Network Working Group F. Templin, Ed.
Internet-Draft Boeing Research & Technology
Intended status: Informational February 14, 2014
Expires: August 18, 2014

6rd Tunnel MTU
draft-foo-v6ops-6rdmtu-03.txt

Abstract

The tunnel MTU on 6rd Provider Edge (PE) and Consumer Edge (CE) routers is currently recommended to be set to 1480. This is to avoid IPv4 fragmentation within the tunnel, but requires the tunnel ingress to drop any IPv6 packet larger than 1480 bytes and return an ICMPv6 Packet Too Big (PTB) message. Concerns for operational issues with both IPv4 and IPv6 Path MTU Discovery point to the possibility of MTU-related black holes when a packet is dropped due to an MTU restriction somewhere in the Internet. Fortunately, the "Internet cell size" is 1500 bytes (i.e., the minimum MTU configured by the vast majority of links in the Internet) so if the 6rd PE router can set a tunnel MTU of at least 1500 bytes the MTU issues are alleviated. This document specifies methods that can be employed to support these larger sizes.

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 August 18, 2014.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal

Provisions Relating to IETF Documents
(http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction

The tunnel MTU on 6rd Provider Edge (PE) and Consumer Edge (CE) routers is currently recommended to be set to 1500 bytes minus the IPv4 header encapsulation overhead minus the encapsulation overhead for any additional encapsulations that may occur on the path [RFC5969]. This is to avoid IPv4 fragmentation within the tunnel [RFC0791], but requires the tunnel ingress to drop any IPv6 packet larger than the tunnel MTU and return an ICMPv6 Packet Too Big (PTB) message [RFC2460]. Concerns for operational issues with both IPv4 and IPv6 Path MTU Discovery [RFC1191][RFC1981] point to the possibility of MTU-related black holes when a packet is dropped due to an MTU restriction somewhere in the Internet. Fortunately, the "Internet cell size" is 1500 bytes (i.e., the minimum MTU configured by the vast majority of links in the Internet) so if the 6rd PE router can set a tunnel MTU of at least 1500 bytes the MTU issues are alleviated. This document specifies methods that can be employed to support these larger sizes.

Pushing the 6rd tunnel MTU to 1500 bytes or larger is met with the challenge that the addition of the IPv4 encapsulation header would cause a 1500 byte IPv6 packet to appear as a 1520 byte IPv4 packet on the wire. This can result in the packet being either fragmented or dropped by an IPv4 router that configures a smaller link MTU, depending on the setting of the "Don't Fragment" (DF) bit in the IPv4 header. Therefore, this document recommends complementary mechanisms to ensure that packets of various sizes can be delivered as long as the underlying IPv4 network can support the larger sizes. The following two sections present the methods used by 6rd PE and CE routers.

2. 6rd Provider Edge (PE) Router MTU Mitigations

The 6rd PE Router employs the following MTU-handling mitigations:

1. Set the 6rd tunnel interface MTU to the maximum of 1500 and the MTU of the underlying IPv4 interface minus the expected encapsulation overhead for the IPv4 header as well as any other encapsulations that may occur on the path.
2. For each 6rd CE, maintain a RATE-LIMIT boolean variable set to TRUE.
3. When the PE sends an IPv6 packet no larger than 1500 bytes minus encapsulation overhead to a CE, encapsulate and set the DF bit to 1.
4. When the PE sends an IPv6 packet larger than 1500 bytes to a CE, encapsulate and set the DF bit to 1. Optionally cache any IPv4 MTU values returned in ICMPv4 packet too big messages that may result.
5. When the PE sends an IPv6 packet larger than 1500 bytes minus the encapsulation overhead but no larger than 1500 bytes, encapsulate and set the DF bit to 0. Send the packet to the CE subject to rate limiting if RATE-LIMIT is TRUE. The packet may be fragmented in the IPv4 network on the path to the CE.
6. Send a 1500 byte IPv6 probe packet to each active CE subject to rate limiting using the neighbor reachability test procedure specified in Section 8 of RFC5969. If the probe succeeds, set RATE-LIMIT for the CE to FALSE.

