavelength Division Multiplexing (DWDM) Networks", RFC 7792, DOI 10.17487/RFC7792, March 2016, <<https://www.rfc-editor.org/info/rfc7792>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

- [RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <<https://www.rfc-editor.org/info/rfc8253>>.
- [RFC8282] Oki, E., Takeda, T., Farrel, A., and F. Zhang, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Inter-Layer MPLS and GMPLS Traffic Engineering", RFC 8282, DOI 10.17487/RFC8282, December 2017, <<https://www.rfc-editor.org/info/rfc8282>>.
- [RFC8306] Zhao, Q., Dhody, D., Ed., Palleti, R., and D. King, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths", RFC 8306, DOI 10.17487/RFC8306, November 2017, <<https://www.rfc-editor.org/info/rfc8306>>.

9.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<https://www.rfc-editor.org/info/rfc4655>>.
- [RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <<https://www.rfc-editor.org/info/rfc4657>>.
- [RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, DOI 10.17487/RFC5920, July 2010, <<https://www.rfc-editor.org/info/rfc5920>>.
- [RFC6123] Farrel, A., "Inclusion of Manageability Sections in Path Computation Element (PCE) Working Group Drafts", RFC 6123, DOI 10.17487/RFC6123, February 2011, <<https://www.rfc-editor.org/info/rfc6123>>.
- [RFC6163] Lee, Y., Ed., Bernstein, G., Ed., and W. Imajuku, "Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs)", RFC 6163, DOI 10.17487/RFC6163, April 2011, <<https://www.rfc-editor.org/info/rfc6163>>.

- [RFC7025] Otani, T., Ogaki, K., Caviglia, D., Zhang, F., and C. Margaria, "Requirements for GMPLS Applications of PCE", RFC 7025, DOI 10.17487/RFC7025, September 2013, <<https://www.rfc-editor.org/info/rfc7025>>.
- [RFC7449] Lee, Y., Ed., Bernstein, G., Ed., Martensson, J., Takeda, T., Tsuritani, T., and O. Gonzalez de Dios, "Path Computation Element Communication Protocol (PCEP) Requirements for Wavelength Switched Optical Network (WSO) Routing and Wavelength Assignment", RFC 7449, DOI 10.17487/RFC7449, February 2015, <<https://www.rfc-editor.org/info/rfc7449>>.

Appendix A. LOAD-BALANCING Usage for SDH Virtual Concatenation

For example a request for one co-signaled $n \times$ VC-4 TE-LSP will not use the LOAD-BALANCING. In case the VC-4 components can use different paths, the BANDWIDTH with object type TBA-2 will contain a traffic specification indicating the complete $n \times$ VC-4 traffic specification and the LOAD-BALANCING the minimum co-signaled VC-4. For an SDH network, a request to have a TE-LSP group with 10 VC-4 containers, each path using at minimum 2 \times VC-4 containers, can be represented with a BANDWIDTH object with OT=TBA-2, Bw Spec Type set to 4, the content of the Generalized Bandwidth is ST=6, RCC=0, NCC=0, NVC=10, MT=1. The LOAD-BALANCING, OT=TBA-4 with Bw Spec Type set to 4, Max-LSP=5, Min Bandwidth Spec is (ST=6, RCC=0, NCC=0, NVC=2, MT=1). The PCE can respond with a response with maximum 5 paths, each of them having a BANDWIDTH OT=TBA-2 and Generalized Bandwidth matching the Min Bandwidth Spec from the LOAD-BALANCING object of the corresponding request.

Authors' Addresses

Cyril Margaria (editor)
Juniper

Email: cmargaria@juniper.net

Oscar Gonzalez de Dios (editor)
Telefonica Investigacion y Desarrollo
C/ Ronda de la Comunicacion
Madrid 28050
Spain

Phone: +34 91 4833441
Email: oscar.gonzalezdedios@telefonica.com

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

Email: zhangfatai@huawei.com

PCE Working Group
Internet-Draft
Intended status: Experimental
Expires: June 9, 2016

D. Dhody
U. Palle
Huawei Technologies
R. Casellas
CTTC
December 7, 2015

Domain Subobjects for Path Computation Element (PCE) Communication
Protocol (PCEP).
draft-ietf-pce-pcep-domain-sequence-12

Abstract

The ability to compute shortest constrained Traffic Engineering Label Switched Paths (TE LSPs) in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks across multiple domains has been identified as a key requirement. In this context, a domain is a collection of network elements within a common sphere of address management or path computational responsibility such as an Interior Gateway Protocol (IGP) area or an Autonomous System (AS). This document specifies a representation and encoding of a Domain-Sequence, which is defined as an ordered sequence of domains traversed to reach the destination domain to be used by Path Computation Elements (PCEs) to compute inter-domain constrained shortest paths across a predetermined sequence of domains. This document also defines new subobjects to be used to encode domain identifiers.

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 June 9, 2016.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Scope	4
1.2. Requirements Language	4
2. Terminology	5
3. Detail Description	6
3.1. Domains	6
3.2. Domain-Sequence	6
3.3. Domain-Sequence Representation	7
3.4. Include Route Object (IRO)	7
3.4.1. Subobjects	8
3.4.1.1. Autonomous system	8
3.4.1.2. IGP Area	9
3.4.2. Update in IRO specification	10
3.4.3. IRO for Domain-Sequence	10
3.4.3.1. PCC Procedures	11
3.4.3.2. PCE Procedures	11
3.5. Exclude Route Object (XRO)	12
3.5.1. Subobjects	13
3.5.1.1. Autonomous system	13
3.5.1.2. IGP Area	14
3.6. Explicit Exclusion Route Subobject (EXRS)	15
3.7. Explicit Route Object (ERO)	16
4. Examples	16
4.1. Inter-Area Path Computation	16
4.2. Inter-AS Path Computation	18
4.2.1. Example 1	19
4.2.2. Example 2	21
4.3. Boundary Node and Inter-AS-Link	23
4.4. PCE Serving multiple Domains	24
4.5. P2MP	24
4.6. Hierarchical PCE	26

5. Other Considerations	26
5.1. Relationship to PCE Sequence	26
5.2. Relationship to RSVP-TE	26
6. IANA Considerations	27
6.1. New Subobjects	27
7. Security Considerations	27
8. Manageability Considerations	28
8.1. Control of Function and Policy	28
8.2. Information and Data Models	28
8.3. Liveness Detection and Monitoring	29
8.4. Verify Correct Operations	29
8.5. Requirements On Other Protocols	29
8.6. Impact On Network Operations	29
9. Acknowledgments	29
10. References	30
10.1. Normative References	30
10.2. Informative References	31
Authors' Addresses	33

1. Introduction

A Path Computation Element (PCE) may be used to compute end-to-end paths across multi-domain environments using a per-domain path computation technique [RFC5152]. The backward recursive path computation (BRPC) mechanism [RFC5441] also defines a PCE-based path computation procedure to compute inter-domain constrained path for (G)MPLS TE LSPs. However, both per-domain and BRPC techniques assume that the sequence of domains to be crossed from source to destination is known, either fixed by the network operator or obtained by other means. Also for inter-domain point-to-multi-point (P2MP) tree computation, [RFC7334] assumes the domain-tree is known in priori.

The list of domains (Domain-Sequence) in point-to-point (P2P) or a domain tree in point-to-multipoint (P2MP) is usually a constraint in inter-domain path computation procedure.

The Domain-Sequence (the set of domains traversed to reach the destination domain) is either administratively predetermined or discovered by some means like H-PCE.

[RFC5440] defines the Include Route Object (IRO) and the Explicit Route Object (ERO). [RFC5521] defines the Exclude Route Object (XRO) and the Explicit Exclusion Route Subobject (EXRS). The use of Autonomous System (AS) (albeit with a 2-Byte AS number) as an abstract node representing a domain is defined in [RFC3209]. In the current document, we specify new subobjects to include or exclude domains including IGP area or an Autonomous Systems (4-Byte as per [RFC6793]).

Further, the domain identifier may simply act as delimiter to specify where the domain boundary starts and ends in some cases.

This is a companion document to Resource ReserVation Protocol - Traffic Engineering (RSVP-TE) extensions for the domain identifiers [DOMAIN-SUBOBJ].

1.1. Scope

The procedures described in this document are experimental. The experiment is intended to enable research for the usage of Domain-Sequence at the PCEs for inter-domain paths. For this purpose this document specifies new domain subobjects as well as how they incorporate with existing subobjects to represent a Domain-Sequence.

The experiment will end two years after the RFC is published. At that point, the RFC authors will attempt to determine how widely this has been implemented and deployed.

This document does not change the procedures for handling existing subobjects in PCEP.

The new subobjects introduced by this document will not be understood by legacy implementations. If a legacy implementation receives one of the subobjects that it does not understand in a PCEP object, the legacy implementation will behave as described in Section 3.4.3. Therefore, it is assumed that this experiment will be conducted only when both the PCE and the PCC form part of the experiment. It is possible that a PCC or PCE can operate with peers some of which form part of the experiment and some that do not. In this case, since no capabilities exchange is used to identify which nodes can use these extensions, manual configuration should be used to determine which peerings form part of the experiment.

When the result of implementation and deployment are available, this document will be updated and refined, and then be moved from Experimental to Standard Track.

1.2. Requirements Language

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

2. Terminology

The following terminology is used in this document.

ABR: OSPF Area Border Router. Routers used to connect two IGP areas.

AS: Autonomous System.

ASBR: Autonomous System Boundary Router.

BN: Boundary Node, Can be an ABR or ASBR.

BRPC: Backward Recursive Path Computation

Domain: As per [RFC4655], any collection of network elements within a common sphere of address management or path computational responsibility. Examples of domains include Interior Gateway Protocol (IGP) area and Autonomous System (AS).

Domain-Sequence: An ordered sequence of domains traversed to reach the destination domain.

ERO: Explicit Route Object

H-PCE: Hierarchical PCE

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

IRO: Include Route Object

IS-IS: Intermediate System to Intermediate System.

OSPF: Open Shortest Path First.

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

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

P2MP: Point-to-Multipoint

P2P: Point-to-Point

RSVP: Resource Reservation Protocol

TE LSP: Traffic Engineering Label Switched Path.

XRO: Exclude Route Object

3. Detail Description

3.1. Domains

[RFC4726] and [RFC4655] define domain as a separate administrative or geographic environment within the network. A domain could be further defined as a zone of routing or computational ability. Under these definitions a domain might be categorized as an AS or an IGP area. Each AS can be made of several IGP areas. In order to encode a Domain-Sequence, it is required to uniquely identify a domain in the Domain-Sequence. A domain can be uniquely identified by area-id or AS number or both.

3.2. Domain-Sequence

A Domain-Sequence is an ordered sequence of domains traversed to reach the destination domain.

A Domain-Sequence can be applied as a constraint and carried in a path computation request to PCE(s). A Domain-Sequence can also be the result of a path computation. For example, in the case of Hierarchical PCE (H-PCE) [RFC6805], Parent PCE could send the Domain-Sequence as a result in a path computation reply.

In a P2P path, the domains listed appear in the order that they are crossed. In a P2MP path, the domain tree is represented as a list of Domain-Sequences.

A Domain-Sequence enables a PCE to select the next domain and the PCE serving that domain to forward the path computation request based on the domain information.

Domain-Sequence can include Boundary Nodes (ABR or ASBR) or Border links (Inter-AS-links) to be traversed as an additional constraint.

Thus a Domain-Sequence can be made up of one or more of -

- o AS Number
- o Area ID
- o Boundary Node ID

- o Inter-AS-Link Address

These are encoded in the new subobjects defined in this document as well as the existing subobjects to represent a Domain-Sequence.

Consequently, a Domain-Sequence can be used:

1. by a PCE in order to discover or select the next PCE in a collaborative path computation, such as in BRPC [RFC5441];
2. by the Parent PCE to return the Domain-Sequence when unknown; this can then be an input to the BRPC procedure [RFC6805];
3. by a Path Computation Client (PCC) or a PCE, to constrain the domains used in inter-domain path computation, explicitly specifying which domains to be expanded or excluded;
4. by a PCE in the per-domain path computation model [RFC5152] to identify the next domain.

3.3. Domain-Sequence Representation

Domain-Sequence appears in PCEP messages, notably in -

- o Include Route Object (IRO): As per [RFC5440], IRO can be used to specify a set of network elements to be traversed to reach the destination, which includes subobjects used to specify the Domain-Sequence.
- o Exclude Route Object (XRO): As per [RFC5521], XRO can be used to specify certain abstract nodes, to be excluded from whole path, which includes subobjects used to specify the Domain-Sequence.
- o Explicit Exclusion Route Subobject (EXRS): As per [RFC5521], EXRS can be used to specify exclusion of certain abstract nodes (including domains) between a specific pair of nodes. EXRS are a subobject inside the IRO.
- o Explicit Route Object (ERO): As per [RFC5440], ERO can be used to specify a computed path in the network. For example, in the case of H-PCE [RFC6805], a Parent PCE can send the Domain-Sequence as a result, in a path computation reply using ERO.

3.4. Include Route Object (IRO)

As per [RFC5440], IRO (Include Route Object) can be used to specify that the computed path needs to traverse a set of specified network elements or abstract nodes.

3.4.1. Subobjects

Some subobjects are defined in [RFC3209], [RFC3473], [RFC3477] and [RFC4874], but new subobjects related to Domain-Sequence are needed.

This document extends the support for 4-Byte AS numbers and IGP Areas.

Type	Subobject
TBD1	Autonomous system number (4 Byte)
TBD2	OSPF Area id
TBD3	ISIS Area id

Note: The twins of these subobjects are carried in RSVP-TE messages as defined in [DOMAIN-SUBOBJ].

3.4.1.1. Autonomous system

[RFC3209] already defines 2 byte AS number.

To support 4 byte AS number as per [RFC6793] following subobject is defined:

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
+	+	+	+
L	Type	Length	Reserved
+	+	+	+
	AS-ID (4 bytes)		
+	+	+	+

L: The L bit is an attribute of the subobject as defined in [RFC3209] and usage in IRO subobject updated in [IRO-UPDATE].

Type: (TBD1 by IANA) indicating a 4-Byte AS Number.

Length: 8 (Total length of the subobject in bytes).

Reserved: Zero at transmission, ignored at receipt.

AS-ID: The 4-Byte AS Number. Note that if 2-Byte AS numbers are in use, the low order bits (16 through 31) MUST be used and the high order bits (0 through 15) MUST be set to zero.

3.4.1.2. IGP Area

Since the length and format of Area-id is different for OSPF and ISIS, following two subobjects are defined:

For OSPF, the area-id is a 32 bit number. The subobject is encoded as follows:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|L|      Type      |      Length      |      Reserved      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     OSPF Area Id (4 bytes)                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

L: The L bit is an attribute of the subobject as defined in [RFC3209] and usage in IRO subobject updated in [IRO-UPDATE].

Type: (TBD2 by IANA) indicating a 4-Byte OSPF Area ID.

Length: 8 (Total length of the subobject in bytes).

Reserved: Zero at transmission, ignored at receipt.

OSPF Area Id: The 4-Byte OSPF Area ID.

For IS-IS, the area-id is of variable length and thus the length of the Subobject is variable. The Area-id is as described in IS-IS by ISO standard [ISO10589]. The subobject is encoded as follows:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|L|      Type      |      Length      |      Area-Len      |      Reserved      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     IS-IS Area ID                                     |
//                                     //
|                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

L: The L bit is an attribute of the subobject as defined in [RFC3209] and usage in IRO subobject updated in [IRO-UPDATE].

Type: (TBD3 by IANA) indicating IS-IS Area ID.

Length: Variable. The Length MUST be at least 8, and MUST be a multiple of 4.

Area-Len: Variable (Length of the actual (non-padded) IS-IS Area Identifier in octets; Valid values are from 1 to 13 inclusive).

Reserved: Zero at transmission, ignored at receipt.

IS-IS Area Id: The variable-length IS-IS area identifier. Padded with trailing zeroes to a four-byte boundary.

3.4.2. Update in IRO specification

[RFC5440] describes IRO as an optional object used to specify network elements to be traversed by the computed path. It further states that the L bit of such subobject has no meaning within an IRO. It also did not mention if IRO is an ordered or un-ordered list of subobjects.

An update to IRO specification [IRO-UPDATE] makes IRO as an ordered list, as well as support for loose bit (L-bit) is added.

The use of IRO for Domain-Sequence, assumes the updated specification for IRO, as per [IRO-UPDATE].

3.4.3. IRO for Domain-Sequence

The subobject type for IPv4, IPv6, and unnumbered Interface ID can be used to specify Boundary Nodes (ABR/ASBR) and Inter-AS-Links. The subobject type for the AS Number (2 or 4 Byte) and the IGP Area are used to specify the domain identifiers in the Domain-Sequence.

The IRO can incorporate the new domain subobjects with the existing subobjects in a sequence of traversal.

Thus an IRO, comprising subobjects, that represents a Domain-Sequence, defines the domains involved in an inter-domain path computation, typically involving two or more collaborative PCEs.

A Domain-Sequence can have varying degrees of granularity. It is possible to have a Domain-Sequence composed of, uniquely, AS identifiers. It is also possible to list the involved IGP areas for a given AS.

In any case, the mapping between domains and responsible PCEs is not defined in this document. It is assumed that a PCE that needs to obtain a "next PCE" from a Domain-Sequence is able to do so (e.g. via administrative configuration, or discovery).

3.4.3.1. PCC Procedures

A PCC builds an IRO to encode the Domain-Sequence, so that the cooperating PCEs could compute an inter-domain shortest constrained path across the specified sequence of domains.

A PCC may intersperse Area and AS subobjects with other subobjects without change to the previously specified processing of those subobjects in the IRO.

3.4.3.2. PCE Procedures

If a PCE receives an IRO in a Path Computation request (PCReq) message that contains the subobjects defined in this document, that it does not recognize, it will respond according to the rules for a malformed object as per [RFC5440]. The PCE MAY also include the IRO in the PCErr message as per [RFC5440].

The interpretation of Loose bit (L bit) is as per section 4.3.3.1 of [RFC3209] (as per [IRO-UPDATE]).

In a Path Computation reply (PCRep), PCE MAY also supply IRO (with Domain-Sequence information) with the NO-PATH object indicating that the set of elements (domains) of the request's IRO prevented the PCEs from finding a path.

The following processing rules apply for Domain-Sequence in IRO -

- o When a PCE parses an IRO, it interprets each subobject according to the AS number associated with the preceding subobject. We call this the "current AS". Certain subobjects modify the current AS, as follows.
 - * The current AS is initialized to the AS number of the PCC.
 - * If the PCE encounters an AS subobject, then it updates the current AS to this new AS number.
 - * If the PCE encounters an Area subobject, then it assumes that the area belongs to the current AS.
 - * If the PCE encounters an IP address that is globally routable, then it updates the current AS to the AS that owns this IP address. This document does not define how the PCE learns which AS owns the IP address.
 - * If the PCE encounters an IP address that is not globally routable, then it assumes that it belongs to the current AS.

- * If the PCE encounters an unnumbered link, then it assumes that it belongs to the current AS.
- o When a PCE parses an IRO, it interprets each subobject according to the Area ID associated with the preceding subobject. We call this the "current Area". Certain subobjects modify the current Area, as follows.
 - * The current Area is initialized to the Area ID of the PCC.
 - * If the current AS is changed, the current Area is reset and need to be determined again by current or subsequent subobject.
 - * If the PCE encounters an Area subobject, then it updates the current Area to this new Area ID.
 - * If the PCE encounters an IP address that belongs to a different area, then it updates the current Area to the Area that has this IP address. This document does not define how the PCE learns which Area has the IP address.
 - * If the PCE encounters an unnumbered link that belongs to a different area, then it updates the current Area to the Area that has this link.
 - * Otherwise, it assumes that the subobject belongs to the current Area.
- o In case the current PCE is not responsible for the path computation in the current AS or Area, then the PCE selects the "next PCE" in the domain-sequence based on the current AS and Area.

Note that it is advised that, PCC should use AS and Area subobject while building the domain-sequence in IRO and avoid using other mechanism to change the "current AS" and "current Area" as described above.

3.5. Exclude Route Object (XRO)

The Exclude Route Object (XRO) [RFC5521] is an optional object used to specify exclusion of certain abstract nodes or resources from the whole path.

3.5.1. Subobjects

Some subobjects to be used in XRO as defined in [RFC3209], [RFC3477], [RFC4874], and [RFC5520], but new subobjects related to Domain-Sequence are needed.

This document extends the support for 4-Byte AS numbers and IGP Areas.

Type	Subobject
TBD1	Autonomous system number (4 Byte)
TBD2	OSPF Area id
TBD3	ISIS Area id

Note: The twins of these subobjects are carried in RSVP-TE messages as defined in [DOMAIN-SUBOBJ].

3.5.1.1. Autonomous system

The new subobjects to support 4 byte AS and IGP (OSPF / ISIS) Area MAY also be used in the XRO to specify exclusion of certain domains in the path computation procedure.

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								
+-----+																																							

The X-bit indicates whether the exclusion is mandatory or desired.

0: indicates that the AS specified MUST be excluded from the path computed by the PCE(s).

1: indicates that the AS specified SHOULD be avoided from the inter-domain path computed by the PCE(s), but MAY be included subject to PCE policy and the absence of a viable path that meets the other constraints.

All other fields are consistent with the definition in Section 3.4.

3.5.1.2. IGP Area

Since the length and format of Area-id is different for OSPF and ISIS, following two subobjects are defined:

For OSPF, the area-id is a 32 bit number. The subobject is encoded as follows:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|X|      Type      |      Length      |      Reserved      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     OSPF Area Id (4 bytes)                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

The X-bit indicates whether the exclusion is mandatory or desired.

0: indicates that the OSFF Area specified MUST be excluded from the path computed by the PCE(s).

1: indicates that the OSFF Area specified SHOULD be avoided from the inter-domain path computed by the PCE(s), but MAY be included subject to PCE policy and the absence of a viable path that meets the other constraints.

All other fields are consistent with the definition in Section 3.4.

For IS-IS, the area-id is of variable length and thus the length of the subobject is variable. The Area-id is as described in IS-IS by ISO standard [ISO10589]. The subobject is encoded as follows:

```

      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
+-----+-----+-----+-----+-----+-----+-----+-----+
|X|      Type      |      Length      | Area-Len      | Reserved      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     IS-IS Area ID                                     |
//                                     //
|                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

The X-bit indicates whether the exclusion is mandatory or desired.

0: indicates that the ISIS Area specified MUST be excluded from the path computed by the PCE(s).

1: indicates that the ISIS Area specified SHOULD be avoided from the inter-domain path computed by the PCE(s), but MAY be included subject to PCE policy and the absence of a viable path that meets the other constraints.

All other fields are consistent with the definition in Section 3.4.

All the processing rules are as per [RFC5521].

Note that, if a PCE receives an XRO in a PCReq message that contains subobjects defined in this document, that it does not recognize, it will respond according to the rules for a malformed object as per [RFC5440].

IGP Area subobjects in the XRO are local to the current AS. In case of multi-AS path computation to exclude an IGP area in a different AS, IGP Area subobject should be part of Explicit Exclusion Route Subobject (EXRS) in the IRO to specify the AS in which the IGP area is to be excluded. Further policy may be applied to prune/ignore Area subobjects in XRO after "current AS" change during path computation.

3.6. Explicit Exclusion Route Subobject (EXRS)

EXRS [RFC5521] is used to specify exclusion of certain abstract nodes between a specific pair of nodes.

The EXRS subobject can carry any of the subobjects defined for inclusion in the XRO, thus the new subobjects to support 4 byte AS and IGP (OSPF / ISIS) Area can also be used in the EXRS. The meanings of the fields of the new XRO subobjects are unchanged when the subobjects are included in an EXRS, except that scope of the exclusion is limited to the single hop between the previous and subsequent elements in the IRO.

The EXRS subobject should be interpreted in the context of the current AS and current Area of the preceding subobject in the IRO. The EXRS subobject does not change the current AS or current Area. All other processing rules are as per [RFC5521].

Note that, if a PCE that supports the EXRS in an IRO, parses an IRO, and encounters an EXRS that contains subobjects defined in this document, that it does not recognize, it will act according to the setting of the X-bit in the subobject as per [RFC5521].

3.7. Explicit Route Object (ERO)

The Explicit Route Object (ERO) [RFC5440] is used to specify a computed path in the network. PCEP ERO subobject types correspond to RSVP-TE ERO subobject types as defined in [RFC3209], [RFC3473], [RFC3477], [RFC4873], [RFC4874], and [RFC5520]. The subobjects related to Domain-Sequence are further defined in [DOMAIN-SUBOBJ].

The new subobjects to support 4 byte AS and IGP (OSPF / ISIS) Area can also be used in the ERO to specify an abstract node (a group of nodes whose internal topology is opaque to the ingress node of the LSP). Using this concept of abstraction, an explicitly routed LSP can be specified as a sequence of domains.

In case of Hierarchical PCE [RFC6805], a Parent PCE can be requested to find the Domain-Sequence. Refer example in Section 4.6. The ERO in reply from parent PCE can then be used in Per-Domain path computation or BRPC.

If a PCC receives an ERO in a PCRep message that contains subobject defined in this document, that it does not recognize, it will respond according to the rules for a malformed object as per [RFC5440].

4. Examples

The examples in this section are for illustration purposes only; to highlight how the new subobjects could be encoded. They are not meant to be an exhaustive list of all possible usecases and combinations.

4.1. Inter-Area Path Computation

In an inter-area path computation where the ingress and the egress nodes belong to different IGP areas within the same AS, the Domain-Sequence could be represented using a ordered list of Area subobjects.

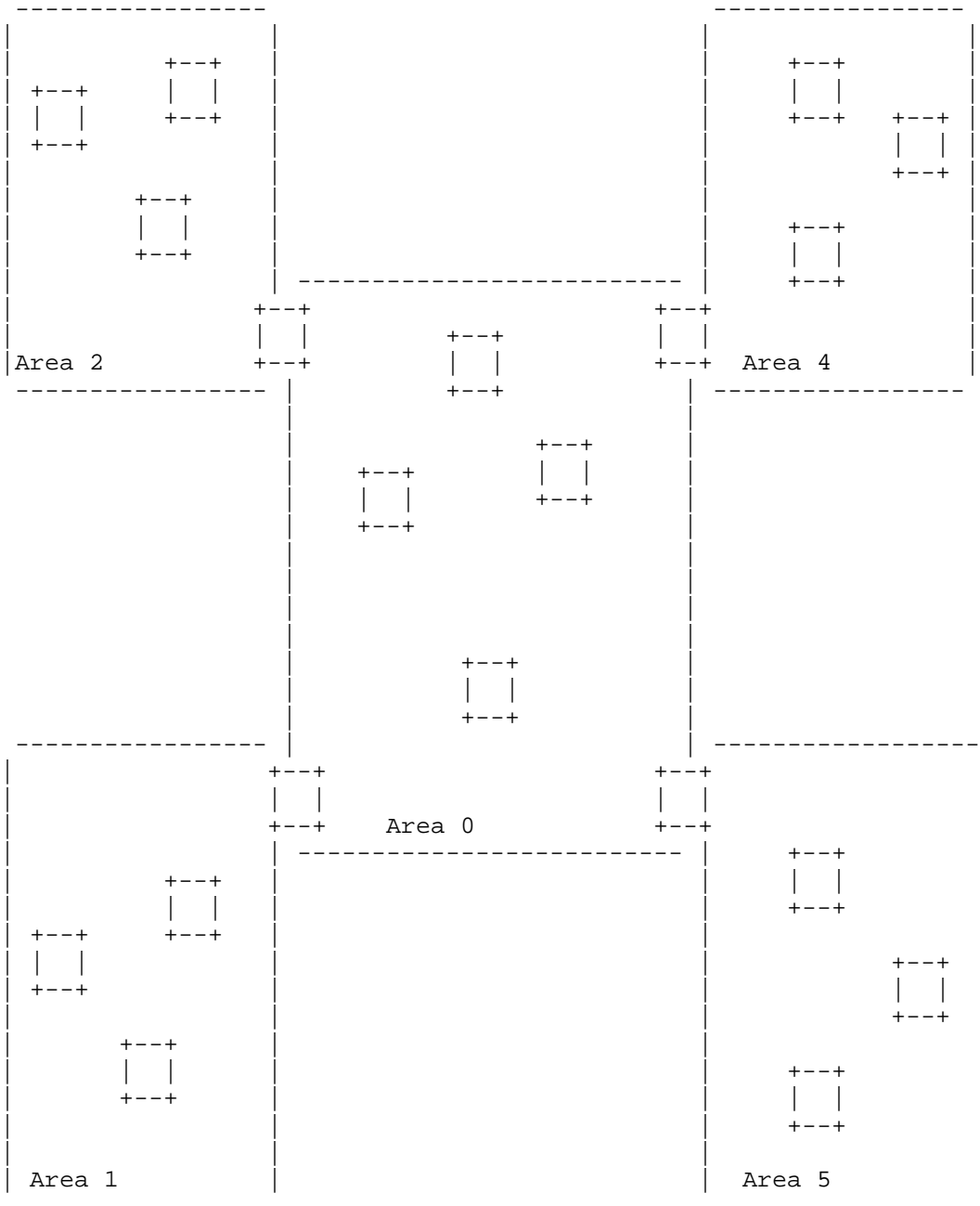


Figure 1: Inter-Area Path Computation

AS Number is 100.

If the ingress is in Area 2, egress in Area 4 and transit through Area 0. Some possible way a PCC can encode the IRO:

-----+	-----+	-----+
IRO	Sub	Sub
Object	Object	Object
Header	Area 0	Area 4
-----+	-----+	-----+

or

-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub
Object	Object	Object	Object
Header	Area 2	Area 0	Area 4
-----+	-----+	-----+	-----+

or

-----+	-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub	Sub
Object	Object AS	Object	Object	Object
Header	100	Area 2	Area 0	Area 4
-----+	-----+	-----+	-----+	-----+

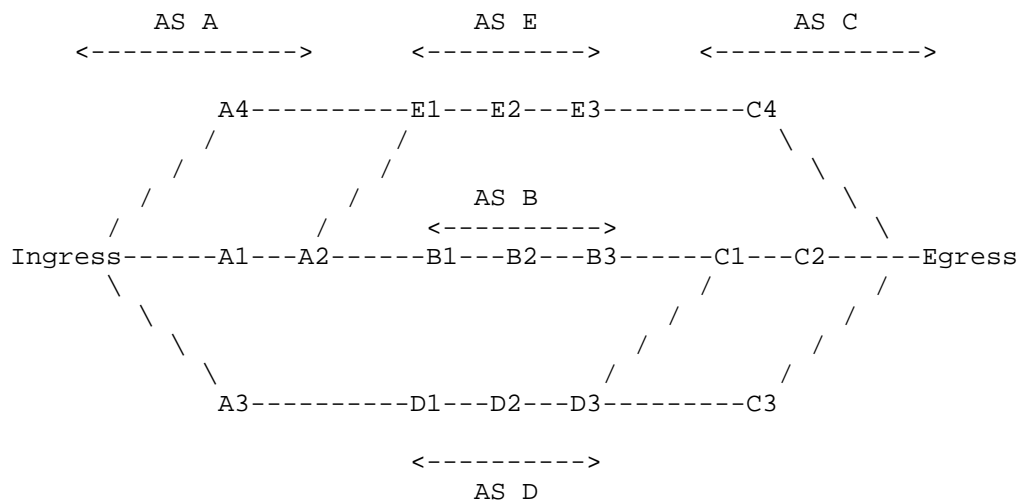
The Domain-Sequence can further include encompassing AS information in the AS subobject.

4.2. Inter-AS Path Computation

In inter-AS path computation, where ingress and egress belong to different AS, the Domain-Sequence could be represented using an ordered list of AS subobjects. The Domain-Sequence can further include decomposed area information in the Area subobject.

4.2.1. Example 1

As shown in Figure 2, where AS has a single area, AS subobject in the domain-sequence can uniquely identify the next domain and PCE.



- * All AS have one area (area 0)

Figure 2: Inter-AS Path Computation

If the ingress is in AS A, egress in AS C and transit through AS B. Some possible way a PCC can encode the IRO:

-----+	-----+	-----+
IRO	Sub	Sub
Object	Object	Object
Header	AS B	AS C
-----+	-----+	-----+

or

-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub
Object	Object	Object	Object
Header	AS A	AS B	AS C
-----+	-----+	-----+	-----+

or

-----+	-----+	-----+	-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub	Sub	Sub	Sub
Object	Object	Object	Object	Object	Object	Object
Header	AS A	Area 0	AS B	Area 0	AS C	Area 0
-----+	-----+	-----+	-----+	-----+	-----+	-----+

Note that to get a domain disjoint path, the ingress could also request the backup path with -

-----+	-----+
XRO	Sub
Object	Object
Header	AS B
-----+	-----+

As described in Section 3.4.3, domain subobject in IRO changes the domain information associated with the next set of subobjects; till you encounter a subobject that changes the domain too. Consider the following IRO:

-----+	-----+	-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub	Sub	Sub
Object	Object	Object	Object	Object	Object
Header	AS B	IP	IP	AS C	IP
-----+	-----+	B1	B3	-----+	C1
-----+	-----+	-----+	-----+	-----+	-----+

On processing subobject "AS B", it changes the AS of the subsequent subobjects till we encounter another subobject "AS C" which changes the AS for its subsequent subobjects.

Consider another IRO:

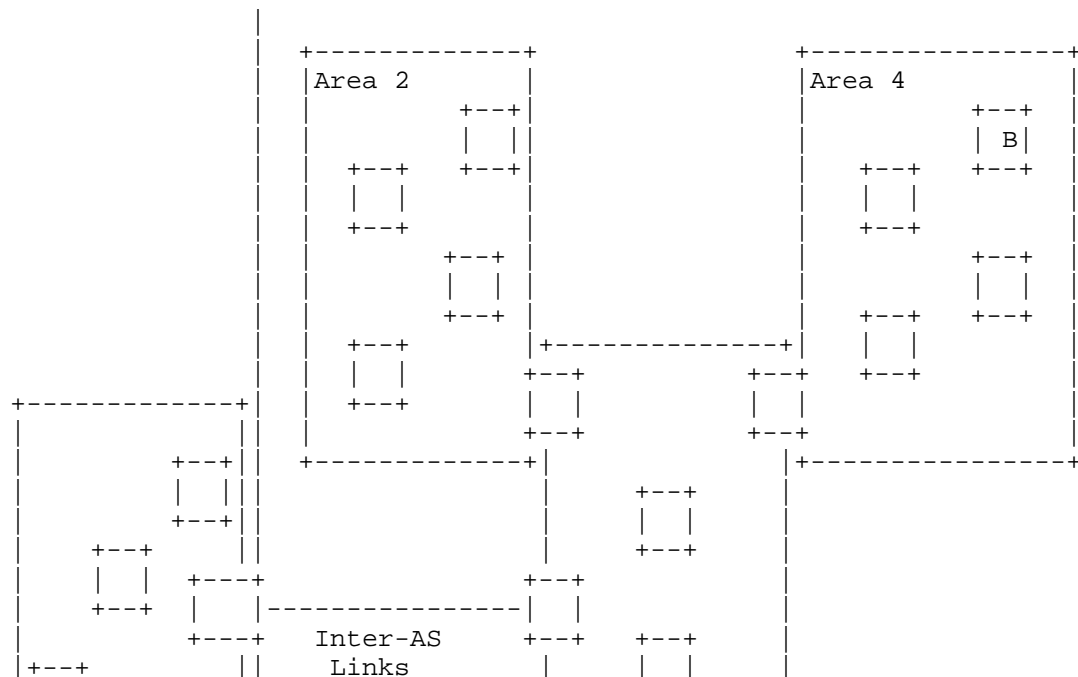
+-----+	+-----+	+-----+	+-----+	+-----+
IRO	Sub	Sub	Sub	Sub
Object	Object	Object	Object	Object
Header	AS D	IP	IP	IP
		D1	D3	C3
+-----+	+-----+	+-----+	+-----+	+-----+

Here as well, on processing "AS D", it changes the AS of the subsequent subobjects till you encounter another subobject "C3" which belong in another AS and changes the AS for its subsequent subobjects.

Further description for the Boundary Node and Inter-AS-Link can be found in Section 4.3.

4.2.2. Example 2

In Figure 3, AS 200 is made up of multiple areas.



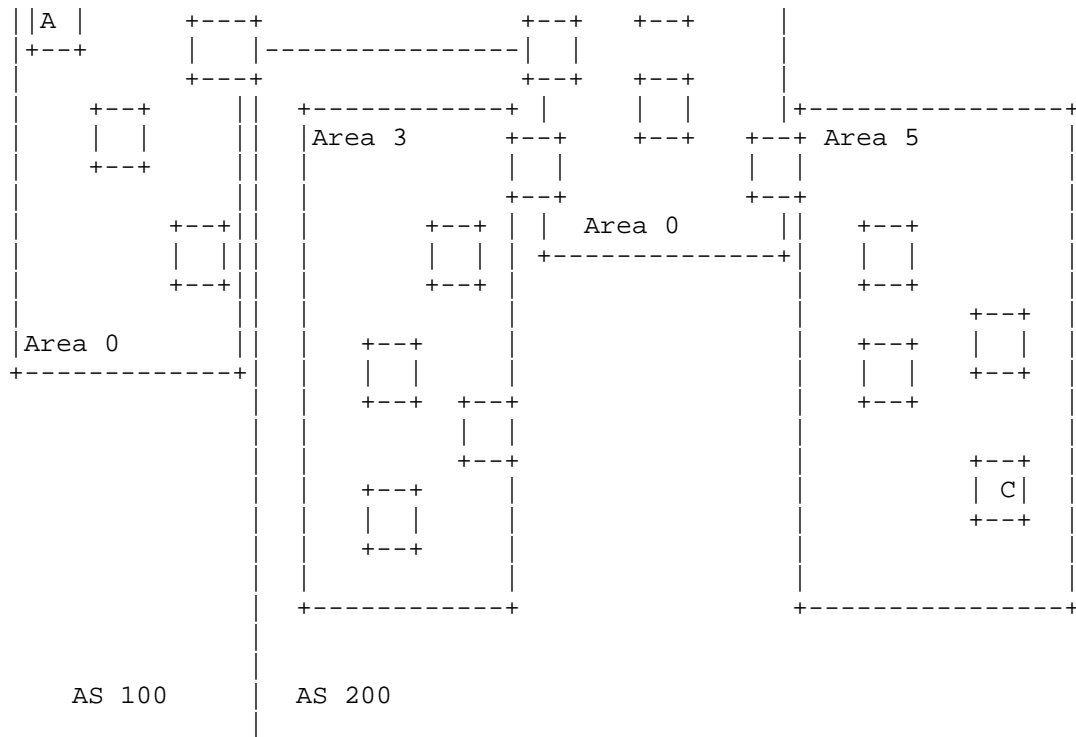


Figure 3: Inter-AS Path Computation

For LSP (A-B), where ingress A is in (AS 100, Area 0), egress B in (AS 200, Area 4) and transit through (AS 200, Area 0). Some possible way a PCC can encode the IRO:

IRO	Sub	Sub	Sub
Object	Object	Object	Object
Header	AS 200	Area 0	Area 4

or

IRO	Sub	Sub	Sub	Sub	Sub
Object	Object	Object	Object	Object	Object
Header	AS 100	Area 0	AS 200	Area 0	Area 4

For LSP (A-C), where ingress A is in (AS 100, Area 0), egress C in (AS 200, Area 5) and transit through (AS 200, Area 0). Some possible way a PCC can encode the IRO:

-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub
Object	Object	Object	Object
Header	AS 200	Area 0	Area 5
-----+	-----+	-----+	-----+

or

-----+	-----+	-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub	Sub	Sub
Object	Object	Object	Object	Object	Object
Header	AS 100	Area 0	AS 200	Area 0	Area 5
-----+	-----+	-----+	-----+	-----+	-----+

4.3. Boundary Node and Inter-AS-Link

A PCC or PCE can include additional constraints covering which Boundary Nodes (ABR or ASBR) or Border links (Inter-AS-link) to be traversed while defining a Domain-Sequence. In which case the Boundary Node or Link can be encoded as a part of the Domain-Sequence.

Boundary Nodes (ABR / ASBR) can be encoded using the IPv4 or IPv6 prefix subobjects usually the loopback address of 32 and 128 prefix length respectively. An Inter-AS link can be encoded using the IPv4 or IPv6 prefix subobjects or unnumbered interface subobjects.

For Figure 1, an ABR (say 203.0.113.1) to be traversed can be specified in IRO as:

-----+	-----+	-----+	-----+	-----+
IRO	Sub	Sub	Sub	Sub
Object	Object	Object	Object	Object
Header	Area 2	IPv4	Area 0	Area 4
		203.0.		
		112.1		
-----+	-----+	-----+	-----+	-----+

For Figure 3, an inter-AS-link (say 198.51.100.1 - 198.51.100.2) to be traversed can be specified as:

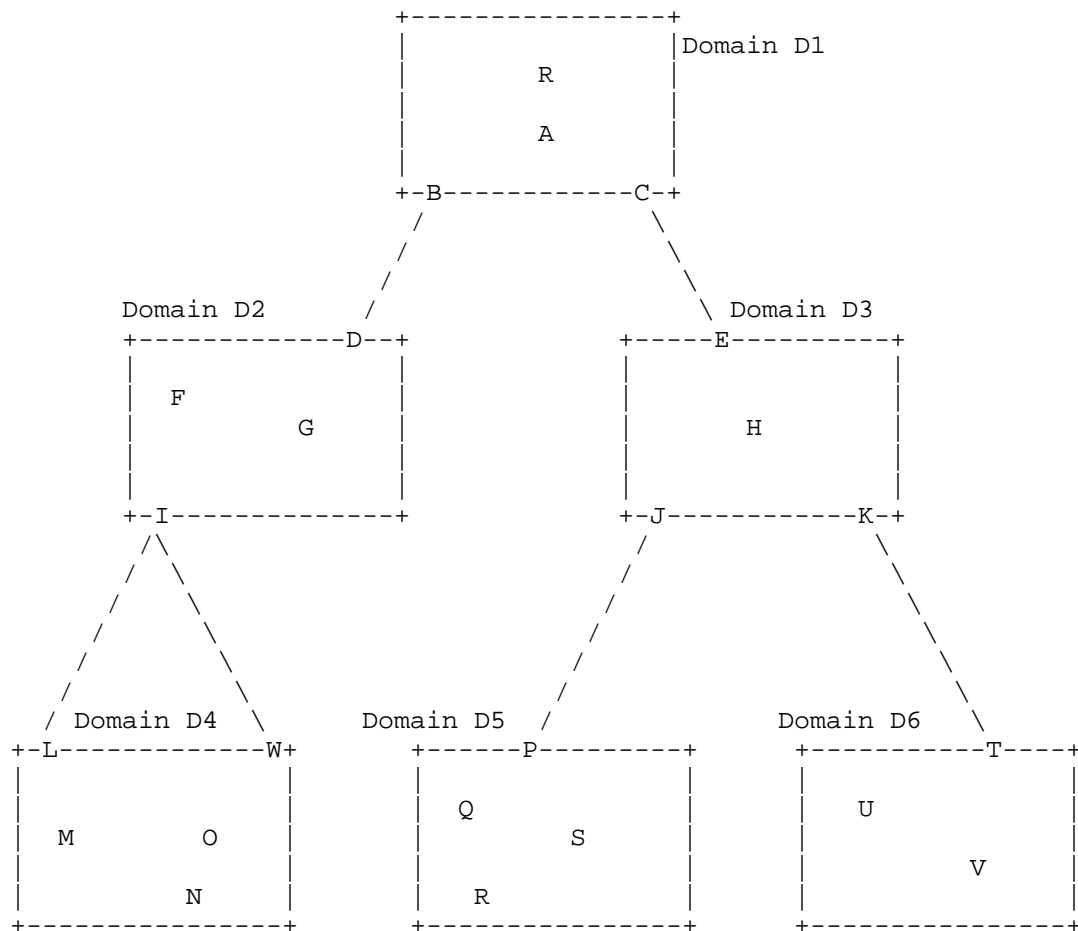
+-----+	+-----+	+-----+	+-----+
IRO	Sub	Sub	Sub
Object	Object AS	Object	Object AS
Header	100	IPv4	200
		198.51.	
		100.2	
+-----+	+-----+	+-----+	+-----+

4.4. PCE Serving multiple Domains

A single PCE can be responsible for multiple domains; for example PCE function deployed on an ABR could be responsible for multiple areas. A PCE which can support adjacent domains can internally handle those domains in the Domain-Sequence without any impact on the other domains in the Domain-Sequence.

4.5. P2MP

[RFC7334] describes an experimental inter-domain P2MP path computation mechanism where the path domain tree is described as a series of Domain-Sequences, an example is shown in the below figure:



The domain tree can be represented as a series of domain-sequence -

- o Domain D1, Domain D3, Domain D6
- o Domain D1, Domain D3, Domain D5
- o Domain D1, Domain D2, Domain D4

The domain sequence handling described in this document could be applied to P2MP path domain tree.

4.6. Hierarchical PCE

In case of H-PCE [RFC6805], the parent PCE can be requested to determine the Domain-Sequence and return it in the path computation reply, using the ERO. . For the example in section 4.6 of [RFC6805], the Domain-Sequence can possibly appear as:

ERO Object Header	Sub Object Domain 1	Sub Object Domain 2	Sub Object Domain 3
-------------------------	---------------------------	---------------------------	---------------------------

or

ERO Object Header	Sub Object BN 21	Sub Object Domain 3
-------------------------	------------------------	---------------------------

5. Other Considerations

5.1. Relationship to PCE Sequence

Instead of a Domain-Sequence, a sequence of PCEs MAY be enforced by policy on the PCC, and this constraint can be carried in the PCReq message (as defined in [RFC5886]).

Note that PCE-Sequence can be used along with Domain-Sequence in which case PCE-Sequence MUST have higher precedence in selecting the next PCE in the inter-domain path computation procedures.

5.2. Relationship to RSVP-TE

[RFC3209] already describes the notion of abstract nodes, where an abstract node is a group of nodes whose internal topology is opaque to the ingress node of the LSP. It further defines a subobject for AS but with a 2-Byte AS Number.

[DOMAIN-SUBOBJ] extends the notion of abstract nodes by adding new subobjects for IGP Areas and 4-byte AS numbers. These subobjects can

be included in Explicit Route Object (ERO), Exclude Route object (XRO) or Explicit Exclusion Route Subobject (EXRS) in RSVP-TE.

In any case subobject type defined in RSVP-TE are identical to the subobject type defined in the related documents in PCEP.

6. IANA Considerations

6.1. New Subobjects

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" at <http://www.iana.org/assignments/pcep>. Within this registry IANA maintains two sub-registries:

- o IRO Subobjects (see IRO Subobjects at <http://www.iana.org/assignments/pcep>)
- o XRO Subobjects (see XRO Subobjects at <http://www.iana.org/assignments/pcep>)

Upon approval of this document, IANA is requested to make identical additions to these registries as follows:

Subobject Type	Reference
TBD1 4 byte AS number	[This I.D.][DOMAIN-SUBOBJ]
TBD2 OSPF Area ID	[This I.D.][DOMAIN-SUBOBJ]
TBD3 IS-IS Area ID	[This I.D.][DOMAIN-SUBOBJ]

Further upon approval of this document, IANA is requested to add a reference to this document to the new RSVP numbers that are registered by [DOMAIN-SUBOBJ].

7. Security Considerations

The protocol extensions defined in this document do not substantially change the nature of PCEP. Therefore, the security considerations set out in [RFC5440] apply unchanged. Note that further security considerations for the use of PCEP over TCP are presented in [RFC6952].

This document specifies a representation of Domain-Sequence and new subobjects, which could be used in inter-domain PCE scenarios as explained in [RFC5152], [RFC5441], [RFC6805], [RFC7334] etc. The security considerations set out in each of these mechanisms remain unchanged by the new subobjects and Domain-Sequence representation in this document.

But the new subobjects do allow finer and more specific control of the path computed by a cooperating PCE(s). Such control increases the risk if a PCEP message is intercepted, modified, or spoofed because it allows the attacker to exert control over the path that the PCE will compute or to make the path computation impossible. Consequently, it is important that implementations conform to the relevant security requirements of [RFC5440]. These mechanisms include:

- o Securing the PCEP session messages using TCP security techniques (Section 10.2 of [RFC5440]). PCEP implementations SHOULD also consider the additional security provided by the TCP Authentication Option (TCP-AO) [RFC5925] or [PCEPS].
- o Authenticating the PCEP messages to ensure the message is intact and sent from an authorized node (Section 10.3 of [RFC5440]).
- o PCEP operates over TCP, so it is also important to secure the PCE and PCC against TCP denial-of-service attacks. Section 10.7.1 of [RFC5440] outlines a number of mechanisms for minimizing the risk of TCP-based denial-of-service attacks against PCEs and PCCs.
- o In inter-AS scenarios, attacks may be particularly significant with commercial as well as service-level implications.

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

8. Manageability Considerations

8.1. Control of Function and Policy

The exact behaviour with regards to desired inclusion and exclusion of domains MUST be available for examination by an operator and MAY be configurable. Manual configurations is needed to identify which PCEP peers understand the new domain subobjects defined in this document.

8.2. Information and Data Models

A MIB module for management of the PCEP is being specified in a separate document [RFC7420]. This document does not imply any new extension to the current MIB module.

8.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in [RFC5440].

8.4. Verify Correct Operations

Mechanisms defined in this document do not imply any new operation verification requirements in addition to those already listed in [RFC5440].

8.5. Requirements On Other Protocols

In case of per-domain path computation [RFC5152], where the full path of an inter-domain TE LSP cannot be, or is not determined at the ingress node, a signaling message can use the domain identifiers. The Subobjects defined in this document SHOULD be supported by RSVP-TE. [DOMAIN-SUBOBJ] extends the notion of abstract nodes by adding new subobjects for IGP Areas and 4-byte AS numbers.

Apart from this, mechanisms defined in this document do not imply any requirements on other protocols in addition to those already listed in [RFC5440].

8.6. Impact On Network Operations

The mechanisms described in this document can provide the operator with the ability to exert finer and more specific control of the path computation by inclusion or exclusion of domain subobjects. There may be some scaling benefit when a single domain subobject may substitute for many subobjects and can reduce the overall message size and processing.

Backward compatibility issues associated with the new subobjects arise when a PCE does not recognize them, in which case PCE responds according to the rules for a malformed object as per [RFC5440]. For successful operations the PCEs in the network would need to be upgraded.

9. Acknowledgments

Authors would like to especially thank Adrian Farrel for his detailed reviews as well as providing text to be included in the document.

Further, we would like to thank Pradeep Shastry, Suresh Babu, Quintin Zhao, Fatai Zhang, Daniel King, Oscar Gonzalez, Chen Huaimo,

Venugopal Reddy, Reeja Paul, Sandeep Boina, Avantika Sergio Belotti and Jonathan Hardwick for their useful comments and suggestions.

Thanks to Jonathan Hardwick for shepherding this document.

Thanks to Deborah Brungard for being the Responsible AD.

Thanks to Amanda Baber for IANA Review.

Thanks to Joel Halpern for Gen-ART Review.

Thanks to Klaas Wierenga for SecDir Review.

Thanks to Spencer Dawkins and Barry Leiba for comments during the IESG Review.

10. References

10.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<http://www.rfc-editor.org/info/rfc3209>>.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, DOI 10.17487/RFC3473, January 2003, <<http://www.rfc-editor.org/info/rfc3473>>.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, DOI 10.17487/RFC3477, January 2003, <<http://www.rfc-editor.org/info/rfc3477>>.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<http://www.rfc-editor.org/info/rfc5440>>.

- [RFC5441] Vasseur, JP., Ed., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, DOI 10.17487/RFC5441, April 2009, <<http://www.rfc-editor.org/info/rfc5441>>.
- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, DOI 10.17487/RFC5521, April 2009, <<http://www.rfc-editor.org/info/rfc5521>>.
- [RFC6805] King, D., Ed. and A. Farrel, Ed., "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, DOI 10.17487/RFC6805, November 2012, <<http://www.rfc-editor.org/info/rfc6805>>.
- [ISO10589] ISO, "Intermediate system to Intermediate system routing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002, 1992.
- [IRO-UPDATE] Dhody, D., "Update to Include Route Object (IRO) specification in Path Computation Element communication Protocol (PCEP. (draft-ietf-pce-iro-update-02)", May 2015.
- [DOMAIN-SUBOBJ] Dhody, D., Palle, U., Kondreddy, V., and R. Casellas, "Domain Subobjects for Resource ReserVation Protocol - Traffic Engineering (RSVP-TE). (draft-ietf-teas-rsvp-te-domain-subobjects-05)", November 2015.

10.2. Informative References

- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <<http://www.rfc-editor.org/info/rfc4655>>.
- [RFC4726] Farrel, A., Vasseur, J., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, DOI 10.17487/RFC4726, November 2006, <<http://www.rfc-editor.org/info/rfc4726>>.

- [RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, DOI 10.17487/RFC4873, May 2007, <<http://www.rfc-editor.org/info/rfc4873>>.
- [RFC4874] Lee, CY., Farrel, A., and S. De Cnodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4874, DOI 10.17487/RFC4874, April 2007, <<http://www.rfc-editor.org/info/rfc4874>>.
- [RFC5152] Vasseur, JP., Ed., Ayyangar, A., Ed., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, DOI 10.17487/RFC5152, February 2008, <<http://www.rfc-editor.org/info/rfc5152>>.
- [RFC5520] Bradford, R., Ed., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, DOI 10.17487/RFC5520, April 2009, <<http://www.rfc-editor.org/info/rfc5520>>.
- [RFC5886] Vasseur, JP., Ed., Le Roux, JL., and Y. Ikejiri, "A Set of Monitoring Tools for Path Computation Element (PCE)-Based Architecture", RFC 5886, DOI 10.17487/RFC5886, June 2010, <<http://www.rfc-editor.org/info/rfc5886>>.
- [RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <<http://www.rfc-editor.org/info/rfc5925>>.
- [RFC6793] Vohra, Q. and E. Chen, "BGP Support for Four-Octet Autonomous System (AS) Number Space", RFC 6793, DOI 10.17487/RFC6793, December 2012, <<http://www.rfc-editor.org/info/rfc6793>>.
- [RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, DOI 10.17487/RFC6952, May 2013, <<http://www.rfc-editor.org/info/rfc6952>>.
- [RFC7334] Zhao, Q., Dhody, D., King, D., Ali, Z., and R. Casellas, "PCE-Based Computation Procedure to Compute Shortest Constrained Point-to-Multipoint (P2MP) Inter-Domain Traffic Engineering Label Switched Paths", RFC 7334, DOI 10.17487/RFC7334, August 2014, <<http://www.rfc-editor.org/info/rfc7334>>.

- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<http://www.rfc-editor.org/info/rfc7420>>.
- [PCEPS] Lopez, D., Dios, O., Wu, W., and D. Dhody, "Secure Transport for PCEP", draft-ietf-pce-pceps-06 (work in progress), November 2015.

Authors' Addresses

Dhruv Dhody
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560037
India

EMail: dhruv.ietf@gmail.com

Udayasree Palle
Huawei Technologies
Divyashree Techno Park, Whitefield
Bangalore, Karnataka 560037
India

EMail: udayasree.palle@huawei.com

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

EMail: ramon.casellas@cttc.es

PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 27, 2015

A. Koushik
Brocade Communications Inc.
E. Stephan
Orange
Q. Zhao
Huawei Technology
D. King
Old Dog Consulting
J. Hardwick
Metaswitch
October 24, 2014

Path Computation Element Communications Protocol (PCEP) Management
Information Base (MIB) Module
draft-ietf-pce-pcep-mib-11

Abstract

This memo defines a portion of the Management Information Base (MIB) for use with network management protocols in the Internet community. In particular, it describes managed objects for modeling of Path Computation Element communications Protocol (PCEP) for communications between a Path Computation Client (PCC) and a Path Computation Element (PCE), or between two PCEs.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

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

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

This Internet-Draft will expire on April 27, 2015.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
1.1. Requirements Language	3
1.2. Terminology	3
2. The Internet-Standard Management Framework	3
3. PCEP MIB Module Architecture	4
3.1. pcePcepEntityTable	4
3.2. pcePcepPeerTable	5
3.3. pcePcepSessTable	5
3.4. PCEP Notifications	6
3.5. Relationship to other MIB modules	6
3.6. Illustrative example	6
4. Object Definitions	7
4.1. PCE-PCEP-MIB	7
5. Security Considerations	48
6. IANA Considerations	49
7. Acknowledgement	49
8. References	49
8.1. Normative References	49
8.2. Informative References	50
Appendix A. Contributors	51
Appendix B. PCEP MIB Module Example	51
B.1. Contents of PCEP MIB module at PCE2	51
B.2. Contents of PCEP MIB module at PCCb	59

1. Introduction

The Path Computation Element (PCE) defined in [RFC4655] is an entity that is capable of computing a network path or route based on a network graph, and applying computational constraints. A Path Computation Client (PCC) may make requests to a PCE for paths to be computed.

PCEP is the communication protocol between a PCC and PCE and is defined in [RFC5440]. PCEP interactions include path computation requests and path computation replies as well as notifications of specific states related to the use of a PCE in the context of Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering (TE).

This memo defines a portion of the Management Information Base (MIB) for use with network management protocols in the Internet community. In particular, it defines a MIB module that can be used to monitor PCEP interactions between a PCC and a PCE, or between two PCEs.

The scope of this document is to provide a MIB module for the PCEP base protocol defined in [RFC5440]. Extensions to the PCEP base protocol are beyond the scope for this document.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY" and "OPTIONAL" in this document are to be interpreted as described in BCP 14, RFC 2119 [RFC2119].

1.2. Terminology

This document uses the terminology defined in [RFC4655] and [RFC5440]. In particular, it uses the following acronyms.

- o Path Computation Request message (PCReq).
- o Path Computation Reply message (PCRep).
- o Notification message (PCNtf).
- o Error message (PCErr).
- o Request Parameters object (RP).
- o Synchronization Vector object (SVEC).
- o Explicit Route object (ERO).

This document uses the term "PCEP entity" to refer to a local PCEP speaker, "peer" to refer to a remote PCEP speaker and "PCEP speaker" where it is not necessary to distinguish between local and remote.

2. The Internet-Standard Management Framework

For a detailed overview of the documents that describe the current Internet-Standard Management Framework, please refer to section 7 of RFC 3410 [RFC3410].

Managed objects are accessed via a virtual information store, termed the Management Information Base or MIB. MIB objects are generally accessed through the Simple Network Management Protocol (SNMP). Objects in the MIB are defined using the mechanisms defined in the Structure of Management Information (SMI). This memo specifies a MIB module that is compliant to the SMIV2, which is described in STD 58, RFC 2578 [RFC2578], STD 58, RFC 2579 [RFC2579], and STD 58, RFC 2580 [RFC2580].

3. PCEP MIB Module Architecture

The PCEP MIB module contains the following information:

- a. PCE and PCC local entity status (see `pcePcepEntityTable`).
- b. PCEP peer information (see `pcePcepPeerTable`).
- c. PCEP session information (see `pcePcepSessTable`).
- d. Notifications to indicate PCEP session changes.

The PCEP MIB module is limited to "read-only" access except for `pcePcepNotificationsMaxRate`, which is used to throttle the rate at which the implementation generates notifications.

3.1. `pcePcepEntityTable`

The PCEP MIB module may contain status information for multiple logical local PCEP entities. There are several scenarios in which there may be more than one local PCEP entity, including the following.

- o A physical router, which is partitioned into multiple virtual routers, each with its own PCC.
- o A PCE device which front-ends a cluster of compute resources, each with a different set of capabilities that are accessed via different IP addresses.

The `pcePcepEntityTable` contains one row for each local PCEP entity. Each row is read-only and contains current status information plus the PCEP entity's running configuration.

The `pcePcepEntityTable` is indexed by `pcePcepEntityIndex`, which also acts as the primary index for the other tables in this MIB module.

3.2. `pcePcepPeerTable`

The `pcePcepPeerTable` contains one row for each peer that the local PCEP entity knows about. Each row is read-only and contains information to identify the peer, the running configuration relating to that peer and statistics that track the messages exchanged with that peer and its response times.

A PCEP speaker is identified by its IP address. If there is a PCEP speaker in the network that uses multiple IP addresses then it looks like multiple distinct peers to the other PCEP speakers in the network.

The `pcePcepPeerTable` is indexed first by `pcePcepEntityIndex`, then by `pcePcepPeerAddrType` and `pcePcepPeerAddr`. This indexing structure allows each local PCEP entity to report its own set of peers.

Since PCEP sessions can be ephemeral, the `pcePcepPeerTable` tracks a peer even when no PCEP session currently exists to that peer. The statistics contained in `pcePcepPeerTable` are an aggregate of the statistics for all successive sessions to that peer.

To limit the quantity of information that is stored, an implementation MAY choose to discard a row from the `pcePcepPeerTable` if and only if no PCEP session exists to the corresponding peer.

3.3. `pcePcepSessTable`

The `pcePcepSessTable` contains one row for each PCEP session that the PCEP entity (PCE or PCC) is currently participating in. Each row is read-only and contains the running configuration that is applied to the session, plus identifiers and statistics for the session.

The statistics in `pcePcepSessTable` are semantically different from those in `pcePcepPeerTable` since the former apply to the current session only, whereas the latter are the aggregate for all sessions that have existed to that peer.

Although [RFC5440] forbids there from being more than one active PCEP session between a given pair of PCEP entities at any one time, there is a window during session establishment where the `pcePcepSessTable` may contain two rows for a given peer, one representing a session initiated by the local PCEP entity and one representing a session initiated by the peer. If either of these sessions reaches active state, then the other is discarded.

The pcePcepSessTable is indexed first by pcePcepEntityIndex, then by pcePcepPeerAddrType and pcePcepPeerAddr, and finally by pcePcepSessInitiator. This indexing structure allows each local PCEP entity to report its own set of active sessions. The pcePcepSessInitiator index allows two rows to exist transiently for a given peer, as discussed above.

3.4. PCEP Notifications

The PCEP MIB module contains notifications for the following conditions.

- a. pcePcepSessUp: PCEP Session has gone up.
- b. pcePcepSessDown: PCEP Session has gone down.
- c. pcePcepSessLocalOverload: Local PCEP entity has sent an overload PCNtf on this session.
- d. pcePcepSessLocalOverloadClear: Local PCEP entity has sent an overload-cleared PCNtf on this session.
- e. pcePcepSessPeerOverload: Peer has sent an overload PCNtf on this session.
- f. pcePcepSessPeerOverloadClear: Peer has sent an overload-cleared PCNtf on this session.

3.5. Relationship to other MIB modules

The PCEP MIB module imports the following textual conventions from the INET-ADDRESS-MIB defined in RFC 4001 [RFC4001]:

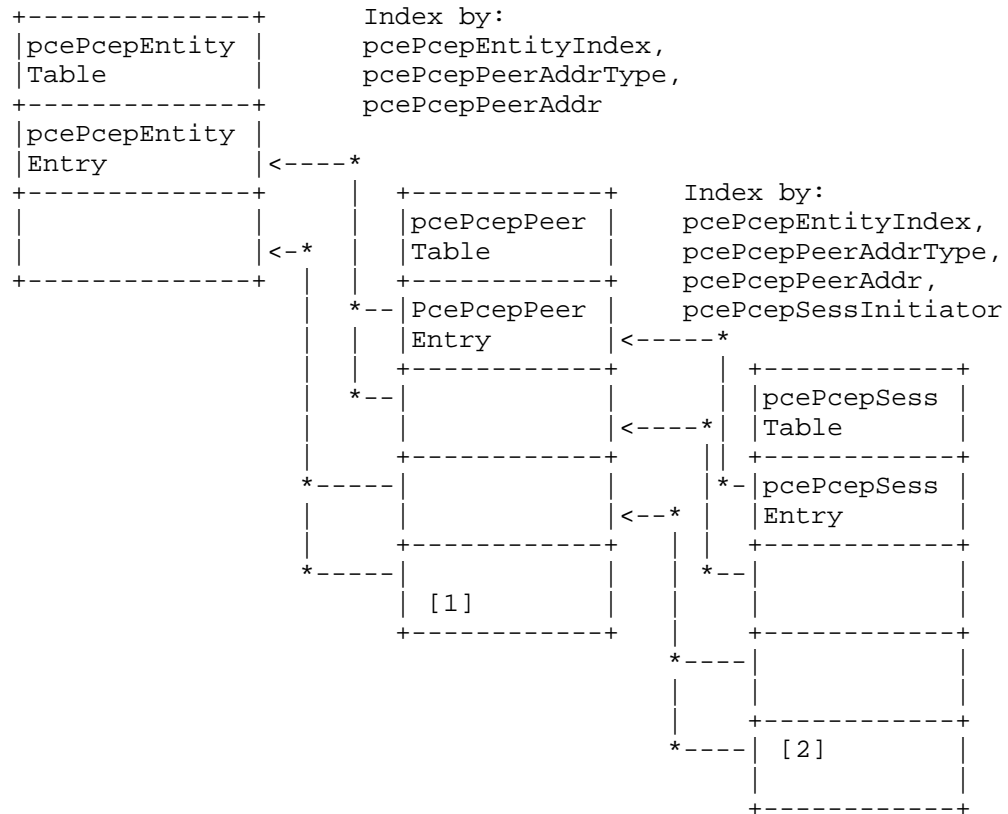
- o InetAddressType
- o InetAddress

PCEP relies on existing protocols which have specialized MIB objects to monitor their own activities. Consequently this document considers that the monitoring of underlying protocols is out of scope of the PCEP MIB module.

3.6. Illustrative example

The following diagram illustrates the relationships between the pcePcepEntityTable, pcePcepPeerTable and pcePcepSessTable.

Index by:
pcePcepEntityIndex



[1]: A peer entry with no current session

[2]: Two sessions exist during a window in session initialization.

4. Object Definitions

4.1. PCE-PCEP-MIB

PCE-PCEP-MIB DEFINITIONS ::= BEGIN

IMPORTS

MODULE-IDENTITY,
OBJECT-TYPE,
mib-2,
NOTIFICATION-TYPE,
Unsigned32,

