PCE Working Group
Internet-Draft
Intended status: Standards Track
Expires: December 26, 2014

Q. Wu
D. Dhody
Huawei
S. Previdi
Cisco Systems, Inc
June 24, 2014

Extensions to Path Computation Element Communication Protocol (PCEP) for handling the Link Bandwidth Utilization draft-wu-pce-pcep-link-bw-utilization-03

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.

The Link bandwidth utilization (the total bandwidth of a link in current use for the forwarding) is an important factor to consider during path computation. [OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] define mechanisms that distribute this information via OSPF and ISIS respectively. This document describes extensions to PCEP to use them as new constraints during 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 26, 2014.

Copyright Notice

Table of Contents

1. Introduction

The link bandwidth utilization based on real time traffic along the path is becoming critical during path computation in some networks. Thus it is important that the link bandwidth utilization is factored in during path computation. A PCC can request a PCE to provide a path such that it selects under-utilized links. This document extends PCEP [RFC5440] for this purpose.

The Traffic Engineering Database (TED) as populated by the Interior Gateway Protocol (IGP) contains the Maximum bandwidth, the Maximum reservable bandwidth and the Unreserved bandwidth ([RFC3630] and [RFC3784]). [OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] further populate the Residual bandwidth, the Available bandwidth and the Utilized bandwidth.

The links in the path MAY be monitored for changes in the link bandwidth utilization, re-optimization of such path MAY be further requested.

[OSPF-TE-EXPRESS] and [ISIS-TE-EXPRESS] also include parameters related to link latency, latency variation and packet loss. [PCE-SERVICE-AWARE] 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).

LBU: Link Bandwidth Utilization. (See Section 3.)

LRBU: Link Reserved Bandwidth Utilization. (See Section 4.)

MRUP: Maximum Reserved Under-Utilized Path. (See Section 6.2.)

MUP: Maximum Under-Utilized Path. (See Section 6.2.)

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].)

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 the network node) that is capable of computing a network path or the route based on a network graph and applying computational constraints.

PCEP: Path Computation Element Communication Protocol.

RSVP: Resource Reservation Protocol

TE LSP: Traffic Engineering Label Switched Path.

3. 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-EXPRESS] and [ISIS-TE-EXPRESS]). 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. 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.

5. PCEP Requirements

The following requirements associated with the bandwidth utilization are identified for PCEP:

1. The PCE supporting this document MUST have the capability to compute end-to-end path with the bandwidth utilization constraints. It MUST also support the combination of the bandwidth utilization constraint with the existing constraints (cost, hop-limit...).

2. 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.

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

4. PCEs are not required to support the bandwidth utilization constraint. Therefore, it MUST be possible for a PCE to reject a PCReq message with a reason code that indicates no support for the bandwidth utilization constraint.

5. PCEP SHOULD provide a mechanism to handle the bandwidth utilization constraint in multi-domain (e.g., Inter-AS, Inter-Area or Multi-Layer) environment.

6. PCEP Extensions

This section defines extensions to PCEP [RFC5440] to meet requirements outlined in Section 5. The proposed solution is used to consider the bandwidth utilization during path computation.

6.1. BU Object

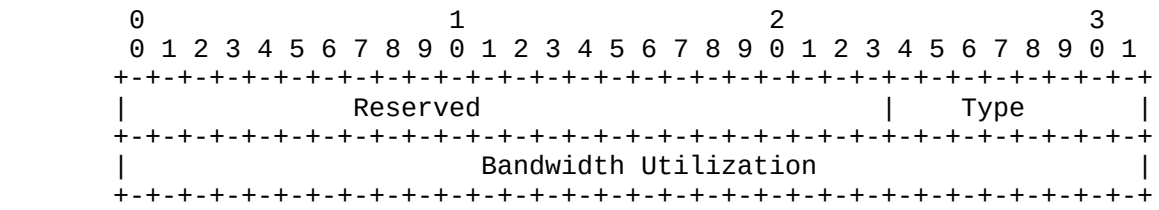
The BU (the Bandwidth Utilization) is used to indicate the upper limit of the acceptable link bandwidth utilization percentage.

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

BU Object-Class is TBD.

BU Object-Type is 1.

The format of the BU object body is as follows:



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 3 and Section 4). 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.

6.1.1. 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].

