

Network Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: 22 September 2025

Y. Zhu  
China Telecom  
Z. Hu  
S. Peng  
Huawei Technologies  
R. Mwehaire  
MTN Uganda Ltd.  
21 March 2025

Signaling Maximum Transmission Unit (MTU) using BGP-LS  
draft-ietf-idr-bgp-ls-link-mtu-09

## Abstract

Segment Routing (SR) leverages the source routing paradigm, which can be directly applied to the MPLS architecture and IPv6 architecture, with a new type of routing header. Since multiple labels or SRv6 SIDs are pushed in the packets, it is more likely that the packet size exceeds the path MTU of SR tunnel. However, SR uses the IGP as the routing protocol and does not support the negotiation of the Path MTU.

BGP Link State (BGP-LS) describes a mechanism by which link-state and TE information can be collected from networks and shared with external components using the BGP routing protocol. This document specifies the extensions to BGP-LS to carry MTU information of link. The centralized controller (PCE/SDN) calculates the Path MTU while completing the service path calculation based on the information transmitted by the BGP-LS.

## 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 RFC 2119 [RFC2119].

## 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 September 2025.

## Copyright Notice

Copyright (c) 2025 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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

## Table of Contents

1. Introduction . . . . .	2
2. Terminology . . . . .	4
3. Deploying scenarios . . . . .	4
4. BGP-LS Extensions for Link MTU . . . . .	5
5. IANA Considerations . . . . .	6
6. Security Considerations . . . . .	6
7. Acknowledgements . . . . .	7
8. Contributors . . . . .	7
9. References . . . . .	7
9.1. Normative References . . . . .	7
9.2. Informative References . . . . .	7
Authors' Addresses . . . . .	9

## 1. Introduction

[RFC9552] describes the implementation mechanism of BGP-LS by which link-state and TE information can be collected from networks and shared with external components using the BGP routing protocol [RFC4271]. BGP-LS allows the necessary Link-State Database (LSDB) and Traffic Engineering Database (TEDB) information to be collected from the IGP within the network, filtered according to configurable policy, and distributed to the PCE as necessary.

The appropriate MTU size guarantees efficient data transmission. If the MTU size is too small and the packet size is large, fragmentation may occur too much and packets are discarded by the QoS queue. If the MTU configuration is too large, packet transmission may be slow. Path MTU is the maximum length of a packet that can pass through a path without fragmentation. [RFC1191] describes a technique for dynamically discovering the maximum transmission unit (MTU) of an arbitrary internet path.

The traditional MPLS tunneling technology has signaling for establishing a path. [RFC3988] defines the mechanism for automatically discovering the Path MTU of LSPs. For a certain FEC, the LSR compares the MTU advertised by all downstream devices with the MTU of the FEC output interface in the local device, and calculates the minimum value for the upstream device.

[RFC3209] specify the mechanism of MTU signaling in RSVP-TE. The ingress node of the RSVP-TE tunnel sends a Path message to the downstream device. The Adspec object in the Path message carries the MTU. Each node along the tunnel receives a Path message, compares the MTU value in the Adspec object with the interface MTU value and MPLS MTU configured on the physical output interface of the local tunnel, obtains the minimum MTU value, and puts it into the newly constructed Path message and continues to send it to the downstream equipment. Thus, the MTU carried in the Path message received by the Egress node is the minimum value of the path MTU. The Egress node brings the negotiated Path MTU back to the Ingress node through the Resv message.

Segment Routing (SR) described in [RFC8402] leverages the source routing paradigm. Segment Routing can be directly applied to the MPLS architecture with no change on the data plane [RFC8660] and applied to the IPv6 architecture with a new type of routing header called the SR header (SRH) [RFC8754]. [RFC9085] defines SR extensions to BGP-LS and specifies the TLVs and sub-TLVs for advertising SR information. Based on the SR information reported by the BGP-LS, the SDN can calculate the end-to-end explicit SR-TE paths or SR Policies.

Nevertheless, Segment Routing is a tunneling technology based on the IGP protocol as the routing protocol, and there is no additional signaling for establishing the path. so the Segment Routing tunnel cannot currently support the negotiation mechanism of the MTU. Multiple labels or SRv6 SIDs are pushed in the packets. This causes the length of the packets encapsulated in the Segment Routing tunnel to increase during packet forwarding. This is more likely to cause packet size exceed the traditional MPLS packet size.

