Workgroup: IPsecme
Internet-Draft:
draft-liu-ipsecme-ikev2-mtu-dect-02
Published: 13 May 2022
Intended Status: Standards Track
Expires: 14 November 2022
Authors: D. Liu, Ed. D. Migault R. Liu C. Zhang
Ericsson Ericsson Ericsson Ericsson

IKEv2 IPv4 Downstream Fragmentation Notification Extension

Abstract

This document defines the IKEv2 IPv4 Downstream Fragmentation Notification Extension which enables a receiving security gateway to notify the sending receiving gateway that downstream fragmentation is ongoing. The sending gateway MAY take action to avoid such fragmentation to occur.

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 14 November 2022.

Table of Contents

1. Introduction

This document considers two security gateways interconnecting two security domains using IPsec/ESP over an untrusted IPv4 network.

As per [RFC0791], IPv4 packets crossing the untrusted network may be non fragmentable (by setting their Don't Fragment bit to 1), to prevent the fragmentation by any downstream node. In that case, when an incoming packet is larger than the accepted Maximum Transmission Unit (MTU), the packet is dropped and an ICMPv4 message Packet Too Big (PTB) [RFC0792] is returned to the sending address. The ICMPv4 PTB message is a Destination Unreachable message with Code equal to 4 and was augmented by [RFC1191] to indicate the acceptable MTU. Unfortunately, one cannot rely on such procedure as in practice some downstream router do not check the MTU and as such do not send ICMPv4 messages. In addition, when ICMv4 message are sent these message are unprotected, and may be blocked by firewalls or ignored. This results in IPv4 packets being dropped without the security gateways being aware of it which is also designated as black holing.

To prevent this situation, IPv4 packets are often fragmentable with their DF bit set to 0. In this case, as described in [RFC0792], when a packet size exceeds its MTU, the node fragments the incoming packet in multiple fragments. The inconvenient is that the receiving security gateway will have to re-assembled the multiple fragments to rebuilt an ESP packet before being able to apply the IPsec decapsulation. Fragments reassembling comes requires additional resources which under heavy load results in service degradations. Firstly, fragment reassemble requires the security gateway to handle states for undefinite time. Then, as detailed in [RFC4963], [RFC6864] or [RFC8900], the 16-bit IPv4 identification field that is not large enough to prevent duplication making fragmentation not

sufficiently robust at high data rates. Such service degradation could be avoided by being able to indicate the sending gateway to send packets of a smaller size.

This document defines IKEv2 IPv4 Downstream Fragmentation Notification Extension so a receiving security gateway can notify the sending receiving gateway that downstream fragmentation is ongoing. Similarly to ICMPv4 PTB [RFC0792], the notification carries an indication of an acceptable MTU value, so the sending gateway reduces the MTU of its packets. This includes indicating the MTU to the source host of the inner packet, fragmenting the inner IPv4 packet, performing source fragmentation.

This mechanism follows the [RFC8900] that recommends each layer handles fragmentation at their layer and to reduce the reliance on IP fragmentation to the greatest degree possible. This document does not describes a Path MTU Discovery (PMTUD) procedure [RFC1191] nor an Execute Packetization Layer PMTUD (PLMTUD) [RFC4821] procedure.

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

3. IPv4 Downstream Fragmentation Support Negotiation

During an IKEv2 negotiation, the initiator and the responder indicate their support for notifying an IPv4 Downstream Fragmentation by exchanging the IP4_DOWNSTREAM_FRAGMENTATION_SUPPORTED notifications. This notification MUST be sent in the IKE_AUTH exchange (in case of multiple IKE_AUTH exchanges - in the first IKE_AUTH message from initiator and in the last IKE_AUTH message from responder). If both the initiator and the responder send this notification during the IKE_AUTH exchange, peers may notify each other when IPv4 Downstream Fragmentation is observed. Upon receiving such notifications, the peers may take the necessary actions to prevent such fragmentation to occur.

```
Initiator                          Responder
-------------------------------------------------------------------
HDR, SA, KEi, Ni -->
                          <-- HDR, SA, KEr, Nr
HDR, SK {IDi, AUTH,
     SA, TSi, TSr,
     N(IP4_DOWNSTREAM_FRAGMENTATION_SUPPORTED)} -->
                          <-- HDR, SK {IDr, AUTH,
                              SA, TSi, TSr,
                              N(IP4_DOWNSTREAM_FRAGMENTATION_SUPPORTED)}
```

4. IPv4 Downstream Fragmentation Notification

Section 4.1 indicates how the receiving security gateway detects downstream fragmentation, the MTU to be used and notifies the sending security gateway with IP4_DOWNSTREAM_FRAGMENTATION notification. Section 4.2 details how the sending security gateway reduces its MTU upon receiving a IP4_DOWNSTREAM_FRAGMENTATION notification.

