PCE Working Group
Internet-Draft
Updates: 5440 (if approved)
Intended status: Standards Track
Expires: 16 August 2025

S. Peng
C. Li
Huawei Technologies
L. Han
China Mobile
L. Ndifor
MTN Cameroon
S. Sidor
Cisco

12 February 2025

Support for Path MTU (PMTU) in the Path Computation Element (PCE) Communication Protocol (PCEP)

draft-ietf-pce-pcep-pmtu-07

Abstract

The Path Computation Element (PCE) provides path computation functions in support of traffic engineering in Multiprotocol Label Switching (MPLS) and Generalized MPLS (GMPLS) networks.

The Source Packet Routing in Networking (SPRING) architecture describes how Segment Routing (SR) can be used to steer packets through an IPv6 or MPLS network using the source routing paradigm. A Segment Routed Path can be derived from a variety of mechanisms, including an IGP Shortest Path Tree (SPT), explicit configuration, or a Path Computation Element (PCE).

Since the SR does not require signaling, the path maximum transmission unit (MTU) information for the SR path is unavailable at the headend. This document specifies the extension to PCE Communication Protocol (PCEP) to carry path MTU as a new metric type in the PCEP messages for SR, but not limited to it.

This document also updates RFC 5440 to allow metric bounds to be minimum as needed in the case of path MTU.

Status of This Memo

This Internet-Draft will expire on 16 August 2025.

Copyright Notice

Table of Contents

1. Introduction 3
2. Requirements Language 4
3. Terminology 4
4. PCEP Extention 4
  4.1. Extensions to METRIC Object 4
    4.1.1. Update to RFC 5440 5
  4.2. Stateful PCE and PCE Initiated LSPs 6
  4.3. Segment Routing 6
    4.3.1. Multi-Path Handling 7
    4.3.2. Path MTU Adjustment 7
    4.3.3. MSD 7
  4.4. Other Path Setup Types 7
5. Security Considerations 8
6. IANA Considerations 8
  6.1. METRIC Type 8
  6.2. PCEP Error Type and Value 8
  6.3. Security Considerations 8
7. References 9
  7.1. Normative References 9
  7.2. Informative References 9
Authors' Addresses 11

1. Introduction

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables communication between a Path Computation Client (PCC) and a PCE, or between PCE and PCE, for the purpose of computation of Multiprotocol Label Switching (MPLS) as well as Generalized MPLS (GMPLS) Traffic Engineering Label Switched Path (TE LSP) characteristics.

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of TE LSPs within and across PCEP sessions in compliance with [RFC4657]. It includes mechanisms to effect LSP State Synchronization between PCCs and PCEs, delegation of control over LSPs to PCEs, and PCE control of timing and sequence of path computations within and across PCEP sessions. The model of operation where LSPs are initiated from the PCE is described in [RFC8281].

As per [RFC8402], with Segment Routing (SR), a node steers a packet through an ordered list of instructions, called segments. A segment can represent any instruction, topological or service-based. A segment can have a semantic local to an SR node or global within an SR domain. SR enables the enforcement of a flow through a specific path and service chain while ensuring that the per-flow state is maintained only at the ingress node of the SR domain. Segments can be derived from different components: IGP, BGP, Services, Contexts, Locators, etc. The SR architecture can be applied to the MPLS forwarding plane without any change, in which case an SR path corresponds to an MPLS Label Switching Path (LSP). The SR is applied to the IPv6 forwarding plane using SRH. An SR path can be derived from an IGP Shortest Path Tree (SPT), but SR-TE paths may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool or a PCE and provisioned on the ingress node.

As per [RFC8664], it is possible to use a stateful PCE for computing one or more SR-TE paths taking into account various constraints and objective functions. Once a path is chosen, the stateful PCE can initiate an SR-TE path on a PCC using PCEP extensions specified in [RFC8281] using the SR specific PCEP extensions specified in [RFC8664]. [RFC8664] specifies PCEP extensions for supporting an SR-TE LSP for MPLS data plane. [RFC9603] extend PCEP to support SR for IPv6 data plane.