This document specifies the extension to BGP Link State (BGP-LS) to carry link maximum transmission unit (MTU) messages.

## 2. 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.

## 3. Deploying scenarios

This document suggests a solution to extension to BGP Link State (BGP-LS) to carry maximum transmission unit (MTU) messages. The MTU information of the link is acquired through the process of collecting link state and TE information by BGP-LS. Concretely, a router maintains one or more databases for storing link-state information about nodes and links in any given area. The router's BGP process can retrieve topology from these IGP, BGP and other sources, and

distribute it to a consumer, either directly or via a peer BGP speaker (typically a dedicated Route Reflector). [RFC7176] specifies a possible way of using the IS-IS mechanism and extensions for link MTU Sub-TLV. In the case of inter-AS scenario (e.g., BGP EPE), the link MTU of the inter-AS link can be collected via BGP-LS directly.

As per [RFC9552], the collection of link-state and TE information and its distribution to consumers is shown in the following figure.

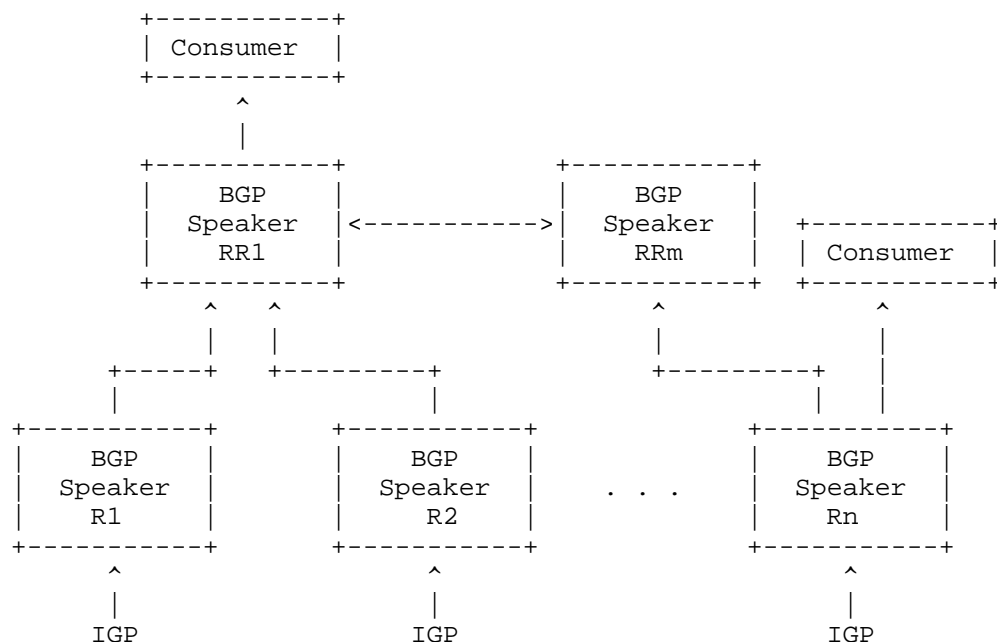


Figure 1: Collection of Link-State and TE Information

Please note that this signaled MTU may be different from the actual MTU, which is usually from configuration mismatches in a control plane and a data plane component.

#### 4. BGP-LS Extensions for Link MTU

[RFC9552] defines the BGP-LS NLRI that can be a Node NLRI, a Link NLRI or a Prefix NLRI. The corresponding BGP-LS attribute is a Node Attribute, a Link Attribute or a Prefix Attribute. [RFC9552] defines the TLVs that map link-state information to BGP-LS NLRI and the BGP-LS attribute. Therefore, according to this document, a new sub-TLV is added to the Link Attribute TLV. It is an independent attribute TLV that can be used for the link NLRI advertised with all the Protocol IDs.