4.1. Sending Downstream Fragmentation Notification

As defined in [RFC0792] IPv4 fragmentation can be handled by any node, that is the host as well as any router on path. Figure 1 shows the IPv4 Header as described in [RFC0791] section 3.1 to illustrate the different fields involved.

A sending gateway supporting the IPv4 Downstream Fragmentation extension and performing fragmentation at the source, SHOULD set the DF bit to 1 on each ESP fragment to avoid any further (Downstream) fragmentation. As a result, a received IPv4 ESP packet with its DF bit set to 0 is a suspected of being fragmented by a downstream router. The receiving security gateway records the corresponding Total Length field as a potential ongoing MTU on any initial fragment. An initial fragment is an ESP packets with the More Fragments (MF) bit is set to 1, and Fragment Offset set to 0.

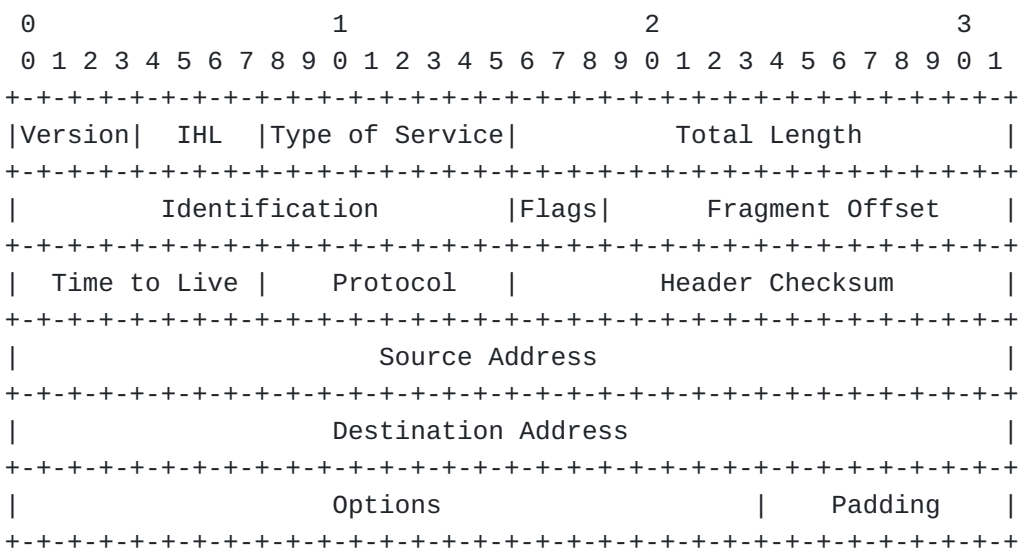

Figure 1: IPv4 Header

Based on internal heuristics, the receiving security gateway MAY decide to inform the sending security gateway that more than expected refragmentation operations are observed. Such heuristics include, for example, a threshold for number of initial fragment received, a threshold for a certain rate of initial fragments. Such thresholds are also expected to be combined with a timer or a counter of already sent IP4_DOWNSTREAM_FRAGMENTATION notifications to avoid overloading the sending gateways with such notifications. It is expected that the time between two such notifications increases with the number of notifications. The receiving security gateway determines a recommended MTU value to be used by the sending gateway. The recommended MTU SHOULD be one of the potential ongoing MTU observed from IPv4 ESP packets that have been correctly authenticated. The recommended MTU SHOULD be greater than some minimal values. [RFC0791] specifies the IPv4 minimum MTU is 68 octets, but greater values are likely to be more realistic. Once the appropriated MTU has been selected, the receiving security gateway sends the sending gateway a IP4_DOWNSTREAM_FRAGMENTATION notification to the sending gateway as described below:

```
Receiving Security Gateway                Sending Security Gateway
-------------------------------------------------------------------
HDR SK { N(IP4_DOWNSTREAM_FRAGMENTATION)} -->
```

4.2. Handling Downstream Fragmentation Notification

Upon receiving a IP4_DOWNSTREAM_FRAGMENTATION notification, the sending node checks the proposed MTU is greater than a minimum acceptable value as well as as lower than the one currently in use with the SAs associated to the IKEv2_SA. If such criteria are not met, the notification is ignored, otherwise the sending security

gateway SHOULD try to reduce its message MTU using one or a combination of the actions described below:

