Workgroup: Interdomain Routing Working Group
Internet-Draft:
draft-ietf-idr-sr-policy-path-mtu-08
Published: 19 October 2023
Intended Status: Standards Track
Expires: 21 April 2024
Authors: C. Li, Huawei Technologies
Y. Zhu, China Telecom
A. Sawaf, Saudi Telecom Company
Z. Li, Huawei Technologies

Segment Routing Path MTU in BGP

Abstract

Segment Routing is a source routing paradigm that explicitly indicates the forwarding path for packets at the ingress node. An SR policy is a set of candidate SR paths consisting of one or more segment lists with necessary path attributes. However, the path maximum transmission unit (MTU) information for SR path is not available in the SR policy since the SR does not require signaling. This document defines extensions to BGP to distribute path MTU information within SR policies.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Segment routing (SR) [RFC8402] is a source routing paradigm that explicitly indicates the forwarding path for packets at the ingress node. The ingress node steers packets into a specific path according to the Segment Routing Policy (SR Policy) as defined in [I-D.ietf-spring-segment-routing-policy]. In order to distribute SR policies to the headend, [I-D.ietf-idr-segment-routing-te-policy] specifies a mechanism by using BGP.

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 needs 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. Likewise, SRv6 pakcets will be dropped if the packet size is larger than path MTU, since IPv6 packet can not be fragmented on transmission [RFC8200] .

The host may discover the PMTU by Path MTU Discovery (PMTUD) [RFC8201] or other mechanisms. But the ingress still needs to examine the packet size for dropping too large packets to avoid malicious traffic or error traffic. Also, the packet size may exceeds the PMTU because of the new encapsulation of SR-MPLS or SRv6 packet at the ingress.

In order to check whether the Packet size exceeds the PMTU or not, the ingress node needs to know the Path MTU associated to the forwarding path. However, the path maximum transmission unit (MTU) information for SR path is not available since the SR does not require signaling.

This document defines extensions to BGP to distribute path MTU information within SR policies. The Link MTU information can be obtained via BGP-LS [I-D.zhu-idr-bgp-ls-path-mtu] or some other means. With the Link MTU, the controller can compute the PMTU and convey the information via the BGP SR policy.

2. Terminology

This memo makes use of the terms defined in [RFC8402] and [RFC3209].

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 other payload) and the label stack but does not include any lower- headers. A link may be an interface (such as Ethernet or Packet-o 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 the egress(es), over each valid (forwarding) path. This size inclu the IP header and data (or other payload) and any part of the labe stack that was received by the ingress LSR before it placed the pa into the LSP (this part of the label stack is considered part of t payload for this LSP). The size does not include any lower-level headers.

Note that: The PMTU value may be modified by subtracting some over introduced by protection mechanism, like TI-LFA. Therefore, the va of PMTU dilivered to the ingress node MAY be smaller than the mini link MTU of all the links in a path between a source node and a destination node.

2.1. 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. SR Policy for Path MTU

As defined in [I-D.ietf-idr-segment-routing-te-policy] , the SR policy encoding structure is as follows:

```
SR Policy SAFI NLRI: <Distinguisher, Policy-Color, Endpoint>
Attributes:
   Tunnel Encaps Attribute (23)
      Tunnel Type: SR Policy
          Binding SID
          Preference
          Priority
          Policy Name
          Explicit NULL Label Policy (ENLP)
          Segment List
              Weight
              Segment
              Segment
              ...
          ...
```

As introduced in Section 1, each SR path has it's path MTU. SR policy with SR path MTU information is expressed as below:

```
SR Policy SAFI NLRI: <Distinguisher, Policy-Color, Endpoint>
Attributes:
   Tunnel Encaps Attribute (23)
      Tunnel Type: SR Policy
          Binding SID
          Preference
          Priority
          Policy Name
          Explicit NULL Label Policy (ENLP)
          Segment List
              Weight
              Path MTU
              Segment
              Segment
              ...
          ...
```

3.1. Path MTU Sub-TLV

A Path MTU sub-TLV is an Optional sub-TLV. When it appears, it must appear only once at most within a Segment List sub-TLV. If multiple

Path MTU sub-TLVs appear within a Segment List sub-TLV, the NLRI MUST be treated as a malformed NLRI.

As per [I-D.ietf-idr-segment-routing-te-policy], when the error determined allows for the router to skip the malformed NLRI(s) and continue processing of the rest of the update message, then it MUST handle such malformed NLRIs as 'Treat-as-withdraw'. This document does not define new error handling rules for Path MTU sub-TLV, and the error handling rules defined in [I-D.ietf-idr-segment-routing-te-policy] apply to this document.

A Path MTU sub-TLV is associated with an SR path specified by a segment list sub-TLV or a path segment [I-D.ietf-spring-mpls-path-segment] [I-D.ietf-spring-srv6-path-segment]. The Path MTU sub-TLV has the following format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |    Length     |            RESERVED           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                            Path MTU                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1. Path MTU sub-TLV

Where:

Type: to be assigned by IANA.

