PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: June 10, 2015

D. Dhody
Q. Wu
Huawei
V. Manral
Ionos Network
Z. Ali
Cisco Systems
K. Kumaki
KDDI Corporation
December 7, 2014

Extensions to the Path Computation Element Communication Protocol (PCEP) to compute service aware Label Switched Path (LSP).

draft-ietf-pce-pcep-service-aware-06

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. The Link Bandwidth Utilization (the total bandwidth of a link in current use for the forwarding) is also an important factor to consider during path computation.

IGP Traffic Engineering (TE) Metric extensions describes mechanisms with which network performance information is distributed via OSPF and IS-IS respectively. 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. This document describes the extension to PCEP to carry Latency, Latency Variation, Packet Loss, and Link Bandwidth Utilization as constraints for end to end path computation.

Status of This Memo

Table of Contents

1. Introduction

Real time network performance is becoming a critical in the path computation in some networks. Mechanisms to measure Latency, Latency-Variation, and Packet Loss in an MPLS network are described in [RFC6374]. Further, there exist mechanisms to measure these network performance metrics 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 set up.

Link bandwidth utilization based on real time traffic along the path is also becoming critical during path computation in some networks. Thus it is important that the Link bandwidth utilization is factored in during path computation itself.

Traffic Engineering Database (TED) is populated with network performance information like link latency, latency variation, and packet loss through [OSPF-TE-METRIC-EXT] or [ISIS-TE-METRIC-EXT]. 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.

[OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT] include parameters related to bandwidth (Residual bandwidth, Available bandwidth and Utilized bandwidth); this document also describes extensions to PCEP to consider them.

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.

LBU: Link Bandwidth Utilization. (See Section 4.2.1.)

LRBU: Link Reserved Bandwidth Utilization. (See Section 4.2.2.)

MPLP: Minimum Packet Loss Path. (See Section 4.3.)

MRUP: Maximum Reserved Under-Utilized Path. (See Section 4.3.)

MUP: Maximum Under-Utilized Path. (See Section 4.3.)

OF: Objective Function. A set of one or more optimization criteria used for the computation of a single path (e.g., path cost minimization) or for the synchronized computation of a set of paths (e.g., aggregate bandwidth consumption minimization, etc). (See [RFC5541].)

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.

RSVP: Resource Reservation Protocol

TE: Traffic Engineering.

3. PCEP Requirements

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

1. PCE supporting this draft MUST have the capability to compute end-to-end path with latency, latency variation, packet loss, and bandwidth utilization constraints. It MUST also support the combination of network performance constraint (latency, latency variation, loss...) with existing constraints (cost, hop-limit...).

2. PCC MUST be able to request for E2E network performance constraint(s) in PCReq message as the key constraint to be optimized or to suggest boundary condition that should not be crossed.

3. The PCC MUST be able to request for the bandwidth utilization constraint in PCReq message as the upper limit that should not be crossed for each link in the path.

4. The PCC MUST be able to request for these constraint in PCReq message as an Objective function (OF) [RFC5541] to be optimized.

5. PCEs are not required to support service aware path computation. Therefore, it MUST be possible for a PCE to reject a PCReq message with a reason code that indicates no support for service-aware path computation.

6. PCEP SHOULD provide a means to return end to end network performance information of the computed path in a PCRep message.

7. 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.

4.1. Extensions to METRIC Object

The METRIC object is defined in section 7.8 of [RFC5440], comprising of metric-value, metric-type (T field) and flags. This document defines the following optional types for the METRIC object.

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

- A network comprises of a set of N links {Li, (i=1...N)}.

- A path P of a P2P LSP is a list of K links {Lpi,(i=1...K)}.

4.1.1. Latency (Delay) Metric

Link delay metric is defined in [OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT]. 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(Lpi), (i=1...K)}.

This is as per sum of means composition function (section 4.2.5 of [RFC6049]).

* Metric Type T=TBD1: Latency metric

PCC MAY use this latency metric in PCReq message 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 latency during path computation, in this case B flag will be cleared.

PCE MAY use this latency metric in PCRep message along with NO-PATH object in case 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-METRIC-EXT] and [ISIS-TE-METRIC-EXT] 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.

4.1.2. Latency Variation (Jitter) Metric

Link delay variation metric is defined in [OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT]. P2P latency variation metric type of METRIC object in PCEP encodes the sum 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) (average delay variation for link L).

