Network Working Group F. Templin, Ed.
Internet-Draft Boeing Research & Technology
Intended status: Informational February 27, 2015
Expires: August 31, 2015

IPv6 Path MTU Interactions With Link Adaptation
draft-templin-6man-linkadapt-02.txt

Abstract

IPv6 intentionally deprecates fragmentation by routers in the network. Instead, links with restricting Maximum Transmission Units (MTUs) must either drop each too-large packet and return an ICMPv6 Packet Too Big (PTB) message or perform link-specific fragmentation and reassembly (also known as "link adaptation") at a layer below IPv6. This latter category of links is often performance-challenged to accommodate steady-state link adaptation. This document therefore proposes an update to the base IPv6 specification to better accommodate links that require link-specific adaptation.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

IPv6 intentionally deprecates fragmentation by routers in the network. Instead, links with restricting Maximum Transmission Units (MTUs) must either drop each too-large packet and return an ICMPv6 Packet Too Big (PTB) message or perform link-specific fragmentation and reassembly (also known as "link adaptation") at a layer below IPv6. This latter category of links is often performance-challenged to accommodate steady-state link adaptation. This document therefore proposes an update to the base IPv6 specification to better accommodate links that require link-specific adaptation.

2. Problem Statement

The current "Internet cell size" is effectively 1500 bytes, i.e., the minimum MTU configured by the vast majority of links in the Internet. IPv6 constrains this even further by specifying a minimum link MTU of 1280 bytes [RFC2460]. However, due to operational issues with Path MTU Discovery (PMTUD) [RFC1981] these sizes can often only be accommodated when links with smaller link-layer segment sizes are configured to perform link adaptation.

Unfortunately, link adaptation can present a significant burden to the link endpoints, i.e., especially when the link supports high data rates and/or is located nearer the "middle" of the network instead of nearer the "edge". An alternative therefore is to ask the originating IPv6 node to either reduce the size of the packets it sends or perform host-based fragmentation, in which case reassembly would be performed by the final destination.

In addition to the above considerations, it is becoming more and more evident that PMTUD uncertainties can be encountered even when there are no links in the path that must perform link adaptation. This is due to the fact that the PTB messages required for PMTUD can be lost due to network filters that block ICMPv6 messages [RFC2923][WAND][SIGCOMM]. Originating IPv6 node are therefore advised to take precautions to avoid path MTU related failure modes.

This document updates the IPv6 protocol specification [RFC2460] to better accommodate paths with various MTUs as described in the following sections.

3. Link Adaptation Signaling and Accommodation

Section 5 of [RFC2460] states:

"IPv6 requires that every link in the Internet have an MTU of 1280 octets or greater. On any link that cannot convey a 1280-octet packet in one piece, link-specific fragmentation and reassembly must be provided at a layer below IPv6."

and:

"A node must be able to accept a fragmented packet that, after reassembly, is as large as 1500 octets.".

This document does not propose to change these requirements, but notes that link adaptation can be burdensome for some links to the point that it would be highly desirable to signal the MTU limitation to the IPv6 communication endpoints. In order to accommodate this, when the router at the link ingress performs link adaptation on a packet it should also send an ICMPv6 PTB message back to the original source (subject to rate limiting) with a Next-Hop MTU set to the link adaptation threshold and with Code field set to 1 [RFC4443]. (Note that these PTB messages are advisory in nature and do not necessarily indicate packet loss.)

As a result, the originating IPv6 node may receive this "new kind" of PTB message and should modify its behavior accordingly. This is accomplished by adding a new final paragraph to Section 5 of [RFC2460] as follows:

"In response to an IPv6 packet that is sent to a destination located beyond an IPv6 link that must perform link adaptation, the originating IPv6 node may receive an ICMP Packet Too Big message with Code=1. In that case, the IPv6 node can either reduce the size of subsequent packet it sends or perform IPv6 fragmentation on packets no larger than 1500 bytes by breaking the packet into N roughly equal-length pieces (where N is minimized and the length of each piece is smaller than the Next-Hop MTU). These fragments will be reassembled by the destination."

3.1. Accommodating Legacy Nodes

Legacy IPv6 nodes observe the current final paragraph of Section 5 of [RFC2460]:

"In response to an IPv6 packet that is sent to an IPv4 destination (i.e., a packet that undergoes translation from IPv6 to IPv4), the originating IPv6 node may receive an ICMP Packet Too Big message reporting a Next-Hop MTU less than 1280. In that case, the IPv6 node is not required to reduce the size of subsequent packets to less than 1280, but must include a Fragment header in those packets so that the IPv6-to-IPv4 translating router can obtain a suitable Identification value to use in resulting IPv4 fragments. Note that this means the payload may have to be reduced to 1232 octets (1280 minus 40 for the IPv6 header and 8 for the Fragment header), and smaller still if additional extension headers are used."

For such legacy nodes, the receipt of a PTB message with a Next-Hop MTU less than 1280 will result in the above behavior regardless of the value in the Code field. As a result, a link ingress node that returns this new kind of PTB message may receive future packets containing a Fragment header with the More Fragments (MF) bit and Offset field set to 0. The link ingress node should process these packets as an indication that the originating IPv6 node is a legacy node, and should not send further PTB messages. Instead, the link ingress node should use the fragment header supplied by the source to fragment the original packet to a size that would avoid link adaptation. These fragments are then reassembled by the final destination.

4. IANA Considerations

There are no IANA considerations for this document.

5. Security Considerations

The security considerations for [RFC2460] apply also to this document.

6. Acknowledgments

This method was inspired through discussion on the IETF v6ops and NANOG mailing lists in the May through July 2012 timeframe. Further

discussion occurred on the Intarea list in the February 2015 timeframe.

7. References

7.1. Normative References

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.

7.2. Informative References

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

[RFC2923] Lahey, K., "TCP Problems with Path MTU Discovery", RFC 2923, September 2000.

[RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, March 2007.

[SIGCOMM] Luckie, M. and B. Stasiewicz, "Measuring Path MTU Discovery Behavior", November 2010.

[WAND] Luckie, M., Cho, K., and B. Owens, "Inferring and Debugging Path MTU Discovery Failures", October 2005.

Author's Address

Fred L. Templin (editor)
Boeing Research & Technology
P.O. Box 3707
Seattle, WA 98124
USA

Email: fltemplin@acm.org