1. The security gateway SHOULD request the hosts to update their MTU, so the resulting ESP packet does not exceed the recommended MTU of the IP4_DOWNSTREAM_FRAGMENTATION notification. The resulting MTU of the inner packet is designated as inner MTU [I-D.ietf-intarea-tunnels]. For each incoming inner packet, the security gateway checks the packet length with the inner MTU. When the packet length exceeds the inner MTU, the security gateway SHOULD discard the packet and send back a ICMPv4 PTB [RFC1191] (resp. an ICMPv6 PTB [RFC4443]) if the sender's IP address is an IPv4 (resp. IPv6) address. The expectation is that the sender will adjust its packet size to the inner MTU.

2. If the inner packets have their DF bit set to 0, the security gateway MAY perform inner fragmentation. Note that this assumes the destination node of the inner packet will be able to perform the defragmentation operation which is only mandated by [RFC0791] for IPv4 packets up to 576 bytes. As a result, the security gateway should be aware that fragmentation may not be handled by the destination node.

3. The sending security gateway MAY perform the outer fragmentation so that fragments fit the recommended MTU of the IP4_DOWNSTREAM_FRAGMENTATION notification. When doing so, the security gateway SHOULD set the DF bit to 1, so the receiving security gateway knows fragmentation is performed by the host and does not continue to send IP4_DOWNSTREAM_FRAGMENTATION notification. Note that setting the DF bit to 1 exposes the communication to potential black holing. Note also that this action does not prevent the receiving security gateway to perform refragmentation and as such has limited impact in term of performance gain.

The sending security gateway MAY perform a PMTUD to further verify the MTU value to be used. As network configuration are dynamic, the MTU may change over time, and the sending security gateway SHOULD consider moving back to the initial value of the MTU. Such time is expected to be configured, and might be further defined by PMTUD mechanisms that are outside the scope of this document.

5. Payload Description

Figure 2 illustrates the Notify Payload packet format as described in Section 3.10 of [RFC7296] with a 4 bytes path allowed MTU value as notification data. This format is used for both the

IP4_DOWNSTREAM_FRAGMENTATION_SUPPORTED and
IP4_DOWNSTREAM_FRAGMENTATION notifications.

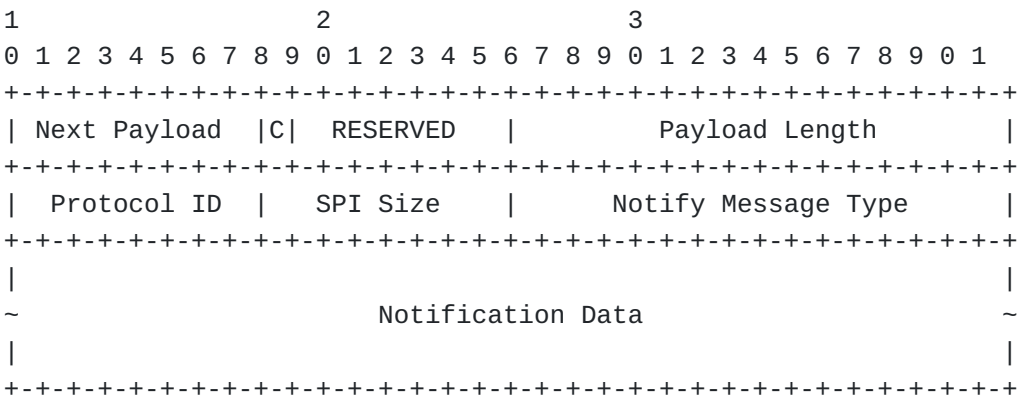

Figure 2: Notify Message Format

The fields Next Payload, Critical Bit, RESERVED and Payload Length are defined in [RFC7296]. Specific fields defined in this document are:

Protocol ID (1 octet): set to zero. SPI Size (1 octet):

set to zero. Notify Message Type (2 octets):

Specifies the type of notification message. It is set to TBD1 by IANA for the IP4_DOWNSTREAM_FRAGMENTATION_SUPPORTED notification or to TBD2 by IANA for the IP4_DOWNSTREAM_FRAGMENTATION notification. Notification Data:

Specifies the data associated to the notification message. It is empty for the IP4_DOWNSTREAM_FRAGMENTATION_SUPPORTED notification or a 4 octets that contains the MTU value for the IP4_DOWNSTREAM_FRAGMENTATION notification - as represented in Figure 3.