- A P2P latency variation metric for the Path P = Sum {DV(Lpi), (i=1...K)}.

Note that the IGP advertisement for link attributes includes average latency variation over a period of time. An implementation, therefore, MAY use sum of the average latency variation of links along a path to derive the average latency variation of the Path. An

implementation MAY also use some enhanced composition function for computing average latency variation of a Path.

* Metric Type T=TBD2: Latency Variation metric

PCC MAY use this latency variation metric in PCReq message 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 latency variation during path computation, in this case B flag will be cleared.

PCE MAY use this latency variation metric in PCRep message along with NO-PATH object in case 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.

4.1.2.1. Latency Variation (Jitter) Metric Value

[OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT] 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.

4.1.3. Packet Loss Metric

[OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT] defines "Unidirectional Link Loss". Packet Loss 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) in percentage.

- A Packet loss in fraction of link L is denoted FPL(L) = PL(L)/100.

- A P2P packet loss metric in percentage for the Path P = (1 - ((1-FPL(Lp1)) * (1-FPL(Lp2)) * .. * (1-FPL(LpK))) * 100 for a path P with link 1 to K.

This is as per the composition function (section 5.1.5 of [RFC6049]).

* Metric Type T=TBD3: Packet Loss metric

PCC MAY use this packet loss metric in PCReq message 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 message along with NO-PATH object in case 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.

4.1.3.1. Packet Loss Metric Value

[OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT] 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.

4.1.4. 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 object.

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].

4.1.5. 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:

- To indicate the metric that MUST be optimized by the path computation algorithm (Latency, Latency Variation or Loss)

- 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.

4.1.5.1. 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=TBD1, C=1, metric-value=0x0000

o Second METRIC object with B=0, T=TBD3, 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 first METRIC object with B=0, T=TBD1, metric-value= computed end to end latency and second METRIC object with B=1, T=TBD3, 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=TBD1, C=1, metric-value=0x0000

o Second METRIC object with B=1, T=TBD3, 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=TBD1, metric-value= computed end to end latency. Additionally, the PCE may insert a second METRIC object with B=1, T=TBD3, metric-value=computed end to end packet loss.

4.2. Bandwidth Utilization

4.2.1. Link Bandwidth Utilization (LBU)

The bandwidth utilization on a link, forwarding adjacency, or bundled link is populated in the TED (Utilized Bandwidth in [OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT]). For a link or forwarding adjacency, the bandwidth utilization represents the actual utilization of the link (i.e., as measured in the router). For a bundled link, the bandwidth utilization is defined to be the sum of

the component link bandwidth utilization. This includes traffic for both RSVP and non-RSVP.

LBU Percentage is described as the (LBU / Maximum bandwidth) * 100.

4.2.2. Link Reserved Bandwidth Utilization (LRBU)

The reserved bandwidth utilization on a link, forwarding adjacency, or bundled link can be calculated from the TED. This includes traffic for only RSVP-TE LSPs.

LRBU can be calculated by using the Residual bandwidth, the Available bandwidth and LBU. The actual bandwidth by non-RSVP TE traffic can be calculated by subtracting the Available Bandwidth from the Residual Bandwidth. Once we have the actual bandwidth for non-RSVP TE traffic, subtracting this from LBU would result in LRBU.

LRBU Percentage is described as the (LRBU / (Maximum reservable bandwidth)) * 100.

4.2.3. BU Object

The BU (the Bandwidth Utilization) is used to indicate the upper limit of the acceptable link bandwidth utilization percentage.

The BU object may be carried within the PCReq message and PCRep messages.

BU Object-Class is TBD4.

BU Object-Type is 1.

The format of the BU 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              Reserved                          |    Type       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                      Bandwidth Utilization                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

BU Object Body Format

Reserved (24 bits): This field MUST be set to zero on transmission and MUST be ignored on receipt.

Type (8 bits): Represents the bandwidth utilization type. Link Bandwidth Utilization (LBU) Type is 1 and Link Reserved Bandwidth Utilization (LRBU) Type is 2.

Bandwidth utilization (32 bits): Represents the bandwidth utilization quantified as a percentage (as described in Section 4.2.1 and Section 4.2.2). The basic unit is 0.000000023%, with the maximum value 4,294,967,295 representing 98.784247785% (4,294,967,295 * 0.000000023%). This value is the maximum Bandwidth utilization percentage that can be expressed.