```
Counter32
    FROM SNMPv2-SMI                -- RFC 2578
TruthValue,
TimeStamp
    FROM SNMPv2-TC                -- RFC 2579
MODULE-COMPLIANCE,
OBJECT-GROUP,
NOTIFICATION-GROUP
    FROM SNMPv2-CONF              -- RFC 2580
InetAddressType,
InetAddress
    FROM INET-ADDRESS-MIB;        -- RFC 4001

pcePcepMIB MODULE-IDENTITY
    LAST-UPDATED
        "201410241200Z" -- 24 October 2014
    ORGANIZATION
        "IETF Path Computation Element (PCE) Working Group"
    CONTACT-INFO
        "Email: pce@ietf.org
        WG charter:
            http://www.ietf.org/html.charters/pce-charter.html"

    DESCRIPTION
        "This MIB module defines a collection of objects for managing
        Path Computation Element communications Protocol (PCEP).

        Copyright (C) The IETF Trust (2014). This version of this
        MIB module is part of RFC YYYY; see the RFC itself for full
        legal notices."
-- RFC Ed.: replace YYYY with actual RFC number & remove this note
    REVISION
        "201410241200Z" -- 24 October 2014
    DESCRIPTION
        "Initial version, published as RFC YYYY."
-- RFC Ed.: replace YYYY with actual RFC number & remove this note
        ::= { mib-2 XXX }
-- RFC Ed.: replace XXX with IANA-assigned number & remove this note

pcePcepNotifications OBJECT IDENTIFIER ::= { pcePcepMIB 0 }
pcePcepObjects        OBJECT IDENTIFIER ::= { pcePcepMIB 1 }
pcePcepConformance    OBJECT IDENTIFIER ::= { pcePcepMIB 2 }

--
-- PCEP Entity Objects
--

pcePcepEntityTable OBJECT-TYPE
```



```

SYNTAX          SEQUENCE OF PcePcepEntityEntry
MAX-ACCESS      not-accessible
STATUS          current
DESCRIPTION
    "This table contains information about local PCEP entities.
    The entries in this table are read-only."
 ::= { pcePcepObjects 1 }

```

```

pcePcepEntityEntry OBJECT-TYPE
    SYNTAX          PcePcepEntityEntry
    MAX-ACCESS      not-accessible
    STATUS          current
    DESCRIPTION
        "This entry represents a local PCEP entity."
    INDEX           { pcePcepEntityIndex }
    ::= { pcePcepEntityTable 1 }

```

```

PcePcepEntityEntry ::= SEQUENCE {
    pcePcepEntityIndex          Unsigned32,
    pcePcepEntityAdminStatus    INTEGER,
    pcePcepEntityOperStatus     INTEGER,
    pcePcepEntityAddrType       InetAddressType,
    pcePcepEntityAddr           InetAddress,
    pcePcepEntityConnectTimer    Unsigned32,
    pcePcepEntityConnectMaxRetry Unsigned32,
    pcePcepEntityInitBackoffTimer Unsigned32,
    pcePcepEntityMaxBackoffTimer Unsigned32,
    pcePcepEntityOpenWaitTimer   Unsigned32,
    pcePcepEntityKeepWaitTimer   Unsigned32,
    pcePcepEntityKeepAliveTimer  Unsigned32,
    pcePcepEntityDeadTimer       Unsigned32,
    pcePcepEntityAllowNegotiation TruthValue,
    pcePcepEntityMaxKeepAliveTimer Unsigned32,
    pcePcepEntityMaxDeadTimer    Unsigned32,
    pcePcepEntityMinKeepAliveTimer Unsigned32,
    pcePcepEntityMinDeadTimer    Unsigned32,
    pcePcepEntitySyncTimer       Unsigned32,
    pcePcepEntityRequestTimer    Unsigned32,
    pcePcepEntityMaxSessions     Unsigned32,
    pcePcepEntityMaxUnknownReqs  Unsigned32,
    pcePcepEntityMaxUnknownMsgs  Unsigned32
}

```

```

pcePcepEntityIndex OBJECT-TYPE
    SYNTAX          Unsigned32
    MAX-ACCESS      not-accessible
    STATUS          current
    DESCRIPTION

```

"This index is used to uniquely identify the PCEP entity."
 ::= { pcePcepEntityEntry 1 }

pcePcepEntityAdminStatus OBJECT-TYPE

SYNTAX INTEGER {
 adminStatusUp(1),
 adminStatusDown(2)
 }

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The administrative status of this PCEP Entity.

This is the desired operational status as currently set by an operator or by default in the implementation. The value of pcePcepEntityOperStatus represents the current status of an attempt to reach this desired status."

::= { pcePcepEntityEntry 2 }

pcePcepEntityOperStatus OBJECT-TYPE

SYNTAX INTEGER {
 operStatusUp(1),
 operStatusDown(2),
 operStatusGoingUp(3),
 operStatusGoingDown(4),
 operStatusFailed(5),
 operStatusFailedPerm(6)
 }

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The operational status of the PCEP entity. Takes one of the following values.

- operStatusUp(1): the PCEP entity is active.
- operStatusDown(2): the PCEP entity is inactive.
- operStatusGoingUp(3): the PCEP entity is activating.
- operStatusGoingDown(4): the PCEP entity is deactivating.
- operStatusFailed(5): the PCEP entity has failed and will recover when possible.
- operStatusFailedPerm(6): the PCEP entity has failed and will not recover without operator intervention."

::= { pcePcepEntityEntry 3 }

pcePcepEntityAddrType OBJECT-TYPE

SYNTAX InetAddressType

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The type of the PCEP entity's Internet address. This object specifies how the value of the pcePcepEntityAddr object should be interpreted. Only values unknown(0), ipv4(1), or ipv6(2) are supported."
 ::= { pcePcepEntityEntry 4 }

pcePcepEntityAddr OBJECT-TYPE

SYNTAX InetAddress

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The local Internet address of this PCEP entity. The type is given by pcePcepEntityAddrType.

If operating as a PCE server, the PCEP entity listens on this address. If operating as a PCC, the PCEP entity binds outgoing TCP connections to this address.

It is possible for the PCEP entity to operate both as a PCC and a PCE Server, in which case it uses this address both to listen for incoming TCP connections and to bind outgoing TCP connections."

::= { pcePcepEntityEntry 5 }

pcePcepEntityConnectTimer OBJECT-TYPE

SYNTAX Unsigned32 (1..65535)

UNITS "seconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The time that the PCEP entity will wait to establish a TCP connection with a peer. If a TCP connection is not established within this time then PCEP aborts the session setup attempt."

::= { pcePcepEntityEntry 6 }

pcePcepEntityConnectMaxRetry OBJECT-TYPE

SYNTAX Unsigned32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The maximum number of times the system tries to establish a TCP connection to a peer before the session with the peer transitions to the idle state.

When the session transitions to the idle state:

- pcePcepPeerSessionExists transitions to false(2)
- the associated PcePcepSessEntry is deleted

```
        - a backoff timer runs before the session is tried again."
 ::= { pcePcepEntityEntry 7 }

pcePcepEntityInitBackoffTimer OBJECT-TYPE
    SYNTAX      Unsigned32 (1..65535)
    UNITS       "seconds"
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The initial back-off time for retrying a failed session
        setup attempt to a peer.

        The back-off time increases for each failed session setup
        attempt, until a maximum back-off time is reached.  The
        maximum back-off time is pcePcepEntityMaxBackoffTimer."
 ::= { pcePcepEntityEntry 8 }

pcePcepEntityMaxBackoffTimer OBJECT-TYPE
    SYNTAX      Unsigned32
    UNITS       "seconds"
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The maximum back-off time for retrying a failed session
        setup attempt to a peer.

        The back-off time increases for each failed session setup
        attempt, until this maximum value is reached.  Session
        setup attempts then repeat periodically without any
        further increase in back-off time."
 ::= { pcePcepEntityEntry 9 }

pcePcepEntityOpenWaitTimer OBJECT-TYPE
    SYNTAX      Unsigned32 (1..65535)
    UNITS       "seconds"
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The time that the PCEP entity will wait to receive an Open
        message from a peer after the TCP connection has come up.
        If no Open message is received within this time then PCEP
        terminates the TCP connection and deletes the associated
        PcePcepSessEntry."
 ::= { pcePcepEntityEntry 10 }

pcePcepEntityKeepWaitTimer OBJECT-TYPE
    SYNTAX      Unsigned32 (1..65535)
    UNITS       "seconds"
```

```
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "The time that the PCEP entity will wait to receive a
    Keepalive or PCErr message from a peer during session
    initialization after receiving an Open message.  If no
    Keepalive or PCErr message is received within this time then
    PCEP terminates the TCP connection and deletes the
    associated PcePcepSessEntry."
 ::= { pcePcepEntityEntry 11 }

pcePcepEntityKeepAliveTimer OBJECT-TYPE
    SYNTAX      Unsigned32 (0..255)
    UNITS       "seconds"
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The keep alive transmission timer that this PCEP entity will
        propose in the initial OPEN message of each session it is
        involved in.  This is the maximum time between two
        consecutive messages sent to a peer.  Zero means that
        the PCEP entity prefers not to send Keepalives at all.

        Note that the actual Keepalive transmission intervals, in
        either direction of an active PCEP session, are determined
        by negotiation between the peers as specified by RFC
        5440, and so may differ from this configured value.  For
        the actually negotiated values (per-session), see
        pcePcepSessKeepaliveTimer and
        pcePcepSessPeerKeepaliveTimer."
    ::= { pcePcepEntityEntry 12 }

pcePcepEntityDeadTimer OBJECT-TYPE
    SYNTAX      Unsigned32 (0..255)
    UNITS       "seconds"
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The dead timer that this PCEP entity will propose in the
        initial OPEN message of each session it is involved in.
        This is the time after which a peer should declare a
        session down if it does not receive any PCEP messages.
        Zero suggests that the peer does not run a dead timer at
        all."
    ::= { pcePcepEntityEntry 13 }

pcePcepEntityAllowNegotiation OBJECT-TYPE
    SYNTAX      TruthValue
```

```
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "Whether the PCEP entity will permit negotiation of session
    parameters."
 ::= { pcePcepEntityEntry 14 }

pcePcepEntityMaxKeepAliveTimer OBJECT-TYPE
SYNTAX        Unsigned32 (0..255)
UNITS         "seconds"
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "In PCEP session parameter negotiation, the maximum value
    that this PCEP entity will accept from a peer for the
    interval between Keepalive transmissions. Zero means that
    the PCEP entity will allow no Keepalive transmission at
    all."
 ::= { pcePcepEntityEntry 15 }

pcePcepEntityMaxDeadTimer OBJECT-TYPE
SYNTAX        Unsigned32 (0..255)
UNITS         "seconds"
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "In PCEP session parameter negotiation, the maximum value
    that this PCEP entity will accept from a peer for the Dead
    timer. Zero means that the PCEP entity will allow not
    running a Dead timer."
 ::= { pcePcepEntityEntry 16 }

pcePcepEntityMinKeepAliveTimer OBJECT-TYPE
SYNTAX        Unsigned32 (0..255)
UNITS         "seconds"
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "In PCEP session parameter negotiation, the minimum value
    that this PCEP entity will accept for the interval between
    Keepalive transmissions. Zero means that the PCEP entity
    insists on no Keepalive transmission at all."
 ::= { pcePcepEntityEntry 17 }

pcePcepEntityMinDeadTimer OBJECT-TYPE
SYNTAX        Unsigned32 (0..255)
UNITS         "seconds"
MAX-ACCESS    read-only
```

```
STATUS      current
DESCRIPTION
    "In PCEP session parameter negotiation, the minimum value
    that this PCEP entity will accept for the Dead timer. Zero
    means that the PCEP entity insists on not running a Dead
    timer."
 ::= { pcePcepEntityEntry 18 }

pcePcepEntitySyncTimer OBJECT-TYPE
SYNTAX      Unsigned32 (0..65535)
UNITS       "seconds"
MAX-ACCESS  read-only
STATUS      current
DESCRIPTION
    "The value of SyncTimer is used in the case of synchronized
    path computation request using the SVEC object.

    Consider the case where a PCReq message is received by a PCE
    that contains the SVEC object referring to M synchronized
    path computation requests. If after the expiration of the
    SyncTimer all the M path computation requests have not been
    received, a protocol error is triggered and the PCE MUST
    cancel the whole set of path computation requests.

    The aim of the SyncTimer is to avoid the storage of unused
    synchronized requests should one of them get lost for some
    reasons (for example, a misbehaving PCC).

    A value of zero is returned if and only if the entity does
    not use the SyncTimer."
 ::= { pcePcepEntityEntry 19 }

pcePcepEntityRequestTimer OBJECT-TYPE
SYNTAX      Unsigned32 (1..65535)
UNITS       "seconds"
MAX-ACCESS  read-only
STATUS      current
DESCRIPTION
    "The maximum time that the PCEP entity will wait for a
    response to a PCReq message."
 ::= { pcePcepEntityEntry 20 }

pcePcepEntityMaxSessions OBJECT-TYPE
SYNTAX      Unsigned32
MAX-ACCESS  read-only
STATUS      current
DESCRIPTION
    "Maximum number of sessions involving this PCEP entity"
```

```
        that can exist at any time."
 ::= { pcePcepEntityEntry 21 }

pcePcepEntityMaxUnknownReqs OBJECT-TYPE
    SYNTAX      Unsigned32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The maximum number of unrecognized requests and replies that
        any session on this PCEP entity is willing to accept per
        minute before terminating the session.

        A PCRep message contains an unrecognized reply if it
        contains an RP object whose request ID does not correspond
        to any in-progress request sent by this PCEP entity.

        A PCReq message contains an unrecognized request if it
        contains an RP object whose request ID is zero."
 ::= { pcePcepEntityEntry 22 }

pcePcepEntityMaxUnknownMsgs OBJECT-TYPE
    SYNTAX      Unsigned32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The maximum number of unknown messages that any session
        on this PCEP entity is willing to accept per minute before
        terminating the session."
 ::= { pcePcepEntityEntry 23 }

--
-- The PCEP Peer Table
--

pcePcepPeerTable OBJECT-TYPE
    SYNTAX      SEQUENCE OF PcePcepPeerEntry
    MAX-ACCESS   not-accessible
    STATUS       current
    DESCRIPTION
        "This table contains information about peers known by
        the local PCEP entity. The entries in this table are
        read-only.

        This table gives peer information that spans PCEP
        sessions. Information about current PCEP sessions can be
        found in the pcePcepSessTable table."
 ::= { pcePcepObjects 2 }
```



```

pcePcepPeerEntry OBJECT-TYPE
    SYNTAX      PcePcepPeerEntry
    MAX-ACCESS  not-accessible
    STATUS      current
    DESCRIPTION
        "Information about a single peer which spans all PCEP
        sessions to that peer."
    INDEX { pcePcepEntityIndex,
            pcePcepPeerAddrType,
            pcePcepPeerAddr }
    ::= { pcePcepPeerTable 1 }

PcePcepPeerEntry ::= SEQUENCE {
    pcePcepPeerAddrType      InetAddressType,
    pcePcepPeerAddr          InetAddress,
    pcePcepPeerRole          INTEGER,
    pcePcepPeerDiscontinuityTime  TimeStamp,
    pcePcepPeerInitiateSession  TruthValue,
    pcePcepPeerSessionExists    TruthValue,
    pcePcepPeerNumSessSetupOK    Counter32,
    pcePcepPeerNumSessSetupFail  Counter32,
    pcePcepPeerSessionUpTime     TimeStamp,
    pcePcepPeerSessionFailTime   TimeStamp,
    pcePcepPeerSessionFailUpTime TimeStamp,
    pcePcepPeerAvgRspTime        Unsigned32,
    pcePcepPeerLWMRspTime        Unsigned32,
    pcePcepPeerHWMRspTime        Unsigned32,
    pcePcepPeerNumPCReqSent       Counter32,
    pcePcepPeerNumPCReqRcvd       Counter32,
    pcePcepPeerNumPCRepSent       Counter32,
    pcePcepPeerNumPCRepRcvd       Counter32,
    pcePcepPeerNumPCErrSent       Counter32,
    pcePcepPeerNumPCErrRcvd       Counter32,
    pcePcepPeerNumPCNtfSent       Counter32,
    pcePcepPeerNumPCNtfRcvd       Counter32,
    pcePcepPeerNumKeepaliveSent   Counter32,
    pcePcepPeerNumKeepaliveRcvd   Counter32,
    pcePcepPeerNumUnknownRcvd     Counter32,
    pcePcepPeerNumCorruptRcvd     Counter32,
    pcePcepPeerNumReqSent         Counter32,
    pcePcepPeerNumSvecSent        Counter32,
    pcePcepPeerNumSvecReqSent     Counter32,
    pcePcepPeerNumReqSentPendRep  Counter32,
    pcePcepPeerNumReqSentEroRcvd  Counter32,
    pcePcepPeerNumReqSentNoPathRcvd Counter32,
    pcePcepPeerNumReqSentCancelRcvd Counter32,
    pcePcepPeerNumReqSentErrorRcvd Counter32,
    pcePcepPeerNumReqSentTimeout  Counter32,

```

```

    pcePcepPeerNumReqSentCancelSent      Counter32,
    pcePcepPeerNumReqSentClosed           Counter32,
    pcePcepPeerNumReqRcvd                 Counter32,
    pcePcepPeerNumSvecRcvd                Counter32,
    pcePcepPeerNumSvecReqRcvd             Counter32,
    pcePcepPeerNumReqRcvdPendRep          Counter32,
    pcePcepPeerNumReqRcvdEroSent           Counter32,
    pcePcepPeerNumReqRcvdNoPathSent       Counter32,
    pcePcepPeerNumReqRcvdCancelSent       Counter32,
    pcePcepPeerNumReqRcvdErrorSent        Counter32,
    pcePcepPeerNumReqRcvdCancelRcvd       Counter32,
    pcePcepPeerNumReqRcvdClosed           Counter32,
    pcePcepPeerNumRepRcvdUnknown          Counter32,
    pcePcepPeerNumReqRcvdUnknown          Counter32
}

pcePcepPeerAddrType OBJECT-TYPE
    SYNTAX      InetAddressType
    MAX-ACCESS  not-accessible
    STATUS      current
    DESCRIPTION
        "The type of the peer's Internet address.  This object
        specifies how the value of the pcePcepPeerAddr object should
        be interpreted.  Only values unknown(0), ipv4(1), or
        ipv6(2) are supported."
    ::= { pcePcepPeerEntry 1 }

pcePcepPeerAddr OBJECT-TYPE
    SYNTAX      InetAddress
    MAX-ACCESS  not-accessible
    STATUS      current
    DESCRIPTION
        "The Internet address of the peer.  The type is given by
        pcePcepPeerAddrType."
    ::= { pcePcepPeerEntry 2 }

pcePcepPeerRole OBJECT-TYPE
    SYNTAX      INTEGER {
                    unknown(0),
                    pcc(1),
                    pce(2),
                    pccAndPce(3)
                }
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The role that this peer took the last time a session was
        established.  Takes one of the following values."

```

- unknown(0): this peer's role is not known.
- pcc(1): this peer is a Path Computation Client (PCC).
- pce(2): this peer is a Path Computation Server (PCE).
- pccAndPce(3): this peer is both a PCC and a PCE."

```
::= { pcePcepPeerEntry 3 }
```

pcePcepPeerDiscontinuityTime OBJECT-TYPE

SYNTAX TimeStamp

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The value of sysUpTime at the time that the information and statistics in this row were last reset."

```
::= { pcePcepPeerEntry 4 }
```

pcePcepPeerInitiateSession OBJECT-TYPE

SYNTAX TruthValue

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"Indicates whether the local PCEP entity initiates sessions to this peer, or waits for the peer to initiate a session."

```
::= { pcePcepPeerEntry 5 }
```

pcePcepPeerSessionExists OBJECT-TYPE

SYNTAX TruthValue

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"Indicates whether a session with this peer currently exists."

```
::= { pcePcepPeerEntry 6 }
```

pcePcepPeerNumSessSetupOK OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of PCEP sessions successfully established with the peer, including any current session. This counter is incremented each time a session with this peer is successfully established."

```
::= { pcePcepPeerEntry 7 }
```

pcePcepPeerNumSessSetupFail OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of PCEP sessions with the peer that have been attempted but failed before being fully established. This counter is incremented each time a session retry to this peer fails."

::= { pcePcepPeerEntry 8 }

pcePcepPeerSessionUpTime OBJECT-TYPE

SYNTAX TimeStamp

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The value of sysUpTime the last time a session with this peer was successfully established."

If pcePcepPeerNumSessSetupOK is zero, then this object contains zero."

::= { pcePcepPeerEntry 9 }

pcePcepPeerSessionFailTime OBJECT-TYPE

SYNTAX TimeStamp

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The value of sysUpTime the last time a session with this peer failed to be established."

If pcePcepPeerNumSessSetupFail is zero, then this object contains zero."

::= { pcePcepPeerEntry 10 }

pcePcepPeerSessionFailUpTime OBJECT-TYPE

SYNTAX TimeStamp

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The value of sysUpTime the last time a session with this peer failed from active."

If pcePcepPeerNumSessSetupOK is zero, then this object contains zero."

::= { pcePcepPeerEntry 11 }

pcePcepPeerAvgRspTime OBJECT-TYPE

SYNTAX Unsigned32

UNITS "milliseconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The average response time for this peer.

If an average response time has not been calculated for this peer then this object has the value zero.

If pcePcepPeerRole is pcc then this field is meaningless and is set to zero."

::= { pcePcepPeerEntry 12 }

pcePcepPeerLWMrspTime OBJECT-TYPE

SYNTAX Unsigned32

UNITS "milliseconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The smallest (low-water mark) response time seen from this peer.

If no responses have been received from this peer then this object has the value zero.

If pcePcepPeerRole is pcc then this field is meaningless and is set to zero."

::= { pcePcepPeerEntry 13 }

pcePcepPeerHWMrspTime OBJECT-TYPE

SYNTAX Unsigned32

UNITS "milliseconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The greatest (high-water mark) response time seen from this peer.

If no responses have been received from this peer then this object has the value zero.

If pcePcepPeerRole is pcc then this field is meaningless and is set to zero."

::= { pcePcepPeerEntry 14 }

pcePcepPeerNumPCReqSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of PCReq messages sent to this peer."

```
 ::= { pcePcepPeerEntry 15 }

pcePcepPeerNumPCReqRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCReq messages received from this peer."
    ::= { pcePcepPeerEntry 16 }

pcePcepPeerNumPCRepSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCRep messages sent to this peer."
    ::= { pcePcepPeerEntry 17 }

pcePcepPeerNumPCRepRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCRep messages received from this peer."
    ::= { pcePcepPeerEntry 18 }

pcePcepPeerNumPCErrSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCErr messages sent to this peer."
    ::= { pcePcepPeerEntry 19 }

pcePcepPeerNumPCErrRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCErr messages received from this peer."
    ::= { pcePcepPeerEntry 20 }

pcePcepPeerNumPCNtfSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCNtf messages sent to this peer."
```

```
 ::= { pcePcepPeerEntry 21 }

pcePcepPeerNumPCNtfRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCNtf messages received from this peer."
    ::= { pcePcepPeerEntry 22 }

pcePcepPeerNumKeepaliveSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of Keepalive messages sent to this peer."
    ::= { pcePcepPeerEntry 23 }

pcePcepPeerNumKeepaliveRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of Keepalive messages received from this peer."
    ::= { pcePcepPeerEntry 24 }

pcePcepPeerNumUnknownRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of unknown messages received from this peer."
    ::= { pcePcepPeerEntry 25 }

pcePcepPeerNumCorruptRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of corrupted PCEP message received from this
        peer."
    ::= { pcePcepPeerEntry 26 }

pcePcepPeerNumReqSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
```

"The number of requests sent to this peer. A request corresponds 1:1 with an RP object in a PCReq message.

This might be greater than pcePcepPeerNumPCReqSent because multiple requests can be batched into a single PCReq message."

::= { pcePcepPeerEntry 27 }

pcePcepPeerNumSvecSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of SVEC objects sent to this peer in PCReq messages. An SVEC object represents a set of synchronized requests."

::= { pcePcepPeerEntry 28 }

pcePcepPeerNumSvecReqSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests sent to this peer that appeared in one or more SVEC objects."

::= { pcePcepPeerEntry 29 }

pcePcepPeerNumReqSentPendRep OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests that have been sent to this peer for which a response is still pending."

::= { pcePcepPeerEntry 30 }

pcePcepPeerNumReqSentEroRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests that have been sent to this peer for which a response with an ERO object was received. Such responses indicate that a path was successfully computed by the peer."

::= { pcePcepPeerEntry 31 }

pcePcepPeerNumReqSentNoPathRcvd OBJECT-TYPE


```
SYNTAX      Counter32
MAX-ACCESS  read-only
STATUS      current
DESCRIPTION
    "The number of requests that have been sent to this peer for
    which a response with a NO-PATH object was received.  Such
    responses indicate that the peer could not find a path to
    satisfy the request."
 ::= { pcePcepPeerEntry 32 }

pcePcepPeerNumReqSentCancelRcvd OBJECT-TYPE
SYNTAX      Counter32
MAX-ACCESS  read-only
STATUS      current
DESCRIPTION
    "The number of requests that were cancelled by the peer with
    a PCNtf message.

    This might be different than pcePcepPeerNumPCNtfRcvd because
    not all PCNtf messages are used to cancel requests, and a
    single PCNtf message can cancel multiple requests."
 ::= { pcePcepPeerEntry 33 }

pcePcepPeerNumReqSentErrorRcvd OBJECT-TYPE
SYNTAX      Counter32
MAX-ACCESS  read-only
STATUS      current
DESCRIPTION
    "The number of requests that were rejected by the peer with a
    PCErr message.

    This might be different than pcePcepPeerNumPCErrRcvd because
    not all PCErr messages are used to reject requests, and a
    single PCErr message can reject multiple requests."
 ::= { pcePcepPeerEntry 34 }

pcePcepPeerNumReqSentTimeout OBJECT-TYPE
SYNTAX      Counter32
MAX-ACCESS  read-only
STATUS      current
DESCRIPTION
    "The number of requests that have been sent to a peer and
    have been abandoned because the peer has taken too long to
    respond to them."
 ::= { pcePcepPeerEntry 35 }

pcePcepPeerNumReqSentCancelSent OBJECT-TYPE
SYNTAX      Counter32
```

```
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "The number of requests that were sent to the peer and
    explicitly canceled by the local PCEP entity sending a
    PCNtf."
 ::= { pcePcepPeerEntry 36 }
```

```
pcePcepPeerNumReqSentClosed OBJECT-TYPE
SYNTAX        Counter32
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "The number of requests that were sent to the peer and
    implicitly canceled when the session they were sent over was
    closed."
 ::= { pcePcepPeerEntry 37 }
```

```
pcePcepPeerNumReqRcvd OBJECT-TYPE
SYNTAX        Counter32
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "The number of requests received from this peer.  A request
    corresponds 1:1 with an RP object in a PCReq message.

    This might be greater than pcePcepPeerNumPCReqRcvd because
    multiple requests can be batched into a single PCReq
    message."
 ::= { pcePcepPeerEntry 38 }
```

```
pcePcepPeerNumSvecRcvd OBJECT-TYPE
SYNTAX        Counter32
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "The number of SVEC objects received from this peer in PCReq
    messages.  An SVEC object represents a set of synchronized
    requests."
 ::= { pcePcepPeerEntry 39 }
```

```
pcePcepPeerNumSvecReqRcvd OBJECT-TYPE
SYNTAX        Counter32
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "The number of requests received from this peer that appeared
    in one or more SVEC objects."
```

```
::= { pcePcepPeerEntry 40 }

pcePcepPeerNumReqRcvdPendRep OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of requests that have been received from this
        peer for which a response is still pending."
    ::= { pcePcepPeerEntry 41 }

pcePcepPeerNumReqRcvdEroSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of requests that have been received from this
        peer for which a response with an ERO object was sent.  Such
        responses indicate that a path was successfully computed by
        the local PCEP entity."
    ::= { pcePcepPeerEntry 42 }

pcePcepPeerNumReqRcvdNoPathSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of requests that have been received from this
        peer for which a response with a NO-PATH object was sent.
        Such responses indicate that the local PCEP entity could
        not find a path to satisfy the request."
    ::= { pcePcepPeerEntry 43 }

pcePcepPeerNumReqRcvdCancelSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of requests received from this peer that were
        cancelled by the local PCEP entity sending a PCNtf message.

        This might be different than pcePcepPeerNumPCNtfSent because
        not all PCNtf messages are used to cancel requests, and a
        single PCNtf message can cancel multiple requests."
    ::= { pcePcepPeerEntry 44 }

pcePcepPeerNumReqRcvdErrorSent OBJECT-TYPE
    SYNTAX      Counter32
```

```
MAX-ACCESS    read-only
STATUS        current
DESCRIPTION
    "The number of requests received from this peer that were
    rejected by the local PCEP entity sending a PCErr message.

    This might be different than pcePcepPeerNumPCErrSent because
    not all PCErr messages are used to reject requests, and a
    single PCErr message can reject multiple requests."
 ::= { pcePcepPeerEntry 45 }

pcePcepPeerNumReqRcvdCancelRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of requests that were received from the peer and
        explicitly canceled by the peer sending a PCNtf."
    ::= { pcePcepPeerEntry 46 }

pcePcepPeerNumReqRcvdClosed OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of requests that were received from the peer and
        implicitly canceled when the session they were received over
        was closed."
    ::= { pcePcepPeerEntry 47 }

pcePcepPeerNumRepRcvdUnknown OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of responses to unknown requests received from
        this peer. A response to an unknown request is a response
        whose RP object does not contain the request ID of any
        request that is currently outstanding on the session."
    ::= { pcePcepPeerEntry 48 }

pcePcepPeerNumReqRcvdUnknown OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of unknown requests that have been received from
        a peer. An unknown request is a request whose RP object
```

```

        contains a request ID of zero."
 ::= { pcePcepPeerEntry 49 }

--
-- The PCEP Sessions Table
--

pcePcepSessTable OBJECT-TYPE
    SYNTAX      SEQUENCE OF PcePcepSessEntry
    MAX-ACCESS   not-accessible
    STATUS       current
    DESCRIPTION
        "A table of PCEP sessions that involve the local PCEP
        entity. Each entry in this table represents a single
        session. The entries in this table are read-only.

        An entry appears in this table when the corresponding PCEP
        session transitions out of idle state. If the PCEP session
        transitions back into idle state then the corresponding
        entry in this table is removed."
 ::= { pcePcepObjects 3 }

pcePcepSessEntry OBJECT-TYPE
    SYNTAX      PcePcepSessEntry
    MAX-ACCESS   not-accessible
    STATUS       current
    DESCRIPTION
        "This entry represents a single PCEP session in which the
        local PCEP entity participates.

        This entry exists only if the corresponding PCEP session has
        been initialized by some event, such as manual user
        configuration, autodiscovery of a peer, or an incoming TCP
        connection."
    INDEX { pcePcepEntityIndex,
            pcePcepPeerAddrType,
            pcePcepPeerAddr,
            pcePcepSessInitiator }
 ::= { pcePcepSessTable 1 }

PcePcepSessEntry ::= SEQUENCE {
    pcePcepSessInitiator          INTEGER,
    pcePcepSessStateLastChange    TimeStamp,
    pcePcepSessState              INTEGER,
    pcePcepSessConnectRetry       Counter32,
    pcePcepSessLocalID            Unsigned32,
    pcePcepSessRemoteID           Unsigned32,
    pcePcepSessKeepaliveTimer     Unsigned32,

```

pcePcepSessPeerKeepaliveTimer	Unsigned32,
pcePcepSessDeadTimer	Unsigned32,
pcePcepSessPeerDeadTimer	Unsigned32,
pcePcepSessKAHoldTimeRem	Unsigned32,
pcePcepSessOverloaded	TruthValue,
pcePcepSessOverloadTime	Unsigned32,
pcePcepSessPeerOverloaded	TruthValue,
pcePcepSessPeerOverloadTime	Unsigned32,
pcePcepSessDiscontinuityTime	TimeStamp,
pcePcepSessAvgRspTime	Unsigned32,
pcePcepSessLWMRspTime	Unsigned32,
pcePcepSessHWMRspTime	Unsigned32,
pcePcepSessNumPCReqSent	Counter32,
pcePcepSessNumPCReqRcvd	Counter32,
pcePcepSessNumPCRepSent	Counter32,
pcePcepSessNumPCRepRcvd	Counter32,
pcePcepSessNumPCErrSent	Counter32,
pcePcepSessNumPCErrRcvd	Counter32,
pcePcepSessNumPCNtfSent	Counter32,
pcePcepSessNumPCNtfRcvd	Counter32,
pcePcepSessNumKeepaliveSent	Counter32,
pcePcepSessNumKeepaliveRcvd	Counter32,
pcePcepSessNumUnknownRcvd	Counter32,
pcePcepSessNumCorruptRcvd	Counter32,
pcePcepSessNumReqSent	Counter32,
pcePcepSessNumSvecSent	Counter32,
pcePcepSessNumSvecReqSent	Counter32,
pcePcepSessNumReqSentPendRep	Counter32,
pcePcepSessNumReqSentEroRcvd	Counter32,
pcePcepSessNumReqSentNoPathRcvd	Counter32,
pcePcepSessNumReqSentCancelRcvd	Counter32,
pcePcepSessNumReqSentErrorRcvd	Counter32,
pcePcepSessNumReqSentTimeout	Counter32,
pcePcepSessNumReqSentCancelSent	Counter32,
pcePcepSessNumReqRcvd	Counter32,
pcePcepSessNumSvecRcvd	Counter32,
pcePcepSessNumSvecReqRcvd	Counter32,
pcePcepSessNumReqRcvdPendRep	Counter32,
pcePcepSessNumReqRcvdEroSent	Counter32,
pcePcepSessNumReqRcvdNoPathSent	Counter32,
pcePcepSessNumReqRcvdCancelSent	Counter32,
pcePcepSessNumReqRcvdErrorSent	Counter32,
pcePcepSessNumReqRcvdCancelRcvd	Counter32,
pcePcepSessNumRepRcvdUnknown	Counter32,
pcePcepSessNumReqRcvdUnknown	Counter32

}

pcePcepSessInitiator OBJECT-TYPE

```

SYNTAX      INTEGER {
                local(1),
                remote(2)
            }
MAX-ACCESS   not-accessible
STATUS       current
DESCRIPTION
    "The initiator of the session, that is, whether the TCP
    connection was initiated by the local PCEP entity or the
    peer.

    There is a window during session initialization where two
    sessions can exist between a pair of PCEP speakers, each
    initiated by one of the speakers.  One of these sessions is
    always discarded before it leaves OpenWait state.  However,
    before it is discarded, two sessions to the given peer
    appear transiently in this MIB module.  The sessions are
    distinguished by who initiated them, and so this field is an
    index for the pcePcepSessTable."
 ::= { pcePcepSessEntry 1 }

pcePcepSessStateLastChange OBJECT-TYPE
    SYNTAX      TimeStamp
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The value of sysUpTime at the time this session entered its
        current state as denoted by the pcePcepSessState object."
 ::= { pcePcepSessEntry 2 }

pcePcepSessState OBJECT-TYPE
    SYNTAX      INTEGER {
                tcpPending(1),
                openWait(2),
                keepWait(3),
                sessionUp(4)
            }
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The current state of the session.

        The set of possible states excludes the idle state since
        entries do not exist in this table in the idle state."
 ::= { pcePcepSessEntry 3 }

pcePcepSessConnectRetry OBJECT-TYPE
    SYNTAX      Counter32

```

MAX-ACCESS read-only
STATUS current
DESCRIPTION
 "The number of times that the local PCEP entity has attempted to establish a TCP connection for this session without success. The PCEP entity gives up when this reaches pcePcepEntityConnectMaxRetry."
 ::= { pcePcepSessEntry 4 }

pcePcepSessLocalID OBJECT-TYPE
SYNTAX Unsigned32 (0..255)
MAX-ACCESS read-only
STATUS current
DESCRIPTION
 "The value of the PCEP session ID used by the local PCEP entity in the Open message for this session.

 If pcePcepSessState is tcpPending then this is the session ID that will be used in the Open message. Otherwise, this is the session ID that was sent in the Open message."
 ::= { pcePcepSessEntry 5 }

pcePcepSessRemoteID OBJECT-TYPE
SYNTAX Unsigned32 (0..255)
MAX-ACCESS read-only
STATUS current
DESCRIPTION
 "The value of the PCEP session ID used by the peer in its Open message for this session.

 If pcePcepSessState is tcpPending or openWait then this field is not used and MUST be set to zero."
 ::= { pcePcepSessEntry 6 }

pcePcepSessKeepaliveTimer OBJECT-TYPE
SYNTAX Unsigned32 (0..255)
UNITS "seconds"
MAX-ACCESS read-only
STATUS current
DESCRIPTION
 "The agreed maximum interval at which the local PCEP entity transmits PCEP messages on this PCEP session. Zero means that the local PCEP entity never sends Keepalives on this session.

 This field is used if and only if pcePcepSessState is sessionUp. Otherwise, it is not used and MUST be set to zero."


```
::= { pcePcepSessEntry 7 }
```

pcePcepSessPeerKeepaliveTimer OBJECT-TYPE

SYNTAX Unsigned32 (0..255)

UNITS "seconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The agreed maximum interval at which the peer transmits PCEP messages on this PCEP session. Zero means that the peer never sends Keepalives on this session.

This field is used if and only if pcePcepSessState is sessionUp. Otherwise, it is not used and MUST be set to zero."

```
::= { pcePcepSessEntry 8 }
```

pcePcepSessDeadTimer OBJECT-TYPE

SYNTAX Unsigned32 (0..255)

UNITS "seconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The DeadTimer interval for this PCEP session."

```
::= { pcePcepSessEntry 9 }
```

pcePcepSessPeerDeadTimer OBJECT-TYPE

SYNTAX Unsigned32 (0..255)

UNITS "seconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The peer's DeadTimer interval for this PCEP session.

If pcePcepSessState is tcpPending or openWait then this field is not used and MUST be set to zero."

```
::= { pcePcepSessEntry 10 }
```

pcePcepSessKAHoldTimeRem OBJECT-TYPE

SYNTAX Unsigned32 (0..255)

UNITS "seconds"

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The keep alive hold time remaining for this session.

If pcePcepSessState is tcpPending or openWait then this field is not used and MUST be set to zero."

```
::= { pcePcepSessEntry 11 }

pcePcepSessOverloaded OBJECT-TYPE
    SYNTAX      TruthValue
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "If the local PCEP entity has informed the peer that it is
         currently overloaded, then this is set to true.  Otherwise,
         it is set to false."
    ::= { pcePcepSessEntry 12 }

pcePcepSessOverloadTime OBJECT-TYPE
    SYNTAX      Unsigned32
    UNITS       "seconds"
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The interval of time that is remaining until the local PCEP
         entity will cease to be overloaded on this session.

         This field is only used if pcePcepSessOverloaded is set to
         true.  Otherwise, it is not used and MUST be set to zero."
    ::= { pcePcepSessEntry 13 }

pcePcepSessPeerOverloaded OBJECT-TYPE
    SYNTAX      TruthValue
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "If the peer has informed the local PCEP entity that it is
         currently overloaded, then this is set to true.  Otherwise,
         it is set to false."
    ::= { pcePcepSessEntry 14 }

pcePcepSessPeerOverloadTime OBJECT-TYPE
    SYNTAX      Unsigned32
    UNITS       "seconds"
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The interval of time that is remaining until the peer will
         cease to be overloaded.  If it is not known how long the
         peer will stay in overloaded state, this field is set to
         zero.

         This field is only used if pcePcepSessPeerOverloaded is set
         to true.  Otherwise, it is not used and MUST be set to
```

```
        zero."
 ::= { pcePcepSessEntry 15 }

pcePcepSessDiscontinuityTime OBJECT-TYPE
    SYNTAX      TimeStamp
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The value of sysUpTime at the time that the statistics in
        this row were last reset."
 ::= { pcePcepSessEntry 16 }

pcePcepSessAvgRspTime OBJECT-TYPE
    SYNTAX      Unsigned32
    UNITS        "milliseconds"
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The average response time for this peer on this session.

        If an average response time has not been calculated for this
        peer then this object has the value zero."
 ::= { pcePcepSessEntry 17 }

pcePcepSessLWMRspTime OBJECT-TYPE
    SYNTAX      Unsigned32
    UNITS        "milliseconds"
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The smallest (low-water mark) response time seen from this
        peer on this session.

        If no responses have been received from this peer then this
        object has the value zero."
 ::= { pcePcepSessEntry 18 }

pcePcepSessHWMRspTime OBJECT-TYPE
    SYNTAX      Unsigned32
    UNITS        "milliseconds"
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The greatest (high-water mark) response time seen from this
        peer on this session.

        If no responses have been received from this peer then this
        object has the value zero."
```

```
 ::= { pcePcepSessEntry 19 }

pcePcepSessNumPCReqSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of PCReq messages sent on this session."
    ::= { pcePcepSessEntry 20 }

pcePcepSessNumPCReqRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of PCReq messages received on this session."
    ::= { pcePcepSessEntry 21 }

pcePcepSessNumPCRepSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of PCRep messages sent on this session."
    ::= { pcePcepSessEntry 22 }

pcePcepSessNumPCRepRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of PCRep messages received on this session."
    ::= { pcePcepSessEntry 23 }

pcePcepSessNumPCErrSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of PCErr messages sent on this session."
    ::= { pcePcepSessEntry 24 }

pcePcepSessNumPCErrRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS  read-only
    STATUS      current
    DESCRIPTION
        "The number of PCErr messages received on this session."
```

```
 ::= { pcePcepSessEntry 25 }

pcePcepSessNumPCNtfSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCNtf messages sent on this session."
    ::= { pcePcepSessEntry 26 }

pcePcepSessNumPCNtfRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of PCNtf messages received on this session."
    ::= { pcePcepSessEntry 27 }

pcePcepSessNumKeepaliveSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of Keepalive messages sent on this session."
    ::= { pcePcepSessEntry 28 }

pcePcepSessNumKeepaliveRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of Keepalive messages received on this session."
    ::= { pcePcepSessEntry 29 }

pcePcepSessNumUnknownRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of unknown messages received on this session."
    ::= { pcePcepSessEntry 30 }

pcePcepSessNumCorruptRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of corrupted PCEP message received on this
```

```
        session."
 ::= { pcePcepSessEntry 31 }

pcePcepSessNumReqSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of requests sent on this session.  A request
        corresponds 1:1 with an RP object in a PCReq message.

        This might be greater than pcePcepSessNumPCReqSent because
        multiple requests can be batched into a single PCReq
        message."
 ::= { pcePcepSessEntry 32 }

pcePcepSessNumSvecSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of SVEC objects sent on this session in PCReq
        messages.  An SVEC object represents a set of synchronized
        requests."
 ::= { pcePcepSessEntry 33 }

pcePcepSessNumSvecReqSent OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of requests sent on this session that appeared in
        one or more SVEC objects."
 ::= { pcePcepSessEntry 34 }

pcePcepSessNumReqSentPendRep OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
    DESCRIPTION
        "The number of requests that have been sent on this session
        for which a response is still pending."
 ::= { pcePcepSessEntry 35 }

pcePcepSessNumReqSentEroRcvd OBJECT-TYPE
    SYNTAX      Counter32
    MAX-ACCESS   read-only
    STATUS       current
```

DESCRIPTION

"The number of successful responses received on this session. A response corresponds 1:1 with an RP object in a PCRep message. A successful response is a response for which an ERO was successfully computed."

::= { pcePcepSessEntry 36 }

pcePcepSessNumReqSentNoPathRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of unsuccessful responses received on this session. A response corresponds 1:1 with an RP object in a PCRep message. An unsuccessful response is a response with a NO-PATH object."

::= { pcePcepSessEntry 37 }

pcePcepSessNumReqSentCancelRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests sent on this session that were cancelled by the peer with a PCNtf message."

This might be different than pcePcepSessNumPCNtfRcvd because not all PCNtf messages are used to cancel requests, and a single PCNtf message can cancel multiple requests."

::= { pcePcepSessEntry 38 }

pcePcepSessNumReqSentErrorRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests sent on this session that were rejected by the peer with a PCErr message."

This might be different than pcePcepSessNumPCErrRcvd because not all PCErr messages are used to reject requests, and a single PCErr message can reject multiple requests."

::= { pcePcepSessEntry 39 }

pcePcepSessNumReqSentTimeout OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests sent on this session that have been sent to a peer and have been abandoned because the peer has taken too long to respond to them."

::= { pcePcepSessEntry 40 }

pcePcepSessNumReqSentCancelSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests sent on this session that were sent to the peer and explicitly canceled by the local PCEP entity sending a PCNtf."

::= { pcePcepSessEntry 41 }

pcePcepSessNumReqRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests received on this session. A request corresponds 1:1 with an RP object in a PCReq message."

This might be greater than pcePcepSessNumPCReqRcvd because multiple requests can be batched into a single PCReq message."

::= { pcePcepSessEntry 42 }

pcePcepSessNumSvecRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of SVEC objects received on this session in PCReq messages. An SVEC object represents a set of synchronized requests."

::= { pcePcepSessEntry 43 }

pcePcepSessNumSvecReqRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests received on this session that appeared in one or more SVEC objects."

::= { pcePcepSessEntry 44 }

pcePcepSessNumReqRcvdPendRep OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests that have been received on this session for which a response is still pending."

::= { pcePcepSessEntry 45 }

pcePcepSessNumReqRcvdEroSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of successful responses sent on this session. A response corresponds 1:1 with an RP object in a PCRep message. A successful response is a response for which an ERO was successfully computed."

::= { pcePcepSessEntry 46 }

pcePcepSessNumReqRcvdNoPathSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of unsuccessful responses sent on this session. A response corresponds 1:1 with an RP object in a PCRep message. An unsuccessful response is a response with a NO-PATH object."

::= { pcePcepSessEntry 47 }

pcePcepSessNumReqRcvdCancelSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests received on this session that were cancelled by the local PCEP entity sending a PCNtf message."

This might be different than pcePcepSessNumPCNtfSent because not all PCNtf messages are used to cancel requests, and a single PCNtf message can cancel multiple requests."

::= { pcePcepSessEntry 48 }

pcePcepSessNumReqRcvdErrorSent OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests received on this session that were rejected by the local PCEP entity sending a PCErr message.

This might be different than pcePcepSessNumPCErrSent because not all PCErr messages are used to reject requests, and a single PCErr message can reject multiple requests."

::= { pcePcepSessEntry 49 }

pcePcepSessNumReqRcvdCancelRcvd OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of requests that were received on this session and explicitly canceled by the peer sending a PCNtf."

::= { pcePcepSessEntry 50 }

pcePcepSessNumRepRcvdUnknown OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of responses to unknown requests received on this session. A response to an unknown request is a response whose RP object does not contain the request ID of any request that is currently outstanding on the session."

::= { pcePcepSessEntry 51 }

pcePcepSessNumReqRcvdUnknown OBJECT-TYPE

SYNTAX Counter32

MAX-ACCESS read-only

STATUS current

DESCRIPTION

"The number of unknown requests that have been received on this session. An unknown request is a request whose RP object contains a request ID of zero."

::= { pcePcepSessEntry 52 }

--- Notifications Configuration

pcePcepNotificationsMaxRate OBJECT-TYPE

SYNTAX Unsigned32

MAX-ACCESS read-write

STATUS current

DESCRIPTION

```
"This variable indicates the maximum number of
notifications issued per second. If events occur
more rapidly, the implementation may simply fail to
emit these notifications during that period, or may
queue them until an appropriate time. A value of 0
means no notifications are emitted and all should be
discarded (that is, not queued)."
```

```
::= { pcePcepObjects 4 }
```

```
---
```

```
--- Notifications
```

```
---
```

```
pcePcepSessUp NOTIFICATION-TYPE
  OBJECTS      {
                pcePcepSessState,
                pcePcepSessStateLastChange
              }
  STATUS        current
  DESCRIPTION   "This notification is sent when the value of
                'pcePcepSessState' enters the 'sessionUp' state."
  ::= { pcePcepNotifications 1 }
```

```
pcePcepSessDown NOTIFICATION-TYPE
  OBJECTS      {
                pcePcepSessState,
                pcePcepSessStateLastChange
              }
  STATUS        current
  DESCRIPTION   "This notification is sent when the value of
                'pcePcepSessState' leaves the 'sessionUp' state."
  ::= { pcePcepNotifications 2 }
```

```
pcePcepSessLocalOverload NOTIFICATION-TYPE
  OBJECTS      {
                pcePcepSessOverloaded,
                pcePcepSessOverloadTime
              }
  STATUS        current
  DESCRIPTION   "This notification is sent when the local PCEP entity enters
                overload state for a peer."
  ::= { pcePcepNotifications 3 }
```

```
pcePcepSessLocalOverloadClear NOTIFICATION-TYPE
  OBJECTS      {
```

```

        pcePcepSessOverloaded
    }
    STATUS current
    DESCRIPTION
        "This notification is sent when the local PCEP entity leaves
        overload state for a peer."
    ::= { pcePcepNotifications 4 }

pcePcepSessPeerOverload NOTIFICATION-TYPE
    OBJECTS {
        pcePcepSessPeerOverloaded,
        pcePcepSessPeerOverloadTime
    }
    STATUS current
    DESCRIPTION
        "This notification is sent when a peer enters overload
        state."
    ::= { pcePcepNotifications 5 }

pcePcepSessPeerOverloadClear NOTIFICATION-TYPE
    OBJECTS {
        pcePcepSessPeerOverloaded
    }
    STATUS current
    DESCRIPTION
        "This notification is sent when a peer leaves overload
        state."
    ::= { pcePcepNotifications 6 }

--
-- Module Conformance Statement
--

pcePcepCompliances
    OBJECT IDENTIFIER ::= { pcePcepConformance 1 }

pcePcepGroups
    OBJECT IDENTIFIER ::= { pcePcepConformance 2 }

--
-- Read-Only Compliance
--

pcePcepModuleReadOnlyCompliance MODULE-COMPLIANCE
    STATUS current
    DESCRIPTION
        "The Module is implemented with support for read-only. In
        other words, only monitoring is available by implementing
```

```
        this MODULE-COMPLIANCE."

