Workgroup: Network Working Group
Internet-Draft: draft-templin-6man-fragrep-00
Published: 25 October 2021
Intended Status: Informational
Expires: 28 April 2022
Authors: F. L. Templin, Ed.
Boeing Research & Technology

IPv6 Fragment Retransmission

Abstract

Internet Protocol version 6 (IPv6) provides a fragmentation and reassembly service for end systems allowing for the transmission of packets that exceed the path MTU. However, loss of just a single fragment requires retransmission of the original packet in its entirety, with the potential for devastating effects on performance. This document specifies an IPv6 fragment retransmission scheme that matches the loss unit to the retransmission unit.

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 28 April 2022.

Table of Contents

1. Introduction

Internet Protocol version 6 (IPv6) [RFC8200] provides a fragmentation and reassembly service similar to that found in IPv4 [RFC0791], with the exception that only the source host (i.e., and not routers on the path) may perform fragmentation. When an IPv6 packet is fragmented, the loss unit (i.e., a single IPv6 fragment) becomes smaller than the retransmission unit (i.e., the entire packet) which under intermittent loss conditions could result in sustained retransmission storms with little or no forward progress.

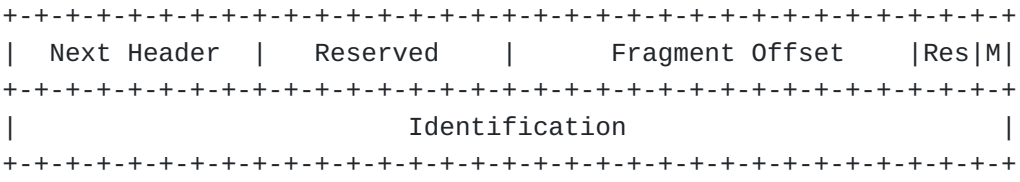
This document proposes IPv6 fragment retransmission service in which the source marks each fragment with an "Ordinal" number, and the destination may request retransmissions of any ordinal fragments not received. This retransmission request service is intended only for short-duration and opportunistic best-effort recovery (i.e., and not true end-to-end reliability). In this way, the service mirrors the Automatic Repeat Request (ARQ) function of common data links by considering an imaginary virtual link that extends from the IPv6 source to destination. The goal therefore is for the destination to quickly obtain missing individual fragments of partial reassemblies before true end-to-end timers would cause retransmission of the entire packet.

2. Terminology

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. IPv6 Fragmentation

IPv6 fragmentation is specified in Section 4.5 of [RFC8200] and is based on an IPv6 Fragment extension header formatted as shown below:



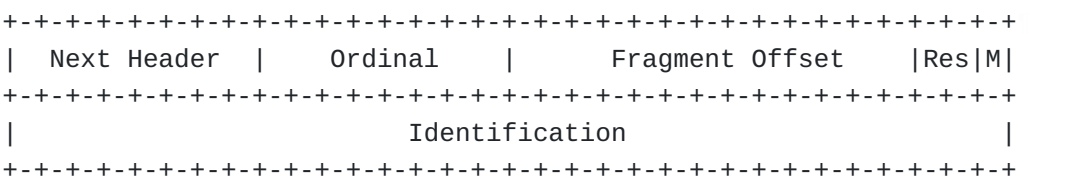
In this format:

- Next Header is a 1-octet IP protocol version of the next header following the Fragment Header.
- Reserved is a 1-octet reserved field set to 0 on transmission and ignored on reception.
- Fragment Offset is a 13-bit field that provides the offset (in 8-octet units) of the data portion that follows from the beginning of the packet.
- Res is a 2-bit field set to 0 on transmission and ignored on reception.
- M is the "more fragments" bit telling whether additional fragments follow.
- Identification is a 32 bit numerical identification value for the entire IPv6 packet. The value is copied into each fragment of the same IPv6 packet.

The fragmentation and reassembly specification in [RFC8200] can be considered as the default method which adheres to the details of that RFC. This document presents an enhanced method that allows for retransmissions of individual fragments.

