

Workgroup: Network Working Group
Internet-Draft: draft-hu-lsr-igp-path-mtu-00
Published: 19 October 2021
Intended Status: Informational
Expires: 22 April 2022
Authors: Z. Hu S. Peng X. Xi
 Huawei Huawei Huawei

IGP Extensions for Path MTU

Abstract

Segment routing (SR) leverages the source routing mechanism. It allows for a flexible definition of end-to-end paths within IGP topologies by encoding paths as sequences of topological sub-paths which are called segments. These segments are advertised by the link-state routing protocols (IS-IS and OSPF). Unlike the MPLS, SR does not have the specific path construction signaling so that it cannot support the Path MTU. This draft provides the necessary IS-IS and OSPF extensions about the Path MTU that need to be used on SR. Here, the term "OSPF" means both OSPFv2 and OSPFv3.

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 BCP 14 [[RFC2119](#)] [[RFC8174](#)] when, and only when, they appear in all capitals, as shown here.

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 <https://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 22 April 2022.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- [1. Introduction](#)
- [2. Terminology](#)
- [3. IGP Extension](#)
 - [3.1. IS-IS Extension](#)
 - [3.2. OSPF Extension](#)
- [4. Acknowledgements](#)
- [5. IANA Considerations](#)
- [6. Security Considerations](#)
- [7. References](#)
- [Authors' Addresses](#)

1. Introduction

Segment routing (SR) leverages the source routing mechanism. SR allows for a flexible definition of end-to-end paths within IGP topologies by encoding paths as sequences of topological sub-paths which are called segments. These segments are advertised by the link-state routing protocols (IS-IS and OSPF). The SR architecture as well as the routing policy is proposed in [RFC8402] and [I-D.ietf-spring-segment-routing-policy]. Two types of segments are defined, Prefix segments and Adjacency segments. Prefix segments represent an ECMP-aware shortest path to a prefix (or a node), as per the state of the IGP topology. Adjacency segments represent a hop over a specific adjacency between two nodes in the IGP. A prefix segment is typically a multi-hop path while an adjacency segment, in most of the cases, is a one-hop path. SR can compute the paths from end to end and without requiring any LDP or RSVP-TE signaling. SR supports per-flow explicit routing while just maintaining per-flow state only at the source node.

SR architecture supports the distributed scenario and the centralized scenario. In the distributed scenario, the segments are

allocated and signaled by IGP or BGP and a node needs to compute the source-routed policy. Some necessary IS-IS and OSPF extensions for SR are proposed in [[RFC8665](#)] [[RFC8666](#)] [[RFC8667](#)]. In a centralized scenario, the SR controller decides which nodes need to steer which packets on which source-routed policies. However, in both conditions, the MTU is not included in the SR policy. As the SR may push more MPLS labels or SRv6 SIDs in the packet header, the packets are more likely to be larger than the minimum MTU in the path compared to the traditional MPLS forwarding process. Unfortunately, with the current mechanisms in SR, the path MTU information cannot be obtained in advance. Therefore it cannot be ensured that the packet size is less than the path MTU which is the minimum link MTU of all the links in a path between a source node and a destination node. The definition of the path MTU is discussed in [[RFC1191](#)] [[RFC8201](#)].

This draft describes the necessary IS-IS and OSPF extensions for obtaining the path MTU to be used on SR. New sub-TLVs are introduced for both the IS-IS and OSPF protocols. With the IGP flooding process in the distributed scenario or the BGP transmission to the controller, the ingress node or the controller is able to compute the path MTU for the SR policy.

2. Terminology

Router: A node that forwards IP packets not explicitly addressed to itself.

Interface: A node's attachment to a link.

Segment: An instruction a node executes on the incoming packet. For example, forward packet according to shortest path to destination or a specific interface, etc..

SR Policy: An ordered list of segments.

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

3. IGP Extension

This document describes IS-IS and OSPF extensions to flood the router interface MTU to each node within an IGP domain. Then the controller or the original node collects all the link MTUs from the routers. So the original node can compute the minimum link MTU of all the links in the path. The source node can limit the packet size less than the path MTU.

3.1. IS-IS Extension

A new sub-TLV called link MTU sub-TLV is defined for TLVs 22, 23, 25, 141, 222, 223 in the Router Information LSP to carry the MTU of the interface associated with the link. Each sub-TLV is encoded as shown in Figure 1.

Type: MTU, 1 byte, TBD.

Length: # of octets in the value field, 1 byte.

Value: The value is the MTU size of a link, 2 bytes.

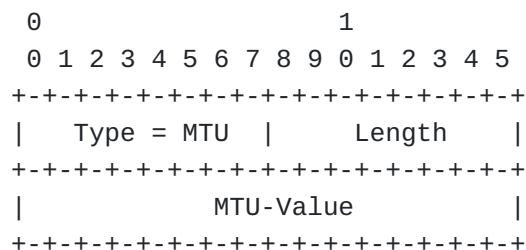


Figure 1: Figure 1: Link MTU Sub-TLV for the IS-IS extension

The use and meaning of these fields are as follows:

Type - A single octet encoding the sub-TLV type. Here the type is 1 octet.

Length - A single octet encoding the total length of the value field of the sub-TLV in octets.

MTU-Value - Two octets encoding the MTU size of the sub-TLV. This field identifies the size of the router interfaces.

This sub-TLV is optional.

This document defines a link MTU sub-TLV for IS-IS extension. The codepoints need to be determined by the IANA.

3.2. OSPF Extension

A new sub-TLV called link MTU sub-TLV is defined in the corresponding LSA as specified for OSPFv2 and OSPFv3 to carry the MTU of the interface associated with the link. Each sub-TLV is encoded as shown in Figure 2.