MODULE -- this module
    MANDATORY-GROUPS {
        pcePcepGeneralGroup,
        pcePcepNotificationsGroup
    }

OBJECT      pcePcepEntityAddrType
SYNTAX      InetAddressType { unknown(0), ipv4(1), ipv6(2) }
DESCRIPTION "Only unknown(0), ipv4(1) and ipv6(2) support
            is required."

OBJECT      pcePcepPeerAddrType
SYNTAX      InetAddressType { unknown(0), ipv4(1), ipv6(2) }
DESCRIPTION "Only unknown(0), ipv4(1) and ipv6(2) support
            is required."

::= { pcePcepCompliances 1 }

-- units of conformance

pcePcepGeneralGroup OBJECT-GROUP
    OBJECTS { pcePcepEntityAdminStatus,
              pcePcepEntityOperStatus,
              pcePcepEntityAddrType,
              pcePcepEntityAddr,
              pcePcepEntityConnectTimer,
              pcePcepEntityConnectMaxRetry,
              pcePcepEntityInitBackoffTimer,
              pcePcepEntityMaxBackoffTimer,
              pcePcepEntityOpenWaitTimer,
              pcePcepEntityKeepWaitTimer,
              pcePcepEntityKeepAliveTimer,
              pcePcepEntityDeadTimer,
              pcePcepEntityAllowNegotiation,
              pcePcepEntityMaxKeepAliveTimer,
              pcePcepEntityMaxDeadTimer,
              pcePcepEntityMinKeepAliveTimer,
              pcePcepEntityMinDeadTimer,
              pcePcepEntitySyncTimer,
              pcePcepEntityRequestTimer,
              pcePcepEntityMaxSessions,
              pcePcepEntityMaxUnknownReqs,
              pcePcepEntityMaxUnknownMsgs,
              pcePcepPeerRole,
              pcePcepPeerDiscontinuityTime,
              pcePcepPeerInitiateSession,
```

pcePcepPeerSessionExists,
pcePcepPeerNumSessSetupOK,
pcePcepPeerNumSessSetupFail,
pcePcepPeerSessionUpTime,
pcePcepPeerSessionFailTime,
pcePcepPeerSessionFailUpTime,
pcePcepPeerAvgRspTime,
pcePcepPeerLWMRspTime,
pcePcepPeerHWMRspTime,
pcePcepPeerNumPCReqSent,
pcePcepPeerNumPCReqRcvd,
pcePcepPeerNumPCRepSent,
pcePcepPeerNumPCRepRcvd,
pcePcepPeerNumPCErrSent,
pcePcepPeerNumPCErrRcvd,
pcePcepPeerNumPCNtfSent,
pcePcepPeerNumPCNtfRcvd,
pcePcepPeerNumKeepaliveSent,
pcePcepPeerNumKeepaliveRcvd,
pcePcepPeerNumUnknownRcvd,
pcePcepPeerNumCorruptRcvd,
pcePcepPeerNumReqSent,
pcePcepPeerNumSvecSent,
pcePcepPeerNumSvecReqSent,
pcePcepPeerNumReqSentPendRep,
pcePcepPeerNumReqSentEroRcvd,
pcePcepPeerNumReqSentNoPathRcvd,
pcePcepPeerNumReqSentCancelRcvd,
pcePcepPeerNumReqSentErrorRcvd,
pcePcepPeerNumReqSentTimeout,
pcePcepPeerNumReqSentCancelSent,
pcePcepPeerNumReqSentClosed,
pcePcepPeerNumReqRcvd,
pcePcepPeerNumSvecRcvd,
pcePcepPeerNumSvecReqRcvd,
pcePcepPeerNumReqRcvdPendRep,
pcePcepPeerNumReqRcvdEroSent,
pcePcepPeerNumReqRcvdNoPathSent,
pcePcepPeerNumReqRcvdCancelSent,
pcePcepPeerNumReqRcvdErrorSent,
pcePcepPeerNumReqRcvdCancelRcvd,
pcePcepPeerNumReqRcvdClosed,
pcePcepPeerNumRepRcvdUnknown,
pcePcepPeerNumReqRcvdUnknown,
pcePcepSessStateLastChange,
pcePcepSessState,
pcePcepSessConnectRetry,
pcePcepSessLocalID,

pcePcepSessRemoteID,
pcePcepSessKeepaliveTimer,
pcePcepSessPeerKeepaliveTimer,
pcePcepSessDeadTimer,
pcePcepSessPeerDeadTimer,
pcePcepSessKAHoldTimeRem,
pcePcepSessOverloaded,
pcePcepSessOverloadTime,
pcePcepSessPeerOverloaded,
pcePcepSessPeerOverloadTime,
pcePcepSessDiscontinuityTime,
pcePcepSessAvgRspTime,
pcePcepSessLWMRspTime,
pcePcepSessHWMRspTime,
pcePcepSessNumPCReqSent,
pcePcepSessNumPCReqRcvd,
pcePcepSessNumPCRepSent,
pcePcepSessNumPCRepRcvd,
pcePcepSessNumPCErrSent,
pcePcepSessNumPCErrRcvd,
pcePcepSessNumPCNtfSent,
pcePcepSessNumPCNtfRcvd,
pcePcepSessNumKeepaliveSent,
pcePcepSessNumKeepaliveRcvd,
pcePcepSessNumUnknownRcvd,
pcePcepSessNumCorruptRcvd,
pcePcepSessNumReqSent,
pcePcepSessNumSvecSent,
pcePcepSessNumSvecReqSent,
pcePcepSessNumReqSentPendRep,
pcePcepSessNumReqSentEroRcvd,
pcePcepSessNumReqSentNoPathRcvd,
pcePcepSessNumReqSentCancelRcvd,
pcePcepSessNumReqSentErrorRcvd,
pcePcepSessNumReqSentTimeout,
pcePcepSessNumReqSentCancelSent,
pcePcepSessNumReqRcvd,
pcePcepSessNumSvecRcvd,
pcePcepSessNumSvecReqRcvd,
pcePcepSessNumReqRcvdPendRep,
pcePcepSessNumReqRcvdEroSent,
pcePcepSessNumReqRcvdNoPathSent,
pcePcepSessNumReqRcvdCancelSent,
pcePcepSessNumReqRcvdErrorSent,
pcePcepSessNumReqRcvdCancelRcvd,
pcePcepSessNumRepRcvdUnknown,
pcePcepSessNumReqRcvdUnknown,
pcePcepNotificationsMaxRate

```

    }
    STATUS current
    DESCRIPTION
        "Objects that apply to all PCEP MIB module implementations."
    ::= { pcePcepGroups 1 }

pcePcepNotificationsGroup NOTIFICATION-GROUP
    NOTIFICATIONS { pcePcepSessUp,
                    pcePcepSessDown,
                    pcePcepSessLocalOverload,
                    pcePcepSessLocalOverloadClear,
                    pcePcepSessPeerOverload,
                    pcePcepSessPeerOverloadClear
    }
    STATUS current
    DESCRIPTION
        "The notifications for a PCEP MIB module implementation."
    ::= { pcePcepGroups 2 }

END

```

5. Security Considerations

The pcePcepNotificationsMaxRate object defined in this MIB module has a MAX-ACCESS clause of read-write. Such objects may be considered sensitive or vulnerable in some network environments. The support for SET operations in a non-secure environment without proper protection can have a negative effect on network operations. In particular, pcePcepNotificationsMaxRate may be used improperly to stop notifications being issued, or to permit a flood of notifications to be sent to the management agent at a high rate.

The readable objects in this MIB module (i.e., objects with a MAX-ACCESS other than not-accessible) may be considered sensitive or vulnerable in some network environments since, collectively, they provide information about the amount and frequency of path computation requests and responses within the network and can reveal some aspects of its configuration. It is thus important to control even GET and/or NOTIFY access to these objects and possibly to even encrypt the values of these objects when sending them over the network via SNMP.

SNMP versions prior to SNMPv3 did not include adequate security. Even if the network itself is secure (for example by using IPsec), there is no control as to who on the secure network is allowed to access and GET/SET (read/change/create/delete) the objects in this MIB module.

Implementations SHOULD provide the security features described by the SNMPv3 framework (see [RFC3410]), and implementations claiming compliance to the SNMPv3 standard MUST include full support for authentication and privacy via the User-based Security Model (USM) [RFC3414] with the AES cipher algorithm [RFC3826]. Implementations MAY also provide support for the Transport Security Model (TSM) [RFC5591] in combination with a secure transport such as SSH [RFC5592] or TLS/DTLS [RFC6353].

Further, deployment of SNMP versions prior to SNMPv3 is NOT RECOMMENDED. Instead, it is RECOMMENDED to deploy SNMPv3 and to enable cryptographic security. It is then a customer/operator responsibility to ensure that the SNMP entity giving access to an instance of this MIB module is properly configured to give access to the objects only to those principals (users) that have legitimate rights to indeed GET or SET (change/create/delete) them.

6. IANA Considerations

The MIB module in this document uses the following IANA-assigned OBJECT IDENTIFIER values recorded in the SMI Numbers registry:

Descriptor	OBJECT IDENTIFIER value
-----	-----
pcePcepMIB	{ mib-2 XXX }

Editor's Note (to be removed prior to publication): the IANA is requested to assign a value for "XXX" under the 'mib-2' subtree and to record the assignment in the SMI Numbers registry. When the assignment has been made, the RFC Editor is asked to replace "XXX" (here and in the MIB module) with the assigned value and to remove this note.

7. Acknowledgement

The authors would like to thank Santanu Mazumder, Meral Shirazipour and Adrian Farrel for their valuable input.

Funding for the RFC Editor function is currently provided by the Internet Society.

8. References

8.1. Normative References

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

- [RFC2578] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Structure of Management Information Version 2 (SMIv2)", STD 58, RFC 2578, April 1999.
- [RFC2579] McCloghrie, K., Ed., Perkins, D., Ed., and J. Schoenwaelder, Ed., "Textual Conventions for SMIv2", STD 58, RFC 2579, April 1999.
- [RFC2580] McCloghrie, K., Perkins, D., and J. Schoenwaelder, "Conformance Statements for SMIv2", STD 58, RFC 2580, April 1999.
- [RFC4001] Daniele, M., Haberman, B., Routhier, S., and J. Schoenwaelder, "Textual Conventions for Internet Network Addresses", RFC 4001, February 2005.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

8.2. Informative References

- [RFC3410] Case, J., Mundy, R., Partain, D., and B. Stewart, "Introduction and Applicability Statements for Internet-Standard Management Framework", RFC 3410, December 2002.
- [RFC3414] Blumenthal, U. and B. Wijnen, "User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)", STD 62, RFC 3414, December 2002.
- [RFC3826] Blumenthal, U., Maino, F., and K. McCloghrie, "The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model", RFC 3826, June 2004.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5591] Harrington, D. and W. Hardaker, "Transport Security Model for the Simple Network Management Protocol (SNMP)", STD 78, RFC 5591, June 2009.
- [RFC5592] Harrington, D., Salowey, J., and W. Hardaker, "Secure Shell Transport Model for the Simple Network Management Protocol (SNMP)", RFC 5592, June 2009.
- [RFC6353] Hardaker, W., "Transport Layer Security (TLS) Transport Model for the Simple Network Management Protocol (SNMP)", STD 78, RFC 6353, July 2011.

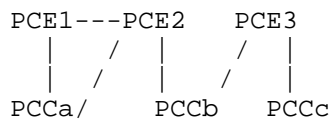
Appendix A. Contributors

Dhruv Dhody
 Huawei Technologies
 Leela Palace
 Bangalore, Karnataka 560008
 India

E-Mail: dhruv.ietf@gmail.com

Appendix B. PCEP MIB Module Example

This example considers the set of PCC / PCE relationships shown in the following figure. The example shows the contents of the PCEP MIB module as read at PCE2 and PCCb.



The IP addresses of the PCE speakers in this diagram are given in the following table.

PCE1	1.1.1.1
PCE2	2.2.2.2
PCE3	3.3.3.3
PCCa	11.11.11.11
PCCb	22.22.22.22
PCCc	33.33.33.33

In this example, the PCEP session between PCCb and PCE3 is currently down.

B.1. Contents of PCEP MIB module at PCE2

At PCE2, there is a single local PCEP entity which has three peers (PCCa, PCCb and PCE1). There is a session active to all of these peers.

The contents of the PCEP MIB module as read at PCE2 are as follows.

```

In pcePcepEntityTable {
    pcePcepEntityIndex          1,
    pcePcepEntityAdminStatus    adminStatusUp(1),
    pcePcepEntityOperStatus     operStatusUp(1),
    pcePcepEntityAddrType       ipv4(1),
    pcePcepEntityAddr           2.2.2.2, -- PCE2
    pcePcepEntityConnectTimer    60,
    pcePcepEntityConnectMaxRetry 5,
    pcePcepEntityInitBackoffTimer 30,
    pcePcepEntityMaxBackoffTimer 3600,
    pcePcepEntityOpenWaitTimer   60,
    pcePcepEntityKeepWaitTimer   60,
    pcePcepEntityKeepAliveTimer  1,
    pcePcepEntityDeadTimer       4,
    pcePcepEntityAllowNegotiation true(1),
    pcePcepEntityMaxKeepAliveTimer 60,
    pcePcepEntityMaxDeadTimer     240,
    pcePcepEntityMinKeepAliveTimer 1,
    pcePcepEntityMinDeadTimer     4,
    pcePcepEntitySyncTimer        60,
    pcePcepEntityRequestTimer     120,
    pcePcepEntityMaxSessions      999,
    pcePcepEntityMaxUnknownReqs   5,
    pcePcepEntityMaxUnknownMsgs   5
}

In pcePcepPeerTable {
    pcePcepPeerAddrType         ipv4(1), --PCE1
    pcePcepPeerAddr             1.1.1.1,
    pcePcepPeerRole              pccAndPce(3),
    pcePcepPeerDiscontinuityTime TimeStamp,
    pcePcepPeerInitiateSession  true(1),
    pcePcepPeerSessionExists     true(1),
    pcePcepPeerNumSessSetupOK    1,
    pcePcepPeerNumSessSetupFail  0,
    pcePcepPeerSessionUpTime     TimeStamp,
    pcePcepPeerSessionFailTime   0,
    pcePcepPeerSessionFailUpTime TimeStamp,
    pcePcepPeerAvgRspTime        0,
    pcePcepPeerLWMRspTime        0,
    pcePcepPeerHWMRspTime        0,
    pcePcepPeerNumPCReqSent      0,
    pcePcepPeerNumPCReqRcvd      0,
    pcePcepPeerNumPCRepSent      0,
    pcePcepPeerNumPCRepRcvd      0,
    pcePcepPeerNumPCErrSent      0,

```

```

pcePcepPeerNumPCErrRcvd          0,
pcePcepPeerNumPCNtfSent          0,
pcePcepPeerNumPCNtfRcvd          0,
pcePcepPeerNumKeepaliveSent      123,
pcePcepPeerNumKeepaliveRcvd      123,
pcePcepPeerNumUnknownRcvd        0,
pcePcepPeerNumCorruptRcvd        0,
pcePcepPeerNumReqSent            0,
pcePcepPeerNumSvecSent           0,
pcePcepPeerNumSvecReqSent        0,
pcePcepPeerNumReqSentPendRep     0,
pcePcepPeerNumReqSentEroRcvd     0,
pcePcepPeerNumReqSentNoPathRcvd  0,
pcePcepPeerNumReqSentCancelRcvd  0,
pcePcepPeerNumReqSentErrorRcvd   0,
pcePcepPeerNumReqSentTimeout     0,
pcePcepPeerNumReqSentCancelSent  0,
pcePcepPeerNumReqSentClosed      0,
pcePcepPeerNumReqRcvd            0,
pcePcepPeerNumSvecRcvd           0,
pcePcepPeerNumSvecReqRcvd        0,
pcePcepPeerNumReqRcvdPendRep     0,
pcePcepPeerNumReqRcvdEroSent     0,
pcePcepPeerNumReqRcvdNoPathSent  0,
pcePcepPeerNumReqRcvdCancelSent  0,
pcePcepPeerNumReqRcvdErrorSent   0,
pcePcepPeerNumReqRcvdCancelRcvd  0,
pcePcepPeerNumReqRcvdClosed      0,
pcePcepPeerNumRepRcvdUnknown     0,
pcePcepPeerNumReqRcvdUnknown     0
},
{
    pcePcepPeerAddrType          ipv4(1),  --PCCa
    pcePcepPeerAddr              11.11.11.11,
    pcePcepPeerRole              pcc(1),
    pcePcepPeerDiscontinuityTime  timeStamp,
    pcePcepPeerInitiateSession   false(0),
    pcePcepPeerSessionExists     true(1),
    pcePcepPeerNumSessSetupOK     1,
    pcePcepPeerNumSessSetupFail   0,
    pcePcepPeerSessionUpTime      timeStamp,
    pcePcepPeerSessionFailTime    0,
    pcePcepPeerSessionFailUpTime  timeStamp,
    pcePcepPeerAvgRspTime         200,
    pcePcepPeerLWMRspTime         100,
    pcePcepPeerHWMRspTime         300,
    pcePcepPeerNumPCReqSent       0,
    pcePcepPeerNumPCReqRcvd       3,

```

```

pcePcepPeerNumPCRepSent          3,
pcePcepPeerNumPCRepRcvd          0,
pcePcepPeerNumPCErrSent          0,
pcePcepPeerNumPCErrRcvd          0,
pcePcepPeerNumPCNtfSent          0,
pcePcepPeerNumPCNtfRcvd          0,
pcePcepPeerNumKeepaliveSent      123,
pcePcepPeerNumKeepaliveRcvd      123,
pcePcepPeerNumUnknownRcvd        0,
pcePcepPeerNumCorruptRcvd        0,
pcePcepPeerNumReqSent            0,
pcePcepPeerNumSvecSent           0,
pcePcepPeerNumSvecReqSent        0,
pcePcepPeerNumReqSentPendRep     0,
pcePcepPeerNumReqSentEroRcvd     0,
pcePcepPeerNumReqSentNoPathRcvd  0,
pcePcepPeerNumReqSentCancelRcvd  0,
pcePcepPeerNumReqSentErrorRcvd   0,
pcePcepPeerNumReqSentTimeout     0,
pcePcepPeerNumReqSentCancelSent  0,
pcePcepPeerNumReqSentClosed      0,
pcePcepPeerNumReqRcvd            3,
pcePcepPeerNumSvecRcvd           0,
pcePcepPeerNumSvecReqRcvd        0,
pcePcepPeerNumReqRcvdPendRep     0,
pcePcepPeerNumReqRcvdEroSent     3,
pcePcepPeerNumReqRcvdNoPathSent  0,
pcePcepPeerNumReqRcvdCancelSent  0,
pcePcepPeerNumReqRcvdErrorSent   0,
pcePcepPeerNumReqRcvdCancelRcvd  0,
pcePcepPeerNumReqRcvdClosed      0,
pcePcepPeerNumRepRcvdUnknown     0,
pcePcepPeerNumReqRcvdUnknown     0
},
{
pcePcepPeerAddrType              ipv4(1), -- PCCb
pcePcepPeerAddr                  22.22.22.22,
pcePcepPeerRole                  pcc(1),
pcePcepPeerDiscontinuityTime     TimeStamp,
pcePcepPeerInitiateSession       true(1),
pcePcepPeerSessionExists         true(1),
pcePcepPeerNumSessSetupOK        1,
pcePcepPeerNumSessSetupFail      0,
pcePcepPeerSessionUpTime         TimeStamp,
pcePcepPeerSessionFailTime       0,
pcePcepPeerSessionFailUpTime     TimeStamp,
pcePcepPeerAvgRspTime            200,
pcePcepPeerLWMRspTime            100,

```

```

pcePcepPeerHWMRspTime          300,
pcePcepPeerNumPCReqSent        0,
pcePcepPeerNumPCReqRcvd        4,
pcePcepPeerNumPCRepSent        4,
pcePcepPeerNumPCRepRcvd        0,
pcePcepPeerNumPCErrSent        0,
pcePcepPeerNumPCErrRcvd        0,
pcePcepPeerNumPCNtfSent        0,
pcePcepPeerNumPCNtfRcvd        0,
pcePcepPeerNumKeepaliveSent    123,
pcePcepPeerNumKeepaliveRcvd    123,
pcePcepPeerNumUnknownRcvd      0,
pcePcepPeerNumCorruptRcvd      0,
pcePcepPeerNumReqSent          0,
pcePcepPeerNumSvecSent         0,
pcePcepPeerNumSvecReqSent      0,
pcePcepPeerNumReqSentPendRep   0,
pcePcepPeerNumReqSentEroRcvd   0,
pcePcepPeerNumReqSentNoPathRcvd 0,
pcePcepPeerNumReqSentCancelRcvd 0,
pcePcepPeerNumReqSentErrorRcvd 0,
pcePcepPeerNumReqSentTimeout   0,
pcePcepPeerNumReqSentCancelSent 0,
pcePcepPeerNumReqSentClosed    0,
pcePcepPeerNumReqRcvd          4,
pcePcepPeerNumSvecRcvd         0,
pcePcepPeerNumSvecReqRcvd      0,
pcePcepPeerNumReqRcvdPendRep   0,
pcePcepPeerNumReqRcvdEroSent    3,
pcePcepPeerNumReqRcvdNoPathSent 1,
pcePcepPeerNumReqRcvdCancelSent 0,
pcePcepPeerNumReqRcvdErrorSent 0,
pcePcepPeerNumReqRcvdCancelRcvd 0,
pcePcepPeerNumReqRcvdClosed    0,
pcePcepPeerNumRepRcvdUnknown   0,
pcePcepPeerNumReqRcvdUnknown   0
}

In pcePcepSessTable {
    pcePcepSessInitiator          local(1), --PCE1
    pcePcepSessStateLastChange    TimeStamp,
    pcePcepSessState              sessionUp(4),
    pcePcepSessConnectRetry       0,
    pcePcepSessLocalID            1,
    pcePcepSessRemoteID           2,
    pcePcepSessKeepaliveTimer     1,
    pcePcepSessPeerKeepaliveTimer 1,
    pcePcepSessDeadTimer          4,

```

```

pcePcepSessPeerDeadTimer          4,
pcePcepSessKAHoldTimeRem          1,
pcePcepSessOverloaded             false(0),
pcePcepSessOverloadTime           0,
pcePcepSessPeerOverloaded         false(0),
pcePcepSessPeerOverloadTime       0,
pcePcepSessDiscontinuityTime      TimeStamp,
pcePcepSessAvgRspTime             0,
pcePcepSessLWMRspTime            0,
pcePcepSessHWMRspTime            0,
pcePcepSessNumPCReqSent           0,
pcePcepSessNumPCReqRcvd           0,
pcePcepSessNumPCRepSent           0,
pcePcepSessNumPCRepRcvd           0,
pcePcepSessNumPCErrSent           0,
pcePcepSessNumPCErrRcvd           0,
pcePcepSessNumPCNtfSent           0,
pcePcepSessNumPCNtfRcvd           0,
pcePcepSessNumKeepaliveSent       123,
pcePcepSessNumKeepaliveRcvd       123,
pcePcepSessNumUnknownRcvd         0,
pcePcepSessNumCorruptRcvd         0,
pcePcepSessNumReqSent             0,
pcePcepSessNumSvecSent            0,
pcePcepSessNumSvecReqSent         0,
pcePcepSessNumReqSentPendRep      0,
pcePcepSessNumReqSentEroRcvd      0,
pcePcepSessNumReqSentNoPathRcvd   0,
pcePcepSessNumReqSentCancelRcvd   0,
pcePcepSessNumReqSentErrorRcvd    0,
pcePcepSessNumReqSentTimeout      0,
pcePcepSessNumReqSentCancelSent   0,
pcePcepSessNumReqRcvd             0,
pcePcepSessNumSvecRcvd            0,
pcePcepSessNumSvecReqRcvd         0,
pcePcepSessNumReqRcvdPendRep      0,
pcePcepSessNumReqRcvdEroSent      0,
pcePcepSessNumReqRcvdNoPathSent   0,
pcePcepSessNumReqRcvdCancelSent   0,
pcePcepSessNumReqRcvdErrorSent    0,
pcePcepSessNumReqRcvdCancelRcvd   0,
pcePcepSessNumRepRcvdUnknown      0,
pcePcepSessNumReqRcvdUnknown      0
},
{
    pcePcepSessInitiator            remote(2), --PCCa
    pcePcepSessStateLastChange      TimeStamp,
    pcePcepSessState                sessionUp(4),

```


pcePcepSessConnectRetry	0,
pcePcepSessLocalID	2,
pcePcepSessRemoteID	1,
pcePcepSessKeepaliveTimer	1,
pcePcepSessPeerKeepaliveTimer	1,
pcePcepSessDeadTimer	4,
pcePcepSessPeerDeadTimer	4,
pcePcepSessKAHoldTimeRem	1,
pcePcepSessOverloaded	false(0),
pcePcepSessOverloadTime	0,
pcePcepSessPeerOverloaded	false(0),
pcePcepSessPeerOverloadTime	0,
pcePcepSessDiscontinuityTime	TimeStamp,
pcePcepSessAvgRspTime	200,
pcePcepSessLWMRspTime	100,
pcePcepSessHWMRspTime	300,
pcePcepSessNumPCReqSent	0,
pcePcepSessNumPCReqRcvd	1,
pcePcepSessNumPCRepSent	1,
pcePcepSessNumPCRepRcvd	0,
pcePcepSessNumPCErrSent	0,
pcePcepSessNumPCErrRcvd	0,
pcePcepSessNumPCNtfSent	0,
pcePcepSessNumPCNtfRcvd	0,
pcePcepSessNumKeepaliveSent	123,
pcePcepSessNumKeepaliveRcvd	123,
pcePcepSessNumUnknownRcvd	0,
pcePcepSessNumCorruptRcvd	0,
pcePcepSessNumReqSent	0,
pcePcepSessNumSvecSent	0,
pcePcepSessNumSvecReqSent	0,
pcePcepSessNumReqSentPendRep	0,
pcePcepSessNumReqSentEroRcvd	0,
pcePcepSessNumReqSentNoPathRcvd	0,
pcePcepSessNumReqSentCancelRcvd	0,
pcePcepSessNumReqSentErrorRcvd	0,
pcePcepSessNumReqSentTimeout	0,
pcePcepSessNumReqSentCancelSent	0,
pcePcepSessNumReqRcvd	3,
pcePcepSessNumSvecRcvd	0,
pcePcepSessNumSvecReqRcvd	0,
pcePcepSessNumReqRcvdPendRep	0,
pcePcepSessNumReqRcvdEroSent	3,
pcePcepSessNumReqRcvdNoPathSent	0,
pcePcepSessNumReqRcvdCancelSent	0,
pcePcepSessNumReqRcvdErrorSent	0,
pcePcepSessNumReqRcvdCancelRcvd	0,
pcePcepSessNumRepRcvdUnknown	0,

```

    pcePcepSessNumReqRcvdUnknown      0
},
{
    pcePcepSessInitiator               remote(2), --PCCb
    pcePcepSessStateLastChange         TimeStamp,
    pcePcepSessState                   sessionUp(4),
    pcePcepSessConnectRetry            0,
    pcePcepSessLocalID                 2,
    pcePcepSessRemoteID                1,
    pcePcepSessKeepaliveTimer          1,
    pcePcepSessPeerKeepaliveTimer      1,
    pcePcepSessDeadTimer               4,
    pcePcepSessPeerDeadTimer           4,
    pcePcepSessKAHoldTimeRem           1,
    pcePcepSessOverloaded              false(0),
    pcePcepSessOverloadTime            0,
    pcePcepSessPeerOverloaded          false(0),
    pcePcepSessPeerOverloadTime        0,
    pcePcepSessDiscontinuityTime       TimeStamp,
    pcePcepSessAvgRspTime              200,
    pcePcepSessLWMRspTime              100,
    pcePcepSessHWMRspTime              300,
    pcePcepSessNumPCReqSent            0,
    pcePcepSessNumPCReqRcvd            4,
    pcePcepSessNumPCRepSent            4,
    pcePcepSessNumPCRepRcvd            0,
    pcePcepSessNumPCErrSent            0,
    pcePcepSessNumPCErrRcvd            0,
    pcePcepSessNumPCNtfSent            0,
    pcePcepSessNumPCNtfRcvd            0,
    pcePcepSessNumKeepaliveSent        123,
    pcePcepSessNumKeepaliveRcvd        123,
    pcePcepSessNumUnknownRcvd          0,
    pcePcepSessNumCorruptRcvd          0,
    pcePcepSessNumReqSent              0,
    pcePcepSessNumSvecSent              0,
    pcePcepSessNumSvecReqSent          0,
    pcePcepSessNumReqSentPendRep       0,
    pcePcepSessNumReqSentEroRcvd       0,
    pcePcepSessNumReqSentNoPathRcvd    0,
    pcePcepSessNumReqSentCancelRcvd    0,
    pcePcepSessNumReqSentErrorRcvd     0,
    pcePcepSessNumReqSentTimeout       0,
    pcePcepSessNumReqSentCancelSent    0,
    pcePcepSessNumReqRcvd              4,
    pcePcepSessNumSvecRcvd             0,
    pcePcepSessNumSvecReqRcvd          0,
    pcePcepSessNumReqRcvdPendRep       0,

```

```

    pcePcepSessNumReqRcvdEroSent      3,
    pcePcepSessNumReqRcvdNoPathSent   1,
    pcePcepSessNumReqRcvdCancelSent   0,
    pcePcepSessNumReqRcvdErrorSent    0,
    pcePcepSessNumReqRcvdCancelRcvd   0,
    pcePcepSessNumRepRcvdUnknown      0,
    pcePcepSessNumReqRcvdUnknown      0
}

```

B.2. Contents of PCEP MIB module at PCCb

At PCCb, there is a single local PCEP entity which has two peers (PCE2 and PCE3). There is a session active to PCE2, but the session to PCE3 is currently down.

The contents of the PCEP MIB module as read at PCCb are as follows.

```

In pcePcepEntityTable {
    pcePcepEntityIndex      1,
    pcePcepEntityAdminStatus adminStatusUp(1),
    pcePcepEntityOperStatus operStatusUp(1),
    pcePcepEntityAddrType   ipv4(1),
    pcePcepEntityAddr       22.22.22.22, -- PCCb
    pcePcepEntityConnectTimer 60,
    pcePcepEntityConnectMaxRetry 5,
    pcePcepEntityInitBackoffTimer 30,
    pcePcepEntityMaxBackoffTimer 3600,
    pcePcepEntityOpenWaitTimer 60,
    pcePcepEntityKeepWaitTimer 60,
    pcePcepEntityKeepAliveTimer 1,
    pcePcepEntityDeadTimer 4,
    pcePcepEntityAllowNegotiation true(1),
    pcePcepEntityMaxKeepAliveTimer 60,
    pcePcepEntityMaxDeadTimer 240,
    pcePcepEntityMinKeepAliveTimer 1,
    pcePcepEntityMinDeadTimer 4,
    pcePcepEntitySyncTimer 60,
    pcePcepEntityRequestTimer 120,
    pcePcepEntityMaxSessions 999,
    pcePcepEntityMaxUnknownReqs 5,
    pcePcepEntityMaxUnknownMsgs 5
}

In pcePcepPeerTable {
    pcePcepPeerAddrType   ipv4(1), --PCE2
    pcePcepPeerAddr       2.2.2.2,
    pcePcepPeerRole        pce(2),
    pcePcepPeerDiscontinuityTime TimeStamp,

```

```

pcePcepPeerInitiateSession      true(1),
pcePcepPeerSessionExists        true(1)),
pcePcepPeerNumSessSetupOK       0,
pcePcepPeerNumSessSetupFail     1,
pcePcepPeerSessionUpTime        TimeStamp,
pcePcepPeerSessionFailTime      TimeStamp,
pcePcepPeerSessionFailUpTime    TimeStamp,
pcePcepPeerAvgRspTime           0,
pcePcepPeerLWMRspTime           0,
pcePcepPeerHWMRspTime           0,
pcePcepPeerNumPCReqSent         4,
pcePcepPeerNumPCReqRcvd         0,
pcePcepPeerNumPCRepSent         0,
pcePcepPeerNumPCRepRcvd         4,
pcePcepPeerNumPCErrSent         0,
pcePcepPeerNumPCErrRcvd         0,
pcePcepPeerNumPCNtfSent         0,
pcePcepPeerNumPCNtfRcvd         0,
pcePcepPeerNumKeepaliveSent     0,
pcePcepPeerNumKeepaliveRcvd     0,
pcePcepPeerNumUnknownRcvd       0,
pcePcepPeerNumCorruptRcvd       0,
pcePcepPeerNumReqSent           4,
pcePcepPeerNumSvecSent          0,
pcePcepPeerNumSvecReqSent       0,
pcePcepPeerNumReqSentPendRep    0,
pcePcepPeerNumReqSentEroRcvd    3,
pcePcepPeerNumReqSentNoPathRcvd 1,
pcePcepPeerNumReqSentCancelRcvd 0,
pcePcepPeerNumReqSentErrorRcvd  0,
pcePcepPeerNumReqSentTimeout    0,
pcePcepPeerNumReqSentCancelSent 0,
pcePcepPeerNumReqSentClosed     0,
pcePcepPeerNumReqRcvd           0,
pcePcepPeerNumSvecRcvd          0,
pcePcepPeerNumSvecReqRcvd       0,
pcePcepPeerNumReqRcvdPendRep    0,
pcePcepPeerNumReqRcvdEroSent    0,
pcePcepPeerNumReqRcvdNoPathSent 0,
pcePcepPeerNumReqRcvdCancelSent 0,
pcePcepPeerNumReqRcvdErrorSent  0,
pcePcepPeerNumReqRcvdCancelRcvd 0,
pcePcepPeerNumReqRcvdClosed     0,
pcePcepPeerNumRepRcvdUnknown    0,
pcePcepPeerNumReqRcvdUnknown    0
},
{
    pcePcepPeerAddrType           ipv4(1),  --PCE3

```

pcePcepPeerAddr	3.3.3.3,
pcePcepPeerRole	pce(2),
pcePcepPeerDiscontinuityTime	TimeStamp,
pcePcepPeerInitiateSession	true(1),
pcePcepPeerSessionExists	false(0),
pcePcepPeerNumSessSetupOK	1,
pcePcepPeerNumSessSetupFail	0,
pcePcepPeerSessionUpTime	TimeStamp,
pcePcepPeerSessionFailTime	TimeStamp,
pcePcepPeerSessionFailUpTime	TimeStamp,
pcePcepPeerAvgRspTime	200,
pcePcepPeerLWMRspTime	100,
pcePcepPeerHWMRspTime	300,
pcePcepPeerNumPCReqSent	4,
pcePcepPeerNumPCReqRcvd	0,
pcePcepPeerNumPCRepSent	0,
pcePcepPeerNumPCRepRcvd	3,
pcePcepPeerNumPCErrSent	0,
pcePcepPeerNumPCErrRcvd	0,
pcePcepPeerNumPCNtfSent	0,
pcePcepPeerNumPCNtfRcvd	0,
pcePcepPeerNumKeepaliveSent	123,
pcePcepPeerNumKeepaliveRcvd	123,
pcePcepPeerNumUnknownRcvd	0,
pcePcepPeerNumCorruptRcvd	0,
pcePcepPeerNumReqSent	4,
pcePcepPeerNumSvecSent	0,
pcePcepPeerNumSvecReqSent	0,
pcePcepPeerNumReqSentPendRep	0,
pcePcepPeerNumReqSentEroRcvd	3,
pcePcepPeerNumReqSentNoPathRcvd	0,
pcePcepPeerNumReqSentCancelRcvd	0,
pcePcepPeerNumReqSentErrorRcvd	0,
pcePcepPeerNumReqSentTimeout	0,
pcePcepPeerNumReqSentCancelSent	0,
pcePcepPeerNumReqSentClosed	1,
pcePcepPeerNumReqRcvd	0,
pcePcepPeerNumSvecRcvd	0,
pcePcepPeerNumSvecReqRcvd	0,
pcePcepPeerNumReqRcvdPendRep	0,
pcePcepPeerNumReqRcvdEroSent	0,
pcePcepPeerNumReqRcvdNoPathSent	0,
pcePcepPeerNumReqRcvdCancelSent	0,
pcePcepPeerNumReqRcvdErrorSent	0,
pcePcepPeerNumReqRcvdCancelRcvd	0,
pcePcepPeerNumReqRcvdClosed	0,
pcePcepPeerNumRepRcvdUnknown	0,
pcePcepPeerNumReqRcvdUnknown	0

```

    }

    In pcePcepSessTable {
        pcePcepSessInitiator                local(1), --PCE2
        pcePcepSessStateLastChange          TimeStamp,
        pcePcepSessState                    sessionUp(4),
        pcePcepSessConnectRetry              0,
        pcePcepSessLocalID                   1,
        pcePcepSessRemoteID                  1,
        pcePcepSessKeepaliveTimer            1,
        pcePcepSessPeerKeepaliveTimer        1,
        pcePcepSessDeadTimer                 4,
        pcePcepSessPeerDeadTimer             4,
        pcePcepSessKAHoldTimeRem             1,
        pcePcepSessOverloaded                false(0),
        pcePcepSessOverloadTime              0,
        pcePcepSessPeerOverloaded            false(0),
        pcePcepSessPeerOverloadTime          0,
        pcePcepSessDiscontinuityTime         TimeStamp,
        pcePcepSessAvgRspTime                200,
        pcePcepSessLWMRspTime                100,
        pcePcepSessHWMRspTime                300,
        pcePcepSessNumPCReqSent              4,
        pcePcepSessNumPCReqRcvd              0,
        pcePcepSessNumPCRepSent              0,
        pcePcepSessNumPCRepRcvd              4,
        pcePcepSessNumPCErrSent              0,
        pcePcepSessNumPCErrRcvd              0,
        pcePcepSessNumPCNtfSent              0,
        pcePcepSessNumPCNtfRcvd              0,
        pcePcepSessNumKeepaliveSent          123,
        pcePcepSessNumKeepaliveRcvd          123,
        pcePcepSessNumUnknownRcvd            0,
        pcePcepSessNumCorruptRcvd            0,
        pcePcepSessNumReqSent                4,
        pcePcepSessNumSvecSent               0,
        pcePcepSessNumSvecReqSent            0,
        pcePcepSessNumReqSentPendRep         0,
        pcePcepSessNumReqSentEroRcvd         3,
        pcePcepSessNumReqSentNoPathRcvd      1,
        pcePcepSessNumReqSentCancelRcvd      0,
        pcePcepSessNumReqSentErrorRcvd       0,
        pcePcepSessNumReqSentTimeout         0,
        pcePcepSessNumReqSentCancelSent      0,
        pcePcepSessNumReqRcvd                0,
        pcePcepSessNumSvecRcvd               0,
        pcePcepSessNumSvecReqRcvd            0,
        pcePcepSessNumReqRcvdPendRep         0,

```

```
    pcePcepSessNumReqRcvdEroSent      0,  
    pcePcepSessNumReqRcvdNoPathSent  0,  
    pcePcepSessNumReqRcvdCancelSent   0,  
    pcePcepSessNumReqRcvdErrorSent    0,  
    pcePcepSessNumReqRcvdCancelRcvd   0,  
    pcePcepSessNumRepRcvdUnknown      0,  
    pcePcepSessNumReqRcvdUnknown      0  
}
```

-- no session to PCE3

Authors' Addresses

Agrahara Kiran Koushik
Brocade Communications Inc.

EMail: kkoushik@brocade.com

Emile Stephan
Orange
2 avenue Pierre Marzin
Lannion F-22307
France

EMail: emile.stephan@orange.com

Quintin Zhao
Huawei Technology
125 Nagog Technology Park
Acton, MA 01719
US

EMail: qzhao@huawei.com

Daniel King
Old Dog Consulting

EMail: daniel@olddog.co.uk

Jonathan Hardwick
Metaswitch
100 Church Street
Enfield EN2 6BQ
UK

EMail: jonathan.hardwick@metaswitch.com

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

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

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

Abstract

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

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

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

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

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

This Internet-Draft will expire on December 21, 2017.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

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

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

1. Introduction

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

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

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

1.1. Requirements Language

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

2. Terminology

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

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

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

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

The following terms are defined in this document:

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

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

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

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

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

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

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

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

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

3. Motivation and Objectives for Stateful PCE

3.1. Motivation

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

3.1.1. Background

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

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

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

3.1.2. Why a Stateful PCE?

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

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

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

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

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

3.1.3. Protocol vs. Configuration

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

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

3.2. Objectives

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

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

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

4. New Functions to Support Stateful PCEs

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

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

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

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

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

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

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

5. Overview of Protocol Extensions

5.1. LSP State Ownership

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

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

Retaining LSP state ownership on the PCC allows for:

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

5.2. New Messages

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

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

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

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

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

Table 1: New Function to Message Mapping

5.3. Error Reporting

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

5.4. Capability Advertisement

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

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

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

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

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

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

5.5. IGP Extensions for Stateful PCE Capabilities Advertisement

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

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

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

Type: 5

Length: Multiple of 4.

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

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

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

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

5.6. State Synchronization

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

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

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

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

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

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

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

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

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

The successful State Synchronization sequence is shown in Figure 1.

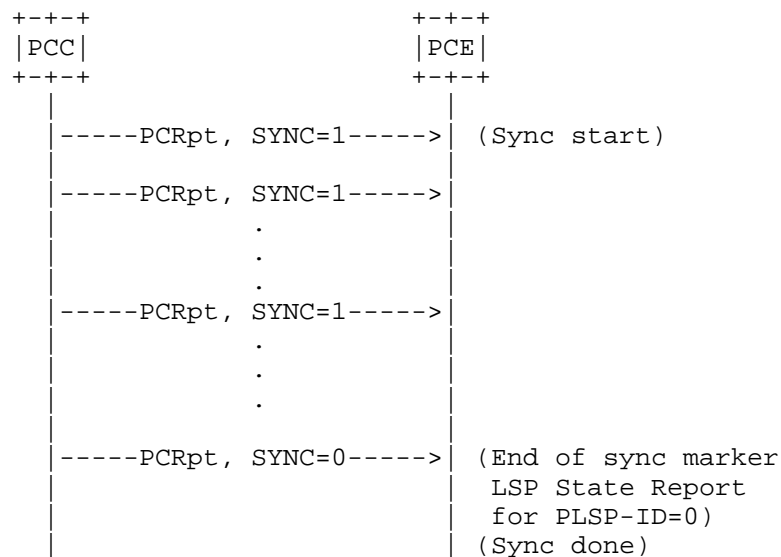


Figure 1: Successful state synchronization

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

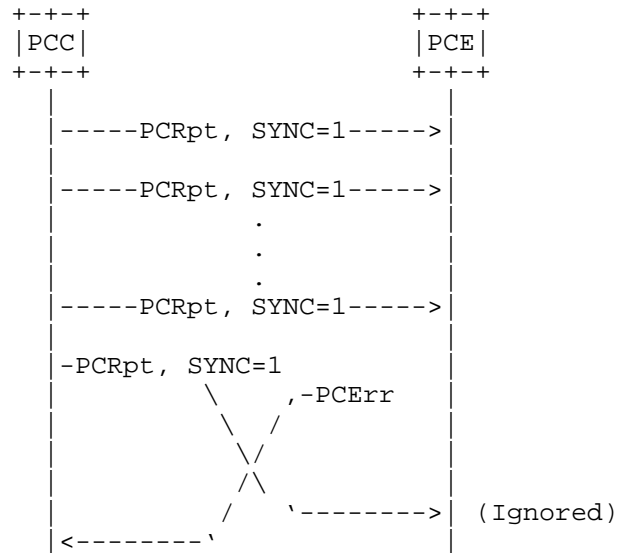


Figure 2: Failed state synchronization (PCE failure)

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

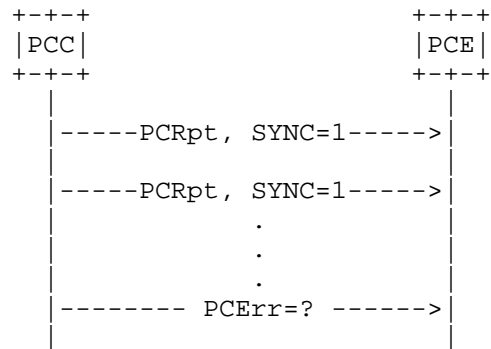


Figure 3: Failed state synchronization (PCC failure)

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

5.7. LSP Delegation

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

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

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

5.7.1. Delegating an LSP

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

The delegation sequence is shown in Figure 4.

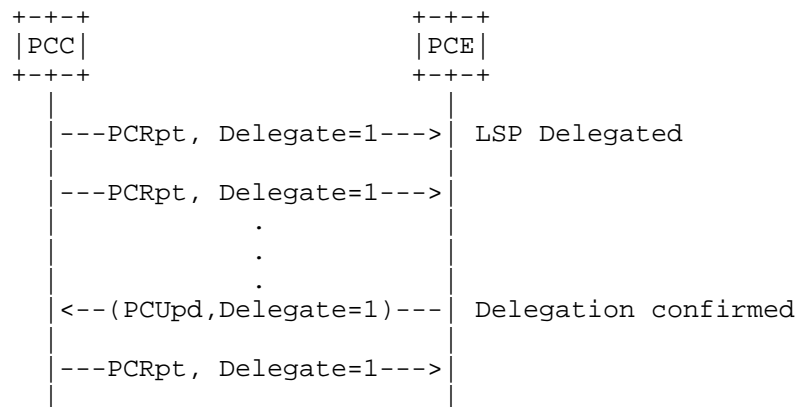


Figure 4: Delegating an LSP

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

5.7.2. Revoking a Delegation

5.7.2.1. Explicit Revocation

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

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

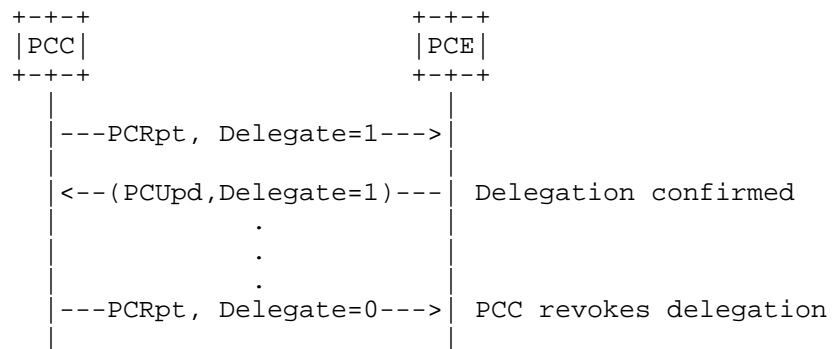


Figure 5: Revoking a Delegation

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

5.7.2.2. Revocation on Redelegating Timeout

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

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

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

5.7.3. Returning a Delegation

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

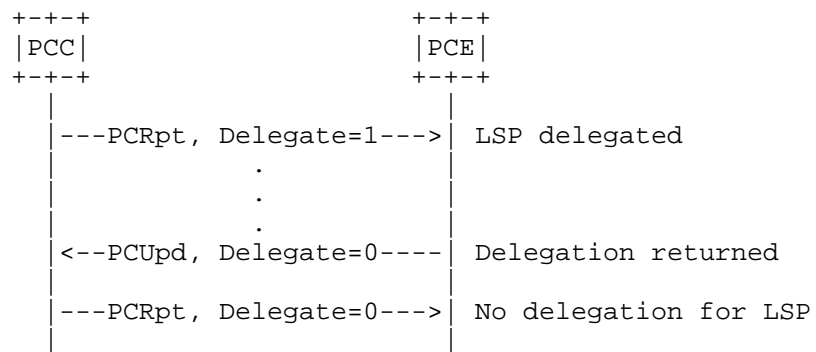


Figure 6: Returning a Delegation

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

5.7.4. Redundant Stateful PCEs

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

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

5.7.5. Redelegation on PCE Failure

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

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

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

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

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

5.8. LSP Operations

5.8.1. Passive Stateful PCE Path Computation Request/Response

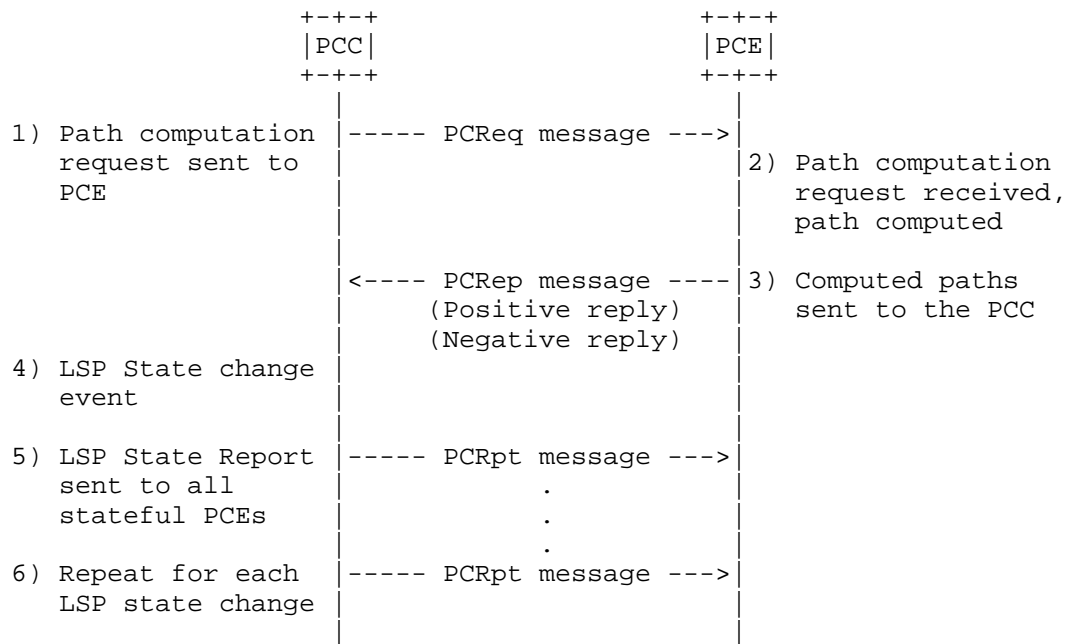


Figure 7: Passive Stateful PCE Path Computation Request/Response

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

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

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

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

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

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

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

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

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

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

5.8.2. Switching from Passive Stateful to Active Stateful

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

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

5.8.3. Active Stateful PCE LSP Update

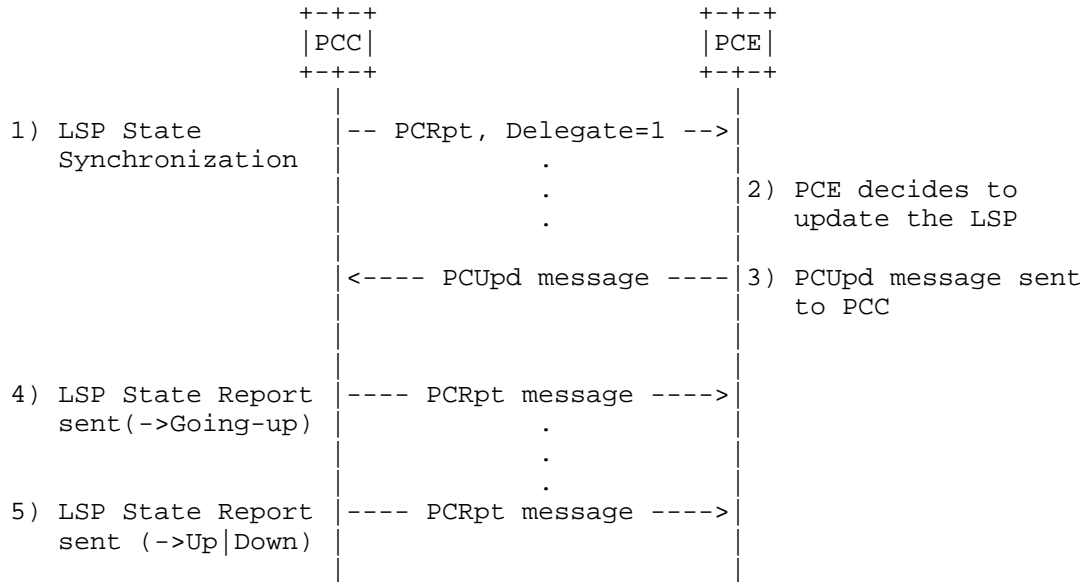


Figure 8: Active Stateful PCE

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

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

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

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

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

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

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

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

5.9. LSP Protection

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

5.10. PCEP Sessions

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

6. PCEP Messages

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

6.1. The PCRpt Message

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

The format of the PCRpt message is as follows:

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

Where:

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

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

Where:

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

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

Where:

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

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

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

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

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

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

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

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

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

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

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

6.2. The PCUpd Message

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

The format of a PCUpd message is as follows:

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

Where:

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

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

Where:

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

Where:

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

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

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

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

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

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

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

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

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

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

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

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

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

6.3. The PCErr Message

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

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

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

```

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

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

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

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

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

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

```

6.4. The PCReq Message

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

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

```

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

```

Where:

```

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

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

```

6.5. The PCRep Message

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

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

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

Where:

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

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

7. Object Formats

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

7.1. OPEN Object

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

7.1.1. Stateful PCE Capability TLV

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



Figure 9: STATEFUL-PCE-CAPABILITY TLV format

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

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

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

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

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

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

7.2. SRP Object

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

SRP Object-Class is 33.

SRP Object-Type is 1.

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

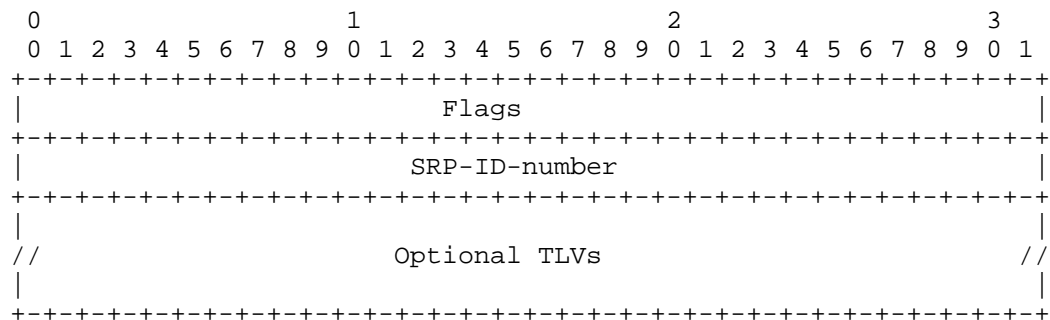


Figure 10: The SRP Object format

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

Flags (32 bits): None defined yet.

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

The values 0x00000000 and 0xFFFFFFFF are reserved.

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

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

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

7.3. LSP Object

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

LSP Object-Class is 32.

LSP Object-Type is 1.

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

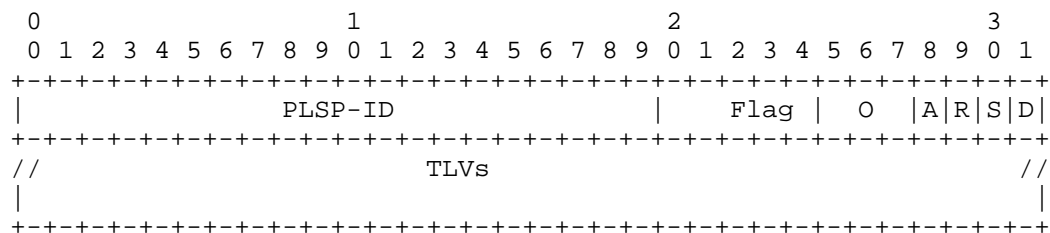


Figure 11: The LSP Object format

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

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

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

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

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

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

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

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

The following values are defined:

0 - DOWN: not active.

1 - UP: signalled.

2 - ACTIVE: up and carrying traffic.

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

4 - GOING-UP: LSP is being signalled.

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

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

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

7.3.1. LSP-IDENTIFIERS TLVs

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

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

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

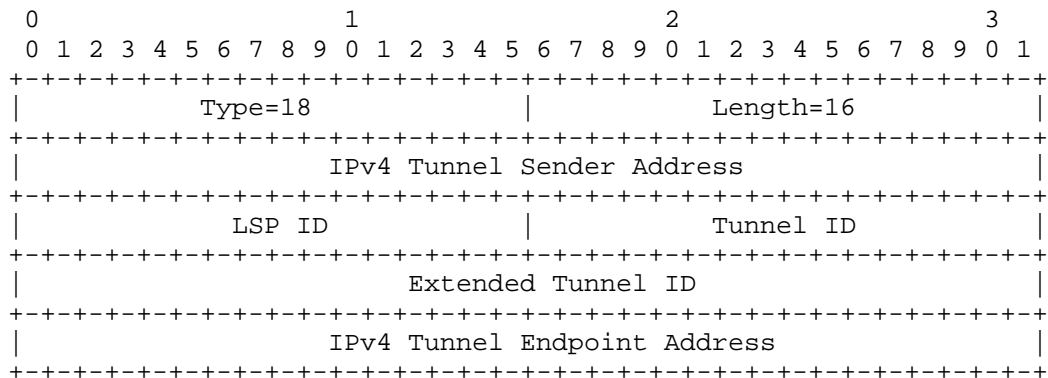


Figure 12: IPV4-LSP-IDENTIFIERS TLV format

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

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

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

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

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

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

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

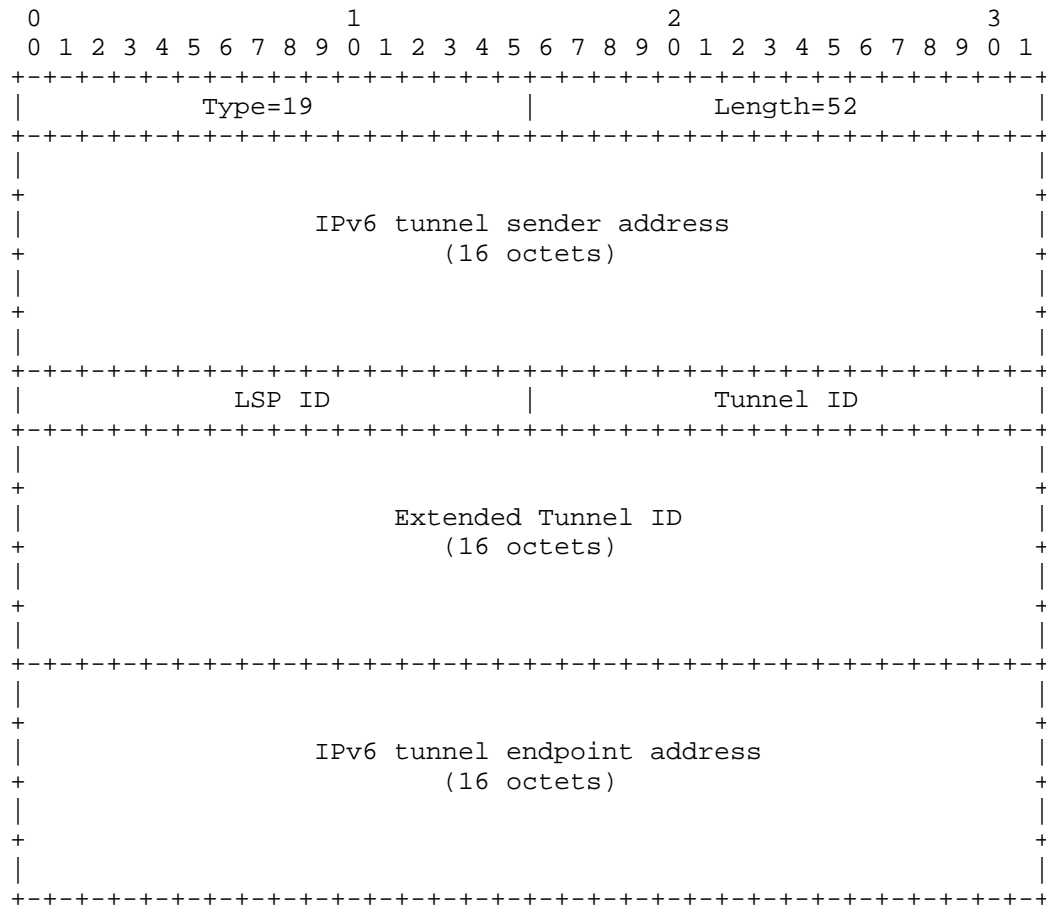


Figure 13: IPV6-LSP-IDENTIFIERS TLV format

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

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

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

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

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

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

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

7.3.2. Symbolic Path Name TLV

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

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

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

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

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

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

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

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

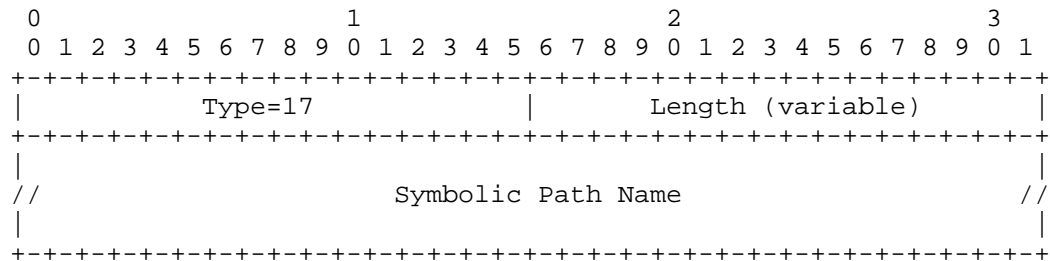


Figure 14: SYMBOLIC-PATH-NAME TLV format

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

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

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

7.3.3. LSP Error Code TLV

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

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

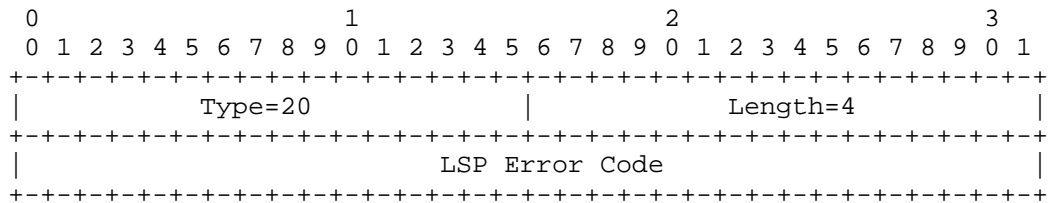


Figure 15: LSP-ERROR-CODE TLV format

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

The following LSP Error Codes are currently defined:

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

7.3.4. RSVP Error Spec TLV

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

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

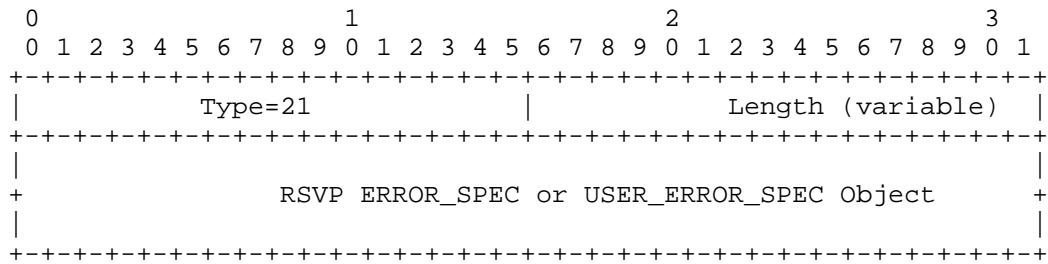


Figure 16: RSVP-ERROR-SPEC TLV format

Type (16 bits): The type is 21.

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

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

8. IANA Considerations

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

8.1. PCE Capabilities in IGP Advertisements

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

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

8.2. PCEP Messages

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

Value	Meaning	Reference
10	Report	This document
11	Update	This document

8.3. PCEP Objects

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

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

8.4. LSP Object

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

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

The following values are defined in this document:

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

8.5. PCEP-Error Object

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

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

8.6. Notification Object

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

Notification-Type	Meaning
4	Stateful PCE resource limit exceeded

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

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

8.7. PCEP TLV Type Indicators

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

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

8.8. STATEFUL-PCE-CAPABILITY TLV

This document requests that a new sub-registry, named "STATEFUL-PCE-CAPABILITY TLV Flag Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the Flag field in the STATEFUL-PCE-CAPABILITY TLV of the PCEP OPEN object (class = 1). New values are to be assigned by Standards Action [RFC5226]. Each bit should be tracked with the following qualities:

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

The following values are defined in this document:

Bit	Description	Reference
31	LSP-UPDATE-CAPABILITY	This document

8.9. LSP-ERROR-CODE TLV

This document requests that a new sub-registry, named "LSP-ERROR-CODE TLV Error Code Field", is created within the "Path Computation Element Protocol (PCEP) Numbers" registry to manage the LSP Error code field of the LSP-ERROR-CODE TLV. This field specifies the reason for failure to update the LSP.

New values are to be assigned by Standards Action [RFC5226]. Each value should be tracked with the following qualities: value, description and defining RFC. The following values are defined in this document:

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

9. Manageability Considerations

All manageability requirements and considerations listed in [RFC5440] apply to PCEP extensions defined in this document. In addition, requirements and considerations listed in this section apply.

9.1. Control Function and Policy

In addition to configuring specific PCEP session parameters, as specified in [RFC5440], Section 8.1, a PCE or PCC implementation MUST allow configuring the stateful PCEP capability and the LSP Update capability. A PCC implementation SHOULD allow the operator to specify multiple candidate PCEs for and a delegation preference for each candidate PCE. A PCC SHOULD allow the operator to specify an LSP delegation policy where LSPs are delegated to the most-preferred online PCE. A PCC MAY allow the operator to specify different LSP delegation policies.

A PCC implementation which allows concurrent connections to multiple PCEs SHOULD allow the operator to group the PCEs by administrative domains and it MUST NOT advertise LSP existence and state to a PCE if the LSP is delegated to a PCE in a different group.

A PCC implementation SHOULD allow the operator to specify whether the PCC will advertise LSP existence and state for LSPs that are not

controlled by any PCE (for example, LSPs that are statically configured at the PCC).

A PCC implementation SHOULD allow the operator to specify both the Redelegating Timeout Interval and the State Timeout Interval. The default value of the Redelegating Timeout Interval SHOULD be set to 30 seconds. An operator MAY also configure a policy that will dynamically adjust the Redelegating Timeout Interval, for example setting it to zero when the PCC has an established session to a backup PCE. The default value for the State Timeout Interval SHOULD be set to 60 seconds.

After the expiration of the State Timeout Interval, the LSP reverts to operator-defined default parameters. A PCC implementation MUST allow the operator to specify the default LSP parameters. To achieve a behavior where the LSP retains the parameters set by the PCE until such time that the PCC makes a change to them, a State Timeout Interval of infinity SHOULD be used. Any changes to LSP parameters SHOULD be done in make-before-break fashion.

LSP Delegation is controlled by operator-defined policies on a PCC. LSPs are delegated individually - different LSPs may be delegated to different PCEs. An LSP is delegated to at most one PCE at any given point in time. A PCC implementation SHOULD support the delegation policy, when all PCC's LSPs are delegated to a single PCE at any given time. Conversely, the policy revoking the delegation for all PCC's LSPs SHOULD also be supported.

A PCC implementation SHOULD allow the operator to specify delegation priority for PCEs. This effectively defines the primary PCE and one or more backup PCEs to which primary PCE's LSPs can be delegated when the primary PCE fails.

Policies defined for stateful PCEs and PCCs should eventually fit in the Policy-Enabled Path Computation Framework defined in [RFC5394], and the framework should be extended to support Stateful PCEs.

9.2. Information and Data Models

The PCEP YANG module [I-D.ietf-pcep-pcep-yang] should include

- o advertised stateful capabilities and synchronization status per PCEP session
- o the delegation status of each configured LSP.

The PCEP MIB [RFC7420] could also be updated to include this information.

9.3. Liveness Detection and Monitoring

PCEP extensions defined in this document do not require any new mechanisms beyond those already defined in [RFC5440], Section 8.3.

9.4. Verifying Correct Operation

Mechanisms defined in [RFC5440], Section 8.4 also apply to PCEP extensions defined in this document. In addition to monitoring parameters defined in [RFC5440], a stateful PCC-side PCEP implementation SHOULD provide the following parameters:

- o Total number of LSP updates
- o Number of successful LSP updates
- o Number of dropped LSP updates
- o Number of LSP updates where LSP setup failed

A PCC implementation SHOULD provide a command to show for each LSP whether it is delegated, and if so, to which PCE.

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

9.5. Requirements on Other Protocols and Functional Components

PCEP extensions defined in this document do not put new requirements on other protocols.

9.6. Impact on Network Operation

Mechanisms defined in [RFC5440], Section 8.6 also apply to PCEP extensions defined in this document.

Additionally, a PCEP implementation SHOULD allow a limit to be placed on the number of LSPs delegated to the PCE and on the rate of PCUpd and PCRpt messages sent by a PCEP speaker and processed from a peer. It SHOULD also allow sending a notification when a rate threshold is reached.

A PCC implementation SHOULD allow a limit to be placed on the rate of LSP Updates to the same LSP to avoid signaling overload discussed in Section 10.3.

10. Security Considerations

10.1. Vulnerability

This document defines extensions to PCEP to enable stateful PCEs. The nature of these extensions and the delegation of path control to PCEs results in more information being available for a hypothetical adversary and a number of additional attack surfaces which must be protected.

The security provisions described in [RFC5440] remain applicable to these extensions. However, because the protocol modifications outlined in this document allow the PCE to control path computation timing and sequence, the PCE defense mechanisms described in [RFC5440] section 7.2 are also now applicable to PCC security.

As a general precaution, it is RECOMMENDED that these PCEP extensions only be activated on authenticated and encrypted sessions across PCEs and PCCs belonging to the same administrative authority, using Transport Layer Security (TLS) [I-D.ietf-pce-pceps], as per the recommendations and best current practices in [RFC7525].

The following sections identify specific security concerns that may result from the PCEP extensions outlined in this document along with recommended mechanisms to protect PCEP infrastructure against related attacks.

10.2. LSP State Snooping

The stateful nature of this extension explicitly requires LSP status updates to be sent from PCC to PCE. While this gives the PCE the ability to provide more optimal computations to the PCC, it also provides an adversary with the opportunity to eavesdrop on decisions made by network systems external to PCE. This is especially true if the PCC delegates LSPs to multiple PCEs simultaneously.

Adversaries may gain access to this information by eavesdropping on unsecured PCEP sessions, and might then use this information in various ways to target or optimize attacks on network infrastructure. For example by flexibly countering anti-DDoS measures being taken to protect the network, or by determining choke points in the network where the greatest harm might be caused.

PCC implementations which allow concurrent connections to multiple PCEs SHOULD allow the operator to group the PCEs by administrative domains and they MUST NOT advertise LSP existence and state to a PCE if the LSP is delegated to a PCE in a different group.

10.3. Malicious PCE

The LSP delegation mechanism described in this document allows a PCC to grant effective control of an LSP to the PCE for the duration of a PCEP session. While this enables PCE control of the timing and sequence of path computations within and across PCEP sessions, it also introduces a new attack vector: an attacker may flood the PCC with PCUpd messages at a rate which exceeds either the PCC's ability to process them or the network's ability to signal the changes, either by spoofing messages or by compromising the PCE itself.

A PCC is free to revoke an LSP delegation at any time without needing any justification. A defending PCC can do this by enqueueing the appropriate PCRpt message. As soon as that message is enqueued in the session, the PCC is free to drop any incoming PCUpd messages without additional processing.

10.4. Malicious PCC

A stateful session also results in an increased attack surface by placing a requirement for the PCE to keep an LSP state replica for each PCC. It is RECOMMENDED that PCE implementations provide a limit on resources a single PCC can occupy. A PCE implementing such a limit MUST send a PCNtf message with notification-type 4 (Stateful PCE resource limit exceeded) and notification-value 1 (Entering resource limit exceeded state) upon receiving an LSP state report causing it to exceed this threshold.

Delegation of LSPs can create further strain on PCE resources and a PCE implementation MAY preemptively give back delegations if it finds itself lacking the resources needed to effectively manage the delegation. Since the delegation state is ultimately controlled by the PCC, PCE implementations SHOULD provide throttling mechanisms to prevent strain created by flaps of either a PCEP session or an LSP delegation.

11. Contributing Authors

Xian Zhang
Huawei Technology
F3-5-B R&D Center
Huawei Industrial Base, Bantian, Longgang District
Shenzhen, Guangdong 518129
P.R.China
EMail: zhang.xian@huawei.com

Dhruv Dhody
Huawei Technology

Leela Palace
Bangalore, Karnataka 560008
INDIA
EMail: dhruv.dhody@huawei.com

Siva Sivabalan
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, Ontario K2K 3E8
Canada
EMail: msiva@cisco.com

12. Acknowledgements

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

We would like to thank Shane Amante, Julien Meuric, Kohei Shiimoto, Paul Schultz and Raveendra Torvi for their comments and suggestions. Thanks also to Jon Hardwick, Oscar Gonzales de Dios, Tomas Janciga, Stefan Kobza, Kexin Tang, Matej Spanik, Jon Parker, Marek Zavodsky, Ambrose Kwong, Ashwin Sampath, Calvin Ying, Mustapha Aissaoui, Stephane Litkowski and Olivier Dugeon for helpful comments and discussions.

13. References

13.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, DOI 10.17487/RFC2205, September 1997, <<http://www.rfc-editor.org/info/rfc2205>>.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<http://www.rfc-editor.org/info/rfc3209>>.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, DOI 10.17487/RFC5088, January 2008, <<http://www.rfc-editor.org/info/rfc5088>>.

- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, DOI 10.17487/RFC5089, January 2008, <<http://www.rfc-editor.org/info/rfc5089>>.
- [RFC5284] Swallow, G. and A. Farrel, "User-Defined Errors for RSVP", RFC 5284, DOI 10.17487/RFC5284, August 2008, <<http://www.rfc-editor.org/info/rfc5284>>.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <<http://www.rfc-editor.org/info/rfc5440>>.
- [RFC5511] Farrel, A., "Routing Backus-Naur Form (RBNF): A Syntax Used to Form Encoding Rules in Various Routing Protocol Specifications", RFC 5511, DOI 10.17487/RFC5511, April 2009, <<http://www.rfc-editor.org/info/rfc5511>>.
- [RFC8051] Zhang, X., Ed. and I. Minei, Ed., "Applicability of a Stateful Path Computation Element (PCE)", RFC 8051, DOI 10.17487/RFC8051, January 2017, <<http://www.rfc-editor.org/info/rfc8051>>.

13.2. Informative References

- [I-D.ietf-pce-gmpls-pcep-extensions]
Margarita, C., Dios, O., and F. Zhang, "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions-11 (work in progress), October 2015.
- [I-D.ietf-pce-pce-initiated-lsp]
Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model", draft-ietf-pce-pce-initiated-lsp-09 (work in progress), March 2017.
- [I-D.ietf-pce-pcep-yang]
Dhody, D., Hardwick, J., Beeram, V., and j. jeffrant@gmail.com, "A YANG Data Model for Path Computation Element Communications Protocol (PCEP)", draft-ietf-pce-pcep-yang-02 (work in progress), March 2017.
- [I-D.ietf-pce-pceps]
Lopez, D., Dios, O., Wu, Q., and D. Dhody, "Secure Transport for PCEP", draft-ietf-pce-pceps-14 (work in progress), May 2017.

- [I-D.ietf-pce-stateful-sync-optimizations]
Crabbe, E., Minei, I., Medved, J., Varga, R., Zhang, X.,
and D. Dhody, "Optimizations of Label Switched Path State
Synchronization Procedures for a Stateful PCE", draft-
ietf-pce-stateful-sync-optimizations-10 (work in
progress), March 2017.
- [MPLS-PC] Chaieb, I., Le Roux, J.L., and B. Cousin, "Improved MPLS-TE
LSP Path Computation using Preemption", Global
Information Infrastructure Symposium, July 2007.
- [MXMN-TE] Danna, E., Mandal, S., and A. Singh, "Practical linear
programming algorithm for balancing the max-min fairness
and throughput objectives in traffic engineering",
INFOCOM, 2012 Proceedings IEEE Page(s): 846-854, 2012.
- [RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J.
McManus, "Requirements for Traffic Engineering Over MPLS",
RFC 2702, DOI 10.17487/RFC2702, September 1999,
<<http://www.rfc-editor.org/info/rfc2702>>.
- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
Label Switching Architecture", RFC 3031,
DOI 10.17487/RFC3031, January 2001,
<<http://www.rfc-editor.org/info/rfc3031>>.
- [RFC3346] Boyle, J., Gill, V., Hannan, A., Cooper, D., Awduche, D.,
Christian, B., and W. Lai, "Applicability Statement for
Traffic Engineering with MPLS", RFC 3346,
DOI 10.17487/RFC3346, August 2002,
<<http://www.rfc-editor.org/info/rfc3346>>.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering
(TE) Extensions to OSPF Version 2", RFC 3630,
DOI 10.17487/RFC3630, September 2003,
<<http://www.rfc-editor.org/info/rfc3630>>.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
Element (PCE)-Based Architecture", RFC 4655,
DOI 10.17487/RFC4655, August 2006,
<<http://www.rfc-editor.org/info/rfc4655>>.
- [RFC4657] Ash, J., Ed. and J. Le Roux, Ed., "Path Computation
Element (PCE) Communication Protocol Generic
Requirements", RFC 4657, DOI 10.17487/RFC4657, September
2006, <<http://www.rfc-editor.org/info/rfc4657>>.

- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, DOI 10.17487/RFC5226, May 2008, <<http://www.rfc-editor.org/info/rfc5226>>.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <<http://www.rfc-editor.org/info/rfc5305>>.
- [RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394, DOI 10.17487/RFC5394, December 2008, <<http://www.rfc-editor.org/info/rfc5394>>.
- [RFC7420] Koushik, A., Stephan, E., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Management Information Base (MIB) Module", RFC 7420, DOI 10.17487/RFC7420, December 2014, <<http://www.rfc-editor.org/info/rfc7420>>.
- [RFC7525] Sheffer, Y., Holz, R., and P. Saint-Andre, "Recommendations for Secure Use of Transport Layer Security (TLS) and Datagram Transport Layer Security (DTLS)", BCP 195, RFC 7525, DOI 10.17487/RFC7525, May 2015, <<http://www.rfc-editor.org/info/rfc7525>>.

Authors' Addresses

Edward Crabbe
Oracle
1501 4th Ave, suite 1800
Seattle, WA 98101
US

Email: edward.crabbe@oracle.com

Ina Minei
Google, Inc.
1600 Amphitheatre Parkway
Mountain View, CA 94043
US

Email: inaminei@google.com

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

Email: jmedved@cisco.com

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

Email: robert.varga@pantheon.tech

Network Working Group
Internet Draft
Intended status: Informational
Expires: April 2015

Y. Lee
Huawei
G. Bernstein
Grotto Networking
Jonas Martensson
Acreo
T. Takeda
NTT
T. Tsuritani
KDDI
O. G. de Dios
Telefonica

October 28, 2014

PCEP Requirements for WSON Routing and Wavelength Assignment

draft-ietf-pce-wson-routing-wavelength-15.txt

Abstract

This memo provides application-specific requirements for the Path Computation Element communication Protocol (PCEP) for the support of Wavelength Switched Optical Networks (WSON). Lightpath provisioning in WSONs requires a routing and wavelength assignment (RWA) process. From a path computation perspective, wavelength assignment is the process of determining which wavelength can be used on each hop of a path and forms an additional routing constraint to optical light path computation. Requirements for PCEP extensions in support of optical impairments will be addressed in a separate document.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 28, 2009.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Conventions used in this document

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

Table of Contents

1. Introduction.....	3
2. WSON RWA Processes & Architecture.....	4
3. Requirements.....	6
3.1. Path Computation Type Option.....	6
3.2. RWA Processing.....	6
3.3. Bulk RWA Path Request/Reply.....	7
3.4. RWA Path Re-optimization Request/Reply.....	7
3.5. Wavelength Range Constraint.....	8
3.6. Wavelength Assignment Preference.....	8
3.7. Signal Processing Capability Restriction.....	8
4. Manageability Considerations.....	9
4.1. Control of Function and Policy.....	9
4.2. Information and Data Models, e.g. MIB module.....	9
4.3. Liveness Detection and Monitoring.....	10
4.4. Verifying Correct Operation.....	10

4.5. Requirements on Other Protocols and Functional Components	10
4.6. Impact on Network Operation	10
5. Security Considerations	10
6. IANA Considerations	11
7. Acknowledgments	11
8. References	11
8.1. Normative References	11
8.2. Informative References	11
Authors' Addresses	12
Intellectual Property Statement	13
Disclaimer of Validity	13

1. Introduction

[RFC4655] defines the PCE-based architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS)-controlled networks at the request of Path Computation Clients (PCCs). A PCC is shown to be any network component that makes such a request and may be for instance an optical switching element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communication Protocol (PCEP) is the communication protocol used between PCC and PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

This document provides a set of application-specific PCEP requirements for support of path computation in Wavelength Switched Optical Networks (WSON). WSON refers to WDM-based optical networks in which switching is performed selectively based on the wavelength of an optical signal.

The path in WSON is referred to as a lightpath. A lightpath may span multiple fiber links and the path should be assigned a wavelength for each link.

A transparent optical network is made up of optical devices that can switch but not convert from one wavelength to another. In a transparent optical network, a lightpath operates on the same wavelength across all fiber links that it traverses. In such case, the lightpath is said to satisfy the wavelength-continuity

constraint. Two lightpaths that share a common fiber link cannot be assigned the same wavelength. To do otherwise would result in both signals interfering with each other. Note that advanced additional multiplexing techniques such as polarization based multiplexing are not addressed in this document since the physical layer aspects are not currently standardized. Therefore, assigning the proper wavelength on a lightpath is an essential requirement in the optical path computation process.

When a switching node has the ability to perform wavelength conversion the wavelength-continuity constraint can be relaxed, and a lightpath may use different wavelengths on different links along its path from origin to destination. It is, however, to be noted that wavelength converters may be limited for cost reasons, while the number of WDM channels that can be supported in a fiber is also limited. As a WSON can be composed of network nodes that cannot perform wavelength conversion, nodes with limited wavelength conversion, and nodes with full wavelength conversion abilities, wavelength assignment is an additional routing constraint to be considered in all lightpath computations.

In this document we first review the processes for routing and wavelength assignment (RWA) used when wavelength continuity constraints are present and then specify requirements for PCEP to support RWA. Requirements for optical impairments will be addressed in a separate document.

The remainder of this document uses terminology from [RFC4655].

2. WSON RWA Processes & Architecture

In [RFC6163] three alternative process architectures were given for performing routing and wavelength assignment. These are shown schematically in Figure 1. R stands for Routing, WA for Wavelength Assignment, and DWA for Distributed Wavelength Assignment.

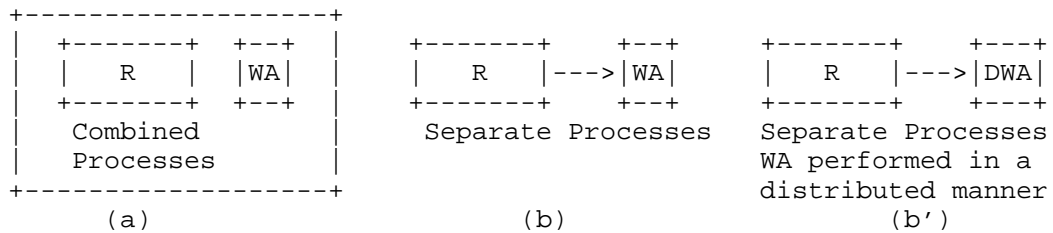


Figure 1. RWA process alternatives

These alternatives have the following properties and impact on PCEP requirements in this document.

(a) Combined Processes (R&WA)

Here path selection and wavelength assignment are performed as a single process. The requirements for PCC-PCE interaction with such a combined RWA process PCE is addressed in this document.

(b) Routing separate from Wavelength Assignment (R+WA)

Here the routing process furnishes one or more potential paths to the wavelength assignment process that then performs final path selection and wavelength assignment. The requirements for PCE-PCE interaction with one PCE implementing the routing process and another implementing the wavelength assignment process are not addressed in this document.

(b') Routing and distributed Wavelength Assignment (R+DWA)

Here a standard path computation (unaware of detailed wavelength availability) takes place, then wavelength assignment is performed along this path in a distributed manner via signaling (RSVP-TE). This alternative is a particular case of R+WA and it should be covered by GMPLS PCEP extensions and does not present new WSON-specific requirements.

In the previous section various process architectures for implementing RWA have been reviewed. Figure 2 shows one typical PCE-based implementation, which is referred to as Combined Process (R&WA). With this architecture, the two processes of routing and wavelength assignment are accessed via a single PCE. This architecture is the base architecture from which the requirements are specified in this document.

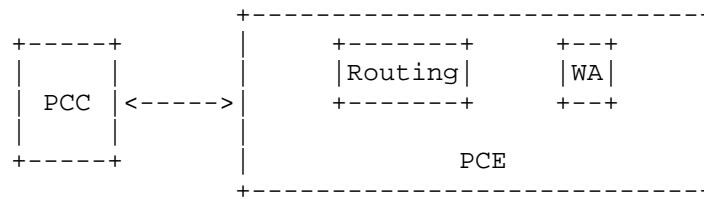


Figure 2. Combined Process (R&WA) architecture

3. Requirements

The requirements for the PCC to PCE interface of Figure 2 are specified in this section.

3.1. Path Computation Type Option

A PCEP request MAY include the path computation type. This can be:

- (i) Both Routing and Wavelength Assignment (RWA),
- (ii) Routing only.

This requirement is needed to differentiate between the currently supported routing with distributed wavelength assignment option and combined RWA. In case of distributed wavelength assignment option, wavelength assignment will be performed at each node of the route.

3.2. RWA Processing

- (a) When the request is a RWA path computation type, the request MUST further include the wavelength assignment options. At the minimum, the following option should be supported:

- (i) Explicit Label Control (ELC) [RFC3473]
- (ii) A set of recommended labels for each hop. The PCC can select the label based on local policy.

Note that option (ii) may also be used in R+WA or R+DWA.

- (b) In case of a RWA computation type, the response MUST include the wavelength(s) assigned to the path and an indication of which label assignment option has been applied (ELC or label set).

- (c) In the case where a valid path is not found, the response MUST include why the path is not found (e.g., network disconnected, wavelength not found, or both, etc.). Note that 'wavelength not found' may include several sub-cases such as wavelength continuity not met, unsupported FEC/Modulation type, etc.

3.3. Bulk RWA Path Request/Reply

Sending simultaneous path requests for "routing only" computation is supported by PCEP specification [RFC5440]. To remain consistent the following requirements are added.

- (a) A PCEP request MUST be able to specify an option for bulk RWA path request. Bulk path request is an ability to request a number of simultaneous RWA path requests.
- (b) The PCEP response MUST include the path and the assigned wavelength assigned for each RWA path request specified in the original bulk request.

3.4. RWA Path Re-optimization Request/Reply

1. For a re-optimization request, the request MUST provide both the path and current wavelength to be re-optimized and MAY include the following options:
 - a. Re-optimize the path keeping the same wavelength(s)
 - b. Re-optimize wavelength(s) keeping the same path
 - c. Re-optimize allowing both the wavelength and the path to change
2. The corresponding response to the re-optimized request MUST provide the re-optimized path and wavelengths even when the request asked for the path or the wavelength to remain unchanged.
3. In case that the new path is not found, the response MUST include why the path is not found (e.g., network disconnected, wavelength not found, or both, etc.). Note that 'wavelength not found' may include several sub-cases such as wavelength continuity not met, unsupported FEC/Modulation type, etc.

3.5. Wavelength Range Constraint

For any RWA computation type request, the requester (PCC) MUST be allowed to specify a restriction on the wavelengths to be used. The requester MAY use this option to restrict the assigned wavelength for explicit label or label set. This restriction may for example come from the tuning ability of a laser transmitter, any optical element, or a policy-based restriction.

Note that the requester (e.g., PCC) is not required to furnish any range restrictions.

3.6. Wavelength Assignment Preference

1. A RWA computation type request MAY include the requester preference for, e.g., random assignment, descending order, ascending order, etc. A response SHOULD follow the requestor preference unless it conflicts with operator's policy.
2. A request for two or more paths MUST allow the requester to include an option constraining the paths to have the same wavelength(s) assigned. This is useful in the case of protection with single transponder (e.g., 1+1 link disjoint paths).

In a network with wavelength conversion capabilities (e.g. sparse 3R regenerators), a request SHOULD be able to indicate whether a single, continuous wavelength should be allocated or not. In other words, the requesting PCC SHOULD be able to specify the precedence of wavelength continuity even if wavelength conversion is available.

3.7. Signal Processing Capability Restriction

Signal processing compatibility is an important constraint for optical path computation. The signal type for an end-to-end optical path must match at source and at destination.

The PCC MUST be allowed to specify the signal type at the endpoints (i.e., at source and at destination). The following signal processing capabilities should be supported at a minimum:

- o Modulation Type List
- o FEC Type List

The PCC MUST also be allowed to state whether transit modification is acceptable for the above signal processing capabilities.

4. Manageability Considerations

Manageability of WSON Routing and Wavelength Assignment (RWA) with PCE must address the following considerations:

4.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- o The ability to send a WSON RWA request.

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- o The support for WSON RWA.
- o The maximum number of bulk path requests associated with WSON RWA per request message.

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

4.2. Information and Data Models, e.g. MIB module

As this document only concerns the requirements to support WSON RWA, no additional MIB module is defined in this document. However, the corresponding solution draft will list the information that should be added to the PCE MIB module defined in [PCEP-MIB].

4.3. Liveness Detection and Monitoring

No new mechanism is defined in this document that implies any new liveness detection and monitoring requirements in addition to those already listed in section 8.3 of [RFC5440].

4.4. Verifying Correct Operation

No new mechanism is defined in this document that implies any new verification requirements in addition to those already listed in section 8.4 of [RFC5440]

4.5. Requirements on Other Protocols and Functional Components

If PCE discovery mechanisms ([RFC5089] and [RFC5088]) were to be extended for technology-specific capabilities, advertising WSON RWA path computation capability should be considered.

4.6. Impact on Network Operation

No new mechanism is defined in this document that implies any new network operation requirements in addition to those already listed in section 8.6 of [RFC5440].

5. Security Considerations

This document has no requirement for a change to the security models within PCEP [RFC5440]. However the additional information distributed in order to address the RWA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

Solutions that address the requirements in this document need to verify that existing PCEP security mechanisms adequately protect the additional network capabilities and must include new mechanisms as necessary.

6. IANA Considerations

This informational document does not make any requests for IANA action.

7. Acknowledgments

The authors would like to thank Adrian Farrel, Cycil Margaria and Ramon Casellas for many helpful comments that greatly improved the contents of this draft.

This document was prepared using 2-Word-v2.0.template.dot.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol", RFC 5440, March 2009.

8.2. Informative References

- [RFC3473] L. Berger, "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.

- [RFC6163] Y. Lee, G. Bernstein, W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, April 2011.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [PCEP-MIB] Koushik, K, et al., "PCE communication protocol(PCEP) Management Information Base", draft-ietf-pce-pcep-mib, work in progress.

Authors' Addresses

Young Lee (Ed.)
Huawei Technologies
5340 Legacy Drive, Building 3
Plano, TX 75245, USA
Phone: (469)277-5838
Email: leeyoung@huawei.com

Greg Bernstein (Ed.)
Grotto Networking
Fremont, CA, USA
Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Jonas Martensson
Acreo
Email:Jonas.Martensson@acreo.se

Tomonori Takeda
NTT Corporation
3-9-11, Midori-Cho
Musashino-Shi, Tokyo 180-8585, Japan
Email: takeda.tomonori@lab.ntt.co.jp

Takehiro Tsuritani
KDDI R&D Laboratories, Inc.
2-1-15 Ohara Kamifukuoka Saitama, 356-8502. Japan
Phone: +81-49-278-7357
Email: tsuri@kddilabs.jp

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
C/ Emilio Vargas 6
Madrid, 28043
Spain
Phone: +34 91 3374013
Email: ogondio@tid.es

PCE Working Group
Internet-Draft
Intended status: Informational
Expires: January 10, 2013

V. Kondreddy
D. Dhody
Huawei Technologies India Pvt
Ltd
July 9, 2012

Applicability of Path Computation Element (PCE) for Fast Reroute (FRR)
Boundary Node protection.
draft-kondreddy-pce-frr-boundary-node-app-00

Abstract

Path computation element (PCE) can be used to compute a label switched path that spans across multiple domains. This document explain the mechanism of Fast Re-Route (FRR) where a point of local repair (PLR) needs to find the appropriate merge point (MP) to do bypass path computation using PCE. In case of boundary node protection when PCE confidentiality (path key) is enabled, new mechanisms are suggested in this document.

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 January 10, 2013.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Requirements Language	3
2. Terminology	3
3. Methods to find MP and calculate the optimal backup path . . .	5
3.1. Intra-domain node protection	5
3.2. Boundary node protection	6
3.2.1. Area Boundary Router (ABR) node protection	6
3.2.2. Autonomous System Border Router (ASBR) node protection	7
3.2.3. Boundary node protection with Path-Key Confidentiality	8
3.2.3.1. Area Boundary Router (ABR) node protection	8
3.2.3.2. Autonomous System Border Router (ASBR) node protection	11
4. Manageability Considerations	11
4.1. Control of Function and Policy	11
4.2. Information and Data Models	11
4.3. Liveness Detection and Monitoring	11
4.4. Verify Correct Operations	11
4.5. Requirements On Other Protocols	11
4.6. Impact On Network Operations	11
5. Security Considerations	12
6. IANA Considerations	12
7. Acknowledgments	12
8. References	12
8.1. Normative References	12
8.2. Informative References	12

1. Introduction

The Path Computation Element (PCE) [RFC4655] can be used to perform complex path computation in large single domain, multi-domain and multi-layered networks. The PCE can also be used to compute a variety of restoration and protection paths and services.

As stated in [RFC4090], there are two independent methods (one-to-one backup and facility backup) of doing fast reroute (FRR). PCE can be used to compute backup path for both the methods. Cooperating PCEs may be used to compute inter-domain backup path.

In case of one to one backup method, the destination MUST be the tail-end of the protected LSP. Whereas for facility backup, destination MUST be the address of the merge point (MP) from the corresponding point of local repair (PLR). The problem of finding the MP using the interface addresses or node-ids present in Record Route Object (RRO) of protected path can be easily solved in the case of a single Interior Gateway Protocol (IGP) area because the PLR has the complete Traffic Engineering Database (TED). Thus, the PLR can unambiguously determine -

- o Is a backup tunnel intersecting a protected TE LSP on a downstream node exists?
- o The MP address regardless of RRO IPv4 or IPv6 sub-objects (interface address or LSR ID).

It is complex for a PLR to find the MP in case of boundary node protection for computing a bypass path because the PLR doesn't have the full TED visibility. When confidentiality (via path key) [RFC5520] is enabled, finding MP is very complex.

This document describes the mechanism to find MP and to setup bypass tunnel to protect a boundary node.

1.1. Requirements Language

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

2. Terminology

The following terminology is used in this document.

ABR: Area Border Router. Router used to connect two IGP areas (Areas in OSPF or levels in IS-IS).

ASBR: Autonomous System Border Router. Router used to connect together ASes of the same or different service providers via one or more inter-AS links.

BN: Boundary Node (BN) a boundary node is either an ABR in the context of inter-area Traffic Engineering or an ASBR in the context of inter-AS Traffic Engineering.

CPS: Confidential Path Segment. A segment of a path that contains nodes and links that the AS policy requires not to be disclosed outside the AS.

CSPF: Constrained Shortest Path First Algorithm.

ERO: Explicit Route Object

FRR: Fast Re-Route

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

IS-IS: Intermediate System to Intermediate System.

LSP: Label Switched Path

LSR: Label Switching Router

MP: Merge Point. The LSR where one or more backup tunnels rejoin the path of the protected LSP downstream of the potential failure.

OSPF: Open Shortest Path First.

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

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

PKS: Path Key Subobject. A subobject of an Explicit Route Object or Record Route Object that encodes a CPS so as to preserve confidentiality.

PLR: Point of Local Repair. The head-end LSR of a backup tunnel or a detour LSP.

RRO: Record Route Object

RSVP: Resource Reservation Protocol

TE: Traffic Engineering.

TED: Traffic Engineering Database.

3. Methods to find MP and calculate the optimal backup path

The Merge Point (MP) address is required at the PLR in order to select a bypass tunnel intersecting a protected Traffic Engineering Label Switched Path (TE LSP) on a downstream LSR.

Some implementations may choose to pre-configure a bypass tunnel on PLR with destination address as MP. MP's Domain to be traversed by bypass path can be administratively configured or learned via some other means (ex Hierarchical PCE (HPCE) [PCE-HIERARCHY-FWK]). Path Computation Client (PCC) on PLR can request its local PCE to compute bypass path from PLR to MP, excluding links and node between PLR and MP. At PLR once primary tunnel is up, a pre-configured bypass tunnel is bound to the primary tunnel, note that multiple bypass tunnel can also exist.

Most implementations may choose to create a bypass tunnel on PLR after primary tunnel is signaled with Record Route Object (RRO) being present in primary path's Resource Reservation Protocol (RSVP) Path Reserve message. MP address has to be determined (described below) to create a bypass tunnel. PCC on PLR can request its local PCE to compute bypass path from PLR to MP, excluding links and node between PLR and MP.

3.1. Intra-domain node protection

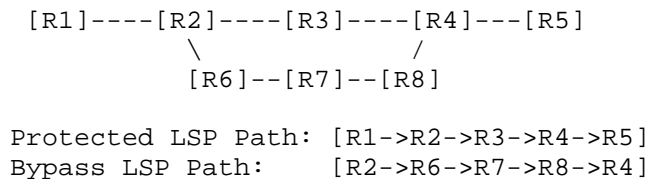


Figure 1: Node Protection for R3

In Figure 1, R2 has to build a bypass tunnel that protects against

the failure of link [R2->R3] and node [R3]. R2 is PLR and R4 is MP in this case. Since, both PLR and MP belong to the same area. The problem of finding the MP using the interface addresses or node-ids can be easily solved. Thus, the PLR can unambiguously determine whether a backup tunnel intersecting a protected TE LSP on a downstream node exists and can also find the MP address regardless of RRO IPv4 or IPv6 sub-objects (interface address or LSR ID).

TED on PLR will have the information of both R2 and R4, which can be used to find MP's TE router IP address and compute optimal backup path from R2 to R4, excluding link [R2->R3] and node [R3].

Thus, RSVP-TE can signal bypass tunnel along the computed path.

3.2. Boundary node protection

3.2.1. Area Boundary Router (ABR) node protection

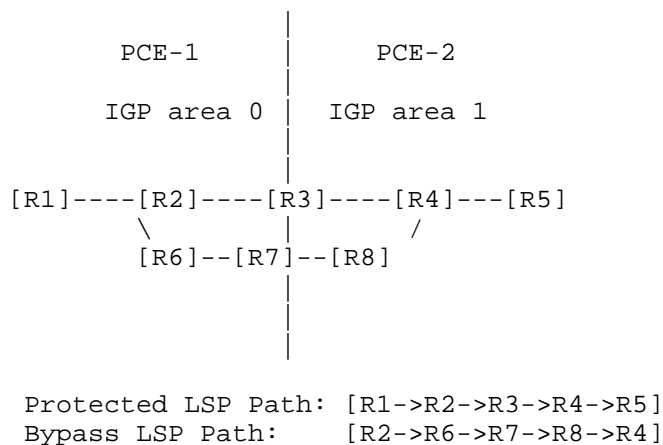


Figure 2: Node Protection for R3 (ABR)

In Figure 2, cooperating PCE(s) (PCE-1 and PCE-2) have computed the primary LSP Path [R1->R2->R3->R4->R5] and provided to R1 (PCC).

R2 has to build a bypass tunnel that protects against the failure of link [R2->R3] and node [R3]. R2 is PLR and R4 is MP. Both PLR and MP are in different area. TED on PLR doesn't have the information of R4.

The problem of finding the MP address in a network with inter-domain TE LSP is solved by inserting a node-id sub-object [RFC4561] in the RRO object carried in the RSVP Path Reserve message. PLR can find

out the MP from the RRO it has received in Path Reserve message from its downstream LSR.

But the computation of optimal backup path from R2 to R4, excluding link [R2->R3] and node [R3] is not possible with running of Constrained Shortest Path First (CSPF) algorithm locally at R2. PCE can be used to compute backup path in this case. R2 acting as PCC on PLR can request PCE-1 to compute bypass path from PLR(R2) to MP(R4), excluding link [R2->R3] and node [R3]. PCE MAY use inter-domain path computation mechanism (like HPCE ([PCE-HIERARCHY-FWK]) etc) when the domain information of MP is unknown at PLR. Further, RSVP-TE can signal bypass tunnel along the computed path.

3.2.2. Autonomous System Border Router (ASBR) node protection

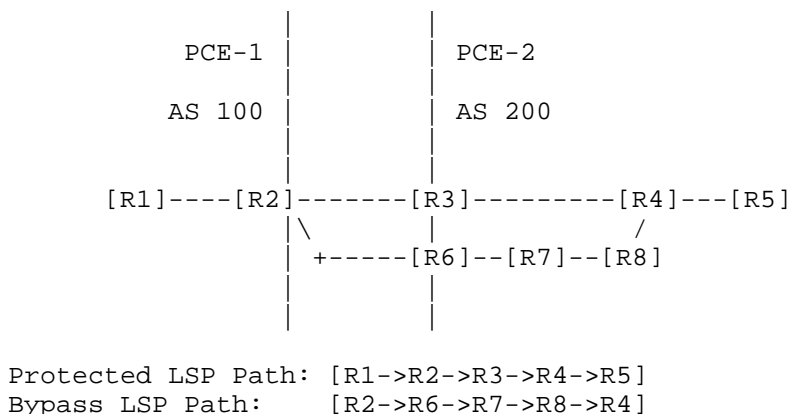


Figure 3: Node Protection for R3 (ASBR)

In Figure 3, Links [R2->R3] and [R2->R6] are inter-AS links. IGP extensions ([RFC5316] and [RFC5392]) describe the flooding of inter-AS TE information for inter-AS path computation. Cooperating PCE(s) (PCE-1 and PCE-2) have computed the primary LSP Path [R1->R2->R3->R4->R5] and provided to R1 (PCC).

R2 is PLR and R4 is MP. Both PLR and MP are in different AS. TED on PLR doesn't have the information of R4.

The address of MP can be found using node-id sub-object [RFC4561] in the RRO object carried in the RSVP Path Reserve message. And Cooperating PCEs could be used to compute the inter-AS bypass path. Thus ASBR boundary node protection is similar to ABR protection.

3.2.3. Boundary node protection with Path-Key Confidentiality

[RFC5520] defines a mechanism to hide the contents of a segment of a path, called the Confidential Path Segment (CPS). The CPS may be replaced by a path-key that can be conveyed in the PCE Communication Protocol (PCEP) and signaled within in a Resource Reservation Protocol TE (RSVP-TE) explicit route object.

[RFC5553] states that, when the signaling message crosses a domain boundary, the path segment that needs to be hidden (that is, a CPS) MAY be replaced in the RRO with a PKS. Note that RRO in Path Reserve message carries the same PKS as originally signaled in the ERO of the Path message.

3.2.3.1. Area Boundary Router (ABR) node protection

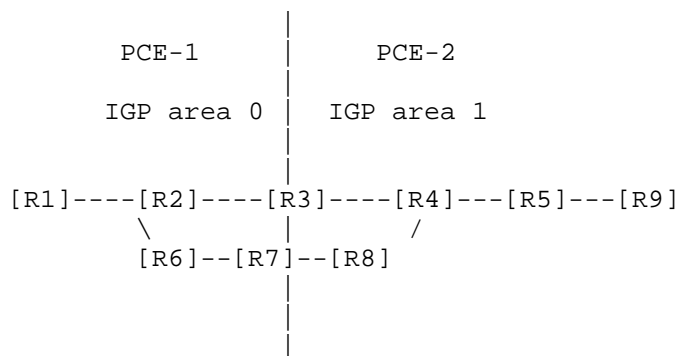


Figure 4: Node Protection for R3 (ABR) and Path-Key

In Figure 4, when path-key is enabled, cooperating PCE(s) (PCE-1 and PCE-2) have computed the primary LSP Path [R1->R2->R3->PKS->R9] and provided to R1 (PCC). Note that there isn't a way to identify the MP when path-key is enabled i.e. using node-id subobject will not work.

3.2.3.1.1. Option 1: New MP Subobject

In Figure 4, on receiving ERO at R3 with PKS, it SHOULD request the PCE identified by PCE-ID in path key subobject to expand the path segment and on receiving RRO it should replace the CPS with the same PKS and append its own RRO subobjects. The boundary node can also append the identity of the MP.

Note that the RRO in Path Reserve Messages can be carried as it is as ERO in Path Message. So if we use the same node-id object as described in [RFC4561] along with PKS it will lead to a looping issue

as explained below -

In Figure 4, at R1 the RRO received will be [R1->R2->R3->R4->PKS->R9] with R4 as the node-id subobject added by R3. if this path is signaled using the ERO, after expansion of PKS at node R3, will lead to presence of R4 twice leading to loop. Note that the issue exists irrespective of the ordering of PKS and Node-id subobject.

Hence there is a need for a new sub-object to identify the MP incase of path-key. This sub-object should be added by the boundary node (R3) during the RRO Path key processing. The PLR can use the MP sub-object to identify the MP. To avoid the looping issue, the immediate upstream LSR (usually PLR) should remove this sub-object from the RRO at the time of RRO processing.

[RFC3209] defines the IPv4 and IPv6 RRO subobjects.

In this document, we define the following new flag:

MP-id: 0x40 (TBA)

When set, this indicates that the address specified in the RRO's IPv4 or IPv6 sub-object is a MP-id address, which refers to the "Router Address" of the MP as defined in [RFC3630], or "Traffic Engineering Router ID" as defined in [RFC5305].

An IPv4 or IPv6 RRO sub-object with the MP-id flag set is also called a MP-id sub-object. The problem of finding an MP address when pathkey is enabled in a network with inter-domain TE LSP is solved by inserting a MP-id sub-object (an RRO "IPv4" and "IPv6" sub-object with the 0x40 flag (TBA) set) in the RRO object carried in the RSVP Path Reserve message.

3.2.3.1.2. Option 2: PCE Path-key Handling

In Figure 4, when path-key is enabled, PCE-2 will replace the segment [R3->R4->R5->R9] with [R3->PKS->R9]. To enable FRR protection with path-key, following change should be done -

- o Scope of confidential path segment is relaxed to immediate downstream node i.e. [R3->R4->PKS->R9] note that R3->R4 MAYBE the address of the interface instead of LSR-ID.
- o Pathkey expansion during signaling would be done at the immediate downstream node of the boundary node. Note that this node must have PCC functionality.

- o This facilitates the insertion of node-id sub-object [RFC4561] in the RRO object from immediate next downstream node of BN in the RSVP Path Reserve message and aids the PLR in previous domain to determine MP

In Figure 4, during RSVP signaling, on receiving ERO at R4 with PKS, it SHOULD request the PCE identified by PCE-ID in path key subobject to expand the path segment and on receiving RRO it should replace the CPS with the same PKS and append its RRO subobjects including Node-id subobject.

On successful signaling of primary tunnel, R2 has to build a bypass tunnel that protects against the failure of link [R2->R3] and node [R3]. Note that by above mechanism PLR (R2) knows the identity of MP (R4) via the RRO Node-id subobject.

Also consider the case of R4 node protection within a single IGP area. R3 is PLR and R5 should become MP.

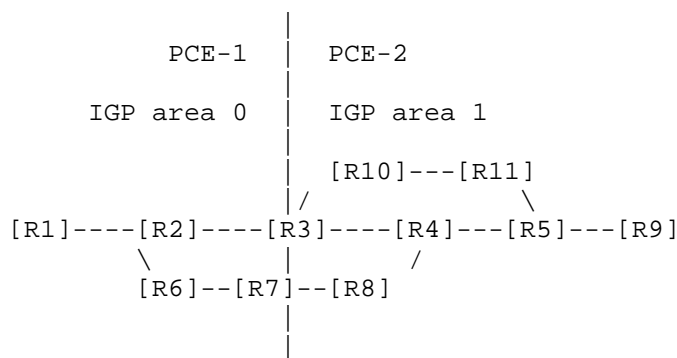


Figure 5: Node Protection for R4

In Figure 5, RRO received in RSVP Path Reserve message at R3 contains [R3->R4->PKS->R9]. Since there is no node-id sub-object in RRO beyond R4, R3 may not be able to find R5 as MP without expansion of PKS. R3 SHOULD request the PCE identified by PCE-ID in PKS to expand the path segment. Note that, the PCE should retain the pathkey for some time as multiple expansion requests will be issued. R3 can now find the identity of MP (R5).

3.2.3.2. Autonomous System Border Router (ASBR) node protection

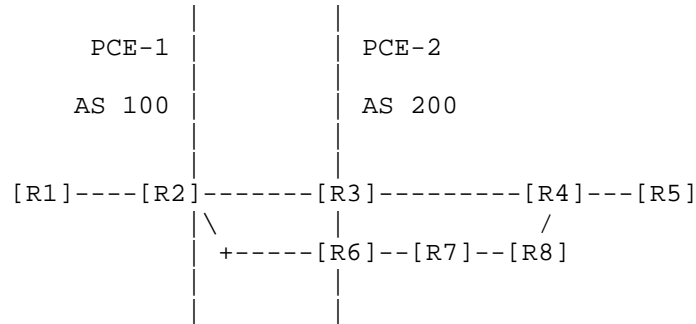


Figure 6: Node Protection for R3 (ASBR)

The address of MP can be found using the same mechanism as explained above. Thus ASBR boundary node protection is similar to ABR protection.

4. Manageability Considerations

4.1. Control of Function and Policy

TBD

4.2. Information and Data Models

TBD

4.3. Liveness Detection and Monitoring

TBD

4.4. Verify Correct Operations

TBD

4.5. Requirements On Other Protocols

TBD

4.6. Impact On Network Operations

TBD

5. Security Considerations

PCE(s) when computes the inter-domain path will generate PKS relaxing the path from BN to next immediate downstream router. (Refer Section 3.2.3.1.2) This relaxation is required to find MP in case of BN node protection. This may be an added security risk. Note that the facility backup method requires that a PLR and its selected MP trust RSVP messages received from each other.

Further analysis must be done.

6. IANA Considerations

TBD

7. Acknowledgments

We would like to thank Sandeep Boina & Reeja Paul for their useful comments and suggestions.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

8.2. Informative References

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC4561] Vasseur, J., Ali, Z., and S. Sivabalan,

- "Definition of a Record Route Object (RRO) Node-Id Sub-Object", RFC 4561, June 2006.
- [RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.
- [RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, December 2008.
- [RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, January 2009.
- [RFC5520] Bradford, R., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, April 2009.
- [RFC5553] Farrel, A., Bradford, R., and JP. Vasseur, "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC 5553, May 2009.
- [PCE-HIERARCHY-FWK] King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS. (draft-ietf-pce-hierarchy-fwk-04)", June 2012.

Authors' Addresses

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

EMail: venugopalreddyk@huawei.com

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

EMail: dhruv.dhody@huawei.com

PCE Working Group
Internet-Draft
Intended status: Informational
Expires: March 9, 2013

V. Kondreddy
D. Dhody
Huawei Technologies India Pvt
Ltd
September 5, 2012

Applicability of Path Computation Element (PCE) for Fast Reroute (FRR)
Boundary Node protection.
draft-kondreddy-pce-frr-boundary-node-app-01

Abstract

Path computation element (PCE) can be used to compute a label switched path that spans across multiple domains. This document explain the mechanism of Fast Re-Route (FRR) where a point of local repair (PLR) needs to find the appropriate merge point (MP) to do bypass path computation using PCE.

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 March 9, 2013.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Requirements Language	3
2. Terminology	3
3. Methods to find MP and calculate the optimal backup path	5
3.1. Intra-domain node protection	5
3.2. Boundary node protection	6
3.2.1. Area Boundary Router (ABR) node protection	6
3.2.2. Autonomous System Border Router (ASBR) node protection	7
3.2.3. Boundary node protection with Path-Key Confidentiality	8
3.2.3.1. Area Boundary Router (ABR) node protection	8
3.2.3.2. Autonomous System Border Router (ASBR) node protection	9
4. Manageability Considerations	9
4.1. Control of Function and Policy	9
4.2. Information and Data Models	9
4.3. Liveness Detection and Monitoring	9
4.4. Verify Correct Operations	9
4.5. Requirements On Other Protocols	9
4.6. Impact On Network Operations	9
5. Security Considerations	10
6. IANA Considerations	10
7. Acknowledgments	10
8. References	10
8.1. Normative References	10
8.2. Informative References	10

1. Introduction

The Path Computation Element (PCE) [RFC4655] can be used to perform complex path computation in large single domain, multi-domain and multi-layered networks. The PCE can also be used to compute a variety of restoration and protection paths and services.

As stated in [RFC4090], there are two independent methods (one-to-one backup and facility backup) of doing fast reroute (FRR). PCE can be used to compute backup path for both the methods. Cooperating PCEs may be used to compute inter-domain backup path.

In case of one to one backup method, the destination MUST be the tail-end of the protected LSP. Whereas for facility backup, destination MUST be the address of the merge point (MP) from the corresponding point of local repair (PLR). The problem of finding the MP using the interface addresses or node-ids present in Record Route Object (RRO) of protected path can be easily solved in the case of a single Interior Gateway Protocol (IGP) area because the PLR has the complete Traffic Engineering Database (TED). Thus, the PLR can unambiguously determine -

- o The MP address regardless of RRO IPv4 or IPv6 sub-objects (interface address or LSR ID).
- o Is a backup tunnel intersecting a protected TE LSP on MP node exists? This is the case where facility backup tunnel already exist either due to another protected TE LSP or it is pre-configured.

It is complex for a PLR to find the MP in case of boundary node protection for computing a bypass path because the PLR doesn't have the full TED visibility. When confidentiality (via path key) [RFC5520] is enabled, finding MP is very complex.

This document describes the mechanism to find MP and to setup bypass tunnel to protect a boundary node.

1.1. Requirements Language

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

2. Terminology

The following terminology is used in this document.

ABR: Area Border Router. Router used to connect two IGP areas (Areas in OSPF or levels in IS-IS).

ASBR: Autonomous System Border Router. Router used to connect together ASes of the same or different service providers via one or more inter-AS links.

BN: Boundary Node (BN) a boundary node is either an ABR in the context of inter-area Traffic Engineering or an ASBR in the context of inter-AS Traffic Engineering.

CPS: Confidential Path Segment. A segment of a path that contains nodes and links that the AS policy requires not to be disclosed outside the AS.

CSPF: Constrained Shortest Path First Algorithm.

ERO: Explicit Route Object

FRR: Fast Re-Route

IGP: Interior Gateway Protocol. Either of the two routing protocols, Open Shortest Path First (OSPF) or Intermediate System to Intermediate System (IS-IS).

IS-IS: Intermediate System to Intermediate System.

LSP: Label Switched Path

LSR: Label Switching Router