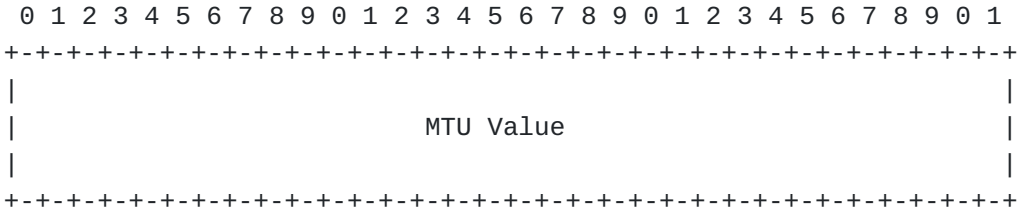

Figure 3: Notification Data for IP4_DOWNSTREAM_FRAGMENTATION

6. IANA Considerations

IANA is requested to allocate two values in the "IKEv2 Notify Message Types - Status Types" registry (available at https://www.iana.org/assignments/ikev2-parameters/ikev2-parameters.xhtml#ikev2-parameters-16) with the following definition:

Value	NOTIFY MESSAGES - STATUS TYPES
TBD1	IP4_DOWNSTREAM_FRAGMENTATION_SUPPORTED
TBD2	IP4_DOWNSTREAM_FRAGMENTATION

7. Security Considerations

This document defines an IKEv2 extension that informs a sending gateway that fragmentation is observed. In addition, an observed MTU value is reported to the sending security gateway. These pieces of information are inferred from a valid ESP packet that is authenticated, and the information is transferred from one security gateway to the other security gateway using the protected IKEv2 channel.

On the other hand, ESP does not provides any protection to the IPv4 header and as such to fragmentation procedure nor related pieces of information defined in Similarly, ICMPv4 PTB messages are not protected either. As a result, the security considerations related to MTU discovery [RFC0791], [RFC8900]. In our case, this includes information such as the DF bit and MF bit of the Flags field as well as the Total Length field from which the MTU is inferred. This is not surprising as fragmentation in the case of IPv4 MAY be performed by any node.[RFC0791], [RFC8900], [RFC4963], [RFC6864], [RFC1191] apply here.

8. Acknowledgements

The authors would like to thank Paul Wouters for his reviews and valuable comments and suggestions.

9. References

9.1. Normative References

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, DOI 10.17487/RFC0791, September 1981, <https://www.rfc-editor.org/info/rfc791>.

[RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, DOI 10.17487/RFC0792, September 1981, <https://www.rfc-editor.org/info/rfc792>.

[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>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <https://www.rfc-editor.org/info/rfc4443>.

[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>.

[RFC6864] Touch, J., "Updated Specification of the IPv4 ID Field", RFC 6864, DOI 10.17487/RFC6864, February 2013, <https://www.rfc-editor.org/info/rfc6864>.

[RFC7296] Kaufman, C., Hoffman, P., Nir, Y., Eronen, P., and T. Kivinen, "Internet Key Exchange Protocol Version 2 (IKEv2)", STD 79, RFC 7296, DOI 10.17487/RFC7296, October 2014, <https://www.rfc-editor.org/info/rfc7296>.

[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>.

[RFC8900] Bonica, R., Baker, F., Huston, G., Hinden, R., Troan, O., and F. Gont, "IP Fragmentation Considered Fragile", BCP 230, RFC 8900, DOI 10.17487/RFC8900, September 2020, <https://www.rfc-editor.org/info/rfc8900>.

9.2. Informative References

[I-D.ietf-intarea-tunnels]
Touch, J. and M. Townsley, "IP Tunnels in the Internet Architecture", Work in Progress, Internet-Draft, draft-ietf-intarea-tunnels-10, 12 September 2019, <https://www.ietf.org/archive/id/draft-ietf-intarea-tunnels-10.txt>.

[RFC4963] Heffner, J., Mathis, M., and B. Chandler, "IPv4 Reassembly Errors at High Data Rates", RFC 4963, DOI 10.17487/RFC4963, July 2007, <https://www.rfc-editor.org/info/rfc4963>.

Authors' Addresses

Daiying Liu (editor)
Ericsson

Email: harold.liu@ericsson.com

Daniel Migault
Ericsson

Email: daniel.migault@ericsson.com

Renwang Liu
Ericsson

Email: renwang.liu@ericsson.com

Congjie Zhang
Ericsson

Email: congjie.zhang@ericsson.com