Workgroup: PCE Working Group
Internet-Draft: draft-ietf-pce-pcep-pmtu-01
Published: 5 July 2022
Intended Status: Standards Track
Expires: 6 January 2023
Authors: S. Peng, Huawei Technologies
C. Li, Huawei Technologies
L. Han, China Mobile
L. Ndifor, MTN Cameroon

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

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 SR path is not available. This document specify the extension to PCE communication protocol (PCEP) to carry path (MTU) in the PCEP messages.

Status of This Memo

Table of Contents

1. Introduction

[RFC5440] describes the Path Computation Element (PCE) Communication Protocol (PCEP). PCEP enables the 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 Generalzied 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 allows to enforce a flow through any path and service chain while maintaining per-flow state 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 IPV6 forwarding plane using SRH. A 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 a SR-TE LSP for MPLS data plane. [I-D.ietf-pce-segment-routing-ipv6] extend PCEP to support SR for IPv6 data plane.

The maximum transmission unit (MTU) is the largest size packet or frame, in bytes, that can be sent in a network. An MTU that is too large might cause retransmissions. Too small an MTU might cause the router to send and handle relatively more header overhead and acknowledgments. When an LSP is created across a set of links with different MTU sizes, the ingress router need to know what the smallest MTU is on the LSP path. If this MTU is larger than the MTU of one of the intermediate links, traffic might be dropped, because MPLS packets cannot be fragmented. Also, the ingress router may not be aware of this type of traffic loss, because the control plane for the LSP would still function normally. [RFC3209] specify the mechanism of MTU signaling in RSVP.

Since the SR does not require signaling, the path MTU information for SR path is not available. This document specify 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 independent of SR.

Note that 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. The PMTU on the other hand determines if the IP fragmentation could be avoided.

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 RFC 2119 [RFC2119] RFC 8174 [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminology

This draft refers to the terms defined in [RFC8201], [RFC4821] and [RFC3988].

MTU: Maximum Transmission Unit, the size in bytes of the largest IP packet, including the IP header and payload, that can be transmitted on a link or path. Note that this could more properly be called the IP MTU, to be consistent with how other standards organizations use the acronym MTU.

Link MTU: The Maximum Transmission Unit, i.e., maximum IP packet size in bytes, that can be conveyed in one piece over a link. Be aware that this definition is different from the definition used by other standards organizations.

For IETF documents, link MTU is uniformly defined as the IP MTU over the link. This includes the IP header, but excludes link layer headers and other framing that is not part of IP or the IP payload.

Be aware that other standards organizations generally define link MTU to include the link layer headers.

For the MPLS data plane, this size includes the IP header and data (or other payload) and the label stack but does not include any lower-layer headers. A link may be an interface (such as Ethernet or Packet-over-SONET), a tunnel (such as GRE or IPsec), or an LSP.

Path: The set of links traversed by a packet between a source node and a destination node.

Path MTU, or PMTU: The minimum link MTU of all the links in a path between a source node and a destination node.

For the MPLS data plane, it is the MTU of an LSP from a given LSR to the egress(es), over each valid (forwarding) path. This size includes the IP header and data (or other payload) and any part of the label stack that was received by the ingress LSR before it placed the packet into the LSP (this part of the label stack is considered part of the payload for this LSP). The size does not include any lower-level headers.

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 bit and C 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 a 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 ask PCE to optimize the path MTU during path computation. In this case, the B bit MUST be cleared.

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. Multi-Path Handling

[I-D.ietf-pce-multipath] extends PCEP to support signaling of multipath information i.e. to all each Candidate-Path to contain multiple Segment-Lists.

The PMTU could be supported per segment list as well. The exact mechanism to support this is left for further revision of this document.

4.3. 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 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. A PCC MAY include the path MTU metric as a bound constraint or to indicate optimization criteria (similar to PCReq).

4.4. 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, [I-D.ietf-pce-segment-routing-ipv6] adds the support for IPv6 data plane in SR.

The new metric type for path MTU is applicable for the SR-TE path and require no additional extensions.

4.5. Path MTU Adjustment

The path MTU metric can be used for both primary and protection path.

The minimal value of the link MTU along the path is collected, based on which minor adjustment is made to cater for overhead introduced by the protection mechanisms such as TI-LFA. The path MTU is the value of the minimum link MTU minus the overhead. In this way, the ingress node can use the path MTU directly.

5. Future Plan

A new SPRING document needs to be published and refered by this document.

6. Security Considerations

This document defines a new METRIC type that do 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 effect. 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 based on TLS is considered a security enhancement and thus is much better suited for the sensitive information.

7. IANA Considerations

This document makes following requests to IANA for action.

7.1. METRIC Type

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

Value	Description	Reference
TBD	Path MTU	This document

8. Acknowledgement

We would like to thank Dhruv Dhody for his contributions for this document.

9. References

9.1. 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>.

9.2. Informative References

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP

Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <https://www.rfc-editor.org/info/rfc3209>.

[RFC3988] Black, B. and K. Kompella, "Maximum Transmission Unit Signalling Extensions for the Label Distribution Protocol", RFC 3988, DOI 10.17487/RFC3988, January 2005, <https://www.rfc-editor.org/info/rfc3988>.

[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>.

[RFC8201] McCann, J., Deering, S., Mogul, J., and R. Hinden, Ed., "Path MTU Discovery for IP version 6", STD 87, RFC 8201, DOI 10.17487/RFC8201, July 2017, <https://www.rfc-editor.org/info/rfc8201>.

[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. Mishra, "PCEP Extensions for Signaling Multipath Information", Work in Progress, Internet-Draft, draft-ietf-pce-multipath-06, 17 May 2022, <https://www.ietf.org/archive/id/draft-ietf-pce-multipath-06.txt>.

[I-D.ietf-pce-segment-routing-ipv6]
Li, C., Negi, M., Sivabalan, S., Koldychev, M., Kaladharan, P., and Y. Zhu, "PCEP Extensions for Segment Routing leveraging the IPv6 data plane", Work in Progress, Internet-Draft, draft-ietf-pce-segment-routing-ipv6-13, 1 April 2022, <https://www.ietf.org/internet-drafts/draft-ietf-pce-segment-routing-ipv6-13.txt>.

[I-D.ietf-idr-bgp-ls-link-mtu] Zhu, Y., Hu, Z., Peng, S., and R. Mwehaire, "Signaling Maximum Transmission Unit (MTU) using BGP-LS", Work in Progress, Internet-Draft, draft-ietf-idr-bgp-ls-link-mtu-03, 30 May 2022, <https://www.ietf.org/archive/id/draft-ietf-idr-bgp-ls-link-mtu-03.txt>.

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