Length: the total length of the value field not including Type and Length fields.

Reserved: 16 bits reserved and MUST be set to 0 on transmission and MUST be ignored on receipt.

Path MTU: 4 bytes value of path MTU in octets. The value can be calculated by a central controller or other devices based on the information that learned via IGP of BGP-LS or other means.

Whenever the path MTU of a physical or logical interface is changed, a new SR policy with new path MTU information should be updated accordingly by BGP.

4. Operations

The document does not bring new operation beyond the description of operations defined in [I-D.ietf-idr-segment-routing-te-policy]. The existing operations defined in

[I-D.ietf-idr-segment-routing-te-policy] can apply to this document directly.

Typically but not limit to, the SR policies carrying path MTU infomation are configured by a controller.

After configuration, the SR policies carrying path MTU infomation will be advertised by BGP update messages. The operation of advertisement is the same as defined in [I-D.ietf-idr-segment-routing-te-policy], as well as the receiption.

The consumer of the SR policies is not the BGP process. The operation of sending information to consumers is out of scope of this document.

5. Implementation Status

[Note to the RFC Editor - remove this section before publication, as well as remove the reference to [RFC7942].

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

5.1. Huawei's Commercial Delivery

The feature has been implemented on Huawei VRP8.

- Organization: Huawei

- Implementation: Huawei's Commercial Delivery implementation based on VRP8.

- Description: The implementation has been done.

*Maturity Level: Product

*Contact: guokeqiang@huawei.com

6. IANA Considerations

This document defines a new Sub-TLV in registries "SR Policy List Sub- TLVs" [I-D.ietf-idr-segment-routing-te-policy]:

Value	Description	Reference
TBA	Path MTU sub-TLV	This document

7. Security Considerations

TBA

8. Contributors

Jun Qiu

Huawei Technologies

China

Email: qiujun8@huawei.com

9. Acknowledgements

Authors would like to thank Ketan Talaulikar, Aijun Wang, Weiqiang Cheng, Huanan Chen, Chongfeng Xie, Stefano Previdi, Taishan Tang, Keqiang Guo, Chen Zhang, Susan Hares, Weiguo Hao, Gong Xia, Bing Yang, Linda Dunbar, Shunwan Zhuang, Huaimo Chen, Mach Chen, Jingring Xie, Zhibo Hu, Jimmy Dong and Jianwei Mao for their proprefessional comments and help.

10. References

10.1. Normative References

[I-D.ietf-idr-segment-routing-te-policy]
Previdi, S., Filsfils, C., Talaulikar, K., Mattes, P., and D. Jain, "Advertising Segment Routing Policies in BGP", Work in Progress, Internet-Draft, draft-ietf-idr-segment-routing-te-policy-25, 26 September 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-idr-segment-routing-te-policy-25>.

[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-22, 22 March 2022, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-segment-routing-policy-22>.

[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>.

[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>.

[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>.

10.2. Informative References

[I-D.ietf-spring-mpls-path-segment] Cheng, W., Li, H., Li, C., Gandhi, R., and R. Zigler, "Path Segment in MPLS Based Segment Routing Network", Work in Progress, Internet-Draft, draft-ietf-spring-mpls-path-segment-16, 12 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-spring-mpls-path-segment-16>.

[I-D.ietf-spring-srv6-path-segment] Li, C., Cheng, W., Chen, M., Dhody, D., and Y. Zhu, "Path Segment for SRv6 (Segment Routing in IPv6)", Work in Progress, Internet-Draft, draft-ietf-spring-srv6-path-segment-07, 19 October 2023, <https://datatracker.ietf.org/api/v1/doc/document/draft-ietf-spring-srv6-path-segment/>.

[I-D.zhu-idr-bgp-ls-path-mtu] Zhu, Y., Hu, Z., Peng, S., and R. Mwehair, "Signaling Maximum Transmission Unit (MTU) using BGP-LS", Work in Progress, Internet-Draft, draft-zhu-idr-bgp-ls-path-mtu-05, 17 November 2020, <https://datatracker.ietf.org/doc/html/draft-zhu-idr-bgp-ls-path-mtu-05>.

[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>.

[RFC7942]
Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <https://www.rfc-editor.org/info/rfc7942>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[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>.

Authors' Addresses

Cheng Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing
100095
China

Email: c.l@huawei.com

YongQing Zhu
China Telecom
109, West Zhongshan Road, Tianhe District.
Guangzhou
China

Email: zhuyq8@chinatelecom.cn

Ahmed El Sawaf
Saudi Telecom Company
Riyadh
Saudi Arabia

Email: aelsawaf.c@stc.com.sa

Zhenbin Li
Huawei Technologies
Huawei Campus, No. 156 Beiqing Rd.
Beijing
100095
China

Email: lizhenbin@huawei.com