The BU object body has a fixed length of 8 bytes.

4.2.3.1. Elements of Procedure

A PCC SHOULD request the PCE to factor in the bandwidth utilization during path computation by including a BU object in the PCReq message.

Multiple BU objects MAY be inserted in a PCReq or a PCRep message for a given request but there MUST be at most one instance of the BU object for each type. If, for a given request, two or more instances of a BU object with the same type are present, only the first instance MUST be considered and other instances MUST be ignored.

BU object MAY be carried in a PCRep message in case of unsuccessful path computation along with a NO-PATH object to indicate the constraints that could not be satisfied.

If the P bit is clear in the object header and PCE does not understand or does not support the bandwidth utilization during path computation it SHOULD simply ignore BU object.

If the P Bit is set in the object header and PCE receives BU object in path request and it understands the BU object, but the PCE is not capable of the bandwidth utilization check during 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 BU object, then the PCE MUST send a PCErr message with a PCEP-ERROR Object Error-Type = 3 (Unknown object) [RFC5440].

4.3. Objective Functions

[RFC5541] defines mechanism to specify an optimization criteria, referred to as objective functions. The new metric types specified in this document MAY continue to use the existing objective functions like Minimum Cost Path (MCP). Latency (Delay) and Latency Variation (Jitter) are well suited to use MCP as an optimization criteria. For Packet Loss following new OF is defined -

o A network comprises a set of N links {Li, (i=1...N)}.

o A path P is a list of K links {Lpi,(i=1...K)}.

o Packet loss of link L is denoted PL(L) in percentage.

o Packet loss in fraction of link L is denoted FPL(L) = PL(L) / 100.

o The Packet loss of a path P (in percentage) is denoted PL(P), where PL(P) = (1 - ((1-FPL(Lp1)) * (1-FPL(Lp2)) * .. * (1-FPL(LpK))) * 100.

Objective Function Code: TBD5

Name: Minimum Packet Loss Path (MPLP)

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

Two additional objective functions -- namely, MUP (the Maximum Under-Utilized Path) and MRUP (the Maximum Reserved Under-Utilized Path). Hence two new objective function codes have to be defined.

Objective functions are formulated using the following additional terminology:

o The Bandwidth Utilization on link L is denoted u(L).

o The Reserved Bandwidth Utilization on link L is denoted ru(L).

o The Maximum bandwidth on link L is denoted M(L).

o The Maximum Reserved bandwidth on link L is denoted R(L).

The description of the two new objective functions is as follows.

Objective Function Code: TBD6

Name: Maximum Under-Utilized Path (MUP)

Description: Find a path P such that (Min {(M(Lpi)- u(Lpi)) / M(Lpi), i=1...K }) is maximized.

Objective Function Code: TBD7

Name: Maximum Reserved Under-Utilized Path (MRUP)

Description: Find a path P such that (Min {(R(Lpi)- ru(Lpi)) / R(Lpi), i=1...K }) is maximized.

These new objective functions are used to optimize paths based on the bandwidth utilization as the optimization criteria.

If the objective function defined in this document are unknown/ unsupported, the procedure as defined in [RFC5541] is followed.

5. PCEP Message Extension

5.1. The PCReq message

The extension to PCReq message are -

- o new metric types using existing METRIC object
- o a new optional BU object
- o new objective functions using existing OF object ([RFC5541])

The format of the PCReq message (with [RFC5541] as a base) is updated as follows:

```
   <PCReq Message> ::= <Common Header>
                       [<svec-list>]
                       <request-list>
   where:
        <svec-list> ::= <SVEC>
                        [<OF>]
                        [<metric-list>]
                        [<svec-list>]

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

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

   and where:
        <bu-list>::=<BU>[<bu-list>]
        <metric-list> ::= <METRIC>[<metric-list>]
```

5.2. The PCRep message

The extension to PCRep message are -

- o new metric types using existing METRIC object
- o a new optional BU object (during unsuccessful path computation, to indicate the bandwidth utilization as a reason for failure)
- o new objective functions using existing OF object ([RFC5541])

The format of the PCRep message (with [RFC5541] as a base) is updated as follows:

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

where:

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

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

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

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

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

and where:

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

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

5.3. Stateful PCE

[STATEFUL-PCE] specifies a set of extensions to PCEP to enable stateful control of MPLS-TE and GMPLS LSPs via PCEP and maintaining of these LSPs at the stateful PCE. It further distinguishes between an active and a passive stateful PCE. A passive stateful PCE uses LSP state information learned from PCCs to optimize path computations but does not actively update LSP state. In contrast, an active stateful PCE utilizes the LSP delegation mechanism to let PCCs relinquish control over some LSPs to the PCE.