3. 6rd Provider Edge (CE) Router MTU Mitigations

The 6rd CE Router employs the following MTU-handling techhniques:

1. Set the 6rd tunnel interface MTU to the maximum of 1280 and the the MTU of the underlying IPv4 interface minus the expected encapsulation overhead for the IPv4 header as well as any other encapsulations that may occur on the path.
2. If the underlying interface has a sufficiently-large MTU, send a 1500 byte IPv6 probe packet to the PE using the neighbor reachability test procedure specified in Section 8 of RFC5969. If the probe succeeds, set the IPv4 MTU for the PE to the MTU of the underlying IPv4 interface; else, set the IPv4 MTU to 1520 minus the expected encapsulation overhead.
3. For each TCP session initiated by an IPv6 host within the CE's LAN, rewrite the Maximum Segment Size (MSS) to 1500 minus the TCP header length minus the IPv6 header length minus the encapsulation overhead for (see: [RFC0879][RFC6691]). As a result, the local IPv6 host and its remote IPv6 correspondent will begin their TCP messages exchanges using IPv6 packets no larger than the minimum tunnel path MTU.
4. When the CE sends an IPv6 packet to the PE, if the encapsulated packet is larger than the IPv4 MTU for the PE drop and return an ICMPv6 Packet Too Big. Else, set the DF bit to 1 and send the packet.
5. For each neighboring CE, maintain a RATE-LIMIT boolean variable set to TRUE.
6. When the CE sends an IPv6 packet no larger than 1500 bytes minus the encapsulation overhead to a neighboring CE, encapsulate and set the DF bit to 1.
7. When the CE sends an IPv6 packet larger than 1500 bytes to a neighboring CE, encapsulate and set the DF bit to 1. Optionally cache any IPv4 MTU values returned in ICMPv4 packet too big messages that may result.
8. When the CE sends an IPv6 packet larger than 1500 bytes minus the encapsulation overhead but no larger than 1500 bytes to a neighboring CE, encapsulate and set the DF bit to 0. Send the packet to the neighboring CE subject to rate limiting if RATE-LIMIT is TRUE.
9. Send a 1500 byte IPv6 probe packet to each active neighboring CE subject to rate limiting using the neighbor reachability test procedure specified in Section 8 of RFC5969. If the probe succeeds, set RATE-LIMIT for the CE to FALSE.

4. Discussion

There are several interrelated aspects to the recommended MTU mitigations. First, the unconditional rewriting of the MSS by CE routers ensures that the initial packets sent by IPv6 correspondents will be no larger than the minimum tunnel path MTU following encapsulation. The IPv6 correspondents can thereafter use

Packetization Layer Path MTU Discovery (PLPMTUD) [RFC4821] to attempt to increase the MSS during the course of the TCP session and thereby take advantage of larger packet sizes when available.

However, not all transport protocols observe the TCP MSS and so the packets of other protocols generated by IPv6 hosts may be larger than would fit in the minimum tunnel path MTU. Since most IPv6 hosts expect to see a minimum MTU of 1500 bytes without any ancillary MTU assurance mitigations, the approach specified here takes special care of packets larger than the minimum tunnel path MTU but no larger than 1500 bytes. Namely, these packets are allowed to undergo IPv4 fragmentation on the path from the PE to a CE or on the path from a CE to another CE. Since sustained fragmentation at high data rates is dangerous however [RFC4963][RFC6864] packets in this size range must only be admitted into the tunnel subject to rate limiting so that reassembly misassociations do not occur. Meanwhile, packets larger than 1500 bytes are admitted into the tunnel unconditionally on a "best effort" basis with the understanding that these packets may be dropped silently.

Using these methods, CE routers may need to perform a small amount of IPv4 reassembly. PE routers on the other hand will never be asked to perform reassembly.

5. IANA Considerations

There are no IANA considerations for this document.

6. Security Considerations

The security considerations for 6rd apply also to this document.

7. Acknowledgments

This method was inspired through many years of discussion on IETF lists and other forums on the topic of tunnel MTU.

8. References

8.1. Normative References

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

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

[RFC5969] Townsley, W. and O. Troan, "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd) -- Protocol Specification", RFC 5969, August 2010.

8.2. Informative References

[RFC0879] Postel, J., "TCP maximum segment size and related topics", RFC 879, November 1983.

[RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, November 1990.

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

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

[RFC4963] Heffner, J., Mathis, M., and B. Chandler, "IPv4 Reassembly Errors at High Data Rates", RFC 4963, July 2007.

[RFC6691] Borman, D., "TCP Options and Maximum Segment Size (MSS)", RFC 6691, July 2012.

[RFC6864] Touch, J., "Updated Specification of the IPv4 ID Field", RFC 6864, February 2013.

Author's Address

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

Email: fltemplin@acm.org