PCE Working Group                                   Francesco Lazzeri
Internet Draft                                     Daniele Ceccarelli
Intended status: Standard Track                              Ericsson
Expires: August 2018                                       Young Lee
                                                          Dhruv Dhody
                                                               Huawei
                                                    February 26, 2018

# Extensions to the Path Computation Element Protocol (PCEP) for residual path bandwidth support

draft-lazzeri-pce-residual-bw-01

## Status of this Memo



## Copyright Notice

## Abstract

The PCEP protocol has objective functions to optimize path attributes like the residual bandwidth. While this is enough for some applications, it's not possible to return the computed values of such attributes to the PCC, or put bounds on them.

This document describes extensions to the PCE Communication Protocol (PCEP) providing new path-related bandwidth metrics allowing a PCE to compute paths taking into account and returning to the PCC information about the remaining bandwidth along the computed paths.

## Table of Contents

## Introduction

The objective of this document is to define an extension to the PCEP [RFC5440] providing information about the bandwidth still available for future reservations on a given path, that is the minimum unreserved bandwidth and the minimum residual bandwidth among all the links of that path.

This is not a new concept to PCEP. In [RFC5541] two objective functions are defined, called minimum load path (MLP) and maximum residual bandwidth path (MBP). Both of them allow to find paths with optimal value of bandwidth-related metrics, defined on a per-link basis, considering the links traversed by that path.

For example, the residual bandwidth of a path is defined as the minimum value of the residual bandwidth on each link in the path. Specifying that OF inside the SVEC object of a PCReq message, the PCE tries and finds the path with the maximum value of the path residual bandwidth.

Unfortunately, being an objective function, MBP can only be used to find a path that optimizes the residual bandwidth, but its value cannot be returned for a path computed with some other objectives (and also when MBP itself is used), or used as a bound.

The same applies to the unreserved bandwidth. The difference between residual and unreserved bandwidth is well described in [RFC7471]:

> "The calculation of Residual Bandwidth is different than that of Unreserved Bandwidth [RFC3630]. Residual Bandwidth subtracts tunnel reservations from Maximum Bandwidth (i.e., the link capacity) [RFC3630] and provides an aggregated remainder across priorities. Unreserved Bandwidth, on the other hand, is subtracted from the Maximum Reservable Bandwidth (the bandwidth that can theoretically be reserved) and provides per priority remainders. Residual Bandwidth and Unreserved Bandwidth [RFC3630] can be used concurrently, and each has a separate use case (e.g., the former can be used for applications like Weighted ECMP while the latter can be used for call admission control)".

Having this information would allow a PCC to reuse a path resulting from a path computation to route additional LSPs without requesting new path computations(with the same end-points and constraints), until the maximum path unreserved bandwidth is taken (or a path deployment fails).

## 1. Requirements for managing the residual bandwidth as a metric

Path computation with optimization of the load or of the residual bandwidth has been defined as important objective functions in [RFC5541].

Managing the unreserved bandwidth (related to the load) and the residual bandwidth of a path as additional metrics, adds the capability to return their value, or putting a bound on their value. This is an added value in distributed PCE applications, like e.g. in ACTN architecture [ACTN-FW] and [PCE-APP]. The following associated key requirements are identified for PCEP:

1. A PCE supporting this draft MUST have the capability to compute end-to-end (E2E) paths with either unreserved bandwidth or with residual bandwidth constraints. It MUST also support the combination of these new constraints with existing constraints, like IGP metric, TE metric, hop limit, and network performance constraints as defined in [RFC5440] and [PCEP-SERV-AWARE].

2. A PCC MUST be able to specify either unreserved bandwidth or residual bandwidth constraints in a Path Computation Request (PCReq) message to be applied during the path computation.

3. A PCC MUST be able to request that a PCE optimizes a path using either unreserved bandwidth or residual bandwidth as objective metric.

4. A PCE that supports this specification is not required to provide unreserved bandwidth or residual bandwidth path computation to any PCC at any time.

   Therefore, it MUST be possible for a PCE to reject a PCReq message with reason codes that indicate unreserved bandwidth or residual bandwidth is not supported. Furthermore, a PCE that does not support this specification will either ignore or reject such requests using pre-existing mechanisms, therefore the requests MUST be identifiable to legacy PCEs and rejections by legacy PCEs MUST be acceptable within this specification.

5. A PCE that supports this specification MUST be able to return unreserved or residual bandwidth information of the computed path in a Path Computation Reply (PCRep) message.