[I-D.ietf-spring-pmtu-sr-policy] specifies the link maximum transmission unit (MTU) and SR Path MTU (SR-PMTU) in the context of SR paths and policies. It also states the motivation, link MTU collection, SR-PMTU Computation, SR-PMTU Enforcement, and handling behaviors on the headend.

Since the SR does not require signaling, the path MTU information for SR path is not available. This document specifies the extension to PCEP to carry path MTU in the PCEP messages. It is assumed that the PCE is aware of the link MTU as part of the Traffic Engineering Database (TED) population. This could be done via IGP, BGP-LS [I-D.ietf-idr-bgp-ls-link-mtu] or some other means. Thus, the PCE can find the path MTU at the time of path computation and include this information as part of the PCEP messages.

Though the key use case for path MTU is SR, the PCEP extension (as specified in this document) creates a new metric type for path MTU, making this a generic extension that can be used for any path setup type (PST) (see Section 4.4).

2. Requirements Language

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

3. Terminology

Link MTU: As per [RFC4821], the Maximum Transmission Unit, i.e., maximum IP packet size in bytes, that can be conveyed in one piece over a link. This includes the IP header but excludes the link layer headers and other framing that is not part of IP or the IP payload. In the case of MPLS, it also includes the label stack, and in case of IPv6, it includes IPv6 extension headers (including SRH).

Path MTU, or PMTU: The minimum link MTU of all the links in a path between a source node and a destination node. In the scope of SR, this is also called SR-PMTU for the SR paths and policies. Note that the link MTU takes the SR overhead (label stack or SRH) into consideration.

4. PCEP Extention

4.1. Extensions to METRIC Object

The METRIC object is defined in Section 7.8 of [RFC5440], comprising metric-value and metric-type (T field), and a flags field, comprising a number of bit flags (B "Bound" bit and C "Computed Metric" bit). This document defines a new type for the METRIC object for Path MTU.

* T = TBD: Path MTU.

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

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

* A Link MTU of link L is denoted M(L).

* A Path MTU metric for the path P = Min {M(Lpi), (i=1...K)}.

The Path MTU metric type of the METRIC object in PCEP represents the minimum of the Link MTU of all links along the path.

When PCE computes the path, it can also find the Path MTU (based on the above criteria) and include this information in the METRIC object with the above metric type in the PCEP message when replying to the PCC. In a Path Computation Reply (PCRep) message, the PCE MAY insert the METRIC object with an Explicit Route Object (ERO) so as to provide the METRIC (path MTU) 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.

Further, a PCC MAY use the Path MTU metric in a Path Computation Request (PCReq) message to request a path meeting the MTU requirement of the path. In this case, the B bit MUST be set to suggest a bound (a maximum) for the Path MTU metric that MUST NOT be exceeded for the PCC to consider the computed path as acceptable. The Path MTU metric must be less than or equal to the value specified in the metric-value field.

A PCC can also use this metric to request that the PCE optimize the path MTU during path computation. In this case, the B bit MUST be cleared.

Note that [RFC5440] allow two metric object instances for optimization (B flag cleared) and thus the Path MTU metric object might be used alongside other metric types as well.

The error handling and processing of the METRIC object is as specified in [RFC5440].

4.1.1. Update to RFC 5440

For the handling of B bit in METRIC Object, [RFC5440] states: "When set in a PCReq message, the metric-value indicates a bound (a maximum) for the path 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."

The new metric type path MTU defined in this document is different. The bound for the path MTU indicates a minimum value instead of maximum. That is when the metric type is set to TBD for path MTU, the metric-value indicates a bound (a minimum path MTU) for the path metric that must not be subceeded for the PCC to consider the computed path as acceptable. The path metric for path MTU must be greater than or equal to the value specified in the metric-value field.

Further, a PCC MAY request that PCE optimizes an individual path computation request to maximize the path MTU of the computed path by clearing the B bit in the METRIC object with metric-type=TBD for path MTU.