MP: Merge Point. The LSR where one or more backup tunnels rejoin the path of the protected LSP downstream of the potential failure.

OSPF: Open Shortest Path First.

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

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

PKS: Path Key Subobject. A subobject of an Explicit Route Object or Record Route Object that encodes a CPS so as to preserve confidentiality.

PLR: Point of Local Repair. The head-end LSR of a backup tunnel or a detour LSP.

RRO: Record Route Object

RSVP: Resource Reservation Protocol

TE: Traffic Engineering.

TED: Traffic Engineering Database.

3. Methods to find MP and calculate the optimal backup path

The Merge Point (MP) address is required at the PLR in order to select a bypass tunnel intersecting a protected Traffic Engineering Label Switched Path (TE LSP) on a downstream LSR.

Some implementations may choose to pre-configure a bypass tunnel on PLR with destination address as MP. MP's Domain to be traversed by bypass path can be administratively configured or learned via some other means (ex Hierarchical PCE (HPCE) [PCE-HIERARCHY-FWK]). Path Computation Client (PCC) on PLR can request its local PCE to compute bypass path from PLR to MP, excluding links and node between PLR and MP. At PLR once primary tunnel is up, a pre-configured bypass tunnel is bound to the primary tunnel, note that multiple bypass tunnel can also exist.

Most implementations may choose to create a bypass tunnel on PLR after primary tunnel is signaled with Record Route Object (RRO) being present in primary path's Resource Reservation Protocol (RSVP) Path Reserve message. MP address has to be determined (described below) to create a bypass tunnel. PCC on PLR can request its local PCE to compute bypass path from PLR to MP, excluding links and node between PLR and MP.

3.1. Intra-domain node protection

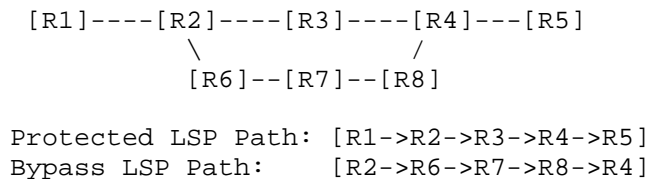


Figure 1: Node Protection for R3

In Figure 1, R2 has to build a bypass tunnel that protects against

the failure of link [R2->R3] and node [R3]. R2 is PLR and R4 is MP in this case. Since, both PLR and MP belong to the same area. The problem of finding the MP using the interface addresses or node-ids can be easily solved. Thus, the PLR can unambiguously find the MP address regardless of RRO IPv4 or IPv6 sub-objects (interface address or LSR ID) and also determine whether a backup tunnel intersecting a protected TE LSP on a downstream node (MP) already exists.

TED on PLR will have the information of both R2 and R4, which can be used to find MP's TE router IP address and compute optimal backup path from R2 to R4, excluding link [R2->R3] and node [R3].

Thus, RSVP-TE can signal bypass tunnel along the computed path.

3.2. Boundary node protection

3.2.1. Area Boundary Router (ABR) node protection

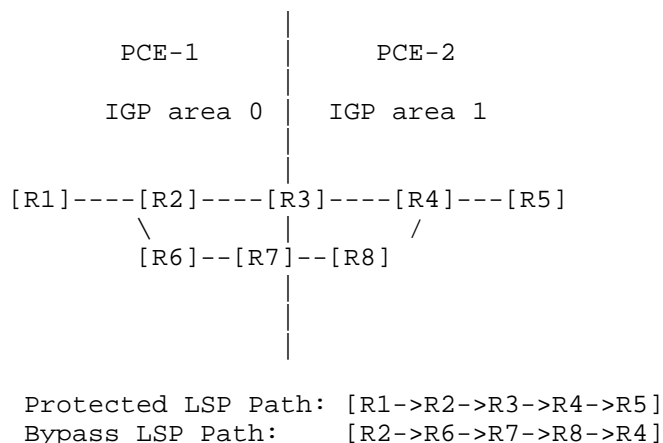


Figure 2: Node Protection for R3 (ABR)

In Figure 2, cooperating PCE(s) (PCE-1 and PCE-2) have computed the primary LSP Path [R1->R2->R3->R4->R5] and provided to R1 (PCC).

R2 has to build a bypass tunnel that protects against the failure of link [R2->R3] and node [R3]. R2 is PLR and R4 is MP. Both PLR and MP are in different area. TED on PLR doesn't have the information of R4.

The problem of finding the MP address in a network with inter-domain TE LSP is solved by inserting a node-id sub-object [RFC4561] in the RRO object carried in the RSVP Path Reserve message. PLR can find

out the MP from the RRO it has received in Path Reserve message from its downstream LSR.

But the computation of optimal backup path from R2 to R4, excluding link [R2->R3] and node [R3] is not possible with running of Constrained Shortest Path First (CSPF) algorithm locally at R2. PCE can be used to compute backup path in this case. R2 acting as PCC on PLR can request PCE-1 to compute bypass path from PLR(R2) to MP(R4), excluding link [R2->R3] and node [R3]. PCE MAY use inter-domain path computation mechanism (like HPCE ([PCE-HIERARCHY-FWK]) etc) when the domain information of MP is unknown at PLR. Further, RSVP-TE can signal bypass tunnel along the computed path.

3.2.2. Autonomous System Border Router (ASBR) node protection

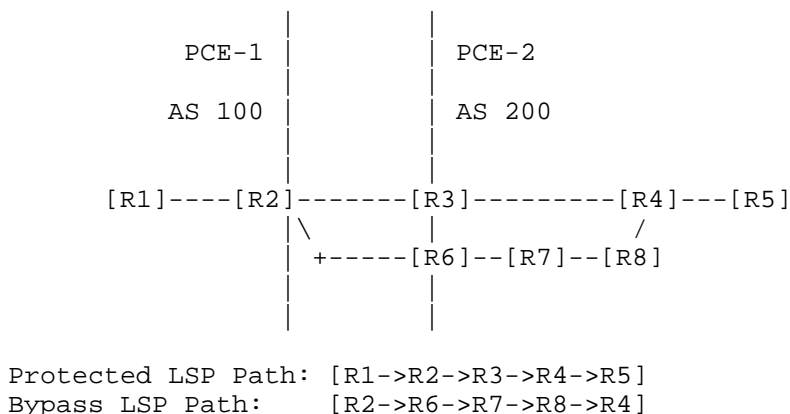


Figure 3: Node Protection for R3 (ASBR)

In Figure 3, Links [R2->R3] and [R2->R6] are inter-AS links. IGP extensions ([RFC5316] and [RFC5392]) describe the flooding of inter-AS TE information for inter-AS path computation. Cooperating PCE(s) (PCE-1 and PCE-2) have computed the primary LSP Path [R1->R2->R3->R4->R5] and provided to R1 (PCC).

R2 is PLR and R4 is MP. Both PLR and MP are in different AS. TED on PLR doesn't have the information of R4.

The address of MP can be found using node-id sub-object [RFC4561] in the RRO object carried in the RSVP Path Reserve message. And Cooperating PCEs could be used to compute the inter-AS bypass path. Thus ASBR boundary node protection is similar to ABR protection.

3.2.3. Boundary node protection with Path-Key Confidentiality

[RFC5520] defines a mechanism to hide the contents of a segment of a path, called the Confidential Path Segment (CPS). The CPS may be replaced by a path-key that can be conveyed in the PCE Communication Protocol (PCEP) and signaled within in a Resource Reservation Protocol TE (RSVP-TE) explicit route object.

[RFC5553] states that, when the signaling message crosses a domain boundary, the path segment that needs to be hidden (that is, a CPS) MAY be replaced in the RRO with a PKS. Note that RRO in Path Reserve message carries the same PKS as originally signaled in the ERO of the Path message.

3.2.3.1. Area Boundary Router (ABR) node protection

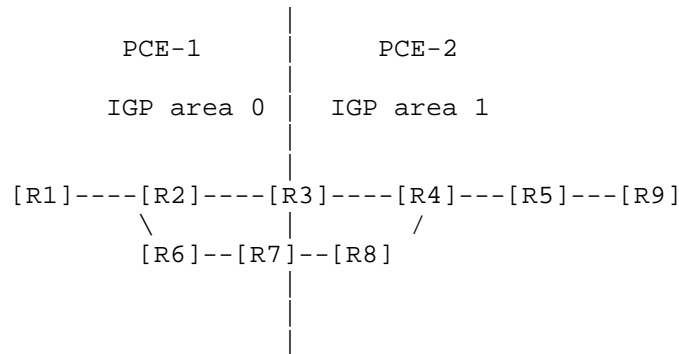


Figure 4: Node Protection for R3 (ABR) and Path-Key

In Figure 4, when path-key is enabled, cooperating PCE(s) (PCE-1 and PCE-2) have computed the primary LSP Path [R1->R2->R3->PKS->R9] and provided to R1 (PCC).

When the ABR node (R3) replaces the CPS with PKS (as originally signaled) during the Path Reserve message handling, it MAY also add the immediate downstream node-id (R4) (so that the PLR (R2) can identify the MP (R4)). Further the PLR (R2) SHOULD remove the MP node-id (R4) before sending the path reserve message upstream to head end router.

Once MP is identified, the backup path computation using PCE is as described earlier. (Section 3.2.1)

3.2.3.2. Autonomous System Border Router (ASBR) node protection

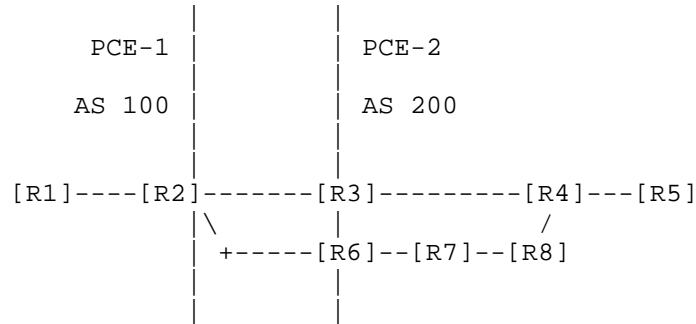


Figure 5: Node Protection for R3 (ASBR)

The address of MP can be found using the same mechanism as explained above. Thus ASBR boundary node protection is similar to ABR protection.

4. Manageability Considerations

4.1. Control of Function and Policy

TBD

4.2. Information and Data Models

TBD

4.3. Liveness Detection and Monitoring

TBD

4.4. Verify Correct Operations

TBD

4.5. Requirements On Other Protocols

TBD

4.6. Impact On Network Operations

TBD

5. Security Considerations

This document does not introduce new security issues. However, MP's node-id is carried as subobject in RRO across domain. This relaxation is required to find MP in case of BN protection. The security considerations pertaining to the [RFC3209], [RFC4090] and [RFC5440] protocols remain relevant.

6. IANA Considerations

TBD

7. Acknowledgments

We would like to thank Sandeep Boina & Reeja Paul for their useful comments and suggestions.

8. References

8.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

8.2. Informative References

- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.
- [RFC4090] Pan, P., Swallow, G., and A. Atlas, "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, May 2005.
- [RFC4561] Vasseur, J., Ali, Z., and S. Sivabalan, "Definition of a Record Route Object (RRO) Node-Id Sub-Object", RFC 4561, June 2006.
- [RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, December 2008.

- [RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, January 2009.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5520] Bradford, R., Vasseur, JP., and A. Farrel, "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, April 2009.
- [RFC5553] Farrel, A., Bradford, R., and JP. Vasseur, "Resource Reservation Protocol (RSVP) Extensions for Path Key Support", RFC 5553, May 2009.
- [PCE-HIERARCHY-FWK] King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS. (draft-ietf-pce-hierarchy-fwk-05)", August 2012.

Authors' Addresses

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

EMail: venugopalreddyk@huawei.com

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

EMail: dhruv.dhody@huawei.com

Network Working Group
Internet Draft
Intended status: Standard Track
Expires: January 2013

Y. Lee
Huawei

G. Bernstein
Grotto Networking

Jonas Martensson
Acreo

T. Takeda
NTT

T. Tsuritani
KDDI

July 6, 2012

PCEP Extensions for WSON Impairments

draft-lee-pce-wson-impairments-04.txt

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on January 6, 2009.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

As an optical signal progresses along its path it may be altered by the various physical processes in the optical fibers and devices it encounters. When such alterations result in signal degradation, these processes are usually referred to as "impairments". These physical characteristics may be important constraints to consider in path computation process in wavelength switched optical networks.

This document provides PCEP extensions to support Impairment Aware Routing and Wavelength Assignment (IA-RWA) in wavelength switched optical networks.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 0.

Table of Contents

1. Introduction.....	3
1.1. WSON RWA Processes (no impairments).....	5
1.2. WSON IA-RWA Processes.....	6
2. WSON PCE Architectures and Requirements.....	7

2.1. RWA PCC to PCE Interface.....	8
2.1.1. A new RWA path request.....	8
2.1.1.1. Signal Quality Measure TLV.....	9
2.1.2. A new RWA path reply.....	11
2.1.2.1. Signal Quality Measure TLV.....	11
2.2. RWA-PCE to IV-PCE Interface.....	13
2.2.1. A new impairment-validated (IV) path request.....	14
2.2.2. A new impairment-validated (IV) path reply.....	14
3. Manageability Considerations.....	14
3.1. Control of Function and Policy.....	14
3.2. Information and Data Models, e.g. MIB module.....	15
3.3. Liveness Detection and Monitoring.....	15
3.4. Verifying Correct Operation.....	15
3.5. Requirements on Other Protocols and Functional Components	15
3.6. Impact on Network Operation.....	16
4. Security Considerations.....	16
5. IANA Considerations.....	16
6. References.....	16
6.1. Normative References.....	16
6.2. Informative References.....	17
Authors' Addresses.....	17
7. Acknowledgments.....	18

1. Introduction

[RFC4655] defines the PCE based architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is shown to be any network component that makes such a request and may be for instance an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communication Protocol (PCEP) is the communication protocol used between PCC and PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

This document provides a set of application-specific PCEP requirements for support of path computation in Wavelength Switched

Optical Networks (WSON) with impairments. WSON refers to WDM based optical networks in which switching is performed selectively based on the wavelength of an optical signal.

The path in WSON is referred to as a lightpath. A lightpath may span multiple fiber links and the path should be assigned a wavelength for each link. A transparent optical network is made up of optical devices that can switch but not convert from one wavelength to another. In a transparent optical network, a lightpath operates on the same wavelength across all fiber links that it traverses. In such case, the lightpath is said to satisfy the wavelength-continuity constraint. Two lightpaths that share a common fiber link can not be assigned the same wavelength. To do otherwise would result in both signals interfering with each other. Note that advanced additional multiplexing techniques such as polarization based multiplexing are not addressed in this document since the physical layer aspects are not currently standardized. Therefore, assigning the proper wavelength on a lightpath is an essential requirement in the optical path computation process.

When a switching node has the ability to perform wavelength conversion the wavelength-continuity constraint can be relaxed, and a lightpath may use different wavelengths on different links along its route from origin to destination. It is, however, to be noted that wavelength converters may be limited due to their relatively high cost, while the number of WDM channels that can be supported in a fiber is also limited. As a WSON can be composed of network nodes that cannot perform wavelength conversion, nodes with limited wavelength conversion, and nodes with full wavelength conversion abilities, wavelength assignment is an additional routing constraint to be considered in all lightpath computation.

One of the most basic questions in communications is whether one can successfully transmit information from a transmitter to a receiver within a prescribed error tolerance, usually specified as a maximum permissible bit error ratio (BER). This generally depends on the nature of the signal transmitted between the sender and receiver and the nature of the communications channel between the sender and receiver. The optical path utilized (along with the wavelength) determines the communications channel.

The optical impairments incurred by the signal along the fiber and at each optical network element along the path determine whether the BER performance or any other measure of signal quality can be met for this particular signal on this particular path. Given the existing standards covering optical characteristics (impairments) and the knowledge of how the impact of impairments may be estimated

along a path, [RFC6566] provides a framework for impairment aware path computation and establishment utilizing GMPLS protocols and the PCE architecture.

Some transparent optical subnetworks are designed such that over any path the degradation to an optical signal due to impairments never exceeds prescribed bounds. This may be due to the limited geographic extent of the network, the network topology, and/or the quality of the fiber and devices employed. In such networks the path selection problem reduces to determining a continuous wavelength from source to destination (the Routing and Wavelength Assignment problem). These networks are discussed in [RFC6163]. In other optical networks, impairments are important and the path selection process must be impairment-aware.

In this document we first review the processes for routing and wavelength assignment (RWA) used when wavelength continuity constraints are present. We then review the processes for optical impairment aware RWA (IA-RWA). Based on selected process models we then specify requirements for PCEP to support IA-RWA. Note that requirements for PCEP to support RWA are specified in a separate document [PCEP-RWA].

The remainder of this document uses terminology from [RFC4655].

1.1. WSON RWA Processes (no impairments)

In [RFC6163] three alternative process architectures were given for performing routing and wavelength assignment. These are shown schematically in Figure 1.

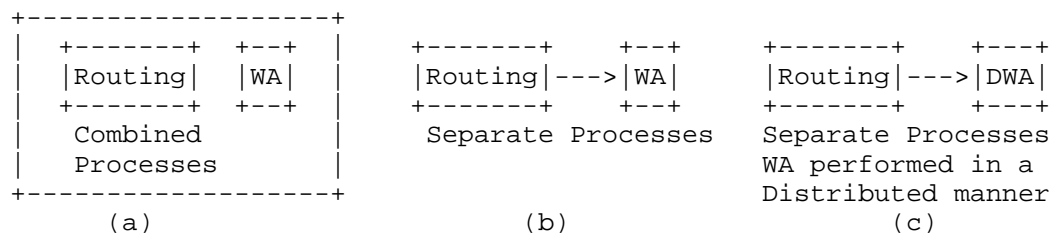


Figure 1

RWA process alter

natives.

Detail description of each alternative can be found in [RFC6163].

2. IV-Candidates + RWA Process - As explained in [RFC6566] separating the impairment validation process from the RWA process maybe necessary to deal with impairment sharing constraints. In this architecture one PCE computes impairment candidates and another PCE uses this information while performing RWA. The requirements for PCE-to-PCE interaction of this architecture will be addressed in this document.
3. Routing + Distributed WA and IV - Here a standard path computation (unaware of detailed wavelength availability or optical impairments) takes place, then wavelength assignment and impairment validation is performed along this path in a distributed manner via signaling (RSVP-TE). This alternative should be covered by existing or emerging GMPLS PCEP extensions and does not present new WSON specific requirements.

2. WSON PCE Architectures and Requirements

In the previous section we reviewed various process architectures for implementing RWA with and without regard for optical impairment. In Figure 3 we reduce these alternatives to two PCE based implementations. As specified in [RFC6566], the PCE in Figure 3(a) should be given the necessary information for RWA and impairment validation, including WSON topology, link wavelength utilization as well as impairment information such as the adjustment range of tunable parameters, etc. Similarly, RWA-PCE should be equipped with all the information other than impairment-related ones which is a necessity for IV-PCE.

In Figure 3(a) we show the three processes of routing, wavelength assignment and impairment validation accessed via a single PCE. The implementation details of the interactions of the processes are not subject to standardization; this document concerns only the PCC to PCE communications.

In Figure 3(b) the impairment validation process is implemented in a separate PCE. Here the RWA-PCE acts as a coordinator and the PCC to RWA-PCE interface will be the same as in Figure 3(a), however in this case we have additional requirements for the RWA-PCE to IV-PCE interface.

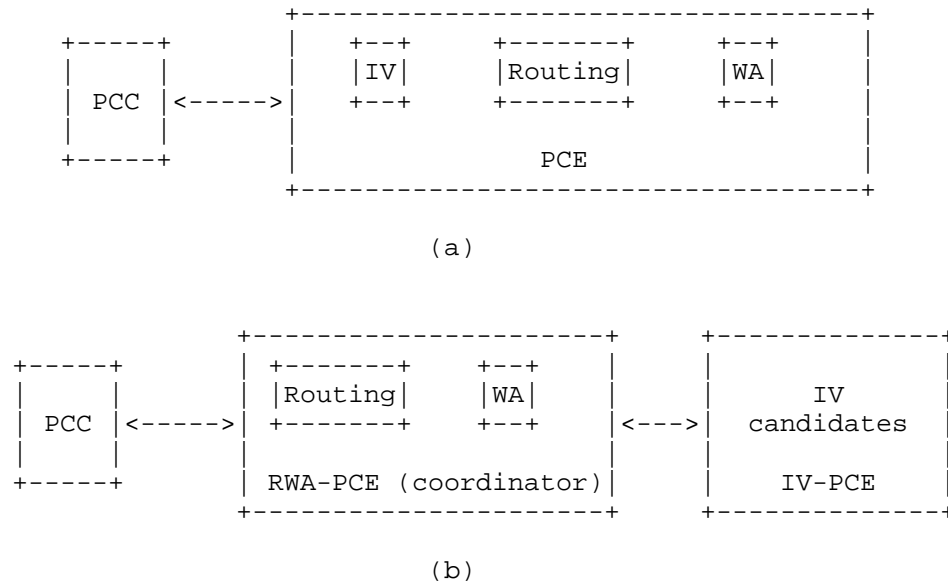


Figure 3

PCE architectures for

IA-RWA.

2.1. RWA PCC to PCE Interface

The PCC to PCE interface of Figure 3(a) and the PCC to RWA-PCE (coordinator) interface of Figure 3(b) are the same and we will cover both in this section. The following requirements for these interfaces are arranged by use cases:

2.1.1. A new RWA path request

The PCReq Message MUST include one or more specific measures of optical signal quality to which all feasible paths should conform:

- o BER limit
- o OSNR + Margin
- o Power
- o PMD
- o Residual Dispersion (RD)
- o Q factor
- o TBD

(Editor's Note: this is not a complete list of optical signal quality measure and subject to further change.)

If the PCReq Message does not include the BER limit and no BER limit information related to the specific path request is provisioned at the PCE then the PCE will return an error specifying that a BER limit must be provided.

"Margin" means "insurance" (e.g. 3~6dB) for suppliers and operators which are set against unpredictable degradation and other degradation not included in the provided estimates such as that due to fiber nonlinearity.

In non-coherent WDM networks, PMD and CD should be carefully considered. However, coherent WDM networks usually have a high tolerance with these two optical signal quality measurements and thus it may not need to be considered.

2.1.1.1. Signal Quality Measure TLV

This TLV represents all impairment constraints that need to be considered by the PCE to calculate a path that passes the requested measure of signal quality for a signal for a given source and destination.

This TLV is repeated one after another until all signal quality types are specified.

The TLV type is TBD.

The TLV data is defined as follow:

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								
P Signal Quality Type										Reserved																													
Threshold																																							

The P bit (1 bit): Indicates if the associated impairment is a path level or not.

The P bit is set to 1 indicates that the associated impairment is a path level. This means that the impairment is associated with the end-to-end path and the threshold must be satisfied on a path level.

The P bit is set to 0 indicates that the associated impairment is a link level. This means the impairment is associated with the link and the threshold must be satisfied on every link of the end-to-end path.

The Signal Quality Type (15 bits): indicates the kind of optical signal quality of interest.

0: reserved

1: BER limit

2: OSNR+ Margin

3: Power

4: PMD

5: CD

6: Q factor

7-up: Reserved for future use

Threshold (32 bits) indicates the threshold (upper or lower) to which the specified signal quality measure must satisfy for the path/link (depending on the P bit).

The reserved bits MUST be set to 0 on transmit and MUST be ignored on receive.

2.1.2. A new RWA path reply

The PCRep Message MUST include the route, wavelengths assigned to the route, and an indicator that says if the path conforms to the required quality or not. Moreover, it should also be able to specify a list of impairment compensation information along the chosen route, i.e., the value or value range of optical signal quality parameter that needs to be adjusted, such as power level, in order to achieve the resultant measure of signal quality as given in Section 2.1.2.1. It is suggested to carry this information in the PCEP ERO object. According to [RFC5440], PCEP ERO object is identical to RSVP-TE ERO object. Therefore, it is suggested to modify the RSVP-TE ERO object to accommodate this need. This will be included in a separate draft in the future.

In the case where a valid path is not found, the PCRep Message MUST include why the path is not found (e.g., no route, wavelength not found, BER failure, etc.)

2.1.2.1. Signal Quality Measure TLV

This TLV represents the result of the requested measure of signal quality for a signal for a given source and destination.

This TLV is repeated one after another until all signal quality types are specified.

The TLV type is TBD.

The TLV data is defined as follow:

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								
P Signal Quality Type										Reserved																													
										Signal Quality Value																													

The P bit (1 bit): Indicates if the associated signal quality measure has passed the threshold or not.

The P bit is set to 1 indicates that the associated signal quality measure has passed the threshold.

The P bit is set to 0 indicates that the associated signal quality measure has failed the threshold.

The Signal Quality Type (15 bits): indicates the kind of optical signal quality of interest.

0: reserved

1: BER limit

2: OSNR_ Margin

3: Power

4: PMD

5: CD

6: Q factor

7-up: Reserved for future use

Signal Quality Value (32 bits) indicates the actual estimated value of the specified signal quality measure for the end-to-end path.

TBD: How to encode link based value needs to be determined in the revision.

The reserved bits MUST be set to 0 on transmit and MUST be ignored on reception.

2.2. RWA-PCE to IV-PCE Interface

In [RFC6566] a sequence diagram for the interaction of the PCC, RWA-PCE and IV-PCE of Figure 3(b) was given and is repeated here in Figure 4. The interface between the PCC and the RWA-PCE (acting as the coordinator) was covered in section 2.1.

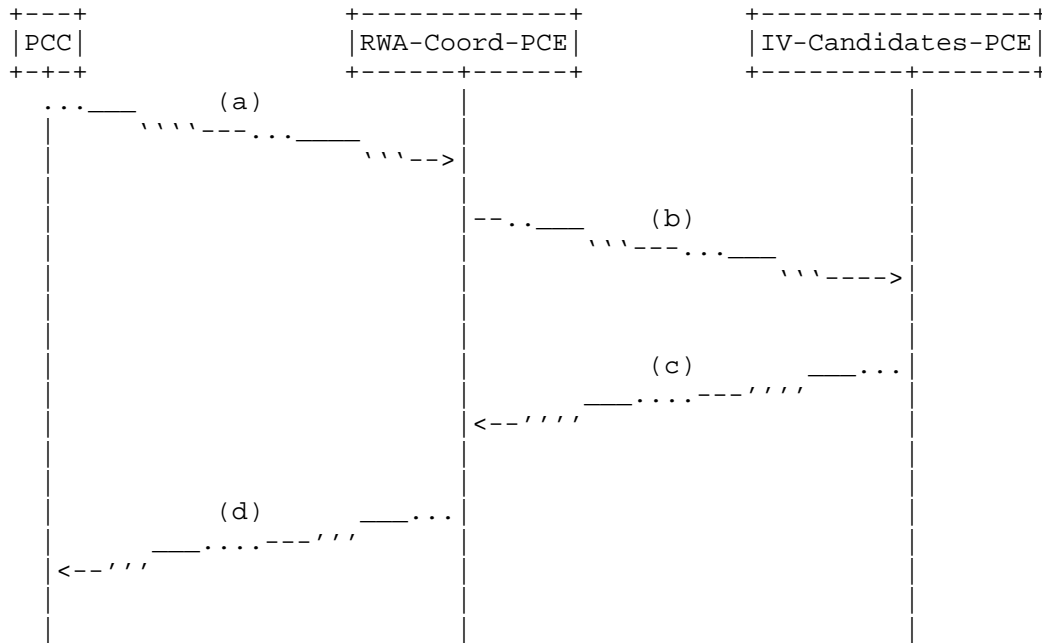


Figure 4 Sequence diagram for the interactions between P
CC, RWA-
Coordinating-PCE and the IV-Candidates-PCE.

The interface between the RWA-Coord-PCE and the IV-Candidates-PCE is specified by the following requirements:

1. The PCReq Message from the RWA-Coord-PCE to the IV-Candidate-PCE MUST include an indicator that more than one (candidate) path between source and destination is desired.
2. The PCReq message from the RWA-Coord-PCE to the IV-Candidates-PCE MUST include a limit on the number of optical impairment qualified paths to be returned by the IV-PCE.

3. The PCReq message from the RWA-Coord-PCE to the IV-Candidates-PCE MAY include wavelength constraints. Note that optical impairments are wavelength sensitive and hence specifying a wavelength constraint may help limit the search for valid paths. This requirement has been already covered in [PCEP-RWA] and is presented here for an illustration purpose.
4. The PCRep Message from the IV-Candidates-PCE to RWA-Coord-PCE MUST include a set of optical impairment qualified paths along with any wavelength constraints on those paths.
5. The PCRep Message from the IV-Candidates-PCE to RWA-Coord-PCE MUST indicate "no path found" in case where a valid path is not found.
6. The PCReq Message from the RWA-Coord-PCE to the IV-Candidate-PCE MAY include one or more specified paths and wavelengths that is to be verified by the IV-PCE. This requirement is necessary when the IV-PCE is allowed to verify specific paths.

Note that once the RWA-Coord-PCE receives the resulting paths from the IV Candidates PCE, then the RWA-Coord-PCE computes RWA for the IV qualified candidate paths and sends the result back to the PCC.

2.2.1. A new impairment-validated (IV) path request

Details on encoding are TBD.

2.2.2. A new impairment-validated (IV) path reply

Details on encoding are TBD.

3. Manageability Considerations

Manageability of WSON Routing and Wavelength Assignment (RWA) with PCE must address the following considerations:

3.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- o The ability to send a WSON IA-RWA request.

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- o The support for WSON IA-RWA.
- o The maximum number of synchronized path requests associated with WSON IA-RWA per request message.
- o A set of WSON IA-RWA specific policies (authorized sender, request rate limiter, etc).

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

3.2. Information and Data Models, e.g. MIB module

Extensions to the PCEP MIB module defined in [PCEP-MIB] should be defined, so as to cover the WSON IA-RWA information introduced in this document. A future revision of this document will list the information that should be added to the MIB module.

3.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in section 8.3 of [RFC5440].

3.4. Verifying Correct Operation

Mechanisms defined in this document do not imply any new verification requirements in addition to those already listed in section 8.4 of [RFC5440]

3.5. Requirements on Other Protocols and Functional Components

The PCE Discovery mechanisms ([RFC5089] and [RFC5088]) may be used to advertise WSON IA-RWA path computation capabilities to PCCs.

3.6. Impact on Network Operation

Mechanisms defined in this document do not imply any new network operation requirements in addition to those already listed in section 8.6 of [RFC5440].

4. Security Considerations

This document has no requirement for a change to the security models within PCEP [PCEP]. However the additional information distributed in order to address the RWA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

5. IANA Considerations

A future revision of this document will present requests to IANA for codepoint allocation.

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol (PCEP)", RFC 5440, March 2009.

6.2. Informative References

- [RFC6163] Lee, Y. and Bernstein, G., W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, April 2011.
- [RFC6566] Lee, Y. and Bernstein, G. (Editors), and D. Li, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6566, March, 2012.
- [PCEP-RWA] Y. Lee, G. Bernstein, J. Martensson, T. Takeda and T. Otani, "PCEP Requirements for WSON Routing and Wavelength Assignment", draft-lee-pce-wson-routing-wavelength, work in progress.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [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.

Authors' Addresses

Young Lee (Ed.)
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075, USA
Phone: (972) 509-5599 (x2240)
Email: ylee@huawei.com

Greg Bernstein (Ed.)
Grotto Networking
Fremont, CA, USA
Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Jonas Martensson
Acreo
Email: Jonas.Martensson@acreo.se

Tomonori Takeda
NTT Corporation
3-9-11, Midori-Cho
Musashino-Shi, Tokyo 180-8585, Japan
Email: takeda.tomonori@lab.ntt.co.jp

Takehiro Tsuritani
2-1-15 Ohara, Fujimino, Saitama, 356-8502, JAPAN
KDDI R&D Laboratories Inc.
Phone: +81-49-278-7806
Email: tsuri@kddilabs.jp

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

Phone: +86-755-28972913
Email: zhang.xian@huawei.com

7. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

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

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- o Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

- o Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- o Neither the name of Internet Society, IETF or IETF Trust, nor the names of specific contributors, may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Network Working Group
Internet Draft
Intended status: Standard Track
Expires: June 16, 2018

Y. Lee
Huawei

G. Bernstein
Grotto Networking

Jonas Martensson
Acreo

T. Takeda
NTT

T. Tsuritani
KDDI

December 17, 2018

PCEP Extensions for WSON Impairments

draft-lee-pce-wson-impairments-08

Abstract

As an optical signal progresses along its path it may be altered by the various physical processes in the optical fibers and devices it encounters. When such alterations result in signal degradation, these processes are usually referred to as "impairments". These physical characteristics may be important constraints to consider in path computation process in wavelength switched optical networks.

This document provides PCEP extensions to support Impairment Aware Routing and Wavelength Assignment (IA-RWA) in wavelength switched optical networks.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on June 3, 2019.

Copyright Notice

Copyright (c) 2018 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.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

Table of Contents

1. Introduction.....	3
1.1. WSON RWA Processes (no impairments).....	5
1.2. WSON IA-RWA Processes.....	6
2. WSON PCE Architectures and Requirements.....	8
2.1. RWA PCC to PCE Interface.....	9
2.1.1. A new RWA path request.....	9
2.1.1.1. Signal Quality Measure TLV.....	10
2.1.2. A new RWA path reply.....	12
2.1.2.1. Signal Quality Measure TLV.....	12
2.2. RWA-PCE to IV-PCE Interface.....	14
2.2.1. A new impairment-validated (IV) path request.....	15
2.2.2. A new impairment-validated (IV) path reply.....	15
3. Manageability Considerations.....	15
3.1. Control of Function and Policy.....	15
3.2. Information and Data Models, e.g. MIB module.....	16
3.3. Liveness Detection and Monitoring.....	16
3.4. Verifying Correct Operation.....	16
3.5. Requirements on Other Protocols and Functional Components.....	16
3.6. Impact on Network Operation.....	17
4. Security Considerations.....	17
5. IANA Considerations.....	17
6. References.....	17
6.1. Normative References.....	17
6.2. Informative References.....	18
Authors' Addresses.....	18
7. Acknowledgments.....	19

1. Introduction

[RFC4655] defines the PCE based architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is shown to be any network component that makes such a request and may be for instance an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communication Protocol (PCEP) is the communication protocol used between PCC and PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

This document provides a set of application-specific PCEP requirements for support of path computation in Wavelength Switched Optical Networks (WSON) with impairments. WSON refers to WDM based optical networks in which switching is performed selectively based on the wavelength of an optical signal.

The path in WSON is referred to as an optical path. An optical path may span multiple fiber links and the path should be assigned a wavelength for each link. A transparent optical network is made up of optical devices that can switch but not convert from one wavelength to another. In a transparent optical network, an optical path operates on the same wavelength across all fiber links that it traverses. In such case, the optical path is said to satisfy the wavelength-continuity constraint. Two optical paths that share a common fiber link can not be assigned the same wavelength. To do otherwise would result in both signals interfering with each other. Note that advanced additional multiplexing techniques such as polarization based multiplexing are not addressed in this document since the physical layer aspects are not currently standardized. Therefore, assigning the proper wavelength on an optical path is an essential requirement in the optical path computation process.

When a switching node has the ability to perform wavelength conversion the wavelength-continuity constraint can be relaxed, and a may use different wavelengths on different links along its route from origin to destination. It is, however, to be noted that wavelength converters may be limited due to their relatively high cost, while the number of WDM channels that can be supported in a fiber is also limited. As a WSON can be composed of network nodes that cannot perform wavelength conversion, nodes with limited wavelength conversion, and nodes with full wavelength conversion abilities, wavelength assignment is an additional routing constraint to be considered in all optical path computation.

One of the most basic questions in communications is whether one can successfully transmit information from a transmitter to a receiver within a prescribed error tolerance, usually specified as a maximum permissible bit error ratio (BER). This generally depends on the nature of the signal transmitted between the sender and receiver and the nature of the communications channel between the sender and

receiver. The optical path utilized (along with the wavelength) determines the communications channel.

The optical impairments incurred by the signal along the fiber and at each optical network element along the path determine whether the BER performance or any other measure of signal quality can be met for this particular signal on this particular path. Given the existing standards covering optical characteristics (impairments) and the knowledge of how the impact of impairments may be estimated along a path, [RFC6566] provides a framework for impairment aware path computation and establishment utilizing GMPLS protocols and the PCE architecture.

Some transparent optical subnetworks are designed such that over any path the degradation to an optical signal due to impairments never exceeds prescribed bounds. This may be due to the limited geographic extent of the network, the network topology, and/or the quality of the fiber and devices employed. In such networks the path selection problem reduces to determining a continuous wavelength from source to destination (the Routing and Wavelength Assignment problem). These networks are discussed in [RFC6163]. In other optical networks, impairments are important and the path selection process must be impairment-aware.

In this document we first review the processes for routing and wavelength assignment (RWA) used when wavelength continuity constraints are present. We then review the processes for optical impairment aware RWA (IA-RWA). Based on selected process models we then specify requirements for PCEP to support IA-RWA. Note that requirements for PCEP to support RWA are specified in a separate document [RFC7449].

The remainder of this document uses terminology from [RFC4655].

1.1. WSON RWA Processes (no impairments)

In [RFC6163] three alternative process architectures were given for performing routing and wavelength assignment. These are shown schematically in Figure 1.

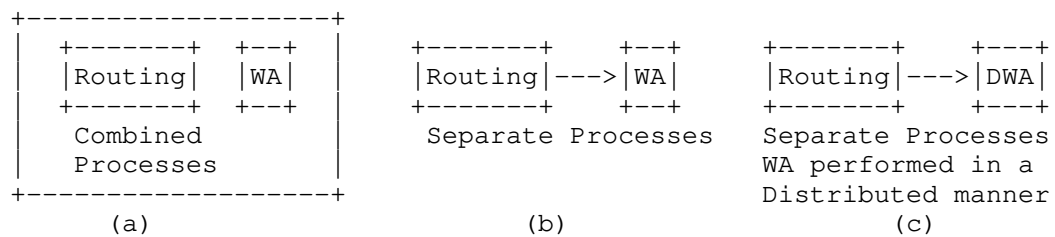


Figure 1

RWA process alte

rnatives.

Detail description of each alternative can be found in [RFC6163].

1.2. WSON IA-RWA Processes

In [RFC6566] impairments were addressed by adding an "impairment validation" (IV) process. For approximate impairment validation three process alternatives were given in [RFC6566] and are shown in Figure 2. Since there are many possible alternative combinations, these are just three examples. Please note that the requirements for all possible architectures can be reduced to the cases in Figure 3 in section 2.

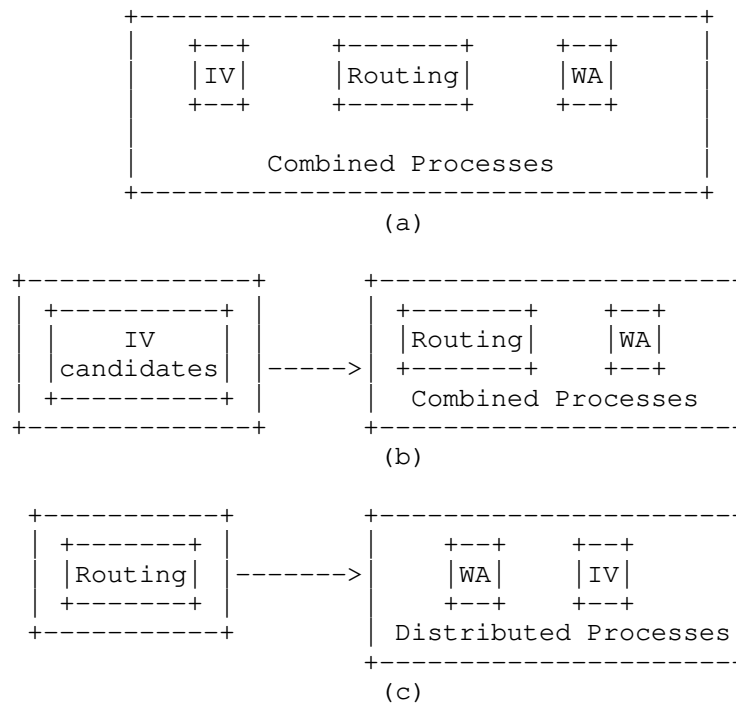


Figure 2 Process flows for the three main architectural alternatives.

These alternatives have the following properties and impact on PCEP requirements in this document.

1. Combined IV and RWA Process - Here the processes of impairment validation, routing and wavelength assignment are aggregated into a single PCE. The requirements for PCC-PCE interaction with such a combined IV-RWA process PCE is addressed in this document.
2. IV-Candidates + RWA Process - As explained in [RFC6566] separating the impairment validation process from the RWA process maybe necessary to deal with impairment sharing constraints. In this architecture one PCE computes impairment candidates and another PCE uses this information while performing RWA. The requirements for PCE-to-PCE interaction of this architecture will be addressed in this document.

3. Routing + Distributed WA and IV - Here a standard path computation (unaware of detailed wavelength availability or optical impairments) takes place, then wavelength assignment and impairment validation is performed along this path in a distributed manner via signaling (RSVP-TE). This alternative should be covered by existing or emerging GMPLS PCEP extensions and does not present new WSON specific requirements.

2. WSON PCE Architectures and Requirements

In the previous section we reviewed various process architectures for implementing RWA with and without regard for optical impairment. In Figure 3 we reduce these alternatives to two PCE based implementations. As specified in [RFC6566], the PCE in Figure 3(a) should be given the necessary information for RWA and impairment validation, including WSON topology, link wavelength utilization as well as impairment information such as the adjustment range of tunable parameters, etc. Similarly, RWA-PCE should be equipped with all the information other than impairment-related ones which is a necessity for IV-PCE.

In Figure 3(a) we show the three processes of routing, wavelength assignment and impairment validation accessed via a single PCE. The implementation details of the interactions of the processes are not subject to standardization; this document concerns only the PCC to PCE communications.

In Figure 3(b) the impairment validation process is implemented in a separate PCE. Here the RWA-PCE acts as a coordinator and the PCC to RWA-PCE interface will be the same as in Figure 3(a), however in this case we have additional requirements for the RWA-PCE to IV-PCE interface.

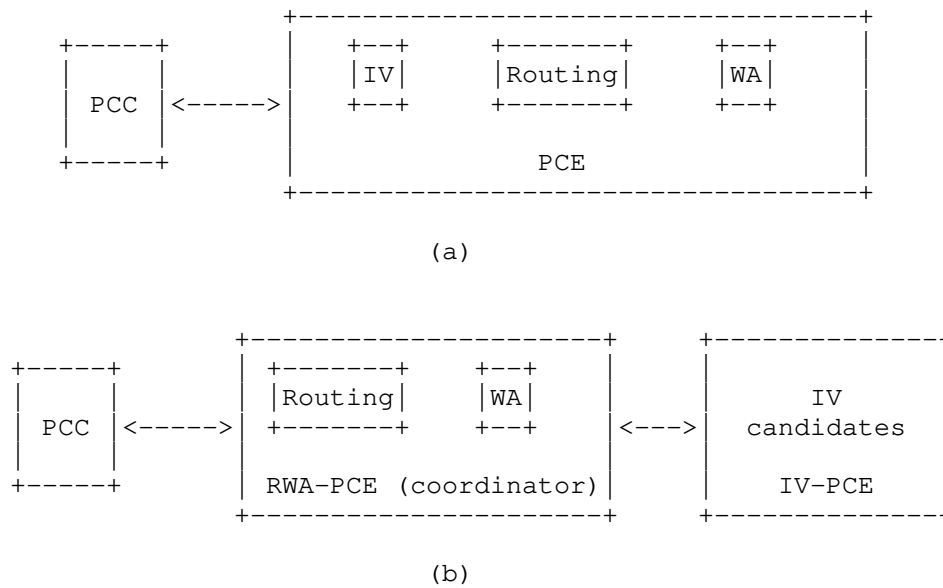


Figure 3

PCE architectures fo

r IA-RWA.

2.1. RWA PCC to PCE Interface

The PCC to PCE interface of Figure 3(a) and the PCC to RWA-PCE (coordinator) interface of Figure 3(b) are the same and we will cover both in this section. The following requirements for these interfaces are arranged by use cases:

2.1.1. A new RWA path request

The PCReq Message MUST include one or more specific measures of optical signal quality to which all feasible paths should conform:

- o BER limit
- o OSNR + Margin
- o Power
- o PMD
- o Residual Dispersion (RD)
- o Q factor
- o TBD

(Editor's Note: this is not a complete list of optical signal quality measure and subject to further change.)

If the PCReq Message does not include the BER limit and no BER limit information related to the specific path request is provisioned at the PCE then the PCE will return an error specifying that a BER limit must be provided.

"Margin" means "insurance" (e.g. 3~6dB) for suppliers and operators which are set against unpredictable degradation and other degradation not included in the provided estimates such as that due to fiber nonlinearity.

In non-coherent WDM networks, PMD and CD should be carefully considered. However, coherent WDM networks usually have a high tolerance with these two optical signal quality measurements and thus it may not need to be considered.

2.1.1.1. Signal Quality Measure TLV

This TLV represents all impairment constraints that need to be considered by the PCE to calculate a path that passes the requested measure of signal quality for a signal for a given source and destination.

This TLV is repeated one after another until all signal quality types are specified.

The TLV type is TBD.

The TLV data is defined as follow:

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								
P										Signal Quality Type										Reserved																			
										Threshold																													

The P bit (1 bit): Indicates if the associated impairment is a path level or not.

The P bit is set to 1 indicates that the associated impairment is a path level. This means that the impairment is associated with the end-to-end path and the threshold must be satisfied on a path level.

The P bit is set to 0 indicates that the associated impairment is a link level. This means the impairment is associated with the link and the threshold must be satisfied on every link of the end-to-end path.

The Signal Quality Type (15 bits): indicates the kind of optical signal quality of interest.

0: reserved

1: BER limit

2: OSNR+ Margin

3: Power

4: PMD

5: CD

6: Q factor

7-up: Reserved for future use

Threshold (32 bits) indicates the threshold (upper or lower) to which the specified signal quality measure must satisfy for the path/link (depending on the P bit).

The reserved bits MUST be set to 0 on transmit and MUST be ignored on receive.

2.1.2. A new RWA path reply

The PCRep Message MUST include the route, wavelengths assigned to the route, and an indicator that says if the path conforms to the required quality or not. Moreover, it should also be able to specify a list of impairment compensation information along the chosen route, i.e., the value or value range of optical signal quality parameter that needs to be adjusted, such as power level, in order to achieve the resultant measure of signal quality as given in Section 2.1.2.1. It is suggested to carry this information in the PCEP ERO object. According to [RFC5440], PCEP ERO object is identical to RSVP-TE ERO object. Therefore, it is suggested to modify the RSVP-TE ERO object to accommodate this need. This will be included in a separate draft in the future.

In the case where a valid path is not found, the PCRep Message MUST include why the path is not found (e.g., no route, wavelength not found, BER failure, etc.)

2.1.2.1. Signal Quality Measure TLV

This TLV represents the result of the requested measure of signal quality for a signal for a given source and destination.

This TLV is repeated one after another until all signal quality types are specified.

The TLV type is TBD.

The TLV data is defined as follow:

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	2	3	4	5	6	7	8	9
P										Signal Quality Type										Reserved																			
										Signal Quality Value																													

The P bit (1 bit): Indicates if the associated signal quality measure has passed the threshold or not.

The P bit is set to 1 indicates that the associated signal quality measure has passed the threshold.

The P bit is set to 0 indicates that the associated signal quality measure has failed the threshold.

The Signal Quality Type (15 bits): indicates the kind of optical signal quality of interest.

0: reserved

1: BER limit

2: OSNR_ Margin

3: Power

4: PMD

5: CD

6: Q factor

7-up: Reserved for future use

Signal Quality Value (32 bits) indicates the actual estimated value of the specified signal quality measure for the end-to-end path.

TBD: How to encode link based value needs to be determined in the revision.

The reserved bits MUST be set to 0 on transmit and MUST be ignored on reception.

2.2. RWA-PCE to IV-PCE Interface

In [RFC6566] a sequence diagram for the interaction of the PCC, RWA-PCE and IV-PCE of Figure 3(b) was given and is repeated here in Figure 4. The interface between the PCC and the RWA-PCE (acting as the coordinator) was covered in section 2.1.

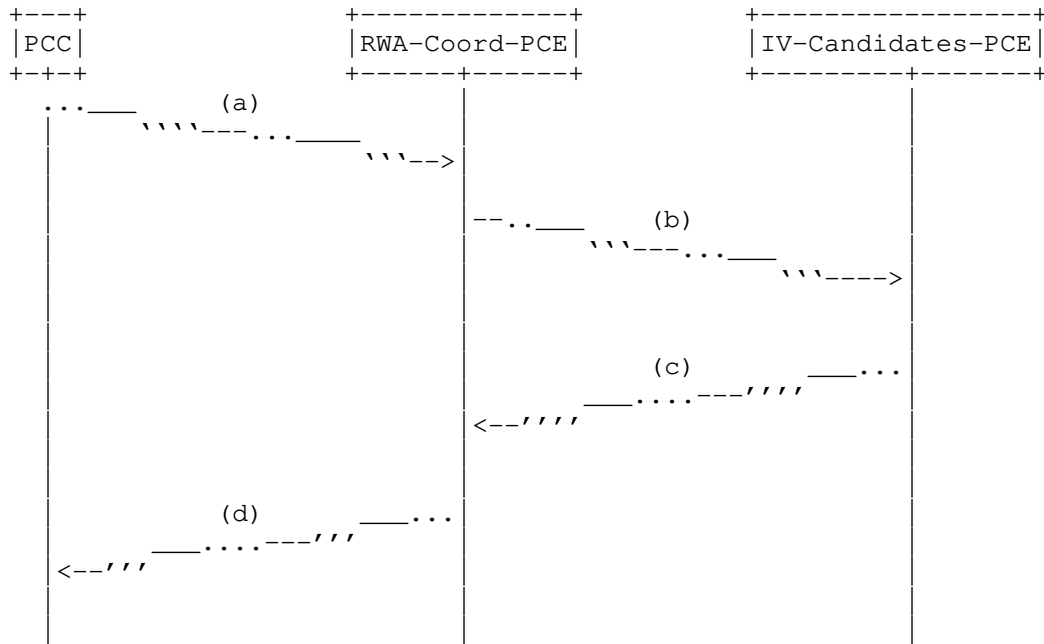


Figure 4 Sequence diagram for the interactions between PCC, RWA-Coordinating-PCE and the IV-Candidates-PCE.

The interface between the RWA-Coord-PCE and the IV-Candidates-PCE is specified by the following requirements:

1. The PCReq Message from the RWA-Coord-PCE to the IV-Candidate-PCE MUST include an indicator that more than one (candidate) path between source and destination is desired.
2. The PCReq message from the RWA-Coord-PCE to the IV-Candidates-PCE MUST include a limit on the number of optical impairment qualified paths to be returned by the IV-PCE.

3. The PCReq message from the RWA-Coord-PCE to the IV-Candidates-PCE MAY include wavelength constraints. Note that optical impairments are wavelength sensitive and hence specifying a wavelength constraint may help limit the search for valid paths. This requirement has been already covered in [RFC7449] and is presented here for an illustration purpose.
4. The PCRep Message from the IV-Candidates-PCE to RWA-Coord-PCE MUST include a set of optical impairment qualified paths along with any wavelength constraints on those paths.
5. The PCRep Message from the IV-Candidates-PCE to RWA-Coord-PCE MUST indicate "no path found" in case where a valid path is not found.
6. The PCReq Message from the RWA-Coord-PCE to the IV-Candidate-PCE MAY include one or more specified paths and wavelengths that is to be verified by the IV-PCE. This requirement is necessary when the IV-PCE is allowed to verify specific paths.

Note that once the RWA-Coord-PCE receives the resulting paths from the IV Candidates PCE, then the RWA-Coord-PCE computes RWA for the IV qualified candidate paths and sends the result back to the PCC.

2.2.1. A new impairment-validated (IV) path request

Details on encoding are TBD.

2.2.2. A new impairment-validated (IV) path reply

Details on encoding are TBD.

3. Manageability Considerations

Manageability of WSON Routing and Wavelength Assignment (RWA) with PCE must address the following considerations:

3.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- o The ability to send a WSON IA-RWA request.

In addition to the parameters already listed in Section 8.1 of [RFC5440], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- o The support for WSON IA-RWA.
- o The maximum number of synchronized path requests associated with WSON IA-RWA per request message.
- o A set of WSON IA-RWA specific policies (authorized sender, request rate limiter, etc).

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

3.2. Information and Data Models, e.g. MIB module

Extensions to the PCEP MIB module defined in [PCEP-MIB] should be defined, so as to cover the WSON IA-RWA information introduced in this document. A future revision of this document will list the information that should be added to the MIB module.

3.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in section 8.3 of [RFC5440].

3.4. Verifying Correct Operation

Mechanisms defined in this document do not imply any new verification requirements in addition to those already listed in section 8.4 of [RFC5440]

3.5. Requirements on Other Protocols and Functional Components

The PCE Discovery mechanisms ([RFC5089] and [RFC5088]) may be used to advertise WSON IA-RWA path computation capabilities to PCCs.

3.6. Impact on Network Operation

Mechanisms defined in this document do not imply any new network operation requirements in addition to those already listed in section 8.6 of [RFC5440].

4. Security Considerations

This document has no requirement for a change to the security models within PCEP [PCEP]. However the additional information distributed in order to address the RWA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

5. IANA Considerations

A future revision of this document will present requests to IANA for codepoint allocation.

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol (PCEP)", RFC 5440, March 2009.

6.2. Informative References

- [RFC6163] Lee, Y. and Bernstein, G., W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, April 2011.
- [RFC6566] Lee, Y. and Bernstein, G. (Editors), and D. Li, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6566, March, 2012.
- [RFC7449] Y. Lee, G. Bernstein, J. Martensson, T. Takeda and T. Otani, "PCEP Requirements for WSON Routing and Wavelength Assignment", RFC 7449, February 2015.
- [RFC5088] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [RFC5089] Le Roux, JL., Ed., Vasseur, JP., Ed., Ikejiri, Y., and R. Zhang, "IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5089, January 2008.
- [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.
- [RFC8174] B. Leiba, "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", RFC 8174, May 2017.

Authors' Addresses

Young Lee (Ed.)
Huawei Technologies
Email: leeyoung@huawei.com

Greg Bernstein (Ed.)
Grotto Networking
Fremont, CA, USA
Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Jonas Martensson
Acreo
Email: Jonas.Martensson@acreo.se

Tomonori Takeda
NTT Corporation
3-9-11, Midori-Cho
Musashino-Shi, Tokyo 180-8585, Japan
Email: takeda.tomonori@lab.ntt.co.jp

Takehiro Tsuritani
2-1-15 Ohara, Fujimino, Saitama, 356-8502, JAPAN
KDDI R&D Laboratories Inc.
Phone: +81-49-278-7806
Email: tsuri@kddilabs.jp

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

Phone: +86-755-28972913
Email: zhang.xian@huawei.com

7. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

Copyright (c) 2018 IETF Trust and the persons identified as authors of the code. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- o Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

- o Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- o Neither the name of Internet Society, IETF or IETF Trust, nor the names of specific contributors, may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Network Working Group
Internet Draft

Y. Lee, Ed.
Huawei Technologies

Intended status: Standard
Expires: October 2012

R. Casellas, Ed.
CTTC

April 30, 2012

PCEP Extension for WSON Routing and Wavelength Assignment

draft-lee-pce-wson-rwa-ext-04.txt

Abstract

This draft provides the Path Computation Element communication Protocol (PCEP) extensions for the support of Routing and Wavelength Assignment (RWA) in Wavelength Switched Optical Networks (WSON). Lightpath provisioning in WSONs requires a routing and wavelength assignment (RWA) process. From a path computation perspective, wavelength assignment is the process of determining which wavelength can be used on each hop of a path and forms an additional routing constraint to optical light path computation.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on October 30, 2012.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Terminology.....	3
2. Requirements Language.....	3
3. Introduction.....	3
4. Encoding of a RWA Path Request.....	6
4.1. Wavelength Assignment (WA) Object.....	6
4.2. Wavelength Restriction Constraint TLV.....	8
4.2.1. Link Identifier sub-TLV.....	11
4.2.2. Wavelength Restriction Field sub-TLV.....	12
4.3. Signal processing capability restrictions.....	12
4.3.1. MODULATION-FORMAT-LIST Restriction TLV.....	13
4.3.2. FEC-LIST Restriction TLV.....	14
4.3.3. Signal Processing Exclusion XRO Sub-Object.....	14
4.3.4. IRO sub-object: signal processing inclusion.....	14
5. Encoding of a RWA Path Reply.....	15
5.1. Error Indicator.....	15
5.2. NO-PATH Indicator.....	16
6. Manageability Considerations.....	16
6.1. Control of Function and Policy.....	16
6.2. Information and Data Models, e.g. MIB module.....	17
6.3. Liveness Detection and Monitoring.....	17
6.4. Verifying Correct Operation.....	17
6.5. Requirements on Other Protocols and Functional Components	17

6.6. Impact on Network Operation.....	17
7. Security Considerations.....	17
8. IANA Considerations.....	18
9. Acknowledgments.....	18
10. References.....	18
10.1. Informative References.....	18
11. Contributors.....	20
Authors' Addresses.....	21
Intellectual Property Statement.....	21
Disclaimer of Validity.....	22

1. Terminology

This document uses the terminology defined in [RFC4655], and [RFC5440].

2. 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].

3. Introduction

[RFC4655] defines the PCE based Architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is said to be any network component that makes such a request and may be, for instance, an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communications Protocol (PCEP) is the communication protocol used between PCC and PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

This document provides the PCEP extensions for the support of Routing and Wavelength Assignment (RWA) in Wavelength Switched Optical Networks (WSON) based on the requirements specified in [PCE-RWA].

WSON refers to WDM based optical networks in which switching is performed selectively based on the wavelength of an optical signal. In this document, it is assumed that wavelength converters require electrical signal regeneration. Consequently, WSONs can be transparent (A transparent optical network is made up of optical devices that can switch but not convert from one wavelength to another, all within the optical domain) or translucent (3R regenerators are sparsely placed in the network).

A LSC Label Switched Path (LSP) may span one or several transparent segments, which are delimited by 3R regenerators (typically with electronic regenerator and optional wavelength conversion). Each transparent segment or path in WSON is referred to as an optical path. An optical path may span multiple fiber links and the path should be assigned the same wavelength for each link. In such case, the optical path is said to satisfy the wavelength-continuity constraint. Figure 1 illustrates the relationship between a LSC LSP and transparent segments (optical paths).

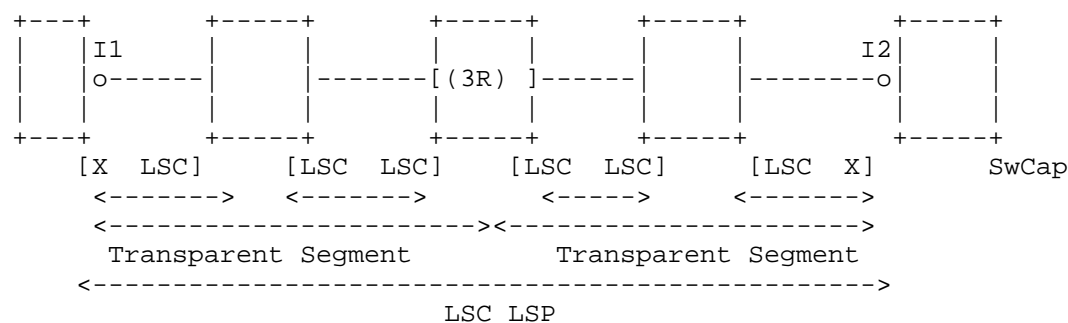


Figure 1 Illustration of a LSC LSP and transparent segments

Note that two optical paths within a WSON LSP need not operate on the same wavelength (due to the wavelength conversion capabilities). Two optical paths that share a common fiber link cannot be assigned the same wavelength. To do otherwise would result in both signals interfering with each other. Note that advanced additional multiplexing techniques such as polarization based multiplexing are

not addressed in this document since the physical layer aspects are not currently standardized. Therefore, assigning the proper wavelength on a lightpath is an essential requirement in the optical path computation process.

When a switching node has the ability to perform wavelength conversion, the wavelength-continuity constraint can be relaxed, and a LSC Label Switched Path (LSP) may use different wavelengths on different links along its route from origin to destination. It is, however, to be noted that wavelength converters may be limited due to their relatively high cost, while the number of WDM channels that can be supported in a fiber is also limited. As a WSON can be composed of network nodes that cannot perform wavelength conversion, nodes with limited wavelength conversion, and nodes with full wavelength conversion abilities, wavelength assignment is an additional routing constraint to be considered in all lightpath computation.

For example, within a translucent WSON, a LSC LSP may be established between interfaces I1 and I2, spanning 2 transparent segments (optical paths) where the wavelength continuity constraint applies (i.e. the same unique wavelength MUST be assigned to the LSP at each TE link of the segment). If the LSC LSP induced a Forwarding Adjacency / TE link, the switching capabilities of the TE link would be [X X] where $X < LSC$ (PSC, TDM, ...).

[Ed note: in general, WSON LSC may not be the only switching layer with switching constraints. From a GMPLS/PCEP perspective, wavelength assignment corresponds to label allocation. This document should align with GMPLS extensions for PCEP. Wavelength restrictions and constraints should be formulated in terms of labels (i.e. LABEL_SET, SUGGESTED_LABEL, UPSTREAM_LABEL, etc.)]

The optical modulation properties, which are also referred to as signal compatibility, are already considered in signaling in [RWA-Encode] and [WSON-OSPF]. In order to improve the signal quality and limit some optical effects several advanced modulation processing are used. Those modulation properties contribute not only to optical signal quality checks but also constrain the selection of sender and receiver, as they should have matching signal processing capabilities. This document includes signal compatibility constraint as part of RWA path computation. That is, the signal processing capabilities (e.g., modulation and FEC) must be compatible between the sender and the receiver of the optical path across all optical elements.

This document, however, does not address optical impairments as part of RWA path computation. See [WSON-Imp] and [RSVP-Imp] for more information on optical impairments and GMPLS.

4. Encoding of a RWA Path Request

Figure 2 shows one typical PCE based implementation, which is referred to as Combined Process (R&WA). With this architecture, the two processes of routing and wavelength assignment are accessed via a single PCE. This architecture is the base architecture from which the requirements have been specified in [PCE-RWA] and the PCEP extensions that are going to be specified in this document based on this architecture.

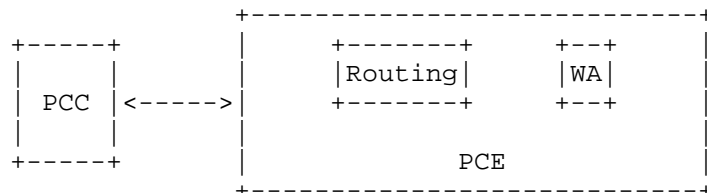


Figure 2 Combined Process (R&WA) architecture

4.1. Wavelength Assignment (WA) Object

The current RP object is used to indicate routing related information in a new path request per [RFC5440]. Since a new RWA path request involves both routing and wavelength assignment, the wavelength assignment related information in the request SHOULD be coupled in the path request.

Wavelength allocation can be performed by the PCE by different means:

- (a) By means of Explicit Label Control, in the sense that one (or two) allocated labels MAY appear after an interface route subobject.
- (b) By means of a Label Set, containing one or more allocated Labels, provided by the PCE.

Option (b) allows distributed label allocation (performed during signaling) to complete wavelength assignment.

Additionally, given a range of potential labels to allocate, the request SHOULD convey the heuristic / mechanism to the allocation.

The format of a PCReq message after incorporating the WA object is as follows:

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

Where:

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

<request> ::= <RP>
              <ENDPOINTS>
              <WA>
              [other optional objects...]
```

If WA object is present in the request, the WA object MUST be encoded after the ENDPOINTS object.