## 2. New metrics definition

### 2.1. Link and Path Unreserved bandwidth

The unreserved bandwidth of a link is the bandwidth available for future allocation on the link at a given priority, that is the difference between the Maximum Reservable Bandwidth of the link and total bandwidth used on that link by LSPs with priority equal or lower (higher value) than the specified priority. In order to define the path unreserved bandwidth, the following concepts and notation need to be introduced:

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

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

- o The maximum reservable bandwidth of the link Li, named Ri.

- o The bandwidth allocated to LSPs at priority p on the link Li is the sum of the bandwidth of all the LSPs passing through the link Li with priority >= p, named Bi(p).

- o The unreserved bandwidth at priority p of the link Li is Ui(p) = Ri - Bi(p)

The path unreserved bandwidth at a given priority k is defined as the minimum value of the unreserved bandwidth at priority k among all the links along the P2P path. Specifically, extending on the above mentioned terminology:

- o Path unreserved bandwidth metric at priority is defined as:
  PU(p) = min {Ui(p), (i=1...K)}

### 2.2. Link and Path Residual bandwidth

The residual bandwidth of a link is the bandwidth physically left free for future allocation on the link. In order to define the path residual bandwidth, the following concepts and notation need to be introduced:

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

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

- The maximum bandwidth of the link Li, named Bi.

- The sum of the bandwidth of all the LSPs passing through the link Li, that is the bandwidth allocated on the link, named Ai.

- The residual bandwidth of the link Li is r(i) = Bi - Ai.

The path residual bandwidth is defined as the minimum value of the residual bandwidth among all the links along the P2P path. Specifically, extending on the above mentioned terminology:

- Path residual bandwidth metric for the P2P path is defined as:
  PB = min {r(Lpi), (i=1...K)}

## 3. PCEP protocol extensions

This section defines PCEP extensions to fulfill the requirements outlined in Section 2. The proposed solution is used to support path unreserved bandwidth and path residual bandwidth as additional metrics of the PCEP protocol.
The METRIC object is defined in section 7.8 of [RFC5440], comprising metric-value, metric-type (T field) and a flags field comprising a number of bit-flags.

This document defines two new types for the METRIC object:

T = TBD1: Path Unreserved Bandwidth

When the T field is set to TBD1, the value of the metric-value field is set to the Path Unreserved Bandwidth for the traffic type and priority requested in the PCReq message.
The same format used by [RFC5440] for the BANDWIDTH object body is used here to represent the value of a path unreserved bandwidth bound or returned value, as shown in the following:

```
       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
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                     Path Unreserved Bandwidth                 |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: PATH UNRESERVED BANDWIDTH value format

Path Unreserved Bandwidth (32 bits): The path unreserved bandwidth is encoded in 32 bits in IEEE floating point format (see [IEEE.754.1985]), expressed in bytes per second.

The PATH UNRESERVED BANDWIDTH value has a fixed length of 4 bytes.

T = TBD2: Path Residual Bandwidth

When the T field is set to TBD2, the value of the metric-value field is set to the Path Residual Bandwidth for the traffic type requested in the PCReq message.

When the T field is set to TBD2, the value of the metric-value field is set to the Path Residual Bandwidth for the traffic type requested in the PCReq message.

The same format used by [RFC5440] for the BANDWIDTH object body is used here to represent the value of a path residual bandwidth bound or returned value, as shown in the following:

```
       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
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                     Path Residual Bandwidth                   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: PATH RESIDUAL BANDWIDTH value format

Path Residual Bandwidth (32 bits): The path residual bandwidth is encoded in 32 bits in IEEE floating point format (see [IEEE.754.1985]), expressed in bytes per second.

The PATH RESIDUAL BANDWIDTH value has a fixed length of 4 bytes.

Editor NOTE: these definitions provide support only of PSC signal type. For other signal types (e.g. ODU, WDM) these fields can be filled with the number of unreserved or residual fixed containers (e.g. 3 ODU0) related to the type of traffic specified in the PCReq. This has to be discussed.

A PCC MAY use the path unreserved or residual bandwidth in a PCReq message to request a path meeting the end to end unreserved or residual bandwidth requirement. In this case, the B bit MUST be set to suggest a bound (a minimum) for the path residual bandwidth metric that must be guaranteed for the PCC to consider the computed path as acceptable. The path unreserved or residual bandwidth metrics must be greater than or equal to the value specified in the metric-value field.
The P bit MAY be set to specify the constraint as mandatory, or MAY be left cleared to specify the bound as optional.

A PCC can also use this metric to ask PCE to optimize (that is maximize) the path residual bandwidth during path computation. In this case, the B bit MUST be cleared.

