PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 10, 2014

D. Dhody
Huawei Technologies
V. Manral
Hewlett-Packard Corp.
Z. Ali
G. Swallow
Cisco Systems
K. Kumaki
KDDI Corporation
July 9, 2013

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

draft-ietf-pce-pcep-service-aware-01

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.

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] describes mechanisms with which network performance information is distributed via OSPF and ISIS 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 and Loss as constraints for end to end path computation.

material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 10, 2014.

Copyright Notice

Table of Contents

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 set up.

Traffic Engineering Database (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. 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. PCC MUST be able to request for 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. 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.

4. PCEP SHOULD provide a means to return end to end network performance information of the computed path in a PCRep 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 {Li, (i=1...N)}.

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

* T=13(IANA): 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. 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(Lpi), (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 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.

4.2.1. 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(Lpi), (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 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.

4.3.1. 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:

- 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.

4.5.1. 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:

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

- 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 first 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 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.

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

- 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 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) = sum {M(Lpi),(i=1...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 PCReq and PCRep Messages.

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 defined in this document can be applied to end to end path computation, in similar manner as existing IGP or TE metric.

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

[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].

6.3. Point-to-Multipoint (P2MP)

This document defines the following optional types for the METRIC object defined in [RFC5440] for P2MP TE LSPs.

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

6.4. 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. 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. This message contains the metric-list as part of attributes, the new metric types defined in this document for network performance parameter MAY be carried to report them to stateful PCEs.

7. IANA Considerations

New metric object T fields have been defined in this document. IANA has made the following allocations from the PCEP "METRIC Object T Field" sub-registry:

Value	Description	Reference
13(TBD)	Latency (delay) metric	[This I.D.]
14(TBD)	Latency Variation (jitter) metric	[This I.D.]
15(TBD)	Packet Loss metric	[This I.D.]
16(TBD)	P2MP latency metric	[This I.D.]
17(TBD)	P2MP latency variation metric	[This I.D.]

8. Security Considerations

This document defines new METRIC types which does not add any new security concerns beyond those discussed in [RFC5440].

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, Reeja Paul, Sandeep Kumar Boina, Suresh Babu, Quintin Zhao, Chen Huaimo and Avantika 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.

[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.

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

[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]", June 2013.

[ISIS-TE-EXPRESS] Previdi, S., Giacalone, S., Ward, D., Drake, J., Atlas, A., and C. Filsfils, "IS-IS Traffic Engineering (TE) Metric Extensions [draft-ietf-isis-te-metric-extensions-00]", June 2013.

[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]", Feb 2013.

[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]", Feb 2013.

[STATEFUL-PCE] Crabbe, E., Medved, J., Minei, I., and R. Varga,, "PCEP Extensions for Stateful PCE (draft-ietf-pce-stateful-pce)", June 2013.

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
Leela Palace
Bangalore, Karnataka 560008
INDIA
EMail: udayasree.palle@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.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