The format of the Wavelength Assignment (WA) object body is as follows:

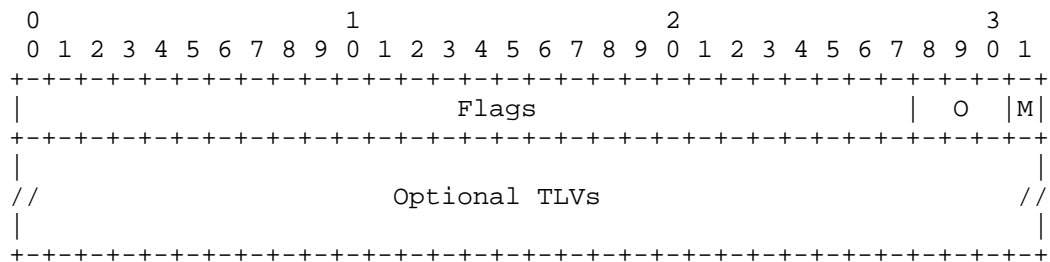


Figure 3 WA Object

o Flags (32 bits)

The following new flags SHOULD be set

- . M (Mode - 1 bit): M bit is used to indicate the mode of wavelength assignment. When M bit is set to 1, this indicates that the label assigned by the PCE must be explicit. That is, the selected way to convey the allocated wavelength is by means of Explicit Label Control (ELC) [RFC4003] for each hop of a computed LSP. Otherwise, the label assigned by the PCE needs not be explicit (i.e., it can be suggested in the form of label set objects in the corresponding response, to allow distributed WA. In such case, the PCE MUST return a Label_Set object in the response if the path is found.

When the distributed WA is applied, some specific wavelength range and/or the maximum number of wavelengths to be returned in the Label Set might be additionally indicated. The optional TLV field will be used for conveying this additional request. The details of this encoding will be provided in a later revision.

- . O (Order - 3 bits): O bit is used to indicate the wavelength assignment constraint in regard to the order of wavelength assignment to be returned by the PCE. This case is only applied when M bit is set to "explicit." The following indicators should be defined:

000 - Reserved

001 - Random Assignment

010 - First Fit (FF) in descending Order

011 - First Fit (FF) in ascending Order

100 - Last Fit (LF) in ascending Order

101 - Last Fit (LF) in descending Order

110 - Unspecified

111 - Reserved

4.2. Wavelength Restriction Constraint TLV

For any request that contains a wavelength assignment, the requester (PCC) MUST be able to specify a restriction on the wavelengths to be used. This restriction is to be interpreted by the PCE as a constraint on the tuning ability of the origination laser transmitter or on any other maintenance related constraints. Note that if the LSP LSC spans different segments, the PCE MUST have

mechanisms to know the tunability restrictions of the involved wavelength converters / regenerators, e.g. by means of the TED either via IGP or NMS. Even if the PCE knows the tunability of the transmitter, the PCC MUST be able to apply additional constraints to the request.

[Ed note: Which PCEP Object will home this TLV is yet to be determined. Since this involves the end-point, The END-POINTS Object might be a good candidate to encode this TLV, which will be provided in a later revision.]

[Ed note: The current encoding assumes that tunability restriction applied to link-level.]

The TLV type is TBD, recommended value is TBD. This TLV MAY appear more than once to be able to specify multiple restrictions.

The TLV data is defined as follows:

<Wavelength Restriction Constraint> ::=

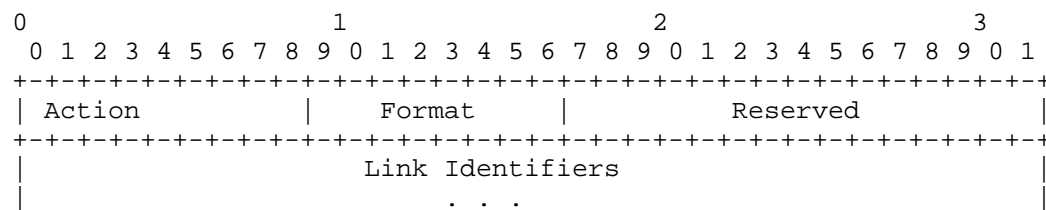
<Action> <Format> <Reserved>

(<Link Identifiers> <Wavelength Restriction>)...

Where

<Link Identifiers> ::=

<Unnumbered IF ID> | <IPv4 Address> | <IPv6 Address>



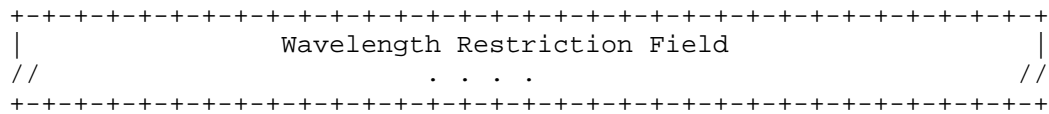


Figure 4 Wavelength Restriction

- o Action: 8 bits

- . 0 - Inclusive List indicates that one or more link identifiers are included in the Link Set. Each identifies a separate link that is part of the set.
- . 1 - Inclusive Range indicates that the Link Set defines a range of links. It contains two link identifiers. The first identifier indicates the start of the range (inclusive). The second identifier indicates the end of the range (inclusive). All links with numeric values between the bounds are considered to be part of the set. A value of zero in either position indicates that there is no bound on the corresponding portion of the range. Note that the Action field can be set to 0 when unnumbered link identifier is used.

Note that "interfaces" such as those discussed in the Interfaces MIB [RFC2863] are assumed to be bidirectional.

- o Format: The format of the link identifier (8 bits)

- . 0 -- Unnumbered Link Identifier
- . 1 -- Local Interface IPv4 Address
- . 2 -- Local Interface IPv6 Address
- . Others TBD.

Note that all link identifiers in the same list must be of the same type.

- o Reserved: Reserved for future use (16 bits)

o Link Identifiers: Identifies each link ID for which restriction is applied. The length is dependent on the link format. See the following section for Link Identifier encoding.

4.2.1. Link Identifier sub-TLV

The link identifier field can be an IPv4, IPv6 or unnumbered interface ID.

<Link Identifier> ::=

<IPv4 Address> | <IPv6 Address> | <Unnumbered IF ID>

The encoding of each case is as follows:

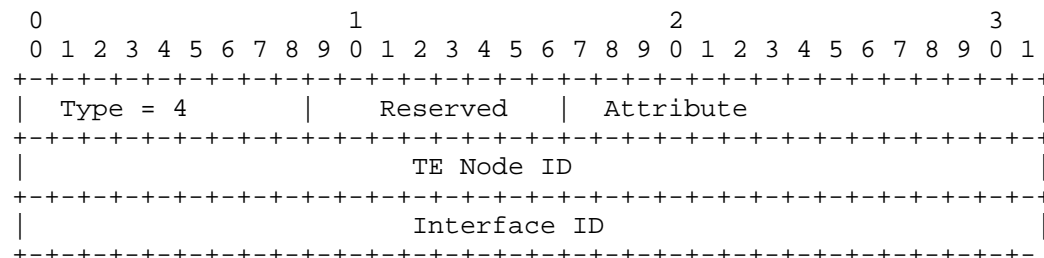
IPv4 prefix Sub-TLV

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	2	3	4	5	6	7	8	9
Type = 1										IPv4 address (4 bytes)																													
IPv4 address (continued)										Prefix Length										Attribute																			

IPv6 prefix Sub-TLV

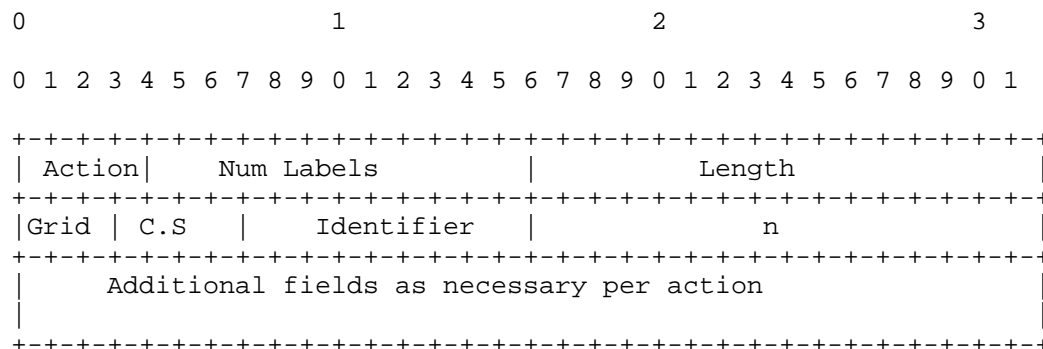
0										1										2										3									
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
Type = 2										IPv6 address (16 bytes)																													
IPv6 address (continued)																																							
IPv6 address (continued)																																							
IPv6 address (continued)																																							
IPv6 address (continued)										Prefix Length										Attribute																			

Unnumbered Interface ID Sub-TLV



4.2.2. Wavelength Restriction Field sub-TLV

The Wavelength Restriction Field of the wavelength restriction TLV is encoded as a Label Set field as specified in [GEN-Encode] section 2.2, as shown below, with base label encoded as a 32 bit LSC label, defined in [RFC6205]. See [RFC6205] for a description of Grid, C.S, Identifier and n, as well as [GEN-Encode] for the details of each action.



4.3. Signal processing capability restrictions

Path computation for WSON include the check of signal processing capabilities, those capability MAY be provided by the IGP, however this is not a MUST. Moreover, a PCC should be able to indicate additional restrictions for those signal compatibility, either on the endpoint or any given link.

The supported signal processing capabilities are the one described in [RWA-Info]:

- . Modulation Type List
- . FEC Type List
- . Bit rate
- . Client signal

The Bit-rate restriction is already expressed in [PCEP-GMPLS] in the GENERALIZED-BANDWIDTH object.

The client signal information can be expressed using the REQ-ADAP-CAP object from the [PCEP-Layer].

In order to support the Modulation and FEC information two new TLV are introduced as endpoint-restriction in the END-POINTS type Generalized endpoint:

- . Modulation restriction TLV
- . FEC restriction TLV.

The END-POINTS type generalized endpoint is extended as follow:

```
<endpoint-restrictions> ::= <LABEL-REQUEST>
                               <label-restriction-list>
                               [<signal-compatibility-restriction>...]
```

Where

```
signal-compatibility-restriction ::=
    <MODULATION-FORMAT-LIST>|<FEC-LIST>
```

The MODULATION-FORMAT-LIST and FEC-LIST TLV are described in the following sections.

4.3.1. MODULATION-FORMAT-LIST Restriction TLV

The format follows the definition from [WSON-Encode] section 5.2.

4.3.2. FEC-LIST Restriction TLV

The format follows the definition from [WSON-Encode] section 5.3.

4.3.3. Signal Processing Exclusion XRO Sub-Object

The PCC/PCE should be able to exclude particular types of signal processing along the path in order to handle client restriction or multi-domain path computation.

In order to support the exclusion a new XRO sub-object is defined: the signal processing exclusion:

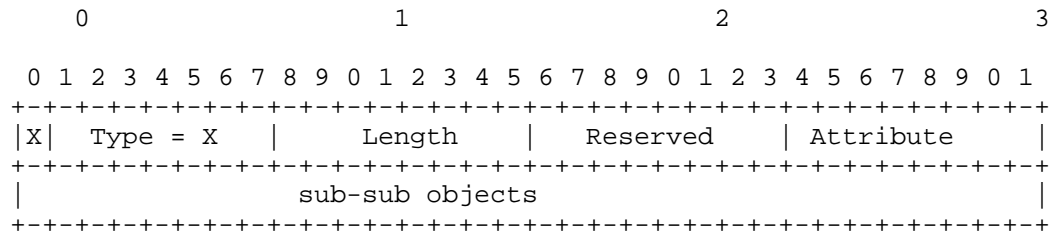


Figure 5 Signaling Processing XRO Sub-Object

The Attribute field indicates how the exclusion sub-object is to be interpreted. The Attribute can only be 0 (Interface) or 1 (Node).

The sub-sub objects are encoded as in RSVP signaling definition [WSON-Sign].

4.3.4. IRO sub-object: signal processing inclusion

Similar to the XRO sub-object the PCC/PCE should be able to include particular types of signal processing along the path in order to handle client restriction or multi-domain path computation.

This is supported by adding the sub-object "processing" defined for ERO in [WSON-Sign] to the PCEP IRO object.

5. Encoding of a RWA Path Reply

The ERO is used to encode the path of a TE LSP through the network. The ERO is carried within a given path of a PCEP response, which is in turn carried in a PCRep message to provide the computed TE LSP if the path computation was successful. The preferred way to convey the allocated wavelength is by means of Explicit Label Control (ELC) [RFC4003].

In order to encode wavelength assignment, the Wavelength Assignment (WA) Object needs to be employed to be able to specify wavelength assignment. Since each segment of the computed optical path is associated with wavelength assignment, the WA Object should be aligned with the ERO object.

Encoding details will be provided further revisions and will be aligned as much as possible with [WSON-Sign].

5.1. Error Indicator

To indicate errors associated with the RWA request, a new Error Type (TDB) and subsequent error-values are defined as follows for inclusion in the PCEP-ERROR Object:

A new Error-Type (TDB) and subsequent error-values are defined as follows:

- . Error-Type=TBD; Error-value=1: if a PCE receives a RWA request and the PCE is not capable of processing the request due to insufficient memory, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=TDB) and an Error-value(Error-value=1). The PCE stops processing the request. The corresponding RWA request MUST be cancelled at the PCC.
- . Error-Type=TBD; Error-value=2: if a PCE receives a RWA request and the PCE is not capable of RWA computation, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=15) and an Error-value (Error-value=2). The PCE stops processing the request. The corresponding RWA computation MUST be cancelled at the PCC.

5.2. NO-PATH Indicator

To communicate the reason(s) for not being able to find RWA for the path request, the NO-PATH object can be used in the PCRep message. The format of the NO-PATH object body is defined in [RFC5440]. The object may contain a NO-PATH-VECTOR TLV to provide additional information about why a path computation has failed.

Two new bit flags are defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object.

- . Bit TDB: When set, the PCE indicates no feasible route was found that meets all the constraints associated with RWA.
- . Bit TDB: When set, the PCE indicates that no wavelength was assigned to at least one hop of the route in the response.
- . Bit TDB: When set, the PCE indicate that no path was found satisfying the signal compatibility constraints.

6. Manageability Considerations

Manageability of WSON Routing and Wavelength Assignment (RWA) with PCE must address the following considerations:

6.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [PCEP], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- . The ability to send a WSON RWA request.

In addition to the parameters already listed in Section 8.1 of [PCEP], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- . The support for WSON RWA.
- . A set of WSON RWA specific policies (authorized sender, request rate limiter, etc).

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

6.2. Information and Data Models, e.g. MIB module

Extensions to the PCEP MIB module defined in [PCEP-MIB] should be defined, so as to cover the WSON RWA information introduced in this document. A future revision of this document will list the information that should be added to the MIB module.

6.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in section 8.3 of [RFC5440].

6.4. Verifying Correct Operation

Mechanisms defined in this document do not imply any new verification requirements in addition to those already listed in section 8.4 of [RFC5440].

6.5. Requirements on Other Protocols and Functional Components

The PCE Discovery mechanisms ([RFC5089] and [RFC5088]) may be used to advertise WSON RWA path computation capabilities to PCCs.

6.6. Impact on Network Operation

Mechanisms defined in this document do not imply any new network operation requirements in addition to those already listed in section 8.6 of [RFC5440].

7. Security Considerations

This document has no requirement for a change to the security models within PCEP [PCEP]. However the additional information distributed in order to address the RWA problem represents a disclosure of

network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

8. IANA Considerations

A future revision of this document will present requests to IANA for codepoint allocation.

9. Acknowledgments

The authors would like to thank Adrian Farrel for many helpful comments that greatly improved the contents of this draft.

This document was prepared using 2-Word-v2.0.template.dot.

10. References

10.1. Informative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.
- [RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", RFC 4003, February 2005.
- [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.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol", RFC 5440, March 2009.
- [PCEP-GMPLS] Margaria, et al., "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions, work in progress.
- [PCEP-Layer] Oki, Takeda, Le Roux, and Farrel, "Extensions to the Path Computation Element communication Protocol (PCEP) for Inter-Layer MPLS and GMPLS Traffic Engineering", draft-ietf-pce-inter-layer-ext, work in progress.
- [RFC6163] Lee, Y. and Bernstein, G. (Editors), and W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, March 2011.
- [PCE-RWA] Lee, Y., et. al., "PCEP Requirements for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-routing-wavelength, work in progress.
- [RFC6205] Tomohiro, O. and D. Li, "Generalized Labels for Lambda-Switching Capable Label Switching Routers", RFC 6205, January, 2011.
- [WSON-Sign] Bernstein et al, "Signaling Extensions for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signaling, work in progress.
- [WSON-OSPF] Lee and Bernstein, "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signal-compatibility-ospf, work in progress.
- [RWA-Info] Bernstein and Lee, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info, work in progress.
- [RWA-Encode] Bernstein and Lee, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode, work in progress.

- [GEN-Encode] Bernstein and Lee, "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode, work in progress.
- [WSON-Imp] Y. Lee, G. Bernstein, D. Li, G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSON) with Impairments", draft-ietf-ccamp-wson-impairments, work in progress.
- [RSVP-Imp] agraz, "RSVP-TE Extensions in Support of Impairment Aware Routing and Wavelength Assignment in Wavelength Switched Optical Networks WSONs)", draft-agraz-ccamp-wson-impairment-rsvp, work in progress.
- [OSPF-Imp] Bellagamba, et al., "OSPF Extensions for Wavelength Switched Optical Networks (WSON) with Impairments", draft-eb-ccamp-ospf-wson-impairments, work in progress.

11. Contributors

Authors' Addresses

Young Lee, Editor
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075, USA
Phone: (972) 509-5599 (x2240)
Email: leeyoung@huawei.com

Ramon Casellas, Editor
CTTC PMT Ed B4 Av. Carl Friedrich Gauss 7
08860 Castelldefels (Barcelona)
Spain
Phone: (34) 936452916
Email: ramon.casellas@cttc.es

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

Cyril Margaria
Nokia Siemens Networks
St Martin Strasse 76
Munich, 81541
Germany
Phone: +49 89 5159 16934
Email: cyril.margaria@nsn.com

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
C/ Emilio Vargas 6
Madrid, 28043
Spain
Phone: +34 91 3374013
Email: ogondio@tid.es

Greg Bernstein
Grotto Networking
Fremont, CA, USA
Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet Draft

Y. Lee, Ed.
Huawei Technologies

Intended status: Standard
Expires: August 2013

R. Casellas, Ed.
CTTC

February 6, 2013

PCEP Extension for WSON Routing and Wavelength Assignment

draft-lee-pce-wson-rwa-ext-05.txt

Abstract

This draft provides the Path Computation Element communication Protocol (PCEP) extensions for the support of Routing and Wavelength Assignment (RWA) in Wavelength Switched Optical Networks (WSON). Lightpath provisioning in WSONs requires a routing and wavelength assignment (RWA) process. From a path computation perspective, wavelength assignment is the process of determining which wavelength can be used on each hop of a path and forms an additional routing constraint to optical light path computation.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on August 6, 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. Terminology.....	3
2. Requirements Language.....	3
3. Introduction.....	3
4. Encoding of a RWA Path Request.....	6
4.1. Wavelength Assignment (WA) Object.....	6
4.2. Wavelength Restriction Constraint TLV.....	8
4.2.1. Link Identifier sub-TLV.....	11
4.2.2. Wavelength Restriction Field sub-TLV.....	12
4.3. Signal processing capability restrictions.....	12
4.3.1. Signal Processing Exclusion XRO Sub-Object.....	13
4.3.2. IRO sub-object: signal processing inclusion.....	14
5. Encoding of a RWA Path Reply.....	14
5.1. Error Indicator.....	15
5.2. NO-PATH Indicator.....	15
6. Manageability Considerations.....	16
6.1. Control of Function and Policy.....	16
6.2. Information and Data Models, e.g. MIB module.....	16
6.3. Liveness Detection and Monitoring.....	16
6.4. Verifying Correct Operation.....	17
6.5. Requirements on Other Protocols and Functional Components.....	17
6.6. Impact on Network Operation.....	17
7. Security Considerations.....	17

8. IANA Considerations.....	17
9. Acknowledgments.....	17
10. References.....	18
10.1. Informative References.....	18
11. Contributors.....	20
Authors' Addresses.....	21
Intellectual Property Statement.....	21
Disclaimer of Validity.....	22

1. Terminology

This document uses the terminology defined in [RFC4655], and [RFC5440].

2. 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].

3. Introduction

[RFC4655] defines the PCE based Architecture and explains how a Path Computation Element (PCE) may compute Label Switched Paths (LSP) in Multiprotocol Label Switching Traffic Engineering (MPLS-TE) and Generalized MPLS (GMPLS) networks at the request of Path Computation Clients (PCCs). A PCC is said to be any network component that makes such a request and may be, for instance, an Optical Switching Element within a Wavelength Division Multiplexing (WDM) network. The PCE, itself, can be located anywhere within the network, and may be within an optical switching element, a Network Management System (NMS) or Operational Support System (OSS), or may be an independent network server.

The PCE communications Protocol (PCEP) is the communication protocol used between PCC and PCE, and may also be used between cooperating PCEs. [RFC4657] sets out the common protocol requirements for PCEP. Additional application-specific requirements for PCEP are deferred to separate documents.

This document provides the PCEP extensions for the support of Routing and Wavelength Assignment (RWA) in Wavelength Switched

Optical Networks (WSON) based on the requirements specified in [PCE-RWA].

WSON refers to WDM based optical networks in which switching is performed selectively based on the wavelength of an optical signal. In this document, it is assumed that wavelength converters require electrical signal regeneration. Consequently, WSONs can be transparent (A transparent optical network is made up of optical devices that can switch but not convert from one wavelength to another, all within the optical domain) or translucent (3R regenerators are sparsely placed in the network).

A LSC Label Switched Path (LSP) may span one or several transparent segments, which are delimited by 3R regenerators (typically with electronic regenerator and optional wavelength conversion). Each transparent segment or path in WSON is referred to as an optical path. An optical path may span multiple fiber links and the path should be assigned the same wavelength for each link. In such case, the optical path is said to satisfy the wavelength-continuity constraint. Figure 1 illustrates the relationship between a LSC LSP and transparent segments (optical paths).

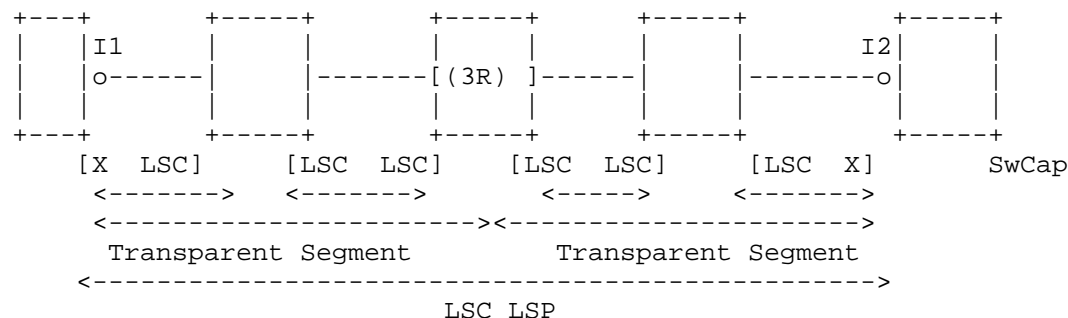


Figure 1 Illustration of a LSC LSP and transparent segments

Note that two optical paths within a WSON LSP need not operate on the same wavelength (due to the wavelength conversion capabilities). Two optical paths that share a common fiber link cannot be assigned the same wavelength. To do otherwise would result in both signals interfering with each other. Note that advanced additional multiplexing techniques such as polarization based multiplexing are not addressed in this document since the physical layer aspects are not currently standardized. Therefore, assigning the proper

wavelength on a lightpath is an essential requirement in the optical path computation process.

When a switching node has the ability to perform wavelength conversion, the wavelength-continuity constraint can be relaxed, and a LSC Label Switched Path (LSP) may use different wavelengths on different links along its route from origin to destination. It is, however, to be noted that wavelength converters may be limited due to their relatively high cost, while the number of WDM channels that can be supported in a fiber is also limited. As a WSON can be composed of network nodes that cannot perform wavelength conversion, nodes with limited wavelength conversion, and nodes with full wavelength conversion abilities, wavelength assignment is an additional routing constraint to be considered in all lightpath computation.

For example, within a translucent WSON, a LSC LSP may be established between interfaces I1 and I2, spanning 2 transparent segments (optical paths) where the wavelength continuity constraint applies (i.e. the same unique wavelength MUST be assigned to the LSP at each TE link of the segment). If the LSC LSP induced a Forwarding Adjacency / TE link, the switching capabilities of the TE link would be [X X] where $X < \text{LSC (PSC, TDM, ...)}$.

This document aligns with GMPLS extensions for PCEP [PCEP-GMPLS] for generic property such as label, label-set and label assignment noting that wavelength is a type of label. Wavelength restrictions and constraints are also formulated in terms of labels per [GEN-ENCODE].

The optical modulation properties, which are also referred to as signal compatibility, are already considered in signaling in [RWA-Encode] and [WSON-OSPF]. In order to improve the signal quality and limit some optical effects several advanced modulation processing are used. Those modulation properties contribute not only to optical signal quality checks but also constrain the selection of sender and receiver, as they should have matching signal processing capabilities. This document includes signal compatibility constraint as part of RWA path computation. That is, the signal processing capabilities (e.g., modulation and FEC) must be compatible between the sender and the receiver of the optical path across all optical elements.

This document, however, does not address optical impairments as part of RWA path computation. See [WSON-Imp] and [RSVP-Imp] for more information on optical impairments and GMPLS.

4. Encoding of a RWA Path Request

Figure 2 shows one typical PCE based implementation, which is referred to as Combined Process (R&WA). With this architecture, the two processes of routing and wavelength assignment are accessed via a single PCE. This architecture is the base architecture from which the requirements have been specified in [PCE-RWA] and the PCEP extensions that are going to be specified in this document based on this architecture.

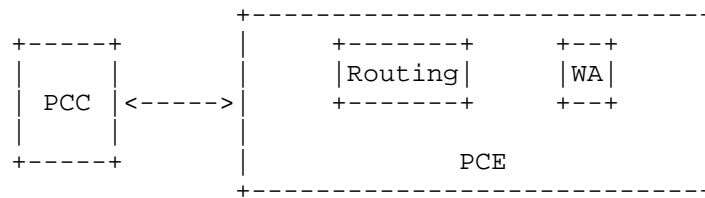


Figure 2 Combined Process (R&WA) architecture

4.1. Wavelength Assignment (WA) Object

The current RP object is used to indicate routing related information in a new path request per [RFC5440]. Since a new RWA path request involves both routing and wavelength assignment, the wavelength assignment related information in the request SHOULD be coupled in the path request.

Wavelength allocation can be performed by the PCE by different means:

- (a) By means of Explicit Label Control, in the sense that one (or two) allocated labels MAY appear after an interface route subobject.
- (b) By means of a Label Set, containing one or more allocated Labels, provided by the PCE.

Option (b) allows distributed label allocation (performed during signaling) to complete wavelength assignment.

Additionally, given a range of potential labels to allocate, the request SHOULD convey the heuristic / mechanism to the allocation.

The format of a PCReq message after incorporating the WA object is as follows:

```
<PCReq Message> ::= <Common Header>
```

```

    [<svec-list>]
    <request-list>

```

Where:

```

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

    <request>::= <RP>

    <ENDPOINTS>

    <WA>

    [other optional objects...]

```

If WA object is present in the request, the WA object MUST be encoded after the ENDPOINTS object.

The format of the Wavelength Assignment (WA) object body is as follows:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Flags                                     | O | M |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     Optional TLVs                                     |
|                                     //                                     //
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Figure 3 WA Object

- o Flags (32 bits)

The following new flags SHOULD be set

M (Mode - 1 bit): M bit is used to indicate the mode of wavelength assignment. When M bit is set to 1, this indicates that the label assigned by the PCE must be explicit. That is, the selected way to convey the allocated wavelength is by means of Explicit Label Control (ELC) [RFC4003] for each hop of a computed LSP. Otherwise, the label assigned by the PCE needs not be explicit (i.e., it can be suggested in the form of label set objects in the corresponding response, to allow distributed WA. In such case, the PCE MUST return a Label Set object as described in Section 2.2 of [Gen-Encode] in the response.

O (Order - 3 bits): O bit is used to indicate the wavelength assignment constraint in regard to the order of wavelength assignment to be returned by the PCE. This case is only applied when M bit is set to "explicit." The following indicators should be defined:

000 - Reserved

001 - Random Assignment

010 - First Fit (FF) in descending Order

011 - First Fit (FF) in ascending Order

100 - Last Fit (LF) in ascending Order

101 - Last Fit (LF) in descending Order

110 - Unspecified

111 - Reserved

4.2. Wavelength Restriction Constraint TLV

For any request that contains a wavelength assignment, the requester (PCC) MUST be able to specify a restriction on the wavelengths to be used. This restriction is to be interpreted by the PCE as a constraint on the tuning ability of the origination laser transmitter or on any other maintenance related constraints. Note that if the LSP LSC spans different segments, the PCE MUST have mechanisms to know the tunability restrictions of the involved wavelength converters / regenerators, e.g. by means of the TED either via IGP or NMS. Even if the PCE knows the tunability of the transmitter, the PCC MUST be able to apply additional constraints to the request.

[Ed note: Which PCEP Object will home this TLV is yet to be determined. Since this involves the end-point, The END-POINTS Object might be a good candidate to encode this TLV, which will be provided in a later revision.]

[Ed note: The current encoding assumes that tunability restriction applied to link-level.]

The TLV type is TBD, recommended value is TBD. This TLV MAY appear more than once to be able to specify multiple restrictions.

The TLV data is defined as follows:

```
<Wavelength Restriction Constraint> ::=
```

```
    <Action> <Format> <Reserved>
```

```
    (<Link Identifiers> <Wavelength Restriction>)...
```

Where

```
<Link Identifiers> ::=
```

```
    <Unnumbered IF ID> | <IPv4 Address> | <IPv6 Address>
```

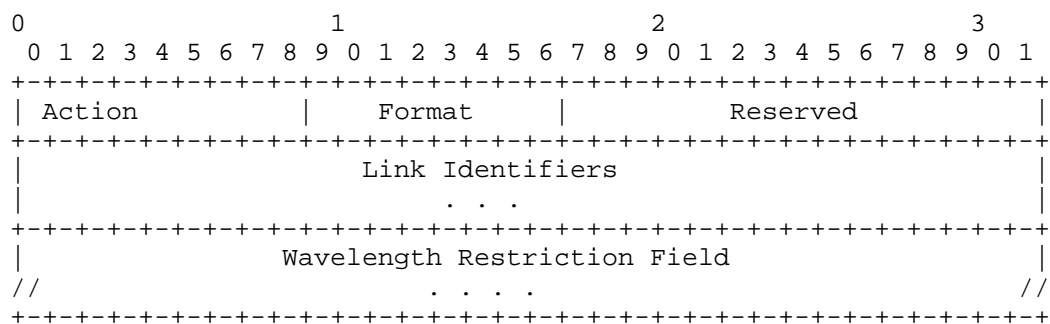


Figure 4 Wavelength Restriction

- o Action: 8 bits

- 0 - Inclusive List indicates that one or more link identifiers are included in the Link Set. Each identifies a separate link that is part of the set.

- 1 - Inclusive Range indicates that the Link Set defines a range of links. It contains two link identifiers. The first identifier indicates the start of the range (inclusive). The second identifier indicates the end of the range (inclusive). All links with numeric values between the bounds are considered to be part of the set. A value of zero in either position indicates that there is no bound on the corresponding portion of the range. Note that the Action field can be set to 0 when unnumbered link identifier is used.

Note that "interfaces" such as those discussed in the Interfaces MIB [RFC2863] are assumed to be bidirectional.

- o Format: The format of the link identifier (8 bits)

- 0 -- Unnumbered Link Identifier
 - 1 -- Local Interface IPv4 Address
 - 2 -- Local Interface IPv6 Address
 - Others TBD.

Note that all link identifiers in the same list must be of the same type.

- o Reserved: Reserved for future use (16 bits)

- o Link Identifiers: Identifies each link ID for which restriction is applied. The length is dependent on the link format. See the following section for Link Identifier encoding.

4.2.1. Link Identifier sub-TLV

The link identifier field can be an IPv4, IPv6 or unnumbered interface ID.

<Link Identifier> ::=

<IPv4 Address> | <IPv6 Address> | <Unnumbered IF ID>

The encoding of each case is as follows:

IPv4 prefix Sub-TLV

```

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 = 1           | IPv4 address (4 bytes)                |
+-----+-----+-----+-----+-----+-----+-----+-----+
| IPv4 address (continued) | Prefix Length | Attribute      |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

IPv6 prefix Sub-TLV

```

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 = 2           | IPv6 address (16 bytes)              |
+-----+-----+-----+-----+-----+-----+-----+-----+
| IPv6 address (continued) |                               |
+-----+-----+-----+-----+-----+-----+-----+-----+
| IPv6 address (continued) |                               |
+-----+-----+-----+-----+-----+-----+-----+-----+
| IPv6 address (continued) |                               |
+-----+-----+-----+-----+-----+-----+-----+-----+
| IPv6 address (continued) | Prefix Length | Attribute      |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

Unnumbered Interface ID Sub-TLV

```

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 = 4      |      Reserved      |      Attribute      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     TE Node ID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Interface ID                                    |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

4.2.2. Wavelength Restriction Field sub-TLV

The Wavelength Restriction Field of the wavelength restriction TLV is encoded as a Label Set field as specified in [GEN-Encode] section 2.2, as shown below, with base label encoded as a 32 bit LSC label, defined in [RFC6205]. See [RFC6205] for a description of Grid, C.S, Identifier and n, as well as [GEN-Encode] for the details of each action.

```

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

+-----+-----+-----+-----+-----+-----+-----+-----+
| Action|      Num Labels      |      Length      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|Grid | C.S |      Identifier  |      n      |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      Additional fields as necessary per action      |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

4.3. Signal processing capability restrictions

Path computation for WSON include the check of signal processing capabilities, those capability MAY be provided by the IGP, however this is not a MUST. Moreover, a PCC should be able to indicate additional restrictions for those signal compatibility, either on the endpoint or any given link.

The supported signal processing capabilities are the one described in [RWA-Info]:

Optical Interface Class List

Bit rate

Client signal

The Bit-rate restriction is already expressed in [PCEP-GMPLS] in the GENERALIZED-BANDWIDTH object.

The client signal information can be expressed using the REQ-ADAP-CAP object from the [PCEP-Layer].

In order to support the Optical Interface Class information a new TLV are introduced as endpoint-restriction in the END-POINTS type Generalized endpoint:

Optical Interface Class List TLV

The END-POINTS type generalized endpoint is extended as follow:

```
<endpoint-restrictions> ::= <LABEL-REQUEST>
                               <label-restriction-list>
                               [<signal-compatibility-restriction>...]
```

Where

```
signal-compatibility-restriction ::=
    <Optical Interface Class List>
```

The encoding for Optical Interface Class List is described in Section 5.2 of [RWA-Encode].

4.3.1. Signal Processing Exclusion XRO Sub-Object

The PCC/PCE should be able to exclude particular types of signal processing along the path in order to handle client restriction or multi-domain path computation.

In order to support the exclusion a new XRO sub-object is defined: the signal processing exclusion:

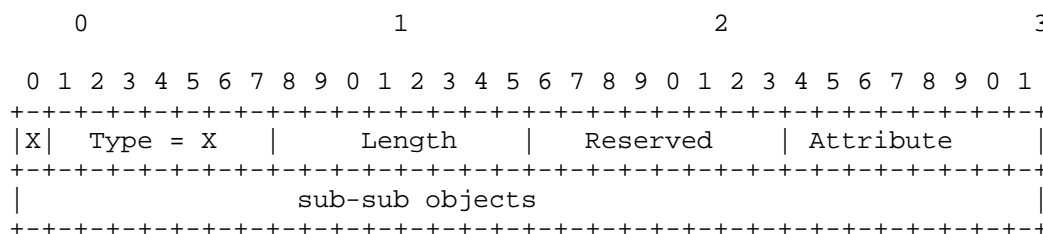


Figure 5 Signaling Processing XRO Sub-Object

The Attribute field indicates how the exclusion sub-object is to be interpreted. The Attribute can only be 0 (Interface) or 1 (Node).

The sub-sub objects are encoded as in RSVP signaling definition [WSON-Sign].

4.3.2. IRO sub-object: signal processing inclusion

Similar to the XRO sub-object the PCC/PCE should be able to include particular types of signal processing along the path in order to handle client restriction or multi-domain path computation.

This is supported by adding the sub-object "processing" defined for ERO in [WSON-Sign] to the PCEP IRO object.

5. Encoding of a RWA Path Reply

The ERO is used to encode the path of a TE LSP through the network. The ERO is carried within a given path of a PCEP response, which is in turn carried in a PCRep message to provide the computed TE LSP if the path computation was successful. The preferred way to convey the allocated wavelength is by means of Explicit Label Control (ELC) [RFC4003].

In order to encode wavelength assignment, the Wavelength Assignment (WA) Object needs to be employed to be able to specify wavelength assignment. Since each segment of the computed optical path is associated with wavelength assignment, the WA Object should be aligned with the ERO object.

Encoding details will be provided further revisions and will be aligned as much as possible with [WSON-Sign] and [LSPA-ERO]

5.1. Error Indicator

To indicate errors associated with the RWA request, a new Error Type (TDB) and subsequent error-values are defined as follows for inclusion in the PCEP-ERROR Object:

A new Error-Type (TDB) and subsequent error-values are defined as follows:

Error-Type=TBD; Error-value=1: if a PCE receives a RWA request and the PCE is not capable of processing the request due to insufficient memory, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=TDB) and an Error-value(Error-value=1). The PCE stops processing the request. The corresponding RWA request MUST be cancelled at the PCC.

Error-Type=TBD; Error-value=2: if a PCE receives a RWA request and the PCE is not capable of RWA computation, the PCE MUST send a PCErr message with a PCEP-ERROR Object (Error-Type=15) and an Error-value (Error-value=2). The PCE stops processing the request. The corresponding RWA computation MUST be cancelled at the PCC.

5.2. NO-PATH Indicator

To communicate the reason(s) for not being able to find RWA for the path request, the NO-PATH object can be used in the PCRep message. The format of the NO-PATH object body is defined in [RFC5440]. The object may contain a NO-PATH-VECTOR TLV to provide additional information about why a path computation has failed.

Two new bit flags are defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object.

Bit TDB: When set, the PCE indicates no feasible route was found that meets all the constraints associated with RWA.

Bit TDB: When set, the PCE indicates that no wavelength was assigned to at least one hop of the route in the response.

Bit TDB: When set, the PCE indicate that no path was found satisfying the signal compatibility constraints.

6. Manageability Considerations

Manageability of WSON Routing and Wavelength Assignment (RWA) with PCE must address the following considerations:

6.1. Control of Function and Policy

In addition to the parameters already listed in Section 8.1 of [PCEP], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCC:

- The ability to send a WSON RWA request.

In addition to the parameters already listed in Section 8.1 of [PCEP], a PCEP implementation SHOULD allow configuring the following PCEP session parameters on a PCE:

- The support for WSON RWA.

- A set of WSON RWA specific policies (authorized sender, request rate limiter, etc).

These parameters may be configured as default parameters for any PCEP session the PCEP speaker participates in, or may apply to a specific session with a given PCEP peer or a specific group of sessions with a specific group of PCEP peers.

6.2. Information and Data Models, e.g. MIB module

Extensions to the PCEP MIB module defined in [PCEP-MIB] should be defined, so as to cover the WSON RWA information introduced in this document. A future revision of this document will list the information that should be added to the MIB module.

6.3. Liveness Detection and Monitoring

Mechanisms defined in this document do not imply any new liveness detection and monitoring requirements in addition to those already listed in section 8.3 of [RFC5440].

6.4. Verifying Correct Operation

Mechanisms defined in this document do not imply any new verification requirements in addition to those already listed in section 8.4 of [RFC5440]

6.5. Requirements on Other Protocols and Functional Components

The PCE Discovery mechanisms ([RFC5089] and [RFC5088]) may be used to advertise WSON RWA path computation capabilities to PCCs.

6.6. Impact on Network Operation

Mechanisms defined in this document do not imply any new network operation requirements in addition to those already listed in section 8.6 of [RFC5440].

7. Security Considerations

This document has no requirement for a change to the security models within PCEP [PCEP]. However the additional information distributed in order to address the RWA problem represents a disclosure of network capabilities that an operator may wish to keep private. Consideration should be given to securing this information.

8. IANA Considerations

A future revision of this document will present requests to IANA for codepoint allocation.

9. Acknowledgments

The authors would like to thank Adrian Farrel for many helpful comments that greatly improved the contents of this draft.

This document was prepared using 2-Word-v2.0.template.dot.

10. References

10.1. Informative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3471] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", RFC 3471, January 2003.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.
- [RFC3477] Kompella, K. and Y. Rekhter, "Signalling Unnumbered Links in Resource ReSerVation Protocol - Traffic Engineering (RSVP-TE)", RFC 3477, January 2003.
- [RFC4003] Berger, L., "GMPLS Signaling Procedure for Egress Control", RFC 4003, February 2005.
- [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.
- [RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) communication Protocol", RFC 5440, March 2009.
- [PCEP-GMPLS] Margaria, et al., "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions, work in progress.
- [LSPA-ERO] Margaria, et al., "LSP Attribute in ERO", draft-margaria-ccamp-lsp-attribute-ero, work in progress.
- [PCEP-Layer] Oki, Takeda, Le Roux, and Farrel, "Extensions to the Path Computation Element communication Protocol (PCEP) for Inter-Layer MPLS and GMPLS Traffic Engineering", draft-ietf-pce-inter-layer-ext, work in progress.

- [RFC6163] Lee, Y. and Bernstein, G. (Editors), and W. Imajuku, "Framework for GMPLS and PCE Control of Wavelength Switched Optical Networks", RFC 6163, March 2011.
- [PCE-RWA] Lee, Y., et. al., "PCEP Requirements for WSON Routing and Wavelength Assignment", draft-ietf-pce-wson-routing-wavelength, work in progress.
- [RFC6205] Tomohiro, O. and D. Li, "Generalized Labels for Lambda-Switching Capable Label Switching Routers", RFC 6205, January, 2011.
- [WSON-Sign] Bernstein et al, "Signaling Extensions for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signaling, work in progress.
- [WSON-OSPF] Lee and Bernstein, "OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signal-compatibility-ospf, work in progress.
- [RWA-Info] Bernstein and Lee, "Routing and Wavelength Assignment Information Model for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-info, work in progress.
- [RWA-Encode] Bernstein and Lee, "Routing and Wavelength Assignment Information Encoding for Wavelength Switched Optical Networks", draft-ietf-ccamp-rwa-wson-encode, work in progress.
- [GEN-Encode] Bernstein and Lee, "General Network Element Constraint Encoding for GMPLS Controlled Networks", draft-ietf-ccamp-general-constraint-encode, work in progress.
- [WSON-Imp] Y. Lee, G. Bernstein, D. Li, G. Martinelli, "A Framework for the Control of Wavelength Switched Optical Networks (WSON) with Impairments", draft-ietf-ccamp-wson-impairments, work in progress.
- [RSVP-Imp] agraz, "RSVP-TE Extensions in Support of Impairment Aware Routing and Wavelength Assignment in Wavelength Switched Optical Networks (WSONs)", draft-agraz-ccamp-wson-impairment-rsvp, work in progress.
- [OSPF-Imp] Bellagamba, et al., "OSPF Extensions for Wavelength Switched Optical Networks (WSON) with Impairments", draft-eb-ccamp-ospf-wson-impairments, work in progress.

11. Contributors

Authors' Addresses

Young Lee, Editor
Huawei Technologies
1700 Alma Drive, Suite 100
Plano, TX 75075, USA
Phone: (972) 509-5599 (x2240)
Email: leeyoung@huawei.com

Ramon Casellas, Editor
CTTC PMT Ed B4 Av. Carl Friedrich Gauss 7
08860 Castelldefels (Barcelona)
Spain
Phone: (34) 936452916
Email: ramon.casellas@cttc.es

Fatai Zhang
Huawei Technologies
Email: zhangfatai@huawei.com

Cyril Margaria
Nokia Siemens Networks
St Martin Strasse 76
Munich, 81541
Germany
Phone: +49 89 5159 16934
Email: cyril.margaria@nsn.com

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
C/ Emilio Vargas 6
Madrid, 28043
Spain
Phone: +34 91 3374013
Email: ogondio@tid.es

Greg Bernstein
Grotto Networking
Fremont, CA, USA
Phone: (510) 573-2237
Email: gregb@grotto-networking.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 17, 2013

F. Zhang, Ed.
Q. Zhao
Huawei
O. Gonzalez de Dios, Ed.
Telefonica I+D
R. Casellas
CTTC
D. King
Old Dog Consulting
July 16, 2012

Extensions to Path Computation Element Communication Protocol (PCEP) for
Hierarchical Path Computation Elements (PCE)
draft-zhang-pce-hierarchy-extensions-02

Abstract

The Hierarchical Path Computation Element (H-PCE) architecture, defined in the companion framework document [I-D.ietf-pce-hierarchy-fwk], facilitates to obtain optimum end-to-end, multi-domain paths when the sequence of domains is not known in advance. Such H-PCE architecture allows the selection of an optimum domain sequence and, through the use of a hierarchical relationship between domains, derive the optimum end-to-end path.

This document defines the Path Computation Element Protocol (PCEP) extensions for the purpose of implementing Hierarchical PCE procedures which are described the aforementioned document.

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 January 17, 2013.

Copyright Notice

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	4
1.1. Terminology	4
1.2. Requirements Language	5
2. PCEP Protocol Extensions Requirements	5
2.1. PCEP Requests	5
2.1.1. PCEP Request Qualifiers	5
2.1.2. New Objective Functions	6
2.1.3. New Metrics	6
2.2. Communication to the parent PCE of the Domain Conectivity information	7
2.3. Parent PCE Capability Discovery	8
2.4. PCE Domain and PCE ID Discovery	8
2.5. Error Case Handling	8
2.6. Determination of destination domain	9
3. PCEP Extensions	9
3.1. Extensions to OPEN object	9
3.1.1. OF Codes	9
3.1.2. OPEN Object Flags	10
3.1.3. Domain-ID TLV	10
3.1.4. PCE-ID TLV	11
3.1.5. Procedures	12
3.2. Extensions to RP object	12
3.2.1. RP Object Flags	12
3.2.2. Domain-ID TLV	12
3.2.3. Procedures	13
3.3. Extensions to Metric object	13
3.4. Extensions to NOTIFICATION object	13
3.4.1. Notification Types	13
3.4.2. Inter-domain Link TLV	14
3.4.3. Inter-domain Node TLV	15
3.4.4. Domain-ID TLV	16
3.4.5. PCE-ID TLV	16

3.4.6. Reachability TLV	16
3.4.7. Procedures	17
3.5. Extensions to PCEP-ERROR object	17
3.5.1. Hierarchy PCE Error-Type	17
3.5.2. Procedures	18
4. Manageability Considerations	18
5. IANA Considerations	18
5.1. Objective Function (OF) codes	18
5.2. OPEN Object Flags	18
5.3. RP Object Flags	18
5.4. PCEP TLVs	18
5.5. PCEP NOTIFICATION types	19
5.6. PCEP PCEP-ERROR types	19
6. Security Considerations	19
7. Contributing Authors	19
8. Acknowledgments	19
9. References	19
9.1. Normative References	19
9.2. Informative References	20
Authors' Addresses	20

1. Introduction

[I-D.ietf-pce-hierarchy-fwk] describes a Hierarchical PCE (H-PCE) architecture which can be used for computing end-to-end paths for inter-domain MPLS Traffic Engineering (TE) and GMPLS Label Switched Paths (LSPs). In the hierarchical PCE architecture, the parent PCE can compute a multi-domain path based on the domain connectivity information and each child PCE is able to compute the intra-domain path based on its domain topology information. The end-to-end domain path computing procedures can be abstracted as follows:

- o A path computation client (PCC) requests its own child PCE the computation of an inter-domain path.
- o The child PCE forwards the request to the parent PCE.
- o The parent PCE computes one or multiple domain paths from the ingress domain to the egress domain.
- o The parent PCE sends the intra-domain path computation requests (between the domain border nodes) to the child PCEs which are responsible for the domains along the domain path(s).
- o The child PCEs return the intra-domain paths to the parent PCE.
- o The parent PCE constructs the end-to-end inter-domain path based on the intra-domain paths
- o The parent PCE returns the inter-domain path to the child PCE.
- o The child PCE forwards the inter-domain path to the PCC.

Alternatively, the parent PCE, instead of building the complete end-to-end path, can reply with the sequence of domains and later standard procedures, like BRPC, can be applied.

This document defines the PCEP extensions for the purpose of implementing Hierarchical PCE procedures, which are described in [I-D.ietf-pce-hierarchy-fwk].

The document also uses a number of editor notes to describe options and alternative solutions. These options and notes will be removed before publication once agreement is reached.

1.1. Terminology

This document uses the terminology defined in [RFC4655], [RFC5440] and the additional terms defined in section 1.4 of

[I-D.ietf-pce-hierarchy-fwk].

1.2. Requirements Language

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

2. PCEP Protocol Extensions Requirements

This section compiles the set of requirements of the extensions needed in the PCEP protocol to support the H-PCE architecture and procedures.

2.1. PCEP Requests

The PCReq messages are used by a PCC or PCE to make a path computation request to a PCE. In order to achieve the full functionality of the H-PCE procedures, some extensions are needed in the pcReq messages:

- o Qualify PCE Requests
- o New Objective Functions
- o New Metrics

2.1.1. PCEP Request Qualifiers

As described in section 5.8.1 of [I-D.ietf-pce-hierarchy-fwk], the H-PCE architecture will introduce new request qualifications as follows:

- o It MUST be possible for a child PCE to indicate that a request it sends to a parent PCE should be satisfied by a domain sequence only, that is, not by a full end-to-end path. This allows the child PCE to initiate a per-domain [RFC5152] or a backward recursive path computation (BRPC) [RFC5441].
- o A parent PCE needs to be able to ask a child PCE whether a particular node address (the destination of an end-to-end path) is present in the domain that the child PCE serves.
- o As stated in [I-D.ietf-pce-hierarchy-fwk], section 5.5, if a PCC knows the egress domain, it can supply this information as the path computation request. It SHOULD be possible to specify the destination domain information in a PCEP request, if it is known.

To meet the above requirements, the PCEP PCReq message should be extended.

2.1.1.2. New Objective Functions

For inter-domain path computation, there are two new objective functions which are defined in section 1.3.1 and 5.1 of [I-D.ietf-pce-hierarchy-fwk]:

- o Minimize the number of domains crossed.
- o Disallow domain re-entry.[Editor's note: Disallow domain re-entry may not be an objective function, but an option in the request]

During the PCEP session establishment procedure, the parent PCE needs to be capable of indicating the objective functions (OF) capability in the Open message. This information can be, in turn, announced by child PCEs and used for selecting the PCE when a PCC want a path that satisfies a certain inter-domain objective function.

When a PCC requests a PCE to compute an inter-domain path, the PCC needs also to be capable of indicating the new objective functions for inter-domain path. Note that a given PCE may act as a regular PCE and as a parent PCE.

For the reasons described above, new OF codes need to be defined for the new inter-domain objective functions. Then the PCE can notify its new inter-domain objective functions to the PCC by carrying them in the OF-list TLV which is carried in the OPEN object. The PCC can specify which objective function code to use, which is carried in the OF object when requesting a PCE to compute an inter-domain path.

The proposed solutions may need to differentiate between the OF code that is requested at the parent level and the OF code that is requested at the intra-domain (child) level.

A parent PCE needs to be able to insure homogeneity when applying OF codes for the intra-domain requests.

2.1.1.3. New Metrics

For inter-domain path computation, there are several path metrics of interest [Editor's note: Current framework only mentions metric objectives. The metric itself should be also defined]:

- o Domain count (number of domains crosses).

- o Border Node count

A PCC may be able to limit the number of domains crossed by applying a limit on the metric.

2.2. Communication to the parent PCE of the Domain Connectivity information

A parent PCE maintains a domain topology map of the child domains and their interconnectivity, as mentioned in section 4.4 of [I-D.ietf-pce-hierarchy-fwk]. Consequently, a parent PCE maintains a Traffic Engineering Database (TED) for the parent domain.