Type: MTU, 2 bytes, TBD.

Length: # of octets in the value field, 2 bytes.

Value: The value is the MTU size of a link, 2 bytes.

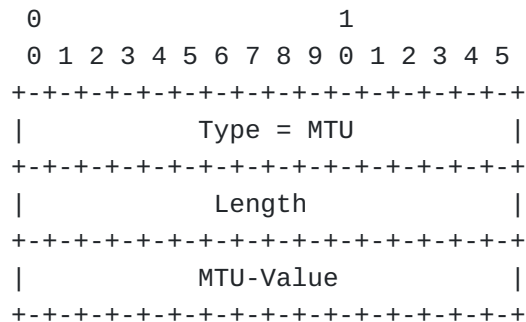


Figure 2: Figure 2: Link MTU Sub-TLV for the OSPF extension

The use and meaning of these fields are as follows:

Type - Two octets encoding the TLV type. Here the type is 2 octets.

For OSPFv2, the link MTU is advertised as an optional sub-TLV of the OSPFv2 Extended Link TLV in the OSPFv2 Extended Link Opaque LSA as defined in [[RFC7684](#)] and the codepoints need to be determined by the IANA.

For OSPFv3, the link MTU is advertised as an optional sub-TLV of the Router-Link TLV in the OSPFv3 E-Router-LSA as defined in [\[RFC8362\]](#) and the codepoints need to be determined by the IANA.

Length - Two octets encoding the total length of the value field of the sub-TLV in octets.

MTU-Value - Two octets encoding the MTU size of the TLV. This field identifies the size of the router interfaces.

If the link MTU sub-TLV is advertised for multiple times for the same link in different OSPFv2 Extended Link Opaque LSAs or OSPFv3 E-Router-LSAs originated by the same OSPF router, the link MTU sub-TLV in the OSPFv2 Extended Link Opaque LSA with the smallest Opaque ID or in the OSPFv3 E-Router-LSA with the smallest Link State ID MUST be used by receiving OSPF routers.

4. Acknowledgements

TBD.

5. IANA Considerations

This document requests that IANA allocates a new sub-TLV type as defined in Section 3.1 from the "Sub-TLVs for TLVs 22, 23, 25, 141, 222, and 223 (Extended IS reachability, IS Neighbor Attribute, L2 Bundle Member Attributes, inter-AS reachability information, MT-ISN, and MT IS Neighbor Attribute TLVs)" registry as specified.

Value	Description	Reference
TBD	IS-IS Link MTU	This document

Figure 3: Figure 3: IS-IS Link MTU

This document requests that IANA allocates a new sub-TLV type as defined in Section 3.2 from the "OSPFv2 Extended Link TLV Sub-TLVs" registry.

Value	Description	Reference
TBD	OSPFv2 Link MTU	This document

Figure 4: Figure 4: OSPFv2 Link MTU

This document requests that IANA allocates a new sub-TLV type as defined in Section 3.2 from the "OSPFv3 Extended LSA Sub-TLVs" registry.

Value	Description	Reference
TBD	OPSFv3 Link MTU	This document

Figure 5: Figure 5: OSPFv3 Link MTU

6. Security Considerations

These extensions to IS-IS and OSPF do not add any new security issues to the existing IGP.

7. References

[I-D.ietf-spring-segment-routing-policy]

Filsfils, C., Talaulikar, K., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", Work in Progress, Internet-Draft, draft-ietf-spring-segment-routing-policy-13, 28 May 2021, <<https://www.ietf.org/archive/id/draft-ietf-spring-segment-routing-policy-13.txt>>.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, DOI 10.17487/RFC1191, November 1990, <<https://www.rfc-editor.org/info/rfc1191>>.

[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>>.

[RFC7684] Psenak, P., Gredler, H., Shakir, R., Henderickx, W., Tantsura, J., and A. Lindem, "OSPFv2 Prefix/Link Attribute Advertisement", RFC 7684, DOI 10.17487/RFC7684, November 2015, <<https://www.rfc-editor.org/info/rfc7684>>.

[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>>.

[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>>.

[RFC8362] Lindem, A., Roy, A., Goethals, D., Reddy Vallem, V., and F. Baker, "OSPFv3 Link State Advertisement (LSA) Extensibility", RFC 8362, DOI 10.17487/RFC8362, April 2018, <<https://www.rfc-editor.org/info/rfc8362>>.

[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>>.

[RFC8665]

Psenak, P., Ed., Previdi, S., Ed., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", RFC 8665, DOI 10.17487/RFC8665, December 2019, <<https://www.rfc-editor.org/info/rfc8665>>.

[RFC8666]

Psenak, P., Ed. and S. Previdi, Ed., "OSPFv3 Extensions for Segment Routing", RFC 8666, DOI 10.17487/RFC8666, December 2019, <<https://www.rfc-editor.org/info/rfc8666>>.

[RFC8667]

Previdi, S., Ed., Ginsberg, L., Ed., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <<https://www.rfc-editor.org/info/rfc8667>>.

Authors' Addresses

Zhibo Hu
Huawei
Huawei Bld., No.156 Beiqing Rd.
Beijing
100095
China

Email: huzhibo@huawei.com

Shuping Peng
Huawei
Huawei Bld., No. 156 Beiqing Rd.
Beijing
100095
China

Email: pengshuping@huawei.com

Xing Xi
Huawei
Huawei Bld., No. 156 Beiqing Rd.
Beijing
100095
China

Email: xixing1@huawei.com