Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 27 September 2024

X. Xu
China Mobile
S. Hegde
Juniper Networks
B. Pismenny
Nvidia
D. Zhang
L. Xia
Huawei
M. Puttaswamy
Juniper Networks

26 March 2024

Encapsulating IPsec ESP in UDP for Load-balancing
draft-xu-ipsecme-esp-in-udp-lb-12

Abstract

IPsec Virtual Private Network (VPN) is widely used by enterprises to interconnect their geographical dispersed branch office locations across the Wide Area Network (WAN) or the Internet, especially in the Software-Defined-WAN (SD-WAN) era. In addition, IPsec is also increasingly used by cloud providers to encrypt IP traffic traversing data center networks and data center interconnect WANs so as to meet the security and compliance requirements, especially in financial cloud and governmental cloud environments. To fully utilize the bandwidth available in the data center network, the data center interconnect WAN or the Internet, load balancing of IPsec traffic over Equal Cost Multi-Path (ECMP) and/or Link Aggregation Group (LAG) is much attractive to those enterprises and cloud providers. This document defines a method to encapsulate IPsec Encapsulating Security Payload (ESP) packets over UDP tunnels for improving load-balancing of IPsec ESP traffic.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

IPsec Virtual Private Network (VPN) is widely used by enterprises to interconnect their geographical dispersed branch office locations across the Wide Area Network (WAN) or the Internet, especially in the Software-Defined-WAN (SD-WAN) era. In addition, IPsec is also increasingly used by cloud providers to encrypt IP traffic traversing data center networks and data center interconnect WANs so as to meet the security and compliance requirements, especially in financial cloud and governmental cloud environments. To fully utilize the bandwidth available in the WAN or the Internet, load balancing of IPsec traffic over Equal Cost Multi-Path (ECMP) and/or Link Aggregation Group (LAG) is much attractive to those enterprises and cloud providers. Although the ESP SPI field within the IPsec packets can be used as the load-balancing key, but it cannot be used by legacy switches and routers.

Since most existing switches within data center networks and core routers within IP WAN or the Internet can already support balancing IP traffic flows based on the hash of the five-tuple of UDP packets, by encapsulating IPsec Encapsulating Security Payload (ESP) packets over UDP tunnels with the UDP source port being used as an entropy field, it will enable existing data center switches and core routers to perform efficient load-balancing of the IPsec ESP traffic without requiring any change to them. Therefore, this specification defines a method of encapsulating IPsec ESP packets over UDP tunnels for improving load-balancing of IPsec ESP traffic.

IPsec VPN gateways are usually implemented in the form of multi-core x86 servers, especially in the public cloud environment. Receive Side Scaling (RSS) is a widely adopted network driver technology which spreads incoming TCP or UDP traffic across multiple CPUs by performing hash function on the network and/or transport layer headers, resulting in increased multi-core efficiency and processor cache utilization. By encapsulating ESP in UDP, it would facilate RSS to distribute the received IPsec traffic more evenly across multiple CPU cores.

Encapsulating ESP in UDP, as defined in this document, can be used in both IPv4 and IPv6 networks. IPv6 flow label has been proposed as an entropy field for load balancing in IPv6 network environment [RFC6438]. However, as stated in [RFC6936], the end-to-end use of flow labels for load balancing is a long-term solution and therefore the use of load balancing using the transport header fields would continue until any widespread deployment is finally achieved. As such, ESP-in-UDP encapsulation would still have a practical application value in the IPv6 networks during this transition timeframe.

Note that the difference between the ESP-in-UDP encapsulation as proposed in this document and the ESP-in-UDP encapsulation as described in [RFC3948] is that the former uses the UDP tunnel for load-balancing improvement purpose and therefore the source port is used as an entropy field while the latter uses the UDP tunnel for NAT traversal purpose and therefore the source port is set to a constant value (i.e., 4500). In addition, the ESP-in-UDP encapsulation as described in this document is applicable to both the tunnel mode ESP encapsulation and the transport mode ESP encapsulation.

There are use cases that do not use NAT traversal such as multi-cloud WAN. ESP-in-UDP encapsulation along with NAT traversal is out of scope in this document.

1.1. 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].

2. Terminology

This memo makes use of the terms defined in [RFC2401]and [RFC2406].