The parent PCE TED may be administratively configured or learnt from information received from the child PCEs. Thus, entities from the child domains (such as the child PCEs) can convey its neighbour information to the parent PCE to maintain the parent TED. One possible option is to use a separate instance of an IGP running within the parent domain in which parent and child PCEs establish an IGP adjacency. Alternatively, as mentioned in section 4.8.4 of [I-D.ietf-pce-hierarchy-fwk], a child PCE may forward the its neighbour domain connectivity (inter-domain links or ABRs) to the parent PCE, for example within PCNTf messages or any other mechanisms, without an IGP adjacency.

There are two types of domain borders for providing the domain connectivity information:

- o Domain border is a TE link, e.g. the inter-AS TE link which connects two ASs.
- o Domain border is a node, e.g. the IGP ABR which connects two IGP areas.

The information that would be exchanged for inter-AS TE links includes:

- o Identifier of advertising child PCE
- o Identifier of PCE's domain
- o Identifier of the link
- o TE properties of the link (metrics, bandwidth)
- o Other properties of the link (technology-specific)

- o Identifier of link end-points
- o Identifier of adjacent domain

For the ABR, the following information needs to be notified to the parent PCE:

- o Identifier of the ABR.
- o Identifier of the IGP Area IDs.

[Editor's Note: Further discussion of the discovery mechanism based on the requirements of section 4.8.4 of [I-D.ietf-pce-hierarchy-fwk] and scope will be discussed in later versions of this document. Note that using PCNtf messages will require PCEP Protocol extensions.]

2.3. Parent PCE Capability Discovery

Parent/Child relationships are likely to be configured. However, as mentioned in [I-D.ietf-pce-hierarchy-fwk], it helps network operations that the parent PCE indicates its H-PCE capability and that the PCC indicates its intention of using parent PCE capabilities. Thus, during the PCEP session establishment procedure, the child PCE needs to be capable of indicating to the parent PCE whether it requests the parent PCE capability or not. Also, during the PCEP session establishment procedure, the parent PCE needs to be capable of indicating whether its parent capability can be provided or not.

2.4. PCE Domain and PCE ID Discovery

A PCE domain is a single domain with an associated PCE. It is possible for a PCE to manage multiple domains. The PCE domain may be an IGP area or AS.

The PCE ID is an IPv4 and/or IPv6 address that is used to reach the parent/child PCE. It is RECOMMENDED to use an address that is always reachable if there is any connectivity to the PCE.

The PCE ID information and PCE domain identifiers may be provided during the PCEP session establishment procedure or the domain connectivity information collection procedure.

2.5. Error Case Handling

A PCE that is capable of acting as a parent PCE might not be configured or willing to act as the parent for a specific child PCE. This fact could be determined when the child sends a PCReq that

requires parental activity (such as querying other child PCEs), and could result in a negative response in a PCEP Error (PCErr) message and indicate the hierarchy PCE error types.

2.6. Determination of destination domain

The PCC that asks for an inter-domain path computation is aware of the identity of the destination node by definition. If the PCC also knows the egress domain to which the destination node belongs to, it can supply the information as part of the path computation request. Otherwise, it must be determined by the parent PCE.

The parent PCE can query the child PCEs to obtain the destination domain, using the PCEP Request Qualifiers mentioned before. Alternatively, the child PCEs can forward in PCNtf the set of reachable addresses of the domain. [Editor's note: This point requires further ellaboration]

3. PCEP Extensions

3.1. Extensions to OPEN object

3.1.1. OF Codes

There are two new OF codes defined here for H-PCE:

- o MTD
 - * Name: Minimize the number of Transit Domains
 - * Objective Function Code: (to be assigned by IANA, recommended 12)
 - * Description: Find a path P such that passes through the least ransit domains.
- o DDR
 - * Name: Disallow Domain Re-entry (DDR)
 - * Objective Function Code: (to be assigned by IANA, recommended 13)
 - * Description: Find a path P such that does not entry a domain more than once.

3.1.2. OPEN Object Flags

There are two OPEN object flags defined here for H-PCE:

- o Parent PCE request bit (to be assigned by IANA, recommended bit 0): if set it means the child PCE wishes to use the peer PCE as a parent PCE.
- o Parent PCE indication bit (to be assigned by IANA, recommended bit 1): if set it means the PCE can be used as a parent PCE by the peer PCE.
- o [Editors Note: It is possible that a parent PCE will also act as a child PCE]

3.1.3. Domain-ID TLV

The type of Domain-ID TLV is to be assigned by IANA (recommended 7). The length is 8 octets. The format of this TLV is defined below:

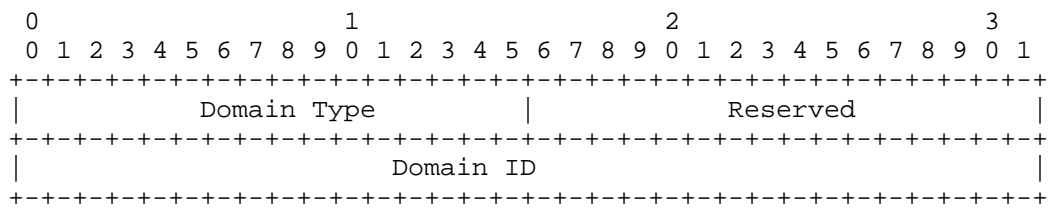


Figure 1: Domain-ID TLV

Domain Type (8 bits): Indicates the domain type. There are two types of domain defined currently:

- o Type=1: the Domain ID field carries an IGP Area ID.
- o Type=2: the Domain ID field carries an AS number.

Domain ID (32 bits): Indicates an IGP Area ID or AS number.

An AS number may be 2 or 4 bytes long. For 2-byte AS numbers, the AS value is left-padded with 0.

[Editor's note: it may be necessary to support 64 bit domain IDs.]

[Editor's note: draft-dhody-pce-pcep-domain-sequence, section 3.2 deals with the encoding of domain sequences, using ERO-subobjects. Work is ongoing to define domain identifiers for OSPF-TE areas, IS-IS area (which are variable sized), 2-byte and 4-byte AS number, and any

other domain that may be defined in the future. It uses RSVP-TE subobject discriminators, rather than new type 1/ type 2. A domain sequence may be encoded as a route object. The "VALUE" part of the TLV could follow common RSVP-TE subobject format:

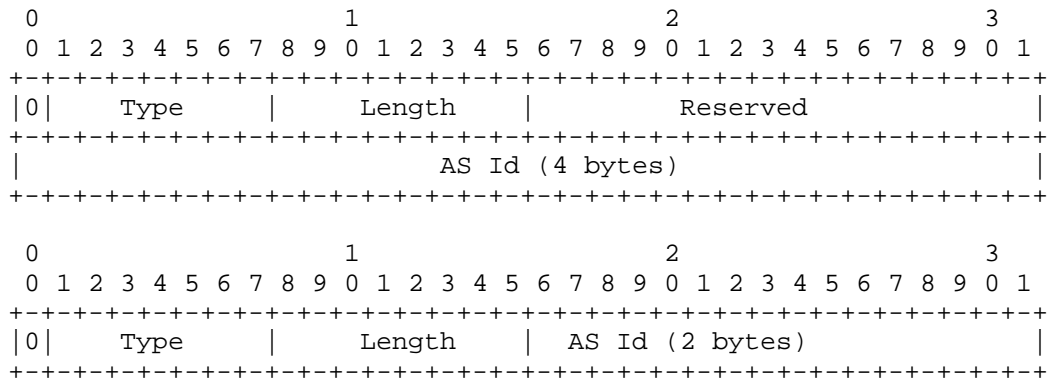


Figure 2: Alternative Domain-ID TLV

3.1.4. PCE-ID TLV

The type of PCE-ID TLV is to be assigned by IANA (recommended 8). The length is 8. The format of this TLV is defined below:

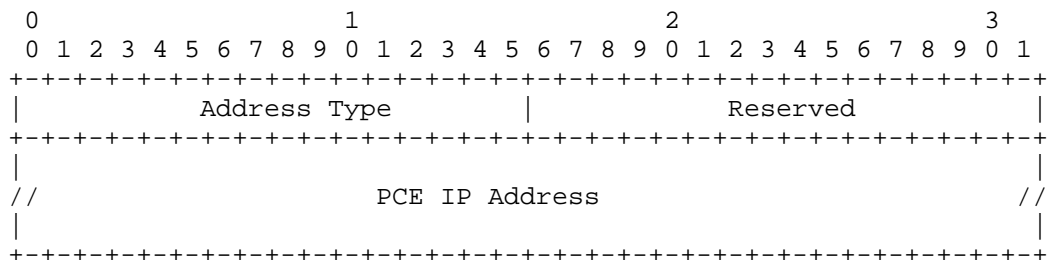


Figure 3: PCE-ID TLV

Address Type (16 bits): Indicates the address type of PCE IP Address. 1 means IPv4 address type, 2 means IPv6 address type.

PCE IP Address: Indicates the reachable address of a PCE.

[Editor's note: [RFC5886] already defines the PCE-ID object. If a semantically equivalent PCE-ID TLV is needed (to avoid modifying message grammars to include the object), it can align with the PCEP object: in any case, the length (4 / 16 bytes) can be used to know whether it is an IPv4 or an IPv6 PCE, the address type is not

needed.]

3.1.5. Procedures

The OF codes defined in this document can be carried in the OF-list TLV of the OPEN object. If the OF-list TLV carries the OF codes, it means that the PCE is capable of implementing the corresponding objective functions. This information can be used for selecting a proper parent PCE when a child PCE wants to get a path that satisfies a certain objective function.

If a child PCE wants to use the peer PCE as a parent, it can set the parent PCE request bit in the OPEN object carried in the Open message during the PCEP session creation procedure. If the peer PCE does not want to provide the parent function to the child PCE, it must send a PCERR message to the child PCE and clear the parent PCE indication bit in the OPEN object.

If the parent PCE can provide the parent function to the peer PCE, it may set the parent PCE indication bit in the OPEN object carried in the Open message during the PCEP session creation procedure.

The PCE may also report its PCE ID and list of domain ID to the peer PCE by specifying them in the PCE-ID TLV and List of Domain-ID TLVs in the OPEN object carried in the Open message during the PCEP session creation procedure.

3.2. Extensions to RP object

3.2.1. RP Object Flags

Domain Path Request bit (to be assigned by IANA, recommended bit 17): if set it means the child PCE wishes to get the domain sequence.

Destination Domain Query bit (to be assigned by IANA, recommended bit 16): if set it means the parent PCE wishes to get the destination domain ID.

3.2.2. Domain-ID TLV

The format of this TLV is defined in Section 3.1.3. This TLV can be carried in an OPEN object to indicate a (list of) managed domains, or carried in a RP object to indicate the destination domain ID when a child PCE responds to the parent PCE's destination domain query by a PCRep message.

[Editors note. In some cases, the Parent PCE may need to allocate a node which is not necessarily the destination node.]

3.2.3. Procedures

If a child PCE only wants to get the domain sequence for a multi-domain path computation from a parent PCE, it can set the Domain Path Request bit in the RP object carried in a PCReq message. The parent PCE which receives the PCReq message tries to compute a domain sequence for it. If the domain path computation succeeds the parent PCE sends a PCRep message which carries the domain sequence in the ERO to the child PCE. The domain sequence is specified as AS or AREA ERO sub-objects (type 32 for AS [RFC3209] or a to-be-defined IGP area type). Otherwise it sends a PCReq message which carries the NO-PATH object to the child PCE.

The parent PCE can set the Destination Domain Query bit in a PCReq message to query the destination (which is specified in the END-POINTS objects) domain ID from a child PCE. If the child PCE knows the destination(s) domain ID, it sends a PCRep message to the parent PCE and specifies the domain ID in the Domain-ID TLV which is carried in the RP object. Otherwise it sends a PCRep message with a NO-PATH object to the parent PCE.

3.3. Extensions to Metric object

There are two new metrics defined here for H-PCE:

- o Domain count (number of domains crosses).
- o Border Node Count

3.4. Extensions to NOTIFICATION object

Because there will not be too many PCEP sessions between the child PCE(s) and parent PCE, it is recommended that the PCEP sessions between them keeping alive all the time. Then the child PCE can report all of the domain connectivity information to the parent PCE when the PCEP session is established successfully. It can also notify the parent PCE to update or delete the domain connectivity information when it detects the changes.

3.4.1. Notification Types

There is a new notification type defined in this document:

- o Domain Connectivity Information notification-type (to be assigned by IANA, recommended 3).
- o Notification-value=1: sent from the parent to the child to query all of the domain connectivity information maintained by the child

PCE.

- o Notification-value=2: sent from the child to the parent to update the domain connectivity information maintained by the child PCE.
- o Notification-value=3: sent from the child to the parent to delete the domain connectivity information maintained by the child PCE.

3.4.2. Inter-domain Link TLV

IGP in each neighbor domain can advertise its inter-domain TE link capabilities [RFC5316],[RFC5392]. This information can be collected by the child PCEs and forwarded to the parent PCE. PCEP Inter-domain Link TLV is used for carrying the inter-domain TE link attributes for this purpose. Each Inter-domain Link TLV can carry the attributes of one inter-domain link at the most.

The type of Inter-domain Link TLV is to be assigned by IANA (recommended 9). The length is variable. The format of this TLV is defined below:

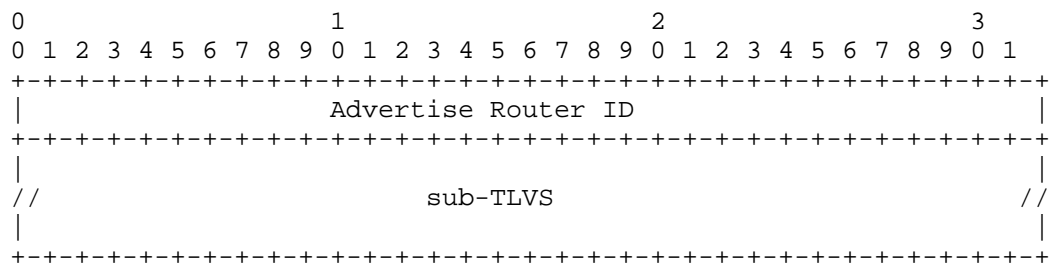


Figure 4: Inter-domain Link TLV

Editor's note: evaluate other possibilities regarding the wrapping and encoding (LSAs / LSUs). Other fields may be needed, such as LSA age (max age methods can be used to "withdraw" or remove a link). Sub-TLVs may need to be defined in the context of a Link TLV (top TLV).

Advertise Router ID (32 bits): indicates the router ID which advertises the TE LSA or LSP.

Sub-TLVs: the OSPF sub-TLVs for a TE link which defined in [RFC5392] and other associated OSPF RFCs. It is noted that if the IGP is IS-IS for the child domain the sub-TLVs must be converted to the OSPF sub-TLVs format when sending this information to the parent PCE through PCEP PCNtf message.

Each inter-domain link is identified by the combination of advertise router ID and the link local IP address or link local unnumbered identifier. The PCNTf message which is used for notifying the parent PCE to update or delete a inter-domain link must contain the information identifies a TE link exclusively.

3.4.3. Inter-domain Node TLV

The Inter-domain Node TLV carries only the two adjacent domain ID and the router (IGP ABR) ID.

he type of Inter-domain Node Information TLV is to be assigned by IANA (recommended 10). The length is variable . The format of this TLV is defined below:

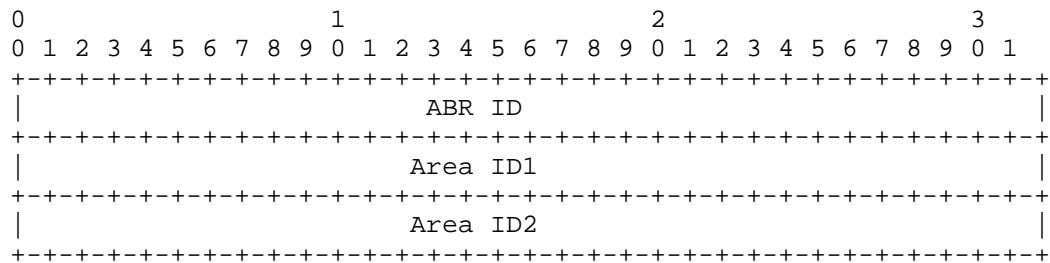


Figure 5: Inter-domain Node TLV

ABR ID (32 bits): indicates the domain border router ID.

Area ID1 and Area ID2 (32 bits): indicates the two neighbor area IDs.

Editor's note (1): a node may be an inter-domain node for more than just 2 areas, the encoding is wrong, unless we explicitly state that this TLV can be repeated and we give an example. Alternatively, we can use the generic concept of "domain id" as introduced earlier, to avoid the restriction of 4 byte areas only.

Editor's note (2): do we homogenize so we also have a Advertising Router ID? would it be different from the ABR id?

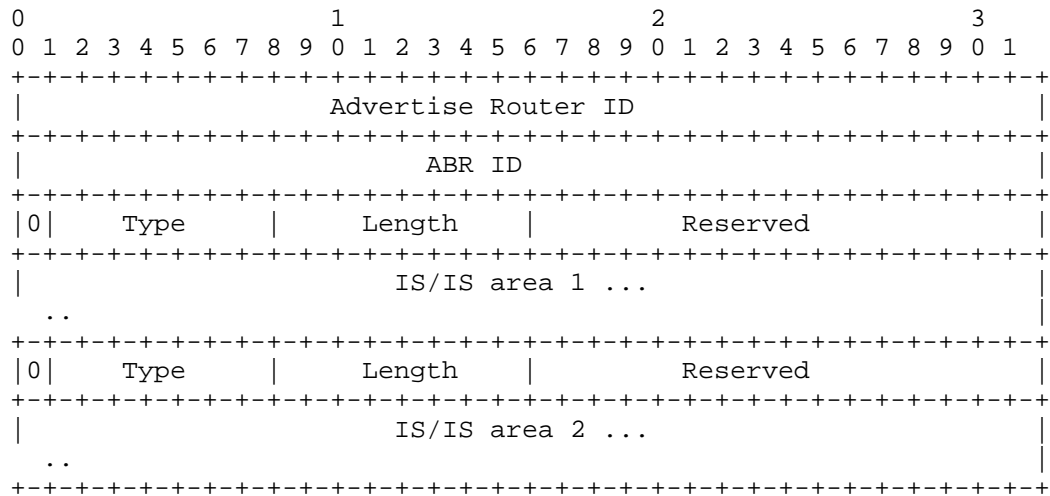


Figure 6: Alternative Inter-domain Node TLV

3.4.4. Domain-ID TLV

The format of this TLV is defined in Section 3.1.3. This TLV can be carried in a NOTIFICATION object to indicate the domain ID of the PCE who sends the PCNtf message.

[Editors note: A PCE may be responsible for several domains, it may be beneficial to use a list of TLVs]

3.4.5. PCE-ID TLV

The format of this TLV is defined in Section 3.1.4. This TLV can be carried in a NOTIFICATION object to indicate the PCE ID of the PCE who sends the PCNtf message.

3.4.6. Reachability TLV

The reachability TLV carries information of the set of end-points reachable in a given domain.

The format of the TLV is a list of IPv4 Prefix, IPv6 Prefix, AS and unnumbered Interface ERO subobjects, as defined in [RFC3209] and [RFC3477]. This TLV can be carried in a NOTIFICATION object to indicate the reachable end-points of the domain of the PCE who sends the PCntf message.

[Editor's note]: If the child PCE represents several domains, the reachability TLV should be sent together with a domain_tlv

3.4.7. Procedures

When a parent PCE establishes a PCEP session with a child PCE successfully, the parent PCE may request the child PCE to report the domain connectivity information. This procedure can be done by sending a PCNTf message from the parent to the child, setting the notification-type to 3 and notification-value to 0 in the NOTIFICATION object.

When a child PCE receives the PCNTf message, it may send all of the domain connectivity information to the parent PCE by the PCNTf message(s). The notification-type is 3 and notification-value is 1 in the NOTIFICATION object. The NOTIFICATION object may carry the inter-domain link TLV and inter-domain node TLV to describe the inter-domain connectivity information. It is noted that if the child PCE does not support this function, it will ignore the received PCNTf message and the parent PCE will not receive the response.

The child PCE can also update the domain connectivity information by re-sending the PCNTf message(s) with the newly information.

When the child PCE detects a deletion of domain connectivity (e.g., the inter-domain link TLV is aged out), it must notify the parent PCE to delete the inter-domain link by sending the PCNTf message. The notification-type is 3 and notification-value is 2 in the NOTIFICATION object.

When a parent PCE establishes a PCEP session with a child PCE successfully, the parent PCE may request the child PCE to report the end-points reachability information of the represented domain. This procedure can be done by sending a PCNTf message from the parent to the child, setting the notification-type to 3 and notification-value to 0 in the NOTIFICATION object.

3.5. Extensions to PCEP-ERROR object

3.5.1. Hierarchy PCE Error-Type

A new PCEP Error-Type is allocated for hierarchy PCE (to be assigned by IANA, recommended 19):

Error-Type	Meaning
19	H-PCE error Error-value=1: parent PCE capability cannot be provided

H-PCE error table

3.5.2. Procedures

When a specific child PCE sends a PCReq to a peer PCE that requires parental activity and the peer PCE does not want to act as the parent for it, the peer PCE should send a PCErr message to the child PCE and specify the error-type (IANA) and error-value (1) in the PCEP-ERROR object.

4. Manageability Considerations

TBD.

5. IANA Considerations

As per [RFC5226], IANA is requested to create/update the following registries

5.1. Objective Function (OF) codes

Value	Meaning	Reference
11	MBN	This document
12	MTD	This document
13	DDR	This document

5.2. OPEN Object Flags

5.3. RP Object Flags

5.4. PCEP TLVs

Value	Meaning	Reference
x	Interdomain Link TLV	This document (section Section 3.3.2)
x	Interdomain Node TLV	This document (section Section 3.3.3)

5.5. PCEP NOTIFICATION types

Type	Value	Meaning
DC Notification	1	query all of the domain connectivity
	2	update domain connectivity information
	3	delete domain connectivity information

5.6. PCEP PCEP-ERROR types

Type	Value	Meaning
H-PCE Error 19	1	parent PCE capability cannot be provided
	2	TBD
	3	TBD

6. Security Considerations

TBD

7. Contributing Authors

Xian Zhang
Huawei
zhang.xian@huawei.com

8. Acknowledgments

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [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.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5152] Vasseur, JP., Ayyangar, A., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.

- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, December 2008.
- [RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, January 2009.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5441] Vasseur, JP., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.
- [RFC5886] Vasseur, JP., Le Roux, JL., and Y. Ikejiri, "A Set of Monitoring Tools for Path Computation Element (PCE)-Based Architecture", RFC 5886, June 2010.

9.2. Informative References

- [I-D.ietf-pce-hierarchy-fwk]
King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS, draft-ietf-pce-hierarchy-fwk-04", June 2012.

Authors' Addresses

Fatai Zhang (editor)
Huawei
Huawei Base, Bantian, Longgang District
Shenzhen, 518129
China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Quintin Zhao
Huawei
125 Nagog Technology Park
Acton, MA 01719
US

Phone:
Email: qzhao@huawei.com

Oscar Gonzalez de Dios (editor)
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid, 28045
Spain

Phone: +34913128832
Email: ogondio@tid.es

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n.7
Castelldefels, Barcelona
Spain

Phone: +34 93 645 29 00
Email: ramon.casellas@cttc.es

Daniel King
Old Dog Consulting
UK

Phone:
Email: adrian@olddog.co.uk

Network Working Group
Internet-Draft
Intended status: Experimental
Expires: January 14, 2014

F. Zhang, Ed.
Q. Zhao
Huawei
O. Gonzalez de Dios, Ed.
Telefonica I+D
R. Casellas
CTTC
D. King
Old Dog Consulting
July 14, 2013

Extensions to Path Computation Element Communication Protocol (PCEP) for
Hierarchical Path Computation Elements (PCE)
draft-zhang-pce-hierarchy-extensions-04

Abstract

The Hierarchical Path Computation Element (H-PCE) architecture, defined in the companion framework document [RFC6805], provides a mechanism to allow the optimum sequence of domains to be selected, and the optimum end-to-end path to be derived through the use of a hierarchical relationship between domains.

This document defines the Path Computation Element Protocol (PCEP) extensions for the purpose of implementing Hierarchical PCE procedures which are described in the aforementioned document. These extensions are experimental and published for examination, discussion, implementation, and evaluation.

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 January 14, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the

document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
1.1. Scope	3
1.2. Terminology	4
1.3. Requirements Language	4
2. Requirements for H-PCE	4
2.1. PCEP Requests	4
2.1.1. Qualification of PCEP Requests	4
2.1.2. Multi-domain Objective Functions	5
2.1.3. Multi-domain Metrics	6
2.2. Parent PCE Capability Discovery	6
2.3. PCE Domain and PCE ID Discovery	6
3. PCEP Extensions (Encoding)	6
3.1. OPEN Object	6
3.1.1. OF Codes	6
3.1.2. OPEN Object Flags	7
3.1.3. Domain-ID TLV	7
3.1.4. PCE-ID TLV	9
3.2. RP object	9
3.2.1. RP Object Flags	9
3.2.2. Domain-ID TLV	9
3.3. Metric Object	10
3.4. PCEP-ERROR Object	10
3.4.1. Hierarchy PCE Error-Type	10
3.5. NO-PATH Object	10
4. H-PCE Procedures	10
4.1. OPEN Procedure between Child PCE and Parent PCE	11
4.2. Procedure to Obtain Domain Sequence	11
5. Error Handling	11
6. Manageability Considerations	12
7. IANA Considerations	12
8. Security Considerations	12
9. Contributing Authors	12
10. Acknowledgments	12
11. Normative References	13
Authors' Addresses	13

1. Introduction

[RFC6805] describes a Hierarchical PCE (H-PCE) architecture which can be used for computing end-to-end paths for inter-domain MPLS Traffic Engineering (TE) and GMPLS Label Switched Paths (LSPs).

Within the hierarchical PCE architecture, the parent PCE is used to compute a multi-domain path based on the domain connectivity information. A child PCE may be responsible for a single domain or multiple domains, it is used to compute the intra-domain path based on its domain topology information.

The H-PCE end-to-end domain path computation procedure is described below:

- o A path computation client (PCC) sends the inter-domain path computation requests to the child PCE responsible for its domain;
- o The child PCE forwards the request to the parent PCE;
- o The parent PCE computes the likely domain paths from the ingress domain to the egress domain;
- o The parent PCE sends the intra-domain path computation requests (between the domain border nodes) to the child PCEs which are responsible for the domains along the domain path;
- o The child PCEs return the intra-domain paths to the parent PCE;
- o The parent PCE constructs the end-to-end inter-domain path based on the intra-domain paths;
- o The parent PCE returns the inter-domain path to the child PCE;
- o The child PCE forwards the inter-domain path to the PCC.

In addition, the parent PCE may be requested to provide only the sequence of domains to a child PCE so that alternative inter-domain path computation procedures, including Per Domain (PD) [RFC5152] and Backwards Recursive Path Computation (BRPC) [RFC5441] may be used.

This document defines the PCEP extensions for the purpose of implementing Hierarchical PCE procedures, which are described in [RFC6805].

1.1. Scope

The following functions are out of scope of this document.

- o Finding end point addresses;
- o Parent Traffic Engineering Database (TED) methods;
- o Domain connectivity;

The document also uses a number of [editor notes] to describe options and alternative solutions. These options and notes will be removed before publication once agreement is reached.

1.2. Terminology

This document uses the terminology defined in [RFC4655], [RFC5440] and the additional terms defined in section 1.4 of [RFC6805].

1.3. Requirements Language

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

2. Requirements for H-PCE

This section compiles the set of requirements of the PCEP protocol to support the H-PCE architecture and procedures.

[RFC6805] identifies high-level requirements of PCEP extensions required to support the hierarchical PCE model.

2.1. PCEP Requests

The PCReq messages are used by a PCC or PCE to make a path computation request to a PCE. In order to achieve the full functionality of the H-PCE procedures, the PCReq message needs to include:

- o Qualification of PCE Requests.
- o Multi-domain Objective Functions (OF).
- o Multi-domain Metrics.

2.1.1. Qualification of PCEP Requests

As described in section 4.8.1 of [RFC6805], the H-PCE architecture introduces new request qualifications, which are:

- o It MUST be possible for a child PCE to indicate that a request it sends to a parent PCE should be satisfied by a domain sequence only, that is, not by a full end-to-end path. This allows the child PCE to initiate a per-domain (PD) [RFC5152] or a backward recursive path computation (BRPC) [RFC5441].
- o As stated in [RFC6805], section 4.5, if a PCC knows the egress domain, it can supply this information as the path computation request. It SHOULD be possible to specify the destination domain information in a PCEP request, if it is known.

2.1.2. Multi-domain Objective Functions

For inter-domain path computation, there are two new objective functions which are defined in section 1.3.1 and 4.1 of [RFC6805]:

- o Minimize the number of domains crossed. A domain can be either an Autonomous System (AS) or an Internal Gateway Protocol (IGP) area depending on the type of multi-domain network hierarchical PCE is applied to.
- o Disallow domain re-entry.[Editor's note: Disallow domain re-entry may not be an objective function, but an option in the request].

During the PCEP session establishment procedure, the parent PCE needs to be capable of indicating the Objective Functions (OF) capability in the Open message. This capability information may then be announced by child PCEs, and used for selecting the PCE when a PCC wants a path that satisfies one or multiple inter-domain objective functions.

When a PCC requests a PCE to compute an inter-domain path, the PCC needs also to be capable of indicating the new objective functions for inter-domain path. Note that a given child PCE may also act as a parent PCE.

For the reasons described previously, new OF codes need to be defined for the new inter-domain objective functions. Then the PCE can notify its new inter-domain objective functions to the PCC by carrying them in the OF-list TLV which is carried in the OPEN object. The PCC can specify which objective function code to use, which is carried in the OF object when requesting a PCE to compute an inter-domain path.

The proposed solution may need to differentiate between the OF code that is requested at the parent level, and the OF code that is requested at the intra-domain (child domain).

A parent PCE MUST be capable of ensuring homogeneity, across domains, when applying OF codes for strict OF intra-domain requests.

2.1.3. Multi-domain Metrics

For inter-domain path computation, there are several path metrics of interest [Editor's note: Current framework only mentions metric objectives. The metric itself should be also defined]:

- o Domain count (number of domains crossed).
- o Border Node count.

A PCC may be able to limit the number of domains crossed by applying a limit on these metrics.

2.2. Parent PCE Capability Discovery

Parent and child PCE relationships are likely to be configured. However, as mentioned in [RFC6805], it would assist network operators if the child and parent PCE could indicate their H-PCE capabilities.

During the PCEP session establishment procedure, the child PCE needs to be capable of indicating to the parent PCE whether it requests the parent PCE capability or not. Also, during the PCEP session establishment procedure, the parent PCE needs to be capable of indicating whether its parent capability can be provided or not.

2.3. PCE Domain and PCE ID Discovery

A PCE domain is a single domain with an associated PCE. Although it is possible for a PCE to manage multiple domains. The PCE domain may be an IGP area or AS.

The PCE ID is an IPv4 and/or IPv6 address that is used to reach the parent/child PCE. It is RECOMMENDED to use an address that is always reachable if there is any connectivity to the PCE.

The PCE ID information and PCE domain identifiers may be provided during the PCEP session establishment procedure or the domain connectivity information collection procedure.

3. PCEP Extensions (Encoding)

3.1. OPEN object

3.1.1. OF Codes

This H-PCE experiment will be carried out using the following OF codes:

- o MTD
 - * Name: Minimize the number of Transit Domains.
 - * Objective Function Code.
 - * Description: Find a path P such that it passes through the lnumber of transit domains.
- o MBN
 - * Name: Minimize the number of border nodes.
 - * Objective Function Code.
 - * Description: Find a path P such that it passes through the least number of border nodes.
- o DDR
 - * Name: Disallow Domain Re-entry (DDR)
 - * Objective Function Code.
 - * Description: Find a path P such that does not entry a domain more than once.

3.1.2. OPEN Object Flags

This H-PCE experiment will also require two OPEN object flags:

- o Parent PCE Request bit (to be assigned by IANA, recommended bit 0): if set, it would signal that the child PCE wishes to use the peer PCE as a parent PCE.
- o Parent PCE Indication bit (to be assigned by IANA, recommended bit 1): if set, it would signal that the PCE can be used as a parent PCE by the peer PCE.

3.1.3. Domain-ID TLV

The Domain-ID TLV for this H-PCE experiment is defined below:

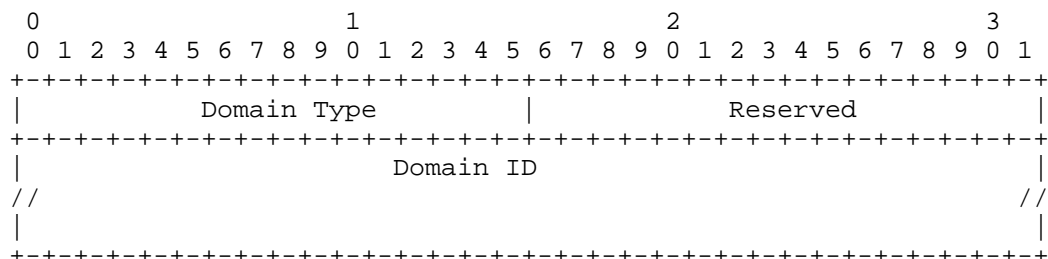


Figure 1: Domain-ID TLV

Domain Type (8 bits): Indicates the domain type. Two types of domain are currently defined:

- o Type=1: the Domain ID field carries an IGP Area ID.
- o Type=2: the Domain ID field carries an AS number.

Domain ID (variable): Indicates an IGP Area ID or AS number. It can be 2 bytes, 4 bytes or 8 bytes long depending on the domain identifier used.

[Editor's note: draft-dhody-pce-pcep-domain-sequence, section 3.2 deals with the encoding of domain sequences, using ERO-subobjects. Work is ongoing to define domain identifiers for OSPF-TE areas, IS-IS area (which are variable sized), 2-byte and 4-byte AS number, and any other domain that may be defined in the future. It uses RSVP-TE subobject discriminators, rather than new type 1/ type 2. A domain sequence may be encoded as a route object. The "VALUE" part of the TLV could follow common RSVP-TE subobject format:

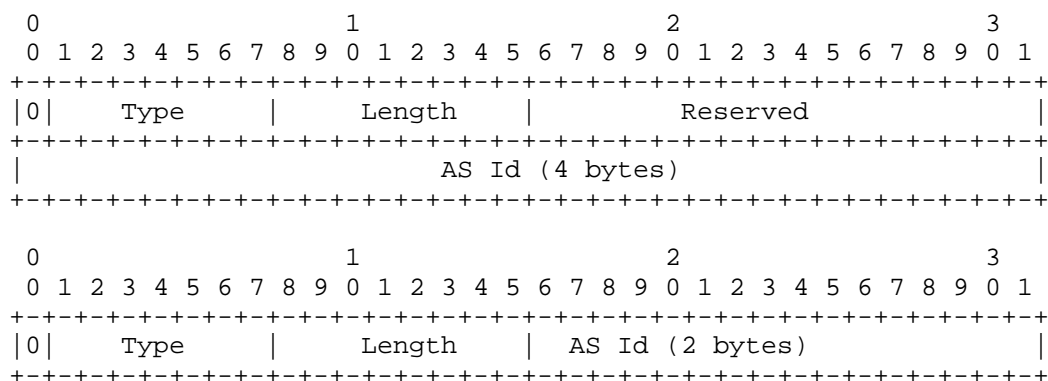


Figure 2: Alternative Domain-ID TLV

3.1.4. PCE-ID TLV

The type of PCE-ID TLV for this H-PCE experiment is defined below:

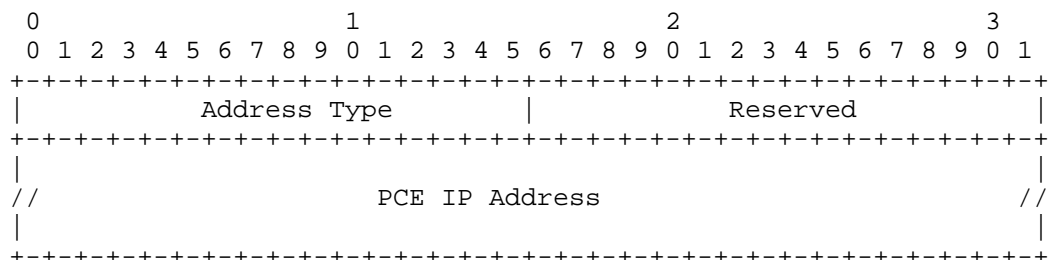


Figure 3: PCE-ID TLV

Address Type (16 bits): Indicates the address type of PCE IP Address. 1 means IPv4 address type, 2 means IPv6 address type.

PCE IP Address: Indicates the reachable address of a PCE.

[Editor's note: [RFC5886] already defines the PCE-ID object. If a semantically equivalent PCE-ID TLV is needed (to avoid modifying message grammars to include the object), it can align with the PCEP object: in any case, the length (4 / 16 bytes) can be used to know whether it is an IPv4 or an IPv6 PCE, the address type is not needed.]

3.2. RP object

3.2.1. RP Object Flags

The following RP object flags are defined for this H-PCE experiment:

- o Domain Path Request bit: if set, it means the child PCE wishes to get the domain sequence.
- o Destination Domain Query bit: if set, it means the parent PCE wishes to get the destination domain ID.

3.2.2. Domain-ID TLV

The format of this TLV is defined in Section 3.1.3. This TLV can be carried in an OPEN object to indicate a (list of) managed domains, or carried in a RP object to indicate the destination domain ID when a child PCE responds to the parent PCE's destination domain query by a PCRep message.

[Editors note. In some cases, the Parent PCE may need to allocate a node which is not necessarily the destination node.]

3.3. Metric Object

There are two new metrics defined in this document for H-PCE:

- o Domain count (number of domains crossed).
- o Border Node Count (number of border nodes crossed).

3.4. PCEP-ERROR object

3.4.1. Hierarchy PCE Error-Type

A new PCEP Error-Type is used for this H-PCE experiment and is defined below:

Error-Type	Meaning
19	H-PCE error Error-value=1: parent PCE capability cannot be provided

H-PCE error table

3.5. NO-PATH Object

To communicate the reason(s) for not being able to find a multi-domain path or domain sequence, the NO-PATH object can be used in the PCRep message. [RFC5440] defines the format of the NO-PATH object. The object may contain a NO-PATH-VECTOR TLV to provide additional information about why a (domain) path computation has failed.

Three new bit flags are defined to be carried in the Flags field in the NO-PATH-VECTOR TLV carried in the NO-PATH Object.

- o Bit 23: When set, the parent PCE indicates that destination domain unknown;
- o Bit 22: When set, the parent PCE indicates unresponsive child PCE(s);
- o Bit 21: When set, the parent PCE indicates no available resource available in one or more domain(s).

4. H-PCE Procedures

4.1. OPEN Procedure between Child PCE and Parent PCE

If a child PCE wants to use the peer PCE as a parent, it can set the parent PCE request bit in the OPEN object carried in the Open message during the PCEP session creation procedure. If the peer PCE does not want to provide the parent function to the child PCE, it must send a PCErr message to the child PCE and clear the parent PCE indication bit in the OPEN object.

If the parent PCE can provide the parent function to the peer PCE, it may set the parent PCE indication bit in the OPEN object carried in the Open message during the PCEP session creation procedure.

The PCE may also report its PCE ID and list of domain ID to the peer PCE by specifying them in the PCE-ID TLV and List of Domain-ID TLVs in the OPEN object carried in the Open message during the PCEP session creation procedure.

The OF codes defined in this document can be carried in the OF-list TLV of the OPEN object. If the OF-list TLV carries the OF codes, it means that the PCE is capable of implementing the corresponding objective functions. This information can be used for selecting a proper parent PCE when a child PCE wants to get a path that satisfies a certain objective function.

When a specific child PCE sends a PCReq to a peer PCE that requires parental activity and the peer PCE does not want to act as the parent for it, the peer PCE should send a PCErr message to the child PCE and specify the error-type (IANA) and error-value (1) in the PCEP-ERROR object.

4.2. Procedure to obtain Domain Sequence

If a child PCE only wants to get the domain sequence for a multi-domain path computation from a parent PCE, it can set the Domain Path Request bit in the RP object carried in a PCReq message. The parent PCE which receives the PCReq message tries to compute a domain sequence for it. If the domain path computation succeeds the parent PCE sends a PCRep message which carries the domain sequence in the ERO to the child PCE. The domain sequence is specified as AS or AREA ERO sub-objects (type 32 for AS [RFC3209] or a to-be-defined IGP area type). Otherwise it sends a PCReq message which carries the NO-PATH object to the child PCE.

5. Error Handling

A PCE that is capable of acting as a parent PCE might not be configured or willing to act as the parent for a specific child PCE.

This fact could be determined when the child sends a PCReq that requires parental activity (such as querying other child PCEs), and could result in a negative response in a PCEP Error (PCErr) message and indicate the hierarchy PCE error types.

Additionally, the parent PCE may fail to find the multi-domain path or domain sequence due to one or more of the following reasons:

- o A child PCE cannot find a suitable path to the egress;
- o The parent PCE do not hear from a child PCE for a specified time;
- o The objective functions specified in the path request cannot be met.

In this case, the parent PCE MAY need to send a negative path computation reply specifying the reason. This can be achieved by including NO-PATH object in the PCRep message. Extension to NO-PATH object is needed to include the aforementioned reasons.

6. Manageability Considerations

TBD.

7. IANA Considerations

Due to the experimental nature of this draft no IANA requests are made.

8. Security Considerations

To be added.

9. Contributing Authors

Xian Zhang
Huawei
zhang.xian@huawei.com

10. Acknowledgments

To be added.

11. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 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.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5152] Vasseur, JP., Ayyangar, A., and R. Zhang, "A Per-Domain Path Computation Method for Establishing Inter-Domain Traffic Engineering (TE) Label Switched Paths (LSPs)", RFC 5152, February 2008.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5441] Vasseur, JP., Zhang, R., Bitar, N., and JL. Le Roux, "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.
- [RFC5886] Vasseur, JP., Le Roux, JL., and Y. Ikejiri, "A Set of Monitoring Tools for Path Computation Element (PCE)-Based Architecture", RFC 5886, June 2010.
- [RFC6805] King, D. and A. Farrel, "The Application of the Path Computation Element Architecture to the Determination of a Sequence of Domains in MPLS and GMPLS", RFC 6805, November 2012.

Authors' Addresses

Fatai Zhang (editor)
Huawei
Huawei Base, Bantian, Longgang District
Shenzhen, 518129
China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Quintin Zhao
Huawei
125 Nagog Technology Park
Acton, MA 01719
US

Phone:
Email: qzhao@huawei.com

Oscar Gonzalez de Dios (editor)
Telefonica I+D
Don Ramon de la Cruz 82-84
Madrid, 28045
Spain

Phone: +34913128832
Email: ogondio@tid.es

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n.7
Castelldefels, Barcelona
Spain

Phone: +34 93 645 29 00
Email: ramon.casellas@cttc.es

Daniel King
Old Dog Consulting
UK

Phone:
Email: daniel@olddog.co.uk

Network Working Group
Internet-Draft
Intended status: Standards Track

Xian Zhang
Young Lee
Huawei
Ramon Casellas
CTTC
Oscar Gonzalez de Dios
Telefonica I+D

Expires: January 07, 2013

July 07, 2012

Path Computation Element (PCE) Protocol Extension for Stateful PCE
Usage in GMPLS Networks

draft-zhang-pce-pcep-stateful-pce-gmpls-00.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on January 07, 2013.

Abstract

The Path Computation Element (PCE) facilitates Traffic Engineering (TE) based path calculation in large, multi-domain, multi-region, or multi-layer networks. PCE can be stateless or stateful. With the LSP state information acquired from the network, a stateful PCE exhibits superiority in facilitating a wide variety of applications, especially in GMPLS networks, such as impairment-aware routing and wavelength assignment in wavelength-switched optical networks (WSO), time-based scheduling applications. This memo provides extensions required for PCE communication protocol (i.e. PCEP) so as to enable the usage of a stateful PCE capability in GMPLS networks. To be more specific, the PCEP extensions specified in this memo include not only new objects but also modification of existing objects in PCEP messages, with regard to stateful PCE usage in GMPLS networks.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

Table of Contents	2
1. Introduction	3
2. PCEP Extensions	4
2.1. Overview	4
2.2. PCEP Extension for Stateful PCE Capability Advertisement and Negotiation	4
2.2.1. PCE Capability Negotiation/Advertisement in Multi-layer Networks	5
2.3. LSP Delegation	5
2.4. PCEP Extensions for LSP Synchronization	6
2.5. PCEP Simplification	6
2.6. Application-specific PCEP extensions for stateful PCE....	7
2.6.1. Time-based Scheduling.....	7
2.6.1.1. PCEP Extension.....	8
2.6.2. RWA in Impairment-aware Wavelength-switched Optical Networks (WSO)	9
3. IANA Considerations	10
3.1. LayerCapability TLV.....	10
3.2. SERVICE-TIME Object.....	10
3.3. Extension to METRIC Object.....	10
4. Manageability Considerations.....	10
5. Security Considerations.....	11
6. References	11
6.1. Normative References.....	11
6.2. Informative References.....	11

7. Contributors' Address.....	12
Authors' Addresses	13

1. Introduction

[RFC 4655] presents the architecture of a Path Computation Element (PCE)-based model for computing Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs). To perform such a constrained computation, a PCE stores the network topology (i.e., TE links and nodes) and resource information (i.e., TE attributes) in its TE Database (TED). To request path computation services to a PCE, [RFC 5440] defines the PCE Communication Protocol (PCEP) for communications between a Path Computation Client (PCC) and a PCE, or between two PCEs. A PCC can initiate a path computation request to a PCE through a Path Computation Request (PCReq) message, and then the PCE will return the computed route to the requesting PCC in response to a previously received PCReq message through a PCEP Path Computation Reply (PCRep) message.

As per [RFC 4655], a PCE can be stateless or stateful. Compared to a stateless PCE, a stateful PCE stores not only the network state, but also the set of computed paths and reserved resources in use in the network. Note that [RFC4655] further specifies that the TED contains link state and bandwidth availability as distributed by IGPs or collected via other means. Even if such information can provide finer granularity and more details, it is not state information in the PCE context and so a model that uses it is still described as a stateless PCE.

Stateful PCE(s) are shown to be helpful in many application scenarios, especially in GMPLS networks, as illustrated in [Stateful-APP]. In order for these applications to be able to exploit the capability of stateful PCE(s), extensions to the PCE communication protocol (i.e., PCEP) are required.

It is expected that there are common aspects for stateful PCE PCEP extension in GMPLS networks with that in MPLS networks [Stateful-PCE]. Therefore, this document focuses on the extensions unique to GMPLS networks while maintains a complete picture of the PCEP extensions required for a stateful PCE. In summary, this draft gives an overview of PCEP extensions necessary for stateful PCE usage in GMPLS networks as well as the details of required PCEP extension unique to stateful PCE usage in GMPLS networks.

2. PCEP Extensions

2.1. Overview

According to the description in [Stateful-APP], a summary of PCEP extensions required for stateful PCE in GMPLS networks is provided as follows:

- o Advertisement and negotiation of stateful PCE capability;
- o LSP synchronization;

These two are fundamental extension requirements needed in order to make a stateful PCE functional. Attention should be paid in terms of the general considerations as discussed in [Stateful-APP]. Since the extensions to these two aspects are straightforward and have already been covered in [Stateful-PCE], we only cover the points that are either relevant to GMPLS or still missing in [Stateful-PCE].

- o LSP Delegation;

As explained in [Stateful-APP], the ability to collect LSP state information should be mandatory. As for PCE's ability to modify the LSP attributes as presented in [Stateful-PCE] as well as how it is enabled (per PCE base, per LSP base, or per NE base?) should be operator-dependent and is for further study.

- o Simplification of the existing PCEP protocol [RFC5440];

Since the LSP state is part of the information that a stateful PCE possesses, some simplifications to PCEP are possible and explained in this draft.

- o Application-specific extensions desired;

A list of examples is provided in [Stateful-APP] and they may require additional extensions or modification of the PCEP protocol. In this draft, we present the PCEP extensions for some typical application examples.

2.2. PCEP Extension for Stateful PCE Capability Advertisement and Negotiation

Whether a PCE has the stateful capability or not can be negotiated during the PCEP session establishment process. It can also be advertised through routing protocols as described in [RFC5088]. In either case, the following additional aspects should also be considered.

2.2.1. PCE Capability Negotiation/Advertisement in Multi-layer Networks

In multi-layer network scenarios where there is a PCE responsible for each layer, then the PCCs should be informed of which PCE they should synchronize their LSP states with as well as send path computation requests to.

A new LayerCapability TLV is defined as shown below to denote to which layer a PCE is in charge of LSP synchronization as well as path computation. It can be included in the OPEN Object if applicable. Alternatively, the extension to current OSPF PCED TLV is needed.

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 (T.B.D.)										Length																													
LSP Enc. Type										Switching Type										Reserved																			
										...																													
LSP Enc. Type										Switching Type										Reserved																			

2.3. LSP Delegation

If a LSP span across multiple domains and the delegation features presented in [Stateful-PCE] is supported, it adds complexity to LSP state synchronization/update. For instance, in a multi-domain networks where one PCE per domain is adopted, a contiguous LSP is setup which spans across multiple domains. Then, each PCE is responsible for synchronizing/updating or is able to modify only part of the LSP within the network for which the PCE is deployed. Moreover, a modification action of a stateful PCE for partial LSP may trigger a chain of LSP updating actions (e.g., informing other PCEs of the modification or requesting other PCEs of additional modification).

This needs to be considered carefully and modification capability specifications might be needed to limit the scope of LSP attribute modification action to avoid conflicts(?).

[Editor Note: this needs clarification and further discussion. The scenario with mixed stateful/stateless PCE might also cause potential issues for LSP delegation ability.]

2.4. PCEP Extensions for LSP Synchronization

For LSP state synchronization of stateful PCE(s) in GMPLS networks, the LSP attributes, such as its bandwidth, associated label as well as protection information etc, should be updated by PCC(s) to PCE LSP database (LSP-DB).

As per [Stateful-PCE], it only covers LSP attributes pertaining to MPLS networks, based on [RFC5440]. Therefore, extensions of PCEP protocol for stateful PCE usage in GMPLS networks are required. The following presents a list of objects/TLVs that should be used by stateful PCE for LSP synchronization purpose when applied in GMPLS networks:

- o GENERALIZED BANDWIDTH
- o PROTECTION ATTRIBUTE
- o Extended Objects to support the inclusion of label sub-object
 - RP
 - IRO
 - XRO

Note that the list above should also be used for path computation requests/replies. Refer to [PCEP-GMPLS] for the details of these objects/TLVs.

2.5. PCEP Simplification

One of the merits mentioned in [Stateful-APP] is its ability to simplify the information exchange between PCC and PCE. To be more specific, with a stateful PCE, it is possible for PCCs to carry only LSP ID information, instead of giving detailed LSP information (such as route), whenever necessary.

Example 1: a PCC (e.g. NMS) requesting for a re-optimization of several LSPs, it can send the request with relevant LSP IDs.

In order to support these, the LSP identifier TLV defined in [Stateful-PCE] can be used in the RP Object to specify the request LSP ID(s). Upon receiving the PCReq message, PCE should be able to

correlate with one or multiple LSPs with their detailed state information and carry out optimization accordingly.

Example 2: in order to set up a LSP which is diversified with one or more specific LSPs, a PCC can send a PCReq with the ID of these LSPs. A stateful PCE should be able to find the corresponding route and resource information so as to meet the constraints set by the requesting PCC.

In order to support this, a new subobject type should be supported, i.e. LSP ID(s). Hence, the LSP identifier TLV defined in [Stateful-PCE] can be used in XRO object for this purpose.

2.6. Application-specific PCEP extensions for stateful PCE

[Editor's Note: this is not a complete list of application-specific PCEP extensions. Suggestions are welcome on expansion on this section.]

2.6.1. Time-based Scheduling

To support time-based scheduling, network operators need to reserve resources in advance according to customers' requests with specified starting time and duration. A simple utilization example of this service is to support scheduled data transmission between data centers or any generic scheduled based services.

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 due to the lack of a time-based mechanism. 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.

Since a stateful PCE needs to collect LSP related information for the whole network, it can naturally support this service with resource usage flexibility (i.e., only excluding the time slot(s) reserved for time-based scheduling requests). Moreover, it can avoid the need to add complexity on network elements in this regard. A stateful PCE should also maintain a database that stores all the reserved information with time reference. This can be achieved either by maintaining a separate database or having all the reserved information with time reference incorporated into LSP-DB. The details of organizing time-based scheduling related information are

subject to network provider's policy and administrative consideration and thus outside of the scope of this document.

2.6.1.1. PCEP Extension

For a PCC to request a path computation for scheduled service, it MUST be able to specify the time-related information, including the starting time and LSP holding time, in PCEP request.

A SERVICE-TIME object is presented as follows to provide the required information (i.e. service starting time and holding time).

The Object-Class is TBD and the Object-Type is 1.

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												
-----Start-Year-----										-----Month-----										-----Day-----																							
-----Hour-----										-----Minute-----										-----Second-----										-----Reserved-----													
-----Duration (in seconds)-----																																											

field	Length	range
-----	-----	-----
Start-Year	16 bit	0..65536
Month	8 bit	1..12
Day	8 bit	1..31
Hour	8 bit	0..23
Minute	8 bit	0..59
Second	8 bit	0..59

The SERVICE-TIME object can be included in a PCEP request as specified in the following manner:

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

<request-list>

Where:

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

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

<request > ::= <RP>

<END-POINTS>

[<LSPA>]

[<BANDWIDTH>]

[<SERVICE-TIME>]

[<metric-list>]

[<RRO> [<BANDWIDTH>]]

[<IRO>]

[<LOAD-BALANCING>]

WHERE:

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

Upon receiving the PCReq message, PCE should compute the path taking into consideration the constraints of the TED, LSP-DB as well as other scheduled service information and return the computed route back to the requesting PCC.

If no path can be found, PCE should return an error message specifying the reason.

2.6.2. RWA in Impairment-aware Wavelength-switched Optical Networks (WSON)

In impairment-aware WSON networks, the routing and wavelength assignment process needs to consider the constraints incurred by the physical impairments. As described in [Stateful-APP], stateful PCE can effectively reduce the control plane overhead by centrally maintaining the impairment-information related to each LSP.

In order to establish an impairment-aware LSP, a path computation request may need to specify the desired values and/or constraints for one or more of the following parameters:

- o Power
- o OSNR
- o PMD
- o T.B.D.

Upon receiving the path computation request, a stateful PCE should take into consideration the explicitly defined constraints as well as those of the existing LSPs, stored in LSP-DB. Furthermore, a PCE may need to reply in PCRep with the actual values of the one or more of the above-mentioned parameters to the requesting PCC as well as the adjustment needed. After receiving the reply message, the PCC can take appropriate actions along the to-be-established paths in tuning its power or changing other impairment-related parameters so as to achieve the desired signal quality.

To support the above-mentioned requirements, the METRIC object defined in [RFC5440] can be exploited with proper extension. A new type value should be added.

T=T.B.D.: Impairment-aware information

Furthermore, as described in [PCE-IA-WSO], there are two types of parameters that can be specified, i.e. path level or link level. If a stateful PCE needs to reply with adjustment needed for path level parameters. Then further extension to the METRIC object is desirable and this will be considered in the future.

3. IANA Considerations

IANA is requested to allocate new Types for the TLV/Object defined in this document.

3.1. LayerCapability TLV

3.2. SERVICE-TIME Object

3.3. Extension to METRIC Object

4. Manageability Considerations

TBD.

5. Security Considerations

TBD.

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.
- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5440] Vasseur, J.-P., and Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5088] Le Roux, JL., Vasseur, J.-P., Ikejiri, Y., Zhang, R., "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.

6.2. Informative References

- [Stateful-APP] Zhang, F., Zhang, X., Lee, Y., Casellas, R., Gonzalez de Dios, O., "Applicability of Stateful Path Computation Element (PCE) ", draft-zhang-pce-stateful-pce-app, work in progress.
- [Stateful-PCE] Crabbe, E., Medved, J., Varga, R., Minei, I., "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce, work in progress.
- [PCE-IA-WSO] Lee, Y., Bernstein G., Takeda, T., Tsuritani, T., "PCEP Extensions for WSON Impairments", draft-lee-pce-wson-impairments, work in progress.
- [PCEP-GMPLS] Margaria, C., Gonzalez de Dios, O., Zhang, F., "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions, work in progress.

7. Contributors' Address

Dhruv Dhody
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruvd@huawei.com

Authors' Addresses

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

Phone: +86-755-28972913
Email: zhang.xian@huawei.com

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

Ramon Casellas
CTTC - Centre Tecnologic de Telecomunicacions de Catalunya
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Phone:
Email: ramon.casellas@cttc.es

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION

HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Full Copyright Statement

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Network Working Group
Internet-Draft
Intended status: Standards Track

Xian Zhang
Young Lee
Fatai Zhang
Huawei
Ramon Casellas
CTTC
Oscar Gonzalez de Dios
Telefonica I+D
Zafar Ali
Cisco Systems

Expires: April 21, 2014

October 21, 2013

Path Computation Element (PCE) Protocol Extensions for Stateful PCE
Usage in GMPLS-controlled Networks

draft-zhang-pce-pcep-stateful-pce-gmpls-03.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 21, 2014.

Abstract

The Path Computation Element (PCE) facilitates Traffic Engineering (TE) based path calculation in large, multi-domain, multi-region, or multi-layer networks. [Stateful-PCE] provides the fundamental PCE communication Protocol (PCEP) extensions needed to support stateful PCE functions, without specifying the technology-specific extensions. This memo provides extensions required for PCEP so as to enable the usage of a stateful PCE capability in GMPLS-controlled networks.

