Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 14, 2013

F. Templin, Ed.
Boeing Research & Technology

July 13, 2012

IPv6 Path MTU Updates
draft-generic-6man-tunfrag-05.txt

Abstract

IPv6 intentionally deprecates fragmentation by routers in the network. Instead, links with restricting 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. A common case that exhibits these link characteristics is seen for IPv6-within-IP tunnels. Additionally, IPv6 nodes can avoid path MTU discovery issues even when no link adaptation is necessary by performing a small amount of fragmentation and/or by probing the path as necessary. This document therefore proposes an update to the base IPv6 specification to better accommodate path MTU issues.

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 http://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 January 14, 2013.

Table of Contents

1. Introduction

IPv6 intentionally deprecates fragmentation by routers in the network. Instead, links with restricting 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. A common case that exhibits these link characteristics is seen for IPv6-within-IP tunnels [I-D.generic-v6ops-tunmtu]. Additionally, IPv6 nodes can avoid path MTU discovery issues even when no link adapation is necessary by performing a small amount of fragmentation and/or by probing the path as necessary. This document therefore proposes an update to the base IPv6 specification to better accommodate path MTU issues.

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 permitted to perform link adaptation. A common example of such links is seen for IPv6-within-IP tunnels.

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 perform fragmentation for the packets it sends, 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 tunnels nor other 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. Considerations for Small MTU Paths

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."

This document does not propose to change this requirement, but notes that link adaptation can be burdensome for some links (e.g., IPv6-within-IP tunnels) to the point that it would be highly desirable to push the fragmentation and reassembly responsibility 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 less than 1280 and with a 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 modifying the final paragraph of Section 5 of [RFC2460] as follows:

"In response to an IPv6 packet that is sent to a destination located beyond an IPv6-within-IP tunnel or an IPv6 link that must perform link adaptation, the originating IPv6 node may receive an ICMP Packet Too Big message reporting a Next-Hop MTU less than 1280 and with Code=1. In that case, the IPv6 node is not required to reduce the size of subsequent packets to less than 1500, but must perform IPv6 fragmentation on those packets using the Next-Hop MTU as the maximum fragment size. These fragments will be reassembled by the destination."

An example tunnel protocol that invokes this behavior appears in: [I-D.templin-intarea-seal].

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 those 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, an IPv6 link that returns this new kind of PTB message may receive future packets up to 1280 bytes in length and containing a Fragment header with MF=0 and Offest=0. The link should process these packets as an indication that the originating IPv6 node is a legacy node, and should not send further PTB messages.

4. Considerations for Medium MTU Paths

Regardless of whether there is a link in the path that performs link adaptation, when an originating IPv6 node receives a PTB message reporting a Next-Hop MTU value between 1280 and 1500 bytes, the node need not reduce the size of the packets it sends but may instead invoke fragmentation for packets up to 1500 bytes using a maximum fragment size of 1280 bytes. These fragments will again be reassembled by the final destination.

A more interesting situation arises when PTB messages are lost on the return path to the originating IPv6 node. Since the node has no way of discerning which paths may exhibit this condition, it may be better served to assume the worst case for all paths and take precautionary measures to avoid silent packet loss.

In one approach, an originating IPv6 node that wishes to ensure that packets between 1281 and 1500 bytes in length will reach the destination can use "proactive fragmentation" to fragment the packets into two pieces that are no larger than 1280 bytes. In a second approach, the node can use Packetization Layer Path MTU Discovery (PLPMTUD) [RFC4821] without fragmentation to verify that packets larger than 1280 bytes are reaching the final destination.

5. Considerations for Large MTU Paths

An originating IPv6 node connected to a link that supports an MTU of 1500 bytes or larger is permitted to send packets larger than 1500 bytes without fragmentation, but should implement [RFC4821] to verify that these larger packets are reaching the final destination.

6. IANA Considerations

There are no IANA considerations for this document.

7. Security Considerations

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

8. Acknowledgments

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

9. References

9.1. 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.

9.2. Informative References

[I-D.generic-v6ops-tunmtu] Templin, F., "Operational Issues with Tunnel Maximum Transmission Unit (MTU)", draft-generic-v6ops-tunmtu-09 (work in progress), July 2012.

[I-D.templin-intarea-seal] Templin, F., "The Subnetwork Encapsulation and Adaptation Layer (SEAL)", draft-templin-intarea-seal-42 (work in progress), December 2011.

[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