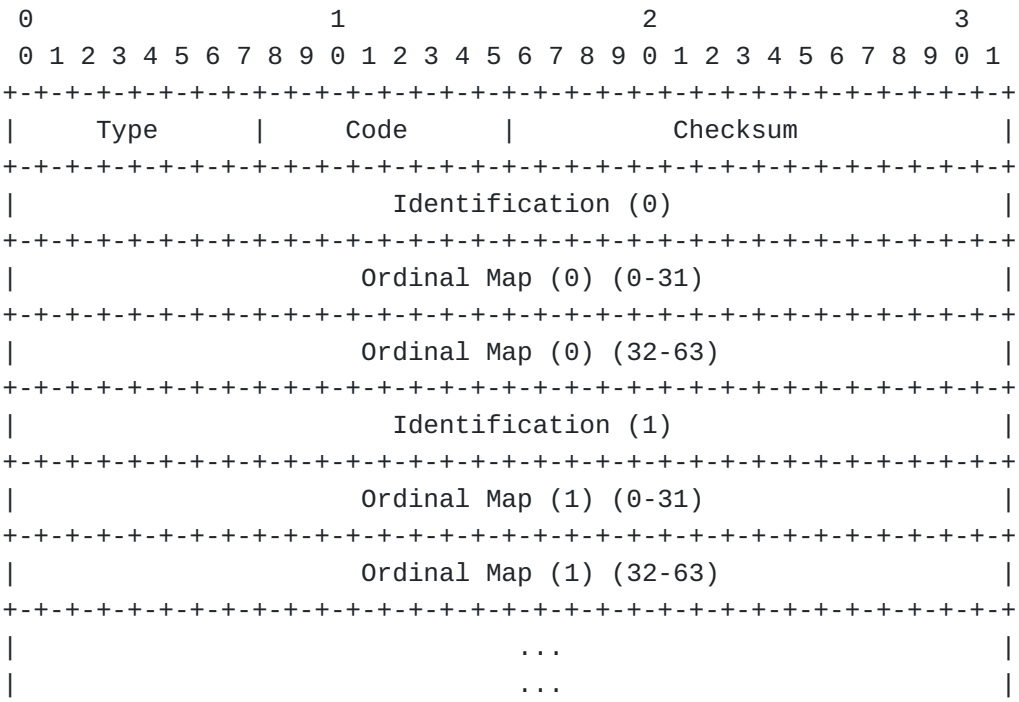
4. IPv6 Fragment Retransmission

Fragmentation implementations that obey this specification write an "Ordinal Number" beginning with 1 and monotonically incrementing for each successive fragment in the one-octet "Reserved" field of the IPv6 Fragment Header. The Reserved field is then renamed as "Ordinal" as shown below:



In particular, when a source that obeys this specification fragments an IPv6 packet it sets the Ordinal value for the first fragment to '1', the Ordinal value for the second fragment to '2', the Ordinal value for the third fragment to '3', etc. up to the total number of IPv6 fragments. When a destination that obeys this specification receives an IPv6 fragment with the Reserved/Ordinal field set to non-zero, it infers that the source participates in the protocol and maintains a checklist of all Ordinal numbered fragments received for a specific Identification number.

If the destination notices one or more Ordinal numbers missing after most other Ordinals for the same Identification have arrived, it can prepare a Fragmentation Report (Fragrep) ICMPv6 message [RFC4443] to send back to the source. The Fragrep message is formatted as follows:



In this format, the destination prepares the Fragrep message as a list of ordered-triples of 4-octet fields. The first field in each triple includes the Identification value for the IPv6 packet that is subject of the report, while the second and third fields include a 64-bit bitmap of the Ordinal values received for this Identification. For example, if the destination receives Ordinals 1, 2, 4, 5, 7, 9, then it sets bitmap bits 0, 1, 3, 4, 6 and 8 to '1'

and sets all other bits to '0'. The destination may include as many ordered triples as necessary without the entire Fragrep message exceeding the minimum IPv6 MTU of 1280 bytes.

After the destination has assembled the Fragrep message, it transmits the message to the IPv6 source. When the source receives the message, it examines each ordered triple to determine the (Identification, Ordinal) pairs that require retransmission. For example, if the source receives an Ordinal bitmap for Identification 0x12345678 with bits 0, 1, 3, 4, 6 and 8 set to '1', it would retransmit Ordinal fragments (0x12345678, 3), (0x12345678, 6) and (0x12345678, 8).

This implies that the source should maintain a cache of recently transmitted fragments for a time period known as the "link persistence interval". Then, if the source receives a Fragrep that requests retransmission of one or more Ordinals, it can retransmit if it still holds the Ordinal in its cache. Otherwise, the Ordinal will incur a cache miss and the original source will eventually retransmit the original packet in its entirety.

Note: The maximum-sized IPv6 packet that can undergo fragmentation is 64KB, and the minimum IPv6 path MTU is 1280B. Assuming the minimum IPv6 path MTU as the nominal size for non-final fragments, the number of Ordinals for each IPv6 packet should be significantly less than the allotted 64 bitmap bits.

5. Implementation Status

TBD.

6. IANA Considerations

A new ICMPv6 Message Type code for "Fragmentation Report (Fragrep)" is requested.

7. Security Considerations

Communications networking security is necessary to preserve confidentiality, integrity and availability.

8. Acknowledgements

This work was inspired by ongoing AERO/OMNI/DTN investigations.

.

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>.

[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>.

[RFC5326] Ramadas, M., Burleigh, S., and S. Farrell, "Licklider Transmission Protocol - Specification", RFC 5326, DOI 10.17487/RFC5326, September 2008, <https://www.rfc-editor.org/info/rfc5326>.

[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>.

[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>.

9.2. Informative References

[FRAG] Mogul, J. and C. Kent, "Fragmentation Considered Harmful, ACM Sigcomm 1987", August 1987.

[I-D.ietf-dtn-bpbis] Burleigh, S., Fall, K., and E. J. Birrane, "Bundle Protocol Version 7", Work in Progress, Internet-Draft, draft-ietf-dtn-bpbis-31, 25 January 2021, <https://www.ietf.org/archive/id/draft-ietf-dtn-bpbis-31.txt>.

[I-D.templin-6man-omni] Templin, F. L. and T. Whyman, "Transmission of IP Packets over Overlay Multilink Network (OMNI) Interfaces", Work in Progress, Internet-Draft, draft-templin-6man-omni-47, 8 September 2021, <https://www.ietf.org/archive/id/draft-templin-6man-omni-47.txt>.

[MPPS] Majkowski, M., "How to Receive a Million Packets Per Second, https://blog.cloudflare.com/how-to-receive-a-million-packets/", June 2015.

[QUIC] Ghedini, A., "Accelerating UDP Packet Transmission for QUIC, https://calendar.perfplanet.com/2019/accelerating-udp-packet-transmission-for-quic/", December 2019.

[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>.

[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>.

[RFC8899] Fairhurst, G., Jones, T., Tüxen, M., Rüngeler, I., and T. Völker, "Packetization Layer Path MTU Discovery for Datagram Transports", RFC 8899, DOI 10.17487/RFC8899, September 2020, <https://www.rfc-editor.org/info/rfc8899>.

[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>.

Author's Address

Fred L. Templin (editor)
Boeing Research & Technology
P.O. Box 3707
Seattle, WA 98124
United States of America

Email: fltemplin@acm.org