A PCE MAY use the path residual bandwidth metric in a PCRep message along with a NO-PATH object in the case where the PCE cannot compute a path meeting this constraint.
A PCE can also use this metric to send the computed path residual bandwidth metric to the PCC.

## 4. Non-Understanding/Non-Support Residual Bandwidth

If a PCE receives a PCReq message containing a METRIC object with type PATH UNRESERVED BANDWIDTH or PATH RESIDUAL BANDWIDTH and the PCE does not understand or support those metric types, and the P bit is clear in the METRIC object header then the PCE SHOULD simply ignore the METRIC object as per the processing specified in [RFC5440].

If the PCE does not understand the new METRIC types, and the P bit is set in the METRIC object header, then the PCE MUST send a PCErr message containing a PCEP-ERROR Object with Error-Type = 4

(Not supported object) and Error-value = 4 (Unsupported parameter) [RFC5440][RFC5541].

If the PCE understands but does not support the new METRIC type, and the P bit is set in the METRIC object header, then the PCE MUST send a PCErr message containing a PCEP-ERROR Object with Error-Type = 4 (Not supported object) with Error-value = TBD3 (Unsupported path unreserved bandwidth constraint) or TBD4 (Unsupported path residual bandwidth constraint).
The path computation request MUST then be cancelled.

If the PCE understands the new METRIC type, but the local policy has been configured on the PCE to not allow network performance constraint, and the P bit is set in the METRIC object header, then the PCE MUST send a PCErr message containing a PCEP-ERROR Object with Error-Type = 5 (Policy violation) with Error-value = TBD5 (Not Allowed path unreserved bandwidth constraint) or TBD6 (Not Allowed path residual bandwidth constraint). The path computation request MUST then be cancelled.

## 4.1. 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 (path unreserved or path residual bandwidth) that MUST be optimized by the path computation algorithm.

- o To indicate a bound on the METRIC (path unreserved or path residual bandwidth) that MUST NOT be exceeded for the path to be considered as acceptable by the PCC.

In a PCRep message, the PCE MAY insert the METRIC object with an Explicit Route Object (ERO) so as to provide the METRIC (residual bandwidth) for the computed path.
The PCE MAY also insert the METRIC object with a 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 the METRIC object as explained in [RFC5440] are applicable to the new metric types as well.

## 5. Procedures

The new metrics defined in this document don't add or change the procedures already defined for PCEP protocol in [RFC5440] and [RFC5541].

In particular, the existing objective function MBP is still usable as appropriate, being equivalent to the usage of the Path Residual Bandwidth metric with the B bit cleared.

The new metric can be used to define new procedures especially in the scope of SDN and ACTN, which are out of the scope of this document.

### 5.1. Use cases

The first use case is the application of the residual bandwidth to simplify the computation of an end-to-end path across a multi-domain network.
The ability of a hierarchy of PCEs to compute accurate end-to-end paths across multiple domains is recognized as an important requirement in many applications.
In particular, this is a key requirement for networks with a centralized path computation function (e.g. hierarchical PCE or SDN. In such scenarios, a hierarchy of PCEs is often implemented, where, as illustrated in [RFC6805], a parent H-PCE coordinates the operations of a set of child (domain) PCEs in order to compute end-to-end paths across the network.
An H-PCE (either stateful or stateless) can make the best of residual bandwidth metrics, using paths from erstwhile path computations to deploy multiple LSPs (having the same end-points and constraints) without additional requests, until either the remaining
In a hierarchical architecture of PCEs, domain PCEs just know the topology of their domains, while the parent PCE has in general detailed information about the managed domains and the relevant inter-domain links, but not necessarily enough information about the internals of each domain, so that it's capable to compute accurately an end-to-end path.

The residual bandwidth information would also be beneficial for implementing  abstractions of the domain topology, building the

abstract connectivity incrementally, based only on really used constraints, as soon as path computation results are returned.
One of the key features of SDN is the support of network abstraction, that is, as described in [RFC7926], the capability of applying policy to a set of information about a network, in order to produce selective information that represents the potential ability to connect across the domain.
The process of abstraction produces a connectivity graph, which can be used by the parent PCE to compute an accurate path based on the abstracted topology. The main issue is that the connectivity graph can be huge, depending on the size of the domain topology and the number of end-points defined on the edge of the domain.
One way to provide similar information is to store the result of path computations requested to the child PCEs (performed by e.g. TE-tunnels "compute only") and try reusing them if possible to save further path computation iterations between parent and child PCEs.
In any case a selection of path computation constraints has to be defined against the abstract topology in order to reduce the number of the abstract links or TE-tunnels exported by the connectivity graph, as it’s impractical to compute or pre-compute all the constraints combinations. It’s also very important to reduce the number of updates of such connectivity information to the parent PCE in order not to flood it with a continuous stream of updates.