The format of the sub-TLV is as shown below.

```
x  TYPE      - TBD
x  LENGTH    - Total length of the value field, it should be 2
x  VALUE     - 2-byte MTU value of the link
```

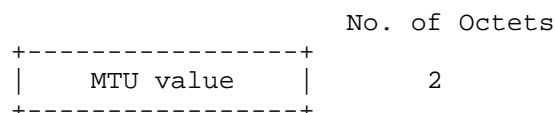


Figure 2: Sub-TLV Format for Link MTU

Whenever there is a change in MTU value represented by Link Attribute TLV, BGP-LS should re-originate the respective TLV with the new MTU value.

## 5. IANA Considerations

This document requests assigning a new code-point from the BGP-LS Link Descriptor and Attribute TLVs registry as specified in section 4.

Value	Description	Reference
TBD	Link MTU	This document

## 6. Security Considerations

Procedures and protocol extensions defined in this document do not affect the base BGP security model. See [RFC6952] for details. The security considerations of the base BGP-LS specification as described in [RFC9552] also apply.

SR operates within a trusted SR domain [RFC8402] and its security considerations also apply to BGP sessions when carrying MTU information. The MTU information advertised to controllers and other applications via BGP-LS is expected to be used entirely within this trusted SR domain.

The general guidance for BGP-LS with respect to the isolation of BGP-LS sessions from BGP sessions for other address-families (refer to security considerations of [RFC9552]) may be used to ensure that the MTU information is not advertised by accident or error to an EBGp peering session outside the SR domain.

Additionally, it may be considered that the export of MTU information, as described in this document, poses a risk to the network's cybersecurity. Specifically, attackers are better able to construct segmented packets for fragmentation attacks through MTU leakage to confuse network devices, consume resources, or perform other attacks by sending a large number of fragmented packets with different offsets and sizes. Thus, it is the responsibility of the network operator to ensure that only trusted nodes (that include both routers and controller applications) are configured to receive such information.

## 7. Acknowledgements

## 8. Contributors

Gang Yan  
Huawei  
China

Email:yangang@huawei.com

Junda Yao  
Huawei  
China

Email:yaojunda@huawei.com

## 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>>.

### 9.2. Informative References

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

[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>>.
- [RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <<https://www.rfc-editor.org/info/rfc4271>>.
- [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>>.
- [RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, DOI 10.17487/RFC6952, May 2013, <<https://www.rfc-editor.org/info/rfc6952>>.
- [RFC7176] Eastlake 3rd, D., Senevirathne, T., Ghanwani, A., Dutt, D., and A. Banerjee, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC 7176, DOI 10.17487/RFC7176, May 2014, <<https://www.rfc-editor.org/info/rfc7176>>.
- [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>>.
- [RFC8402] Filtsfils, 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>>.
- [RFC8660] Bashandy, A., Ed., Filtsfils, C., Ed., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with the MPLS Data Plane", RFC 8660, DOI 10.17487/RFC8660, December 2019, <<https://www.rfc-editor.org/info/rfc8660>>.
- [RFC8754] Filtsfils, C., Ed., Dukes, D., Ed., Previdi, S., Leddy, J., Matsushima, S., and D. Voyer, "IPv6 Segment Routing Header (SRH)", RFC 8754, DOI 10.17487/RFC8754, March 2020, <<https://www.rfc-editor.org/info/rfc8754>>.



- [RFC9085] Previdi, S., Talaulikar, K., Ed., Filsfils, C., Gredler, H., and M. Chen, "Border Gateway Protocol - Link State (BGP-LS) Extensions for Segment Routing", RFC 9085, DOI 10.17487/RFC9085, August 2021, <<https://www.rfc-editor.org/info/rfc9085>>.
- [RFC9552] Talaulikar, K., Ed., "Distribution of Link-State and Traffic Engineering Information Using BGP", RFC 9552, DOI 10.17487/RFC9552, December 2023, <<https://www.rfc-editor.org/info/rfc9552>>.

## Authors' Addresses

Yongqing Zhu  
China Telecom  
109, West Zhongshan Road, Tianhe District.  
Guangzhou  
510000  
China  
Email: zhuyq8@chinatelecom.cn

Zhibo Hu  
Huawei Technologies  
Huawei Bld., No.156 Beiqing Rd.  
Beijing  
100095  
China  
Email: huzhibo@huawei.com

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

Robbins Mwehaire  
MTN Uganda Ltd.  
Uganda  
Email: Robbins.Mwehair@mtn.com