6.2. New Objective Functions

This document defines 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 terminology:

- A network comprises a set of N links {Li, (i=1...N)}.
- A path P is a list of K links {Lpi,(i=1...K)}.
- The Bandwidth Utilization on link L is denoted u(L).
- The Reserved Bandwidth Utilization on link L is denoted ru(L).
- The Maximum bandwidth on link L is denoted M(L).
- 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: TBD

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: TBD

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.

6.3. PCEP Message Extension

6.3.1. The PCReq message

The new optional BU objects MAY be specified in the PCReq message. As per [RFC5541], an OF object specifying a new objective function MAY also be specified.

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>]
```

6.3.2. The PCRep message

The BU objects MAY be specified in the PCRep message, in case of an unsuccessful path computation, to indicate the bandwidth utilization as a reason for failure. The OF object MAY be carried within a PCRep message to indicate the objective function used by the PCE during path computation.

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>]
```

7. Other Considerations

7.1. Reoptimization Consideration

PCC can monitor the link bandwidth utilization of an LSP 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].

7.2. 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 BU object defined in this document can be applied to end to end path computation, in similar manner as existing METRIC object.

All domains should have the same understanding of the BU object for end-to-end inter-domain path computation to make sense.

7.2.1. 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 bandwidth related 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).

7.3. P2MP Consideration

They are currently out of scope of this document.

7.4. 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. 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 6.3.1 and Section 6.3.2 to enable the use of BU object.

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

7.4.1. PCEP Message Extension

7.4.1.1. 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 6.3.2 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.

8. IANA Considerations

IANA assigns values to PCEP parameters in registries defined in [RFC5440]. IANA has made the following additional assignments.

8.1. New PCEP Object

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

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

8.2. 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.]

8.3. Objective Functions

Two new Objective Functions have been defined. IANA has made the following allocations from the PCEP "Objective Function" sub-registry:

Code Point	Name	Reference
TBA	Maximum Under-Utilized Path (MUP)	[This I.D.]
TBA	Maximum Reserved Under-Utilized Path (MRUP)	[This I.D.]

9. Security Considerations

This document defines a new BU object and OF codes which do not add any new security concerns beyond those discussed in [RFC5440].

10. Manageability Considerations

10.1. 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 as specified in Section 6.1.1.

10.2. Information and Data Models

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

10.3. 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].

10.4. Verify Correct Operations

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

10.5. Requirements On Other Protocols

PCE requires the TED to be populated with the bandwidth utilization. This mechanism is described in [OSPF-TE-EXPRESS] or [ISIS-TE-EXPRESS].

10.6. 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].

11. Acknowledgments

We would like to thank Alia Atlas, John E Drake and David Ward for their useful comments and suggestions.

12. References

12.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

12.2. Informative References

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, September 2003.

[RFC3784] Smit, H. and T. Li, "Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)", RFC 3784, June 2004.

[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.

[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.

[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-05 (work in progress), December 2013.

[ISIS-TE-EXPRESS]
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-03 (work in progress), April 2014.

[PCEP-MIB]
Koushik, K., Emile, S., Zhao, Q., King, D., and J. Hardwick, "Path Computation Element Protocol (PCEP) Management Information Base", draft-ietf-pce-pcep-mib-08 (work in progress), April 2014.

[STATEFUL-PCE]
Crabbe, E., Medved, J., Minei, I., and R. Varga, "PCEP Extensions for Stateful PCE", draft-ietf-pce-stateful-pce-08 (work in progress), February 2014.

[PCE-SERVICE-AWARE]
Dhody, D., Manral, V., Ali, Z., Swallow, G., and K. Kumaki, "Extensions to the Path Computation Element Communication Protocol (PCEP) to compute service aware Label Switched Path (LSP).", draft-ietf-pce-pcep-service-aware-04 (work in progress), March 2014.

Appendix A. Contributor Addresses

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

Zafar Ali
Cisco Systems

EMail: zali@cisco.com

Authors' Addresses

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

EMail: sunseawq@huawei.com

Dhruv Dhody
Huawei Technologies
Leela Palace
Bangalore, Karnataka 560008
INDIA

EMail: dhruv.ietf@gmail.com

Stefano Previdi
Cisco Systems, Inc
Via Del Serafico 200
Rome 00191
IT

EMail: sprevidi@cisco.com