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

Table of Contents	2
1. Introduction	3
2. PCEP Extensions	3
2.1. Overview of Requirements.....	3
2.2. Stateful PCE Capability Advertisement	4
2.2.1. PCE Capability Advertisement in Multi-layer Networks	4
2.3. LSP Delegation in GMPLS-controlled Networks	5
2.4. LSP Synchronization in GMPLS-controlled networks.....	6
2.5. Modification of Existing PCEP Messages and Procedures....	7
2.5.1. Use cases	8
2.5.2. Modification for LSP Re-optimization	8
2.5.3. Modification for Route Exclusion	9
2.6. Additional Error Type and Error Values Defined.....	10
3. IANA Considerations	10
4. Manageability Considerations	10
4.1. Requirements on Other Protocols and Functional Components	10
5. Security Considerations.....	11
6. Acknowledgement	11
7. References	11
7.1. Normative References.....	11
7.2. Informative References.....	11
8. Contributors' Address.....	12
Authors' Addresses	13

1. Introduction

[RFC 4655] presents the architecture of a Path Computation Element (PCE)-based model for computing Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs). To perform such a constrained computation, a PCE stores the network topology (i.e., TE links and nodes) and resource information (i.e., TE attributes) in its TE Database (TED). To request path computation services to a PCE, [RFC 5440] defines the PCE communication Protocol (PCEP) for interaction between a Path Computation Client (PCC) and a PCE, or between two PCEs. PCEP as specified in [RFC 5440] mainly focuses on MPLS networks and the PCEP extensions needed for GMPLS-controlled networks are provided in [PCEP-GMPLS].

Stateful PCEs are shown to be helpful in many application scenarios, in both MPLS and GMPLS networks, as illustrated in [Stateful-APP]. In order for these applications to be able to exploit the capability of stateful PCEs, extensions to the PCE communication protocol (i.e., PCEP) are required.

[Stateful-PCE] provides the fundamental extensions needed for stateful PCE to support general functionality, but leaves out the specification for technology-specific objects/TLVs. Complementarily, this document focuses on the extensions that are necessary in order for the deployment of stateful PCEs in GMPLS-controlled networks.

2. PCEP Extensions

2.1. Overview of Requirements

This section notes the main functional requirements for PCEP extensions to support stateful PCE for use in GMPLS-controlled networks, based on the description in [Stateful-APP]. Many requirements are common across a variety of network types (e.g., MPLS-TE networks and GMPLS networks) and the protocol extensions to meet the requirements are already described in [Stateful-PCE]. This document does not repeat the description of those protocol extensions. Other requirements that are also common across a variety of network types do not currently have protocol extensions defined in [Stateful-PCE]. In these cases, this document presents protocol extensions for discussion by the PCE working group and potential inclusion in [Stateful-PCE]. In addition, this document presents protocol extensions for a set of requirements which are specific to the use of a stateful PCE in a GMPLS-controlled network.

The basic requirements are as follows:

- o Advertisement of the stateful PCE capability. This generic requirement is covered in Section 7.1.1 of [Stateful-PCE]. Section 2.2 of this document discusses other potential extensions for this functionality.
- o LSP delegation is already covered in Section 5.5 of [Stateful-PCE]. Section 2.3 of this document provides extension for its application in GMPLS-controlled networks. Moreover, further discussion of some generic details that may need additional consideration is provided.
- o LSP state synchronization. This is a generic requirement already covered in Section 5.4 of [Stateful-PCE]. However, there are further extensions required specifically for GMPLS-controlled networks and discussed in Section 2.4. Reference to LSPs by identifiers is discussed in Section 7.2 of [Stateful-PCE]. This feature can be applied to reduce the data carried in PCEP messages. Use cases and additional Error Codes are necessary, as described in Section 2.5 and 2.6.

2.2. Stateful PCE Capability Advertisement

Whether a PCE has stateful capability or not can be advertised during the PCEP session establishment process. It can also be advertised through routing protocols as described in [RFC5088]. In either case, the following additional aspects should also be considered.

2.2.1. PCE Capability Advertisement in Multi-layer Networks

In multi-layer network scenarios, such as an IP-over-optical network, if there are dedicated PCEs responsible for each layer, then the PCCs should be informed of which PCEs they should synchronize their LSP states with, as well as send path computation requests to. The Layer-Cap TLV defined in [INTER-LAYER] can be used to indicate which layer a PCE is in charge of. (Editor's note: this change is currently not included in the current version of the [INTER-LAYER] draft. It is expected that it will be included in its next version.) This TLV is optional and MAY be carried in the OPEN object. It is RECOMMENDED that a PCC synchronizes its LSP states with the same PCEs that it can use for path computation in a multi-layer network. In a single layer, this TLV MAY not be used. However, if the PCE capability discovery depends on IGP and if an IGP instance spans across multiple layers, this TLV is still needed.

Alternatively, the extension to current OSPF PCED TLV is needed. A new domain-type denoting the layer information can be defined:

domain-type: T.B.D.

When it is carried in PCE-DOMAIN sub-TLV, it denotes the layer for which a PCE is responsible for path computation as well as LSP state synchronization. When carried in the PCE-NEIG-DOMAIN sub-TLV, it denotes its adjacent layers for which a PCE can compute paths and synchronize the LSP states. The DOMAIN-ID information can be represented using the following format, to denote the layer information:

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								
LSP Enc. Type										Switching Type										Reserved																			

2.3. LSP Delegation in GMPLS-controlled Networks

To enable the PCE to control an LSP, the PCUpd message is defined in [Stateful-PCE]. However, the specification of technology specific extensions is not covered. The following defines the <path> descriptor, present in the PCUpd message, that should be used in GMPLS-controlled networks:

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

Where:

```

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

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

```

As explained in [stateful-APP], LSP parameter update controlled by a stateful PCE in a multi-domain network is complex and requires well-defined operational procedures as well as protocol design.

[TBD: protocol extensions]

2.4. LSP Synchronization in GMPLS-controlled networks

For LSP state synchronization of stateful PCEs in GMPLS networks, the LSP attributes, such as its bandwidth, associated route as well as protection information etc, should be updated by PCCs to PCE LSP database (LSP-DB). Note the LSP state synchronization described in this document denotes both the bulk LSP report at the initialization phase as well as the LSP state report afterwards described in [Stateful-PCE].

As per [Stateful-PCE], it does not cover technology-specific specification for state synchronization. Therefore, extensions of PCEP for stateful PCE usage in GMPLS networks are required. For LSP state synchronization, the objects/TLVs that should be used for stateful PCE in GMPLS networks are defined in [PCEP-GMPLS] and are briefly summarized as below:

- o GENERALIZED BANDWIDTH
- o GENERALIZED ENDPOINTS
- o PROTECTION ATTRIBUTE
- o Use of IF_ID_ERROR_SPEC. [Stateful-PCE] section 7.2.2 only considers RSVP_ERROR_SPEC TLVs. GMPLS extends this to also support IF_ID_ERROR_SPEC, for example, to report about failed unnumbered interfaces.
- o Extended objects to support the inclusion of the label and unnumbered links.

Per [Stateful-PCE], the PCRpt message is defined for LSP state synchronization purposes. PCRpt is used by a PCC to report one or more of its LSPs to a stateful PCE. However, the <path> descriptor is technology-specific and left undefined.

For LSP state synchronization in GMPLS-controlled networks, the encoding of the <path> descriptor is defined as follows:

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

Where:

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

[<IRO>]

[<XRO>]

[<metric-list>]

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

The objects included in the <path> descriptor can be found in [RFC5440], [PCE-GMPLS] and [RFC5521].

For all the objects presented in this section, the P and I bit MUST be set to 0 since they are only used by a PCC to report its LSP information.

In GMPLS-controlled networks, the <ERO> object may include a list of the label sub-object for SDH/SONET, OTN and DWDM networks. It may also include a list of unnumbered interface IDs to denote the allocated resource. The <RRO>, <IRO> and <XRO> objects MAY include unnumbered interface IDs and labels for networks such as OTN and WDM networks.

If the LSP being reported is a protecting LSP, the <PROTECTION-ATTRIBUTE> TLV MUST be included in the <LSPA> object to denote its attributes and restrictions. Moreover, if the status of the protecting LSP changes from non-operational to operational, this should be synchronized to the stateful PCE. For example, in 1:1 protection, the combination of S=0, P=1 and O=0 denotes the protecting path is set up already but not used for carrying traffic. Upon the working path failure, the operational status of the aforementioned protecting LSP changes to in-use (i.e., O=1). This information should be synchronized with a stateful PCE through a PCRpt message.

The O bit in the <GENERALIZED-BANDWIDTH> object has no meaning for LSP state synchronization and MUST be set to 0. Furthermore, this object MAY appear twice, one with R set to 1 and the other with R set to 0. This is to denote the asymmetric bandwidth property of the updated bi-directional LSP.

2.5. Modification of Existing PCEP Messages and Procedures

One of the advantages mentioned in [Stateful-APP] is that the stateful nature of a PCE simplifies the information conveyed in PCEP messages, notably between PCC and PCE, since it is possible to refer to PCE managed state for active LSPs. To be more specific, with a

stateful PCE, it is possible to refer to a LSP with a unique identifier in the scope of the PCC-PCEP session and thus use such identifier to refer to that LSP.

2.5.1. Use cases

Use Case 1: Assuming a stateful PCE's LSP-DB is up-to-date, a PCC (e.g. NMS) requesting for a re-optimization of one or several LSPs can send the request with "R" bit set and only provides the relevant LSP unique identifiers.

Upon receiving the PCReq message, PCE should be able to correlate with one or multiple LSPs with their detailed state information and carry out optimization accordingly.

The handling of RP object specified in [RFC5440] is stated as following:

"The absence of an RRO in the PCReq message for a non-zero-bandwidth TE LSP (when the R bit of the RP object is set) MUST trigger the sending of a PCErr message with Error-Type="Required Object Missing" and Error-value="RRO Object missing for re-optimization."

If a PCE has stateful capabilities, and such capabilities have been negotiated and advertised, specific rules given in [RFC5440] may need to be relaxed. In particular, the re-optimization case: if the re-optimization request refers to a given LSP state, and the RRO information is available, the PCE can proceed.

Use Case 2: in order to set up a LSP which has a constraint that its route should not use resources used by one or more existing LSPs, a PCC can send a PCReq with the identifiers of these LSPs. A stateful PCE should be able to find the corresponding route and resource information so as to meet the constraints set by the requesting PCC. Hence, the LSP identifier TLV defined in [Stateful-PCE] can be used in XRO object for this purpose. Note that if the PCC is a node in the network, the constraint LSP ID information will be confined to the LSPs initiated by itself.

2.5.2. Modification for LSP Re-optimization

For re-optimization, upon receiving a path computation request and the "R" bit is set, the stateful PCE SHOULD still perform the re-optimization in the following two cases:

Case 1: the existing bandwidth and route information of the to-be-optimized LSP is provided in the path computation request. This

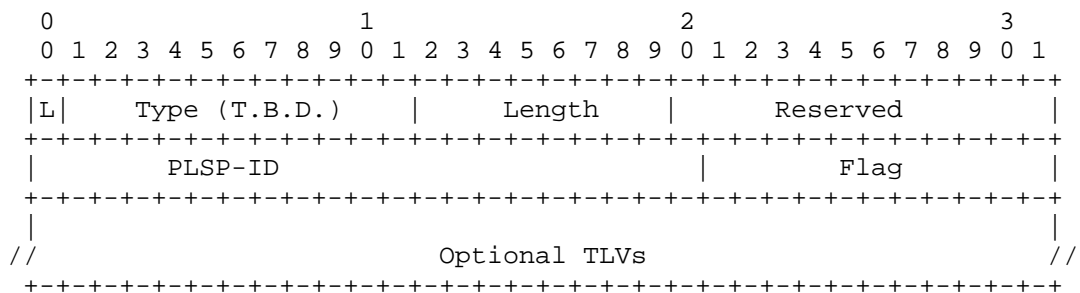
information should be provided via <BANDWIDTH>, <GENERALIZED-BANDWIDTH>, <ERO> objects.

Case 2: the existing bandwidth and route information can be found locally in its LSP-DB. In this case, the PCRep and PCReq messages need to be modified to carry LSP identifiers. The stateful PCE can find this information using the per-node LSP ID together with the PCC's address.

If no LSP state information is available to carry out re-optimization, the stateful PCE should report the error "LSP state information unavailable for the LSP re-optimization" (Error Type = T.B.D., Error value= T.B.D.).

2.5.3. Modification for Route Exclusion

A LSP identifier sub-object is defined and its format as follows:



L bit:

The L bit SHOULD NOT be set, so that the subobject represents a strict hop in the explicit route.

Type:

Subobject Type for a per-node LSP identifier.

Length:

The Length contains the total length of the subobject in bytes, including the Type and Length fields.

PLSP-ID:

This is the identifier given to a LSP and it is unique on a node basis. It is defined in [Stateful-PCE].

Flags:

This field is defined in [Stateful-PCE]. It is not used in this sub-object and should be ignored upon receipt.

Optional TLVs:

Additional TLVs can be defined in the future to provide further information to identify a LSP. In this document, no TLVs are defined.

One or multiple of these sub-objects can be present in the XRO object. When a stateful PCE receives a path computation request carrying this sub-object, it should find relevant information of these LSPs and preclude the resource during the path computation process. If a stateful PCE cannot recognize one or more of the received LSP identifiers, it should reply PCErr saying "the LSP state information for route exclusion purpose cannot be found" (Error-type = T.B.D., Error-value= T.B.D.). Optionally, it may provide with the unrecognized identifier information to the requesting PCC.

2.6. Additional Error Type and Error Values Defined

Error Type Meaning

21(TBD) LSP state information missing

Error-value 1: LSP state information unavailable for the LSP re-optimization

Error-value 2: the LSP state information for route exclusion purpose cannot be found

3. IANA Considerations

IANA is requested to allocate new Types for the TLV/Object defined in this document.T.B.D.

4. Manageability Considerations

The description and functionality specifications presented related to stateful PCEs should also comply with the manageability specifications covered in Section 8 of [RFC4655]. Furthermore, a further list of manageability issues presented in [Stateful-PCE] should also be considered.

Additional considerations are presented in the next sections.

4.1. Requirements on Other Protocols and Functional Components

When the detailed route information is included for LSP state synchronization (either at the initial stage or during LSP state

report process), this require the ingress node of an LSP carry the RRO object in order to enable the collection of such information.

5. Security Considerations

The security issues presented in [RFC5440] and [Stateful-PCE] apply to this document.

6. Acknowledgement

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

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.
- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5440] Vasseur, J.-P., and Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC5088] Le Roux, JL., Vasseur, J.-P., Ikejiri, Y., Zhang, R., "OSPF Protocol Extensions for Path Computation Element (PCE) Discovery", RFC 5088, January 2008.
- [INTER-LAYER] Oki, E., Takeda, Tomonori, Le Roux, JL., Farrel, A., Zhang, F., "Extensions to the Path Computation Element communication Protocol (PCEP) for Inter-Layer MPLS and GMPLS Traffic Engineering", draft-ietf-pce-inter-layer-ext, work in progress.

7.2. Informative References

- [Stateful-APP] Zhang, X., Minei, I., et al "Applicability of Stateful Path Computation Element (PCE) ", draft-ietf-pce-stateful-pce-app, , work in progress.
- [Stateful-PCE] Crabbe, E., Medved, J., Varga, R., Minei, I., "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce, work in progress.

[PCE-IA-WSO] Lee, Y., Bernstein G., Takeda, T., Tsuritani, T.,
"PCEP Extensions for WSO Impairments", draft-lee-pce-
wso-impairments, work in progress.

[PCEP-GMPLS] Margaria, C., Gonzalez de Dios, O., Zhang, F., "PCEP
extensions for GMPLS", draft-ietf-pce-gmpls-pcep-
extensions, work in progress.

8. Contributors' Address

Dhruv Dhody
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruvd@huawei.com

Yi Lin
Huawei Technologies
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
Shenzhen 518129 P.R.China

Phone: +86-755-28972914
Email: yi.lin@huawei.com

Authors' Addresses

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

Phone: +86-755-28972645
Email: zhang.xian@huawei.com

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

Fatai Zhang
Huawei
F3-5-B R&D Center, Huawei Base
Bantian, Longgang District
P.R. China

Phone: +86-755-28972912
Email: zhangfatai@huawei.com

Ramon Casellas
CTTC
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Phone:
Email: ramon.casellas@cttc.es

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Zafar Ali
Cisco Systems
Email: zali@cisco.com

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms,

conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Full Copyright Statement

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.

Network Working Group
Internet-Draft
Intended status: Informational

Fatai Zhang
Xian Zhang
Young Lee
Huawei
Ramon Casellas
CTTC
Oscar Gonzalez de Dios
Telefonica I+D

Expires: January 13, 2013

July 13, 2012

Applicability of Stateful Path Computation Element (PCE)

draft-zhang-pce-stateful-pce-app-01.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on January 13, 2013.

Abstract

The Path Computation Element (PCE) provides a solution for Traffic Engineering (TE) based path calculation in large, multi-domain, multi-region, or multi-layer networks. Depending on whether a PCE keeps information about LSPs and reserved resource usage in the network or not, it can be categorized as either stateful or stateless.

This memo describes general considerations for stateful PCE(s) and examines its applicability through a number of typical scenarios. It shows how stateful PCE(s) can be applied to facilitate these applications. PCEP extensions required for stateful PCE usage are covered in separate document(s).

Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

Table of Contents

Table of Contents	2
1. Introduction	3
2. General Considerations.....	4
2.1. Architectural Considerations.....	4
2.2. LSP State Synchronization.....	5
2.2.1. Single Domain.....	5
2.2.2. Multi-domain.....	6
2.2.3. Multi-layer.....	8
2.3. PCE Survivability/Reliability.....	8
2.4. Delegation and Policy.....	8
2.4.1. Use of Under-construction LSPs Information.....	8
3. Application Scenarios.....	10
3.1. Impairment-Aware Routing and Wavelength Assignment (IA-RWA)	10
3.2. Defragmentation in Flexible Grid Networks	11
3.3. Recovery	12
3.3.1. Protection.....	12
3.3.2. Restoration.....	13
3.4. SRLG Diversity	14
3.5. Maintenance of Virtual Network Topology (VNT)	15
3.6. Global Concurrent Optimization (GCO)	15
3.7. Point-to-Multipoint (P2MP) Application	16
3.8. Time-based Scheduling	16
4. Manageability Considerations	17

4.1. Information and Data Models	17
5. Security Considerations	17
6. References	17
6.1. Normative References	17
6.2. Informative References	18
7. Contributors' Address	20
Authors' Addresses	20

1. Introduction

[RFC 4655] defines the architecture for a Path Computation Element (PCE)-based model for the computation of Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs). To perform such a constrained computation, a PCE stores the network topology (i.e., TE links and nodes) and resource information (i.e., TE attributes) in its TE Database (TED). To request path computation services to a PCE, [RFC 5440] defines the PCE Communication Protocol (PCEP) for communications between a Path Computation Client (PCC) and a PCE, or between two PCEs. A PCC can initiate a path computation request to a PCE through a Path Computation Request (PCReq) message, and then the PCE will return the computed path to the requesting PCC in response to a previously received PCReq message through a PCEP Path Computation Reply (PCRep) message.

As per [RFC 4655], a PCE can be either stateful or stateless. Compared to a stateless PCE, a stateful PCE stores not only the network states, but also the set of computed paths and 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. Note that [RFC4655] further specifies that the TED contains link state and bandwidth availability as distributed by the IGPs or collected via other methods. Even if such information can provide increased granularity and more detail, it is not state information in the PCE context and so a model that uses it is still described as a stateless PCE.

As described in section 6.8 of [RFC 4655], there are many applications which can benefit from stateful PCE(s), e.g.:

- o Minimum perturbation: stateful PCE(s) can minimize the number of existing TE LSPs that are affected and preempted by a higher-priority TE LSP request in a crowded network.
- o Virtual Network Topology (VNT) maintenance: the information of existing LSPs in the higher layer is used as an input for setting up/tearing down the LSPs in the lower layer (i.e., VNT modification).

Besides these scenarios, there are some additional scenarios that should be investigated further. For instance, in impairment-aware Wavelength Switched Optical Networks (WSO) [WSO-Impairment], stateful PCEs could be used to perform Impairment-Aware Routing and Wavelength Assignment (IA-RWA) procedures. In this case, PCE(s) need to know the detailed information of the existing LSPs so that the new LSP(s) will not impact them. Such PCE(s) would maintain the existing LSPs states (e.g., route, wavelength and speed) to perform impairment aware RWA procedures simpler and with less protocol overhead.

[RFC 4655] also discusses potential scalability and synchronization issues in order to implement stateful PCE(s). The main problem pointed out by [RFC 4655] is that a PCE would be constrained if the states of all the TE LSPs in a network are to be maintained by a PCE. Moreover, such state, when there are multiple PCEs, needs to be properly synchronized. These issues are especially relevant in packet networks, such as MPLS-TE networks, given a potentially large number of LSPs. Nonetheless, it is expected that in transport networks, such as OTN networks, the number of the LSPs will be much smaller, which makes stateful PCEs more applicable. Finally, with the increasing power and memory of the hardware platforms that a PCE may run, the number of LSPs that can be managed by a PCE is significantly large. Hence, there is lesser scaling issue for a PCE to store all the LSPs' states, especially for a transport network.

This document presents general considerations for stateful PCE(s) and several examples of its application scenarios. It exhibits the utility of stateful PCE(s) in effective support of these applications to obtain better performance. PCEP extension details are covered in separate documents [stateful-PCEP-mpls], [stateful-PCE-gmpls].

2. General Considerations

2.1. Architectural Considerations

Several PCE architectures are described in Section 5 of [RFC4655]. A stateful PCE needs to maintain a large amount of data and potentially incur in a very high amount of control plane overhead.

Moreover, there might be high computational demands on stateful PCE entities to effectively support the applications listed in Section 3. Therefore, the composite PCE architecture is NOT RECOMMENDED to support stateful PCEs. It does not exclude the possibility that multiple PCEs with different capabilities are included in the network. For example, both stateless and stateful PCEs can co-exist to be in charge of path computation of different types. In all cases, the stateful capability of PCE should be made known within the domain.

2.2. LSP State Synchronization

As suggested by the definition, a stateful PCE maintains two databases for path computation. The first one is the Traffic Engineering Database (TED) which includes the topology and resource in the network. TED can be obtained through participating in routing distribution of TE information or other means as explained in Section 6.7 of [RFC4655].

The other database is the LSP state Database (LSP-DB), in which a PCE stores attributes of all existing LSPs in the network, such as payload signal, switching types and bandwidth/resource usage etc. A stateful PCE should gather the LSP information either from the network management system (NMS) or from the nodes in the network. For a NMS-based PCE, if the PCE is not co-located with the NMS, a standard communication protocol might be needed for LSP state synchronization; otherwise, proprietary APIs can be used. If a PCE relies on network nodes for state synchronization, the strategies may vary depending on the network scenarios in which the PCE is applied to (i.e., single domain, multiple domain or multi-layer networks.) as well as the adoption of PCE computation model.

2.2.1. Single Domain

In a single domain network, LSP state information is maintained locally by the nodes initiating LSP(s). Therefore, PCE(s) should gather the LSP state information either passively or actively from the nodes in the network they have visibility. With a centralized stateful PCE computation model, it is straightforward that all nodes in the domain could communicate with the PCE for its LSP-DB synchronization. As for distributed stateful PCE computation model (i.e., there are multiple stateful PCEs in the network), there are several alternatives for synchronization:

- o Every node can update the PCE LSP-DBs by sending the LSP state information to each of the PCEs in the network separately.

- o Another feasible strategy is to choose one of the PCEs (i.e., a designated PCE) for synchronization with all the nodes in the network and the designated PCE also updates the LSP-DBs of all the other PCE(s).
- o A mixed of these two methods listed above can also be considered in which more than one PCEs (e.g., two PCEs) are chosen to interact directly with nodes in the network for state synchronization while other PCEs are updated via these PCEs.

2.2.2. Multi-domain

In a multi-domain network with a centralized PCE model, the LSP state synchronization is similar to that of a single domain scenario. If there is a stateful PCE responsible for performing path computation within each domain, the LSPs (segments) traversing the domain/layer should be synchronized to the PCE.

As described in [RFC4726], there are four methods to set up a LSP traversing multiple domains: LSP nesting, contiguous LSP, LSP stitching and hybrid methods, respectively. Hence, the ingress nodes of a LSP traversing a domain may exist in another domain (e.g., a contiguous LSP spanning across multiple domains). In this case, the border node of a domain (i.e., an intermediate node of a LSP), could be responsible for synchronizing the LSP segment in the domain to the PCE.

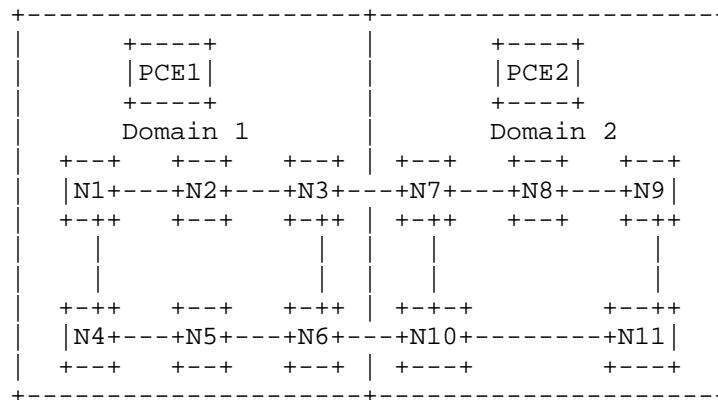


Figure 1: Multi-domain Scenario

Figure 1 shows an example of multi-domain scenario. Suppose a contiguous LSP traverses N1-N2-N3-N7-N8-N9. Then in domain 1, the ingress node of the LSP (i.e., N1) SHOULD synchronize the state of

the LSP segment N1-N2-N3 to PCE1. In domain 2, the border node (i.e., N7) SHOULD synchronize the state of the LSP segment N7-N8-N9 to PCE2.

This approach requires that N7 has a PCEP adjacency with its PCE (PCE2), i.e. setting up a PCEP session, for LSP state synchronization purpose even if no path computation expansions are required. N7 needs to check whether its RSVP-TE upstream node belongs to another domain and notify the PCE when the LSP is released. Note that synchronization may require detailed information of the LSP (e.g., a full record route, the actual reserved resources) which may only be available during Resv message processing.

Alternatively, inter-PCE communication strategy can be adopted for LSP-DB synchronization. For instance, in Figure 1, upon the notification of the setup of LSP N1-N2-N3-N7-N8-N9, PCE1 can establish a PCEP adjacency to inform PCE2 to update its LSP-DB. This method SHOULD be preferred only when PCE1 has sufficient and valid information of the across-domain LSP, such as explicit LSP information. Otherwise, the method in which the border node(s) are in charge of LSP state update is more appropriate. For example, Backward Recursive Path Computation (BRPC) [RFC5441] in conjunction with path-key-based mechanism [RFC5520] can be adopted for inter-domain path computation. If this is the case with the example in Figure 1, PCE1 only acquires a loose LSP path (e.g., N1-N2-N3-N7-KEY1, where KEY1 can be interpreted only by PCE2). Since it depends on the local policy that how long a Path-Key should be stored, KEY1 might not be valid anymore when it is used by PCE1 for PCE2 LSP-DB update notification. In this case, N7 will need to request PCE2 to unlock the Path-Key in order to complete the signaling process. Therefore, it is possible to use N7 instead for updating PCE2 LSP-DB.

Note that a timely synchronization of PCEs and these two databases is a prerequisite to maintaining a good performance of a stateful PCE.

To benefit from stateful PCE, during inter-domain path computation procedure, PCC and cooperating PCEs should try to select stateful PCE when multiple PCEs (stateful and stateless) are available in the domain. This will enable correct end-to-end path computation using of TED and LDP-DB in all domains. In case of unavailability of stateful PCE, stateless PCE can still be used to provide the inter-domain path computation.

The inter-domain LSP synchronization as explained in this section is still applicable if some domain does not have stateful PCE support. All the domains with a stateful PCE present should synchronize their segment at the least.

2.2.3. Multi-layer

In multi-layer scenarios, one node/domain may have multiple switching capabilities. For instance, Optical Transport Network (OTN) nodes may have both of electrical (e.g., ODU1, ODU2, ODU3) and optical switch capabilities. ODU LSPs and wavelength LSPs may be established in an OTN network.

In such networks, a PCE may have the capability of performing single layer path computation or multi-layer path computation. If a stateful PCE has single layer path computation capability, the nodes should be aware of information pertaining to which layer should be synchronized to a specific PCE. Otherwise, the state of the LSPs in all layers should be synchronized to the single stateful PCE.

2.3. PCE Survivability/Reliability

Since a PCE supports a centralized path computation model, its survivability should be carefully considered to ensure its proper operation. If a multiple stateful PCE model is used and these PCEs have a consistent view of the network, they can act as a hot backup for each other. Otherwise, other backup strategies SHOULD be present if only one PCE is deployed in the network to avoid a single point of failure.

2.4. Delegation and Policy

Stateful PCE(s) are still subject to policies when performing path computation based on TED and LSP-DB as well as in what concerns LSP-DB organization and maintenance.

For LSP-DB maintenance, a basic function of stateful PCEs that SHOULD be supported is the ability to keep LSP state information in the network within which they have visibility. OPTIONALLY, a stateful PCE can also extend its ability to support modification of LSP state information. This can be realized by obtaining the temporal LSP state control through negotiation with LSRs (i.e., LSP delegation). Please note that LSP state delegation should comply with the policy imposed by LSP state owner (i.e., LSRs) as well as the policy imposed upon PCE(s).

2.4.1. Use of Under-construction LSPs Information

The TED and/or LSP-DB information retained by a stateful PCE might be out-of-syn. If this is the case, it might cause resource contention when the PCE computes paths based of the out-of-date information. Some sources of the potential TED/LSP-DB inaccuracy are:

- o Control plane link latencies. Such latencies may be increased due to several factors such as:

- a) The time required for a PCC to obtain the paths after a successful computation, requiring several Round-Trip-Times (RTT) as per TCP;

- b) The setup delay;

- c) The time it takes for the PCE to update the local TED given IGP update times;

- o The routing and topology dissemination protocol (i.e. OSPF-TE), which may operate with timers for LSA updates, to avoid excessive control plane overhead.

- o Concurrent requests that arrive during the time window, between a response is sent and the LSP is setup and the topology changes are flooded. Even for very fast networks with low latency, there may be a batched of requests: several path computation requests within a PCReq message or, in dynamic restoration without pre-planning, several LSPs that need to be rerouted so as to avoid a failed link.

- o Local PCE contention, where the PCE needs to concurrently serve path computation requests and update the LSA (e.g. parsing OSPF-TE LSA updates). A PCE implementation may need to find a trade-off, when synchronizing access to the local TED: favor OSPF-TE parsing which means that some path computations are slightly delayed to allow an 'update' to be processed, or give strict priority to computation requests.

In consequence, a PCE may assign the same (or a subset of the same) resources to several requests. Thus, it may result in contention and degraded network performance since it might cause path setup failure and excessive crank-backs.

Therefore, information of the LSPs that are under construction can be used together with the TED and LSP-DB by a stateful PCE to reduce the path blocking and crank-backs issues. For example, the PCE can retain some context from paths it has recently computed so that it avoids suggesting the use of the same resources for other TE LSPs, using heuristics / statistic or forecasting for improved resource (i.e. wavelength) allocation. In other words, a given PCE implementation may decide to perform additional book-keeping and management of resources strategies using the information of under construction LSPs, deploying policies that prevent sub-optimal allocations. For instance, a PCE may compute the mean time used to update the TED based on the previous calculated TE-LSPs and TED

updates. Those kinds of mechanisms may reduce the TED inaccuracy but in all cases they cannot infer the PCC use of the TE-path.

3. Application Scenarios

In this section, several examples exploiting the capabilities of stateful PCE(s) are presented, although the application of stateful PCE(s) is not limited to them. In general, stateful PCE(s) can be deployed for applications where LSP state as well as traffic engineering information in the network are necessary inputs to achieve one or multiple of the following goals:

- o Improving the performance such as reducing network blocking probability, achieving load balancing, improve network resources utilization or increasing the route computation success rate;
- o Reducing the complexity of the relevant procedure(s) associated with the application(s);
- o Lowering resource consumption;

As discussed in [PSU-WSN] and [LCA-Stateless], some of the objectives can be achieved through limited LSP awareness in stateless PCE by exploiting objects defined in existing protocols, such as the SVEC object defined in [RFC5440] and/or XRO object defined in [RFC5521]. These methods are considered as transitional solutions because of two reasons. Firstly, these methods only have local/partial/temporal LSP related information and thus have limited utility in terms of achieving the goals, particularly for objectives set at a network level. Secondly, it might incur a substantial amount of overhead since it requires frequent message exchanges among PCC and PCE entities.

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

In WSON networks [RFC6163], a wavelength-switched LSP traverses one or multiple 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 PCE(s) are used to perform the IA-RWA procedure, it needs 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., [GEN-OSPF] and [WSON-OSPF], 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.

3.2. Defragmentation in Flexible Grid Networks

Traditionally, in Dense Wavelength Division Multiplexing (DWDM) networks, the frequency and channel spacing for a single wavelength allocated to an optical connection is fixed, in terms of a fixed channel spacing grid. With the development of mixed-rate transmission and the increase in the speed of optical signal, the issue of poor optical spectrum usage needs to be addressed. Flexible grid is proposed to solve this problem [G.FLEXIGRID]. In Flexible grid networks, LSPs with different slot widths (such as 12.5G, 25G etc.) can co-exist so as to accommodate the services with different bandwidth requests.

Yet another problem arises in this type of DWDM networks. Since in flexible grid networks LSPs are dynamically allocated and released over time, the optical spectrum resource becomes fragmented. The overall available spectrum resource on a link might be sufficient

for a new LSP request. But if the available spectra are not continuous, the request would be rejected. In order to perform frequency defragmentation procedure, stateful PCE(s) COULD be used, since existing TE LSPs information (i.e., slot width and spectrum location information associated with TE LSPs) is required to accurately assess spectrum resources on the LSPs, and perform defragmentation while ensuring a minimal disruption of the network, e.g., based on active LSP priorities.

[Editor's note: it is not suggested to start PCEP extensions on this application until the data plane technology and the corresponding GMPLS control is mature.]

3.3. Recovery

3.3.1. 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. In either way, the resources bound to backup paths can be shared by different LSPs to improve the overall network efficiency. 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. If such information is required on each network node, extensions to existing signaling or routing protocols are needed in order to carry the necessary information for avoiding allocating shared protection resources for two non-disjoint working LSPs. However, stateful PCE(s) can easily accommodate this need using the information stored in its LSP-DB, without requiring extensions to existing routing protocols.

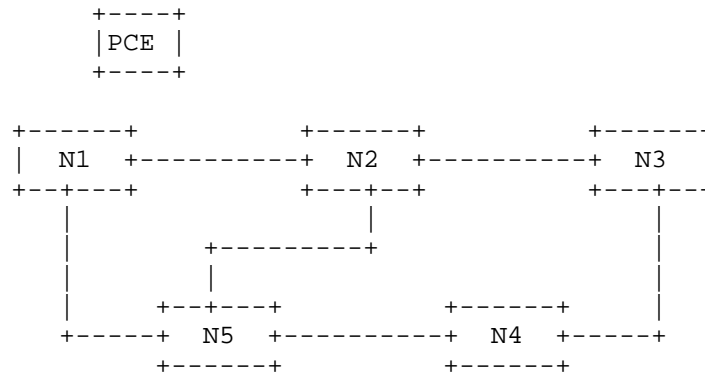


Figure 2: Example Network

For example, in the network depicted in Figure 2, 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 (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.

3.3.2. 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 2: (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.

3.4. SRLG Diversity

A common requirement 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 [RFC 5521]. 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 the PCEa 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 LSPs' identifiers.

3.5. 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 to provide 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. Traditionally, 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 , requiring the cooperation of a NMS or a VNT cooperation with VNTMmanager (VNTM). On the other hand, Therefore, 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.

3.6. Global Concurrent Optimization (GCO)

GCO is introduced in [RFC5557] to calculate multiple paths concurrently so as to improve network resource efficiency. By taking into consideration the network topology as well as existing TE LSPs information, GCO can (re)optimize the entire network simultaneously. Alternatively, GCO can be applied to (re)optimize one or a subset of existing TE LSPs or plan for forthcoming LSP(s) with specific objectives. GCO can also support off-line one-time optimization (i.e., planning) given a traffic matrix and network topology. Due to its complexity and potentially high computational demand, it is recommended to be performed in a centralized way (e.g., based on a management-based PCE).

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 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.

3.7. Point-to-Multipoint (P2MP) Application

Route computation for P2MP application involves selection of branching points together with calculating multiple sub-LSPs with certain objective(s) such as minimizing the overall cost of the P2MP tree. Moreover, egress nodes addition and removal in a P2MP tree necessitates (re)optimization. Besides these, there are also some constraints and policies that make the P2MP tree computation hard, requiring high computation power. Therefore, PCE is proposed to support P2MP application [RFC5671].

If a stateless PCE is used for P2MP calculation or optimization under constraints such as load balancing or path disjointedness, then a large amount of sub-LSP information might need to be exchanged between the PCE and the requesting entities. Moreover, if the requesting entity cannot provide complete information of sub-LSPs pertaining to the P2MP tree, then the performance of stateless PCE will be sub-optimal. On the contrary, a stateful PCE can support the P2MP tree computation/optimization with reduced overhead and improved efficiency.

3.8. Time-based Scheduling

Time-based 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 time-based scheduling requests. In order to support this application, a stateful PCE should also maintain a database that stores all the reserved information with time reference. This can be achieved either by maintaining a separate database or incorporated into LSP-DB. The details of organizing time-based scheduling related information as well as its impact on LSP-DB is subject to network provider's policy and administrative consideration and thus outside of the scope of this document.

4. Manageability Considerations

The description and functionality specifications presented related to stateful PCE(s) should also comply with the manageability specifications covered in Section 8 of [RFC4655]. Furthermore, a further list of manageability issues presented in [Stateful-PCEP-mppls] may also be considered.

4.1. Information and Data Models

A Management Information Base (MIB) module for management of the PCEP is being specified in a separate document [PCEP-MIB]. That MIB module allows examination of individual PCEP messages, in particular requests, responses and errors. The MIB module MUST be extended to include the ability to view stateful PCE PCEP extensions defined in relevant documents.

5. Security Considerations

The security issues presented in [RFC5440] still applies to this document. In addition, the security concerns raised by [Stateful-PCEP-mppls] may also be considered.

6. References

6.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to indicate requirements levels", RFC 2119, March 1997.
- [RFC4655] Farrel, A., Vasseur, J.-P., and Ash, J., "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.

- [RFC5440] Vasseur, J.-P., and Le Roux, JL., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.
- [RFC6163] Lee, Y., Bernstein, G., "Framework for GMPLS and Path Computation Element (PCE) Control of Wavelength Switched Optical Networks (WSONs)", RFC 6163, April, 2011.
- [RFC5521] Oki, E., Farrel, A., "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC5521, April 2009.

6.2. Informative References

- [WSN-Impairment] Lee, Y., Bernstein, G., Li, D., Martinelli, G., "A Framework for the Control of Wavelength Switched Optical Network (WSN) with Impairments", draft-ietf-ccamp-wson-impairments, work in progress.
- [RFC4726] Farrel, A., Vasseur, J.-P., Ayyangar, A., "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.
- [RFC5520] Bradford, R., Vasseur, JP., Farrel, A., "Preserving Topology Confidentiality in Inter-Domain Path Computation Using a Path-Key-Based Mechanism", RFC 5520, April 2009.
- [RFC5441] Vasseur, J.-P., Zhang, R., Bitar, N., Le Roux, JL., "A Backward-Recursive PCE-Based Computation (BRPC) Procedure to Compute Shortest Constrained Inter-Domain Traffic Engineering Label Switched Paths", RFC 5441, April 2009.
- [PSU-WSN] Giorgetti, A, Cugini, G, et al, "Path state-based update of PCE traffic engineering database in wavelength switched optical networks", IEEE Com. Let., June 2010.
- [LCA-Stateless] Gonzalez de Dios, O., et al, "Benefits of limited context awareness in stateless PCE", Optical Fiber Communication Conference, March 2011.
- [WSN-OSPF] Lee, Y., Bernstein, G., "GMPLS OSPF Enhancement for Signal and Network Element Compatibility for Wavelength Switched Optical Networks", draft-ietf-ccamp-wson-signal-compatibility-ospf-07, October 2011.

- [GEN-OSPF] Zhang, Fatai, Lee, Y., Han, Jianrui, Bernstein, G., Xu, Yunbin, "OSPF-TE Extensions for General Network Element Constraints", draft-ietf-ccamp-gmpls-general-constraints-ospf-te-02, September 2011.
- [G.FLEXIGRID] Draft revised G.694.1 version 1.3, Unpublished ITU-T Study Group 15, Question 6.
- [RFC5212] Shiimoto, K., Papadimitriou, D., Le Roux, JL., Vigoureux, M., Brungard, D., "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC 5212, July 2008.
- [RFC5557] Lee, Y., Le Roux, JL., King, D., Oki E., "Path Computation Element Communication Protocol (PCEP) Requirements and Protocol Extensions in Support of Global Concurrent Optimization", RFC 5557, July, 2009.
- [RFC5671] Yasukawa, S., Farrel, A., "Applicability of the Path Computation Element (PCE) to Point-to-Multipoint (P2MP) MPLS and GMPLS Traffic Engineering (TE)", October, 2009.
- [RFC5623] Oki, E., Takeda, T., Le Roux, JL., Farrel, A., "Framework for PCE-Based Inter-Layer MPLS and GMPLS Traffic Engineering", RFC5623, September 2009.
- [stateful-PCEP-mpls] Crabbe, E., Medved, J., Varga, R., Minei, I., "'PCEP Extensions for Stateful PCE'", draft-ietf-pce-stateful-pce, work in progress.
- [stateful-PCEP-gmpls] Zhang, X., Lee, Y., Casellas, R., Gonzalez de Dios, O., "' Path Computation Element (PCE) Protocol Extension for Stateful PCE Usage in GMPLS Networks'", draft-zhang-pce-pcep-stateful-pce-gmpls, work in progress.
- [PCEP-MIB] Kiran Koushik, A S., Stephan, E., Zhao, Q., King, D., "PCE communication protocol (PCEP) Management Information Base", draft-ietf-pce-pcep-mib, work in progress.

7. Contributors' Address

Dhruv Dhody
Huawei Technology
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruvd@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

Authors' Addresses

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

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

Phone: +86-755-28972913
Email: zhang.xian@huawei.com

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

Ramon Casellas
CTTC - Centre Tecnologic de Telecomunicacions de Catalunya
Av. Carl Friedrich Gauss n7
Castelldefels, Barcelona 08860
Spain

Phone:
Email: ramon.casellas@cttc.es

Oscar Gonzalez de Dios
Telefonica Investigacion y Desarrollo
Emilio Vargas 6
Madrid, 28045
Spain

Phone: +34 913374013
Email: ogondio@tid.es

Intellectual Property

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions.

For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of RFC 5378. No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under RFC 5378, shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Full Copyright Statement

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

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Network Working Group
Internet-Draft
Intended status: Informational
Expires: November 25, 2013

X. Zhang, Ed.
Huawei Technologies
I. Minei, Ed.
Juniper Networks, Inc.
May 24, 2013

Applicability of Stateful Path Computation Element (PCE)
draft-zhang-pce-stateful-pce-app-04

Abstract

A stateful Path Computation Element (PCE) maintains information about Label Switched Path (LSP) characteristics and resource usage within a network in order to provide traffic engineering calculations for its associated Path Computation Clients (PCCs). This document describes general considerations for a stateful PCE deployment and examines its applicability and benefits through a number of use cases. Path Computation Element Protocol (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 November 25, 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	6
5.1. Optimization of LSP placement	6
5.1.1. Throughput Maximization and Bin Packing	7
5.1.2. Deadlock	8
5.1.3. Minimum Perturbation	10
5.1.4. Predictability	11
5.2. Auto-bandwidth Adjustment	12
5.3. Bandwidth Scheduling	13
5.4. Recovery	13
5.4.1. Protection	13
5.4.2. Restoration	15
5.4.3. SRLG Diversity	16
5.5. Maintenance of Virtual Network Topology (VNT)	16
5.6. LSP Re-optimization	17
5.7. Resource Defragmentation	17
5.8. Impairment-Aware Routing and Wavelength Assignment (IA-RWA)	18
6. Security Considerations	19
7. Contributing Authors	19
8. Acknowledgements	21
9. References	21
9.1. Normative References	21
9.2. Informative References	21
Appendix A. Editorial notes and open issues	23
Authors' Addresses	23

1. Introduction

[RFC4655] defines the architecture for a Path Computation Element (PCE)-based model for the computation of Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) Traffic Engineering Label Switched Paths (TE LSPs). To perform such a constrained computation, a PCE stores the network topology (i.e., TE links and nodes) and resource information (i.e., TE attributes) in its TE Database (TED). [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 two PCEs, 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. Stateless PCEs have been shown to be useful in many scenarios, including constraint-based path computation in multi-domain/multi-layer networks. Compared to a stateless PCE, a stateful PCE has access to not only the network state, but also to the set of active paths and their reserved resources. Furthermore, a stateful PCE might also retain information regarding LSPs under construction in order to reduce churn and resource contention. This state allows the PCE to compute constrained paths while considering individual LSPs and their interactions. Note that this requires reliable state synchronization mechanisms between the PCE and the network, PCE and PCC, and between cooperating PCEs, with potentially significant control plane overhead and maintenance of a large amount of state data, as explained in [RFC4655].

This document describes how a stateful PCE can be used to solve various problems for MPLS-TE and GMPLS networks, and the benefits it brings to such deployments. Note that alternative solutions relying on stateless PCEs may also be possible for some of these use cases, and will be mentioned for completeness where appropriate.

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 term:

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 referenced 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 within 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 IGPs for PCE discovery.

4. Deployment considerations

This section discusses generic issues with Stateful PCE deployments, and how specific protocol mechanisms can be used to address them.

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 may 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 sets of information for use in path computation. The first 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 is the LSP State Database (LSP-DB), in which a PCE stores attributes of all active LSPs in the network, such as their paths through the network, bandwidth/resource usage, switching types and LSP constraints. The stateful PCE extensions defined in [I-D.ietf-pce-stateful-pce] support population of this database using information received from the network nodes via LSP State 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 Label Edge Router (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. Note that locally recovering the state would still require some degree of resynchronization to ensure that the recovered state is indeed up-to-date.

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.

Some of the use cases below are focused on MPLS-TE deployments, but may also apply to GMPLS. Unless otherwise cited, use cases assume that all LSPs listed exist at the same LSP priority.

The main benefit in the cases below comes from moving away from an asynchronous PCC-driven mode of operation to a model that allows for central control over LSP computations and setup, and focuses specifically on the active stateful PCE model of operation.

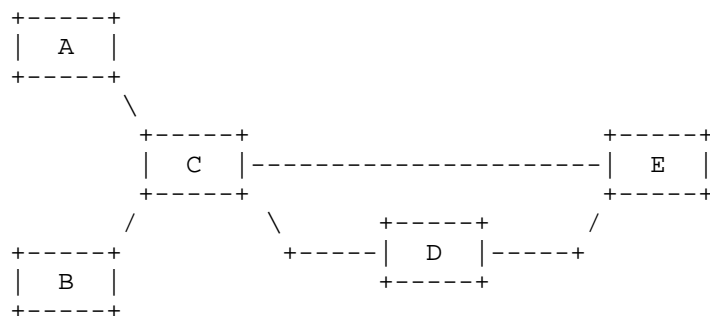


Figure 1: Reference topology 1

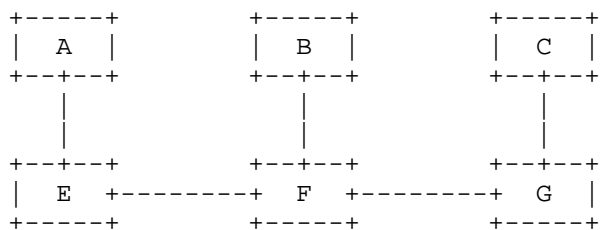


Figure 2: Reference topology 2

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, as will be shown in the example below.

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

5.1.2. Deadlock

This section discusses a use case of cross-LSP impact under degraded operation. 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 dynamic ingress admission control (policing of the traffic volume mapped onto an LSP) at the LER. 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 for stateful PCE, but has not been yet at the time of this writing).

Lack of ingress admission control coupled with the behavior in [RFC3209] may result in LSPs operating out of profile for significant periods of time. It is reasonable to expect that these out-of-profile LSPs will be operating in a degraded state and experience traffic loss, but because they end up sharing common network interfaces with other LSPs operating within their bandwidth reservations, they will end up impacting the operation of the in-profile LSPs, even when there is unused network capacity elsewhere in the network. Furthermore, this behavior will cause information loss in the TED with regards to the actual available bandwidth on the links used by the out-of-profile LSPs, as the reservations on the links no longer reflect the capacity used.

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 resigaled. 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 'Degraded operation' 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: Degraded operation 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 (LSP2) is established, preempting LSP1, potentially causing traffic loss. LSP1 is then reestablished on 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

A stateful PCE can help in this scenario by evaluating both requests at the same time (due to their proximity in time). This will ensure placement of the more important LSP along the shortest path, avoiding the preemption of the lower priority LSP.

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. Predictable results are valuable for being able to simulate the network and reliably test it under various scenarios, especially under various failure modes and planned maintenances when predictable path characteristics are desired under contention for network resources.

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

As can be shown in the example, both LSPs were routed in both cases, but along very different paths. This would be a challenge if reliable simulation of the network was attempted. A stateful PCE can solve this through control over LSP ordering.

5.2. Auto-bandwidth Adjustment

The bandwidth requirement of LSPs often change over time, requiring resizing the LSP. Currently the head-end node performs this function by monitoring the actual bandwidth usage, triggering a recomputation and resignaling when a threshold is reached. This operation is referred as auto-bandwidth adjustment. The head-end node either recomputes the path locally, or it requests a recomputation from a PCE by sending a PCReq message. In the latter case, the PCE computes a new path and provides the new route suggestion. Upon receiving the reply from the PCE, the PCC re-signals the LSP in Shared-Explicit (SE) mode along the newly computed path. If a passive stateful PCE is used, only the new bandwidth information is needed to trigger a path re-computation since the LSP information is already known to the PCE. Note that in this scenario, the head-end node is the one that drives the LSP resizing based on local information, and that the difference between using a stateless and a passive stateful PCE is in the level of optimization of the LSP placement as discussed in the previous section.

A more interesting smart bandwidth adjustment case is one where the LSP resizing decision is done by an external entity, with access to additional information such as historical trending data, application-specific information about expected demands or policy information, as well as knowledge of the actual desired flow volumes. In this case

an active stateful PCE provides an advantage in both the computation with knowledge of all LSPs in the domain and in the ability to trigger bandwidth modification of the LSP.

5.3. Bandwidth Scheduling

Bandwidth scheduling allows network operators to reserve resources in advance according to the agreements with their customers, and allow them to transmit data with specified starting time and duration, for example for a scheduled bulk data replication 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 passive stateful PCE can support this application with better efficiency since it can alleviate the burden of processing on network elements. This requires the PCE to maintain the scheduled LSPs and their associated resource usage, as well as the ability of head-ends to trigger signaling for LSP setup/deletion at the correct time. This approach requires coarse time synchronization between PCEs and PCCs. If an active stateful PCE is available, the PCE can trigger the setup/deletion of scheduled requests in a centralized manner, without modification of existing head-end behaviors.

5.4. Recovery

The recovery use cases discussed in the following sections show how leveraging a stateful PCE can simplify the computation of recovery path(s). In particular, two characteristics of a stateful PCE are used: 1) using information stored in the LSP-DB for determining shared protection resources and 2) performing computations with knowledge of all LSPs in a domain.

5.4.1. 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. A stateless PCE can accommodate this use case by having the PCC pass in this information as a constraint to the path computation request. A stateful PCE can more easily accommodate this need using the information stored in its LSP-DB.

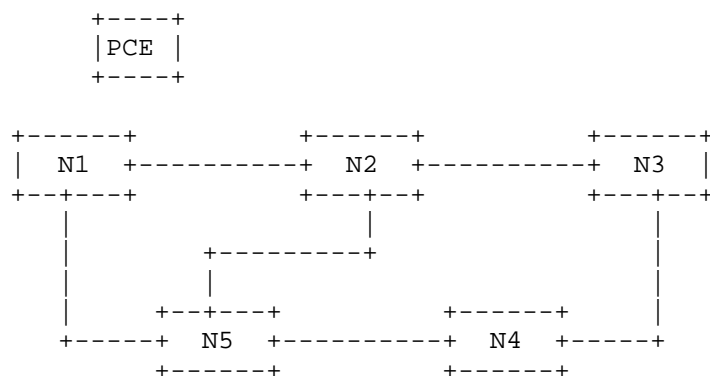


Figure 3: Reference topology 3

For example, in the network depicted in Figure 3, suppose there exists LSP1 with working path LSP1_working following N1->N5 and with backup path LSP1_backup following N1->N2->N5. A request arrives asking for a working and backup path pair to be computed for LSP2, for a request from N2 to N5. If the PCE decides LSP2_working follows N2->N1->N5, then the backup path LSP2_backup should not use the same protection resource with LSP1 since LSP2 shares part of its resource (specifically N1->N5) 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 for LSP2_working.

If a stateless PCE is used, the head node N2 needs to be aware of the existence of LSPs which share the route of LSP2_working and of the details of their protection resources. N2 must pass this information to the PCE as a constraint so as to request a path with SRLG diversity. On the other hand, a stateful PCE can get the LSPs information by itself and can achieve the goal of finding SRLG-diversified protection paths for both LSPs. This is made possible by comparing the LSP resource usage exploiting the LSP DB accessible by the stateful PCE.

5.4.2. 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 the existing signaling protocols.

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.

5.4.3. SRLG Diversity

An alternative way to achieve efficient resilience is to maintain SRLG disjointness between LSPs, irrespective of whether these LSPs share the source and destination nodes or not. 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. A stateless PCE can compute an SRLG-disjoint path by inspecting the TED and precluding the links with the same SRLG values specified in the PCReq message sent by a PCC.

A stateful PCE maintains the updated SRLG information of the established LSPs in a centralized manner. Therefore, the PCC can specify as constraints to the path computation the SRLG disjointness of a set of already established LSPs by only providing the LSP identifiers.

5.5. 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 setting up connections in the lower layer to provide TE links for upper layer. Hence, when a stateless PCE cannot find the route for a request based on the upper layer topology information, it needs to interact with the VNTM and rely on the VNTM to decide whether to set up or remove a TE link or not. On the other hand, a stateful PCE can make the decision of when and how to modify the VNT either to accommodate new LSP requests or to re-optimize resource usage across layers irrespective of the PCE models as described in [RFC5623].

5.6. LSP Re-optimization

In order to make efficient usage of network resources, it is sometimes desirable to re-optimize one or more LSPs dynamically. In the case of a stateless PCE, in order to optimize network resource usage dynamically through online planning, a PCC must send a request to the PCE together with detailed path/bandwidth information of the LSPs that need to be concurrently optimized. This means the PCC must be able to determine when and which LSPs should be optimized. In the case of a stateful PCE, given the LSP state information in the LSP database, the process of dynamic optimization of network resources can be automated without requiring the PCC to supply LSP state information or to trigger the request. Moreover, since a stateful PCE can maintain information for all LSPs that are in the process of being set up and since it may have the ability to control timing and sequence of LSP setup/deletion, the optimization procedures can 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. A 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.

5.7. 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 is rejected. In order to perform the defragmentation procedure, stateful PCEs can be used, since global visibility of LSPs in the network is required to accurately assess resources on the

LSPs, and perform de-fragmentation while ensuring a minimal disruption of the network. This use case cannot be accommodated by a stateless PCE since it does not possess the detailed information of existing LSPs in the network.

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 it is 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 fragmentation.

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

In WSONs [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 beyond those discussed in [I-D.ietf-pce-stateful-pce].

The following topics will be discussed in a future version of this document: whether use of a stateful PCE makes the network more or less secure, and security use cases if any.

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
Email: unknown

8. Acknowledgements

We would like to thank Cyril Margaria, Adrian Farrel and JP Vasseur for the useful comments and discussions.

9. References

9.1. Normative References

- [I-D.ietf-pce-stateful-pce]
Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE",
draft-ietf-pce-stateful-pce-04 (work in progress),
May 2013.
- [RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, August 2006.
- [RFC5440] Vasseur, JP. and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, March 2009.

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-01 (work in progress), May 2013.
- [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]
Margaria, C., Dios, O., and F. Zhang, "PCEP extensions for GMPLS", draft-ietf-pce-gmpls-pcep-extensions-07 (work in progress), May 2013.

progress), October 2012.

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

Dios, O., Casellas, R., Zhang, F., Fu, X., Ceccarelli, D., and I. Hussain, "Framework and Requirements for GMPLS based control of Flexi-grid DWDM networks", draft-ogrcetal-ccamp-flexi-grid-fwk-02 (work in progress), February 2013.

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

Sivabalan, S., Medved, J., and X. Zhang, "IGP Extensions for Stateful PCE Discovery", draft-sivabalan-pce-disco-stateful-01 (work in progress), April 2013.

[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.

[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.

[RFC4427] Mannie, E. and D. Papadimitriou, "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4427, March 2006.

[RFC4657] Ash, J. and J. Le Roux, "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, September 2006.

[RFC5212] Shiimoto, 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.

[RFC5394] Bryskin, I., Papadimitriou, D., Berger, L., and J. Ash, "Policy-Enabled Path Computation Framework", RFC 5394,

December 2008.

- [RFC5521] Oki, E., Takeda, T., and A. Farrel, "Extensions to the Path Computation Element Communication Protocol (PCEP) for Route Exclusions", RFC 5521, April 2009.
- [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.
- [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.

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