The passive stateful PCE implementation MAY use the extension of PCReq and PCRep messages as defined in Section 5.1 and Section 5.2 to enable the use of service aware parameters.

The additional objective functions defined in this document can also be used with stateful PCE.

5.3.1. 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 or delegate control of an LSP. The PCRpt message is extended to support BU object. This optional BU object can specify the upper limit that should not be crossed.

As per [STATEFUL-PCE], 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>

     <path> ::= <ERO><attribute-list>[<RRO>]
```

Where <attribute-list> is extended as per Section 5.2 for BU object.

Thus a BU object can be used to specify the upper limit set at the PCC at the time of LSP delegation to an active stateful PCE.

6. Other Considerations

6.1. Inter-domain Consideration

[RFC5441] describes the Backward-Recursive PCE-Based Computation (BRPC) procedure to compute end to end optimized inter-domain path by cooperating PCEs. The new metric types defined in this document can be applied to end to end path computation, in similar manner as existing IGP or TE metric. The new BU object defined in this document can be applied to end to end path computation, in similar manner as the METRIC object.

All domains should have the same understanding of the METRIC (Latency Variation etc) and BU object 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-METRIC-EXT] and [ISIS-TE-METRIC-EXT], the same properties must be advertised using the mechanism described in [RFC5392] (OSPF) and [RFC5316] (ISIS).

6.1.2. Inter-Layer Consideration

[RFC5623] provides a framework for PCE-Based inter-layer MPLS and GMPLS Traffic Engineering. Lower-layer LSPs that are advertised as TE links into the higher-layer network form a Virtual Network Topology (VNT). The advertisement in higher-layer should include the network performance link properties based on the end to end metric of lower-layer LSP. Note that the new metric defined in this document are applied to end to end path computation, even though the path may cross multiple layers.

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]. Based on the changes in performance parameters in TED, a PCC MAY also issue a reoptimization request.

Further, PCC can also monitor the link bandwidth utilization along the path by monitoring changes in the bandwidth utilization parameters of one or more links on the path in the TED. In case of drastic change, 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. The usage of BU object for P2MP LSP is out of scope of this document.

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 destinations 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 type (T) = TBD8 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 destinations 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 type (T) = TBD9 is to be assigned by IANA.

6.3.3. P2MP Packet Loss Metric

P2MP packet loss metric type of METRIC object in PCEP encodes the path packet loss metric for destination that observes the worst packet loss metric among all destinations 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 packet loss metric of the Path to destination Dest_j is denoted by PLM(Dest_j).

- P2MP packet loss metric for the P2MP tree T = Maximum {PLM(Dest_j), (j=1...M)}.

Value for P2MP packet loss metric type (T) = = TBD10 is to be assigned by IANA.

7. IANA Considerations

7.1. METRIC types

Six new metric types are defined in this document for the METRIC object (specified in [RFC5440]). IANA maintains a registry of metric types in the "METRIC Object T Field" sub-registry of the "Path Computation Element Protocol (PCEP) Numbers" registry.

IANA is requested to make the following allocations:

Value	Description	Reference
TBD1	Latency (delay) metric	[This I.D.]
TBD2	Latency Variation (jitter) metric	[This I.D.]
TBD3	Packet Loss metric	[This I.D.]
TBD8	P2MP latency metric	[This I.D.]
TBD9	P2MP latency variation metric	[This I.D.]
TBD10	P2MP packet loss metric	[This I.D.]

7.2. New PCEP Object

IANA assigned a new object class in the registry of PCEP Objects as follows.

Object Class	Object Type	Name	Reference
TBD4	1	BU	[This I.D.]

7.3. BU Object

IANA created a registry to manage the codespace of the Type field of the METRIC Object.