4.2. Stateful PCE and PCE Initiated LSPs

[RFC8231] specifies a set of extensions to PCEP to enable stateful control of MPLS-TE LSPs via PCEP and the 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 the LSP state. In contrast, an active stateful PCE utilizes the LSP delegation mechanism to update LSP parameters in those PCCs that delegated control over their LSPs to the PCE. [RFC8281] describes the setup, maintenance, and teardown of PCE-initiated LSPs under the stateful PCE model. The document defines the PCInitiate message that is used by a PCE to request a PCC to set up a new LSP.

The new metric type defined in this document can also be used with the stateful PCE extensions. The format of PCEP messages described in [RFC8231] and [RFC8281] uses <intended-attribute-list> and <attribute-list>, respectively, (where the <intended-attribute-list> is the attribute-list defined in Section 6.5 of [RFC5440]).

A PCE MAY include the path MTU metric in PCInitiate or PCUpd message to inform the PCC of the path MTU calculated for the path or as a bound constraint. A PCC MAY include the path MTU metric as a bound constraint or to indicate optimization criteria (similar to PCReq).

4.3. Segment Routing

A Segment Routed path (SR path) can be derived from an IGP Shortest Path Tree (SPT). Segment Routed Traffic Engineering paths (SR-TE paths) may not follow IGP SPT. Such paths may be chosen by a suitable network planning tool and provisioned on the source node of the SR-TE path.

It is possible to use a PCE for computing one or more SR-TE paths taking into account various constraints and objective functions. Once a path is chosen, the PCE can inform an SR-TE path on a PCC using PCEP extensions specified in [RFC8664]. Further, [RFC9603] adds the support for IPv6 data plane in SR.

Refer [I-D.ietf-spring-pmtu-sr-policy] for SR-PMTU considerations.

4.3.1. Multi-Path Handling

[I-D.ietf-spring-pmtu-sr-policy] specify the handling of SR-PMTU at the SR Policy, Candidate paths, and at each segment list level. In PCEP, support for multiple segment list is added in [I-D.ietf-pce-multipath]. The METRIC object is currently encoded at the candidate path level. A future update of the document could investigate and include mechanisms to support SR-PMTU at each segment list level.

4.3.2. Path MTU Adjustment

As per [I-D.ietf-spring-pmtu-sr-policy], it is possible for the headend implementation to take an FRR overhead into consideration when determining if fragmentation would be needed for the SR Path with TI-LFA enabled where the overhead is allowed to be configured by an operator.

4.3.3. MSD

In SR, the term Maximum SID Depth (MSD) [RFC8491] refers to the maximum number of SIDs that an ingress is capable of imposing on a packet. In contrast, the PMTU determines whether IP fragmentation can be avoided, making it an unrelated factor.

4.4. Other Path Setup Types

While SR is the primary use case for the extension defined in this document, the extension in itself is not limited to use within SR.

The PMTU metric type can be used for any path setup type. An implementation MAY choose not to support the use of this metric type for a particular PST as a local policy, in which case it MUST respond to its peer with a PCErr message with the Error-Type = 5 ("Policy Violation") and Error-value = TBD2 ("Metric Type not supported with this PST").

5. Security Considerations

This document defines a new METRIC type that does not add any new security concerns beyond those discussed in [RFC5440] in itself. Some deployments may find the path MTU information to be extra sensitive and could be used to influence path computation and setup with adverse effects. Additionally, snooping of PCEP messages with such data or using PCEP messages for network reconnaissance may give an attacker sensitive information about the operations of the network. Thus, such deployment should employ suitable PCEP security mechanisms like TCP Authentication Option (TCP-AO) [RFC5925] or Transport Layer Security (TLS) [RFC8253]. The procedure is based on TLS is considered a security enhancement and thus is much better suited for sensitive information.

6. IANA Considerations

This document makes the following requests to IANA for action.

6.1. METRIC Type

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" registry group. Within this registry group, IANA maintains a registry for "METRIC Object T Field". IANA is requested to make the following allocation:

Value	Description	Reference
TBD1	Path MTU	This document

6.2. PCEP Error Type and Value

IANA maintains the "Path Computation Element Protocol (PCEP) Numbers" registry group. Within this registry group, IANA maintains a registry for "PCEP-ERROR Object Error Types and Values". IANA is requested to make the following allocation:

Error-Type	Meaning	Error-Value	Reference
5	Policy violation		[RFC5440]
		TBD2: Metric Type not supported with this PST	This document

6.3. Security Considerations

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <https://www.rfc-editor.org/info/rfc5440>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8231] Crabbe, E., Minei, I., Medved, J., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for Stateful PCE", RFC 8231, DOI 10.17487/RFC8231, September 2017, <https://www.rfc-editor.org/info/rfc8231>.