# 6. IANA considerations

## 6.1. METRIC types

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" at <http://www.iana.org/assignments/pcep>.  Within this registry IANA maintains one sub-registry for "METRIC object T field".
Two new metric types are defined in this document for the METRIC object (specified in [RFC5440]).

IANA is requested to make the following allocations:

| Value | Description | Reference |
|---|---|---|
| TBD1 | Path unreserved bandwidth metric | [This I.D.] |

| | | |
|---|---|---|
| TBD2 | Path residual bandwidth metric | [This I.D.] |

## 6.2. New Error-Values

IANA maintains a registry of Error-Types and Error-values for use in PCEP messages.  This is maintained as the "PCEP-ERROR Object Error Types and Values" sub-registry of the "Path Computation Element Protocol (PCEP) Numbers" registry.

IANA is requested to make the following allocations:

Four new Error-values are defined for the Error-Type "Not supported object" (type 4) and "Policy violation" (type 5).

| Error-Type | Meaning and error values | Reference |
|---|---|---|
| 4 | Not supported object | |
| | Error-value=TBD3 Unsupported Path unreserved bandwidth constraint | [This I.D.] |
| | Error-value=TBD4 Unsupported Path residual bandwidth constraint | |
| 5 | Policy violation | |
| | Error-value=TBD5 Not allowed Path unreserved bandwidth constraint | [This I.D.] |
| | Error-value=TBD6 Not allowed Path residual bandwidth constraint | |

## 7. Security Considerations

This document defines new METRIC types, which do not add any new security concerns beyond those discussed in [RFC5440] and [RFC5541] in itself.

In some scenarios, path unreserved bandwidth and path residual bandwidth information could be considered sensitive and could be used to influence path computation and setup with adverse effect.

Snooping of PCEP messages with such data, or using PCEP messages for network reconnaissance, may give an attacker sensitive information about the capabilities of the network. Thus, such deployment should employ suitable PCEP security mechanisms like TCP Authentication Option (TCP-AO) [RFC5925] or [PCEPS].
The Transport Layer Security (TLS) based procedure in [PCEPS] is considered as a security enhancement and thus much better suited for the sensitive residual bandwidth information.

# 8. References

## 8.1. Normative References

[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>.

[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, <http://www.rfc-editor.org/info/rfc5541>.

[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>.

[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.Lopez, D., Dios, O., Wu, W., and D. Dhody, "Secure Transport for PCEP", draft-ietf-pce-pceps-11 (work in progress), January 2017.

[IEEE.754.1985]  IEEE, "Standard for Binary Floating-Point Arithmetic", IEEE 754, August 1985

## 8.2. Informative References

[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, November 2012, <http://www.rfc-editor.org/info/rfc6805>.

[RFC7471] Giacalone S., Ward D., Drake J., Atlas A. and S. Previdi, "OSPF Traffic Engineering (TE) Metric Extensions", RFC 7471, DOI 10.17487/RFC7471, March 2015, <http://www.rfc-editor.org/info/rfc7471>

[RFC7926] Farrel, A. et al., "Problem Statement and Architecture for Information Exchange Between Interconnected Traffic Engineered Networks", RFC 7926, July 2016.

[ACTN-FW] Ceccarelli, D. and Y. Lee, "Framework for Abstraction and Control of Traffic Engineered Networks", draft-ietf-teas-actn-framework-03 (work in progress), February 2017.

[PCE-APP] Dhody, D. Lee, Y. Ceccarelli, D. "Applicability of Path Computation Element (PCE) for  Abstraction and Control of TE Networks (ACTN)" draft-dhody-pce-applicability-actn-02

## 9. Contributors

## Authors' Addresses

Francesco Lazzeri
Ericsson
Via Melen 77
Genova - Italy
Email: francesco.lazzeri@ericsson.com

Daniele Ceccarelli
Ericsson AB
Gronlandsgatan 21
Kista - Sweden
Email: daniele.ceccarelli@ericsson.com

Young Lee
Huawei Technologies
5340 Legacy Drive
Plano, TX 75023, USA
Phone: (469)277-5838
Email: leeyoung@huawei.com

Dhruv Dhody
Huawei Technologies,
Divyashree Technopark, Whitefield
Bangalore, India
Email: dhruv.ietf@gmail.com

## Intellectual Property Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.