Codespace of the T field (Metric Object)

Type	Name	Reference
1	LBU (Link Bandwidth Utilization	[This I.D.]
2	LRBU (Link Residual Bandwidth Utilization	[This I.D.]

7.4. OF Codes

One new Objective Functions have been defined in this document for the OF code (described in [RFC5541]). IANA maintains this registry at "Objective Function" sub-registry of the "Path Computation Element Protocol (PCEP) Numbers" registry.

IANA is requested to make the following allocations:

Code Point	Name	Reference
TBD5	Minimum Packet Loss Path (MPLP)	[This I.D.]
TBD6	Maximum Under-Utilized Path (MUP)	[This I.D.]
TBD7	Maximum Reserved Under-Utilized Path (MRUP)	[This I.D.]

8. Security Considerations

This document defines new METRIC types, a new BU object, and OF codes which does not add any new security concerns beyond those discussed in [RFC5440] and [RFC5541] in itself. Some deployments may find the service aware information like delay and packet loss as extra sensitive and thus should employ suitable PCEP security mechanisms like TCP-AO and [PCEPS].

9. Manageability Considerations

9.1. Control of Function and Policy

The only configurable item is the support of the new constraints on a PCE which MAY be controlled by a policy module. If the new constraints are not supported/allowed on a PCE, it MUST send a PCErr message accordingly.

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, packet loss, and utilized bandwidth. This mechanism is described in [OSPF-TE-METRIC-EXT] and [ISIS-TE-METRIC-EXT].

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 Alia Atlas, John E Drake, David Ward, Young Lee, Venugopal Reddy, Reeja Paul, Sandeep Kumar Boina, Suresh Babu, Quintin Zhao, Chen Huaimo and Avantika for their useful comments and suggestions.

Also the authors gratefully acknowledge reviews and feedback provided by Qin Wu, Alfred Morton and Paul Aitken during performance directorate review.

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.

[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.

[RFC6049] Morton, A. and E. Stephan, "Spatial Composition of Metrics", RFC 6049, January 2011.

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for MPLS Networks", RFC 6374, September 2011.

[OSPF-TE-METRIC-EXT]
Giacalone, S., Ward, D., Drake, J., Atlas, A., and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", draft-ietf-ospf-te-metric-extensions-08 (work in progress), December 2014.

[ISIS-TE-METRIC-EXT]
Previdi, S., Giacalone, S., Ward, D., Drake, J., Atlas, A., Filsfils, C., and W. Wu, "IS-IS Traffic Engineering (TE) Metric Extensions", draft-ietf-isis-te-metric-extensions-04 (work in progress), October 2014.

[PCEP-MIB]
Sreenivasa, K., Emile, S., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Communications Protocol (PCEP) Management Information Base (MIB) Module", draft-ietf-pce-pcep-mib-11 (work in progress), October 2014.

[STATEFUL-PCE]
Crabbe, E., Minei, I., Medved, J., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-10 (work in progress), October 2014.

[PCEPS] Lopez, D., Dios, O., Wu, W., and D. Dhody, "Secure Transport for PCEP", draft-ietf-pce-pceps-02 (work in progress), October 2014.

Appendix A. Contributor Addresses

Clarence Filsfils
Cisco Systems
Email: cfilsfil@cisco.com

Siva Sivabalan
Cisco Systems
Email: msiva@cisco.com

George Swallow
Cisco Systems
Email: swallow@cisco.com

Stefano Previdi
Cisco Systems, Inc
Via Del Serafico 200
Rome 00191
Italy
Email: sprevidi@cisco.com

Udayasree Palle
Huawei Technologies
Leela Palace
Bangalore, Karnataka 560008
India
Email: udayasree.palle@huawei.com

Avantika
Huawei Technologies
Leela Palace
Bangalore, Karnataka 560008
India
Email: avantika.sushilkumar@huawei.com

Xian Zhang
Huawei Technologies
F3-1-B R&D Center, Huawei Base Bantian, Longgang District
Shenzhen, Guangdong 518129
P.R.China
Email: zhang.xian@huawei.com

Authors' Addresses

Dhruv Dhody
Huawei Technologies
Leela Palace
Bangalore, Karnataka 560008
India

EMail: dhruv.ietf@gmail.com

Qin Wu
Huawei Technologies
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

EMail: bill.wu@huawei.com

Vishwas Manral
Ionos Network
4100 Moorpark Av
San Jose, CA
USA

EMail: vishwas.ietf@gmail.com

Zafar Ali
Cisco Systems

EMail: zali@cisco.com

Kenji Kumaki
KDDI Corporation

EMail: ke-kumaki@kddi.com