[RFC8281] Crabbe, E., Minei, I., Sivabalan, S., and R. Varga, "Path Computation Element Communication Protocol (PCEP) Extensions for PCE-Initiated LSP Setup in a Stateful PCE Model", RFC 8281, DOI 10.17487/RFC8281, December 2017, <https://www.rfc-editor.org/info/rfc8281>.

[I-D.ietf-spring-pmtu-sr-policy] Peng, S., Dhody, D., Talaulikar, K., and G. S. Mishra, "Path MTU (PMTU) for Segment Routing (SR) Policy", Work in Progress, Internet-Draft, draft-ietf-spring-pmtu-sr-policy-01, 12 September 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-pmtu-sr-policy-01>.

7.2. Informative References

[RFC4657] Ash, J., Ed. and J.L. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol Generic Requirements", RFC 4657, DOI 10.17487/RFC4657, September 2006, <https://www.rfc-editor.org/info/rfc4657>.

[RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, DOI 10.17487/RFC4821, March 2007, <https://www.rfc-editor.org/info/rfc4821>.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, DOI 10.17487/RFC5925, June 2010, <https://www.rfc-editor.org/info/rfc5925>.

[RFC8253] Lopez, D., Gonzalez de Dios, O., Wu, Q., and D. Dhody, "PCEPS: Usage of TLS to Provide a Secure Transport for the Path Computation Element Communication Protocol (PCEP)", RFC 8253, DOI 10.17487/RFC8253, October 2017, <https://www.rfc-editor.org/info/rfc8253>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8491] Tantsura, J., Chunduri, U., Aldrin, S., and L. Ginsberg, "Signaling Maximum SID Depth (MSD) Using IS-IS", RFC 8491, DOI 10.17487/RFC8491, November 2018, <https://www.rfc-editor.org/info/rfc8491>.

[RFC8664] Sivabalan, S., Filsfils, C., Tantsura, J., Henderickx, W., and J. Hardwick, "Path Computation Element Communication Protocol (PCEP) Extensions for Segment Routing", RFC 8664, DOI 10.17487/RFC8664, December 2019, <https://www.rfc-editor.org/info/rfc8664>.

[I-D.ietf-pce-multipath] Koldychev, M., Sivabalan, S., Saad, T., Beeram, V. P., Bidgoli, H., Yadav, B., Peng, S., and G. S. Mishra, "PCEP Extensions for Signaling Multipath Information", Work in Progress, Internet-Draft, draft-ietf-pce-multipath-12, 8 October 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-pce-multipath-12>.

[RFC9603] Li, C., Ed., Kaladharan, P., Sivabalan, S., Koldychev, M., and Y. Zhu, "Path Computation Element Communication Protocol (PCEP) Extensions for IPv6 Segment Routing", RFC 9603, DOI 10.17487/RFC9603, July 2024, <https://www.rfc-editor.org/info/rfc9603>.

[I-D.ietf-idr-bgp-ls-link-mtu] Zhu, Y., Hu, Z., Peng, S., and R. Mwehair, "Signaling Maximum Transmission Unit (MTU) using BGP-LS", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-ls-link-mtu-08, 17 September 2024, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-bgp-ls-link-mtu-08>.

Authors' Addresses

Shuping Peng
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing
100095
China
Email: pengshuping@huawei.com

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing
100095
China
Email: c.l@huawei.com

Liuyan Han
China Mobile
Beijing
100053
China
Email: hanliuyan@chinamobile.com

Luc-Fabrice Ndifor
MTN Cameroon
Cameroon
Email: Luc-Fabrice.Ndifor@mtn.com

Samuel Sidor
Cisco
Slovakia
Email: ssidor@cisco.com