3. Encapsulation in UDP

ESP-in-UDP encapsulation format is shown as follows:

```
     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
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |    Source Port = Entropy      |        Dest Port = TBD1       |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |           UDP Length          |        UDP Checksum           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    ~                           ESP Packet                          ~
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1: ESP-in-UDP Encapsulation Format

Source Port of UDP:

This field contains a 16-bit entropy value that is generated by the encapsulator to uniquely identify a flow. What constitutes a flow is locally determined by the encapsulator and therefore is outside the scope of this document. What algorithm is actually used by the encapsulator to generate an entropy value is outside the scope of this document.

In case the tunnel does not need entropy, this field of all packets belonging to a given flow SHOULD be set to a randomly selected constant value so as to avoid packet reordering.

To ensure that the source port number is always in the range 49152 to 65535 (Note ports less than 49152 are reserved by IANA to identify specific applications/protocols) which may be required in some cases, instead of calculating a 16-bit hash, the encapsulator SHOULD calculate a 14-bit hash and use those 14 bits as the least significant bits of the source port field while the most significant two bits SHOULD be set to binary 11. That still conveys 14 bits of entropy information which would be enough as well in practice.

Destination Port of UDP:

This field is set to a value (TBD1) allocated by IANA to indicate that the UDP tunnel payload is an ESP packet.

UDP Length:

The usage of this field is in accordance with the current UDP specification [RFC0768].

UDP Checksum:

For IPv4 UDP encapsulation, this field is RECOMMENDED to be set to zero for performance or implementation reasons because the IPv4 header includes a checksum and use of the UDP checksum is optional with IPv4. For IPv6 UDP encapsulation, the IPv6 header does not include a checksum, so this field MUST contain a UDP checksum that MUST be used as specified in [RFC0768] and [RFC2460] unless one of the exceptions that allows use of UDP zero-checksum mode (as specified in [RFC6935]) applies.

ESP Packet:

This field contains one ESP packet.

4. Processing Procedures

This ESP-in-UDP encapsulation causes ESP [RFC2406] packets to be forwarded across IP WAN via "UDP tunnels". When performing ESP-in-UDP encapsulation by an IPsec VPN gateway, ordinary ESP encapsulation procedure is performed and then a formatted UDP header is inserted between ESP header and IP header. The Source Port field of the UDP header is filled with an entropy value which is generated by the IPsec VPN gateway. Upon receiving these UDP encapsulated packets, remote IPsec VPN gateway MUST decapsulate these packets by removing the UDP header and then perform ordinary ESP decapsulation procedure consequently.

Similar to all other IP-based tunneling technologies, ESP-in-UDP encapsulation introduces overheads and reduces the effective Maximum Transmission Unit (MTU) size. ESP-in-UDP encapsulation may also impact Time-to-Live (TTL) or Hop Count (HC) and Differentiated Services (DSCP). Hence, ESP-in-UDP MUST follow the corresponding procedures defined in [RFC2003].

Encapsulators MUST NOT fragment ESP packet, and when the outer IP header is IPv4, encapsulators MUST set the DF bit in the outer IPv4 header. It is strongly RECOMMENDED that IP transit core be configured to carry an MTU at least large enough to accommodate the added encapsulation headers. Meanwhile, it is strongly RECOMMENDED that Path MTU Discovery [RFC1191] [RFC1981] or Packetization Layer Path MTU Discovery (PLPMTUD) [RFC4821] is used to prevent or minimize fragmentation.

5. Congestion Considerations

TBD.

6. Applicability Statements

TBD.

7. Acknowledgements

8. IANA Considerations

One UDP destination port number indicating ESP needs to be allocated by IANA:

Service Name: ESP-in-UDP Transport Protocol(s):UDP
Assignee: IESG <iesg@ietf.org>
Contact: IETF Chair <chair@ietf.org>.
Description: Encapsulate ESP packets in UDP tunnels.
Reference: This document.
Port Number: TBD1 -- To be assigned by IANA.

9. Security Considerations

If source port is generated using inner packet parameters, care should be taken to not reveal those parameters. Including some random bytes along with the inner packet parameters will ensure the information of inner IP header is not revealed.

Because packets are traversing different paths and the ESP sequence number is assigned sequencially by the encapsulator irrespective of the packet flow, the receiver might receive packets out-of-order and end up dropping them as delayed/out-of-order packets. Based on the network speed and load, administrator should be able to adjust the replay window size or entirely disable the replay check.

10. References

10.1. Normative References

[RFC0768] Postel, J., "User Datagram Protocol", STD 6, RFC 768, DOI 10.17487/RFC0768, August 1980, <https://www.rfc-editor.org/info/rfc768>.

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

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, DOI 10.17487/RFC1981, August 1996, <https://www.rfc-editor.org/info/rfc1981>.

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, DOI 10.17487/RFC2003, October 1996, <https://www.rfc-editor.org/info/rfc2003>.

[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>.

[RFC2401] Kent, S. and R. Atkinson, "Security Architecture for the Internet Protocol", RFC 2401, DOI 10.17487/RFC2401, November 1998, <https://www.rfc-editor.org/info/rfc2401>.

[RFC2406] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload (ESP)", RFC 2406, DOI 10.17487/RFC2406, November 1998, <https://www.rfc-editor.org/info/rfc2406>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <https://www.rfc-editor.org/info/rfc2460>.

[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>.

[RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label for Equal Cost Multipath Routing and Link Aggregation in Tunnels", RFC 6438, DOI 10.17487/RFC6438, November 2011, <https://www.rfc-editor.org/info/rfc6438>.

[RFC6935] Eubanks, M., Chimento, P., and M. Westerlund, "IPv6 and UDP Checksums for Tunneled Packets", RFC 6935, DOI 10.17487/RFC6935, April 2013, <https://www.rfc-editor.org/info/rfc6935>.

[RFC6936] Fairhurst, G. and M. Westerlund, "Applicability Statement for the Use of IPv6 UDP Datagrams with Zero Checksums", RFC 6936, DOI 10.17487/RFC6936, April 2013, <https://www.rfc-editor.org/info/rfc6936>.

10.2. Informative References

[RFC3948] Huttunen, A., Swander, B., Volpe, V., DiBurro, L., and M. Stenberg, "UDP Encapsulation of IPsec ESP Packets", RFC 3948, DOI 10.17487/RFC3948, January 2005, <https://www.rfc-editor.org/info/rfc3948>.

Authors' Addresses

Xiaohu Xu
China Mobile
Email: xuxiaohu_ietf@hotmail.com

Shraddha Hegde
Juniper Networks
Email: shraddha@juniper.net

Boris Pismenny
Nvidia
Email: borisp@nvidia.com

Dacheng Zhang
Huawei
Email: dacheng.zhang@huawei.com

Liang Xia
Huawei
Email: frank.xialiang@huawei.com

Mahendra Puttaswamy
Juniper Networks
Email: mpmahendra@juniper.net