Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 26, 2007

F. Templin, Ed.
Boeing Phantom Works

October 23, 2006

Requirements for IP-in-IP Tunnel MTU Assurance
draft-templin-mtuassurance-02.txt

Status of this Memo

Copyright Notice

Abstract

IP-in-IP tunnels present a Maximum Transmission Unit (MTU) to layer 3 via static prearrangements and/or dynamic MTU determination based on layer 2 ICMP messages, but these methods have known operational limitations that can fail to enforce an assured MTU resulting in degraded performance and communications failures. A method for providing an assured MTU to layer 3 over IP-in-IP tunnels is needed.

Table of Contents

1. Introduction

IP-in-IP tunnels span multiple layer 2 network hops yet are seen by layer 3 as ordinary links that must support an assured MTU, e.g., 1280 bytes for the IPv6 minimum MTU. Common tunneling mechanisms (e.g., [RFC2529][RFC3056][RFC3931][RFC4213][RFC4214][RFC4380]) meet this requirement through conservative static prearrangements at the expense of degraded performance and/or communications failures over some paths due to excessive layer 2 network-based fragmentation. Optional dynamic MTU determination methods based on layer 2 ICMP "packet too big" messages are also available, but can result in communication failures due to the unreliable and untrustworthy nature of layer 2 ICMP messages generated by network middleboxes. This document discusses operational issues with the MTU determination schemes used by common tunneling mechanisms and outlines requirements for a new method that can present an assured MTU to layer 3.

2. Problems with Network-Based Fragmentation

Common IP-in-IP tunneling mechanism encapsulators set a static layer 3 tunnel MTU (e.g., 1280 bytes or slightly larger for IPv6) and do not set the DF bit in the layer 2 IP headers of tunneled packets such that packets that are too large to traverse the path before reaching the decapsulator will be fragmented by the network. Unfortunately, network-based IP fragmentation has well-known issues [FRAG][RFC4459][I-D.heffner-frag-harmful] that can result in degraded performance and/or communications failures along some paths. In particular, a) firewalls and NAT boxes typically discard fragments other than the first fragment of fragmented IP datagrams, and b) self-sustaining cyclical reassembly mis-associations due to fragment loss can occur resulting in communications failures.

3. Problems with Path MTU Discovery

IP-in-IP tunneling mechanisms can use Path MTU Discovery by setting the DF bit in the layer 2 IP headers of tunneled packets, but this method relies on layer 2 ICMP "packet too big" messages coming from untrusted network middleboxes along the path. A well-known issue is that ICMP messages are often dropped by firewalls and/or NATs resulting in MTU-related black holes along some paths [RFC2923]. Additionally, the untrusted middlebox paradigm opens the possibility for various spoofing attacks via fabricated ICMP messages inserted by on-path or off-path adversaries. [I-D.ietf-pmtud-method] and [I-D.gont-tcpm-icmp-attacks] discuss possible mitigations for dealing with fabricated ICMP messages, but no mitigations are possible when legitimate middleboxes fail to send/forward the ICMP's.

4. Requirements for IP-in-IP Tunnel MTU Assurance

Due to the operational issues with both layer 2 network-based IP fragmentation and ICMP-based Path MTU discovery, a new mechanism is needed to assure efficient and robust use of the available MTU over IP-in-IP tunnels. In particular, a mechanism is needed to present an assured MTU to layer 3 such that packets no larger than the MTU will be accepted by the tunnel or a suitable layer 3 "packet too big" message will be returned.

The following subsections present requirements for IP-in-IP tunnel MTU assurance:

4.1. Tunnel Endpoint Negotiation

The MTU assurance scheme must provide a means for the encapsulating and decapsulating tunnel endpoints to determine that the scheme is implemented at both ends. When only one (or neither) of the tunnel endpoints implements the scheme, behavior must revert back to that specified by the current tunneling mechanisms.

4.2. Compatible with IP Mechanisms

The MTU assurance scheme must be compatible with both layer 2 network-based IP fragmentation/reassembly and layer 2 ICMP "packet too big" messages from Path MTU Discovery that may occur from within the tunnel. In particular, any packets prepared by the MTU assurance scheme must not be disrupted by any layer 2 network-based IP fragmentation that occurs along the path. An encapsulating node that implements the MTU assurance scheme must also be prepared to deal with any layer 2 ICMP "packet too big" messages it may receive in response to tunneled packets, e.g. as outlined in [I-D.ietf-pmtud-method][I-D.gont-tcpm-icmp-attacks].

4.3. Host-based Segmentation at the Encapsulator

The MTU assurance scheme must provide a means for the encapsulating tunnel endpoint to split layer 3 payloads into segments that are no larger than the tunnel path MTU. The segmentation must occur below layer 3 and prior to layer 2 IP encapsulation.

4.4. Reassembly at the Decapsulator

The MTU assurance scheme must provide a means for the decapsulating tunnel endpoint to reassemble layer 3 payloads that were conveyed in multiple segments from the encapsulator. The reassembly must occur after layer 2 IP reassembly (and prior to layer 3 delivery), since it is possible that the segments may have also incurred fragmentation

along the path.

4.5. Means for Detecting Packet Splicing Errors

The MTU assurance scheme must provide a means for the decapsulating tunnel endpoint to detect packet splicing errors as it reassembles the segments of layer 3 payloads.

4.6. Means for Accommodating Out-of-Order Delivery

The MTU assurance scheme must provide a means for the decapsulating tunnel endpoint to accommodate out-of-order delivery for the segments it receives while reassembling the segments of layer 3 payloads.

4.7. Path Probing by the Encapsulator

The MTU assurance scheme must provide a means for the encapsulator to send "probe" segments used to determine whether segments of a certain size can traverse the tunnel. The scheme should allow for in-of-band path probing (i.e., when the probe segment is a segment of an actual tunneled packet) and must allow for out-of-band path probing.

4.8. Authenticated Probe Response from the Decapsulator

The MTU assurance scheme must provide a means for the decapsulator to send an authenticated probe response message back to the encapsulator to acknowledge the receipt of a probe segment.

4.9. Proactive Path Probing

The MTU assurance scheme should perform proactive path probing to quickly determine the most efficient segment size to use for a particular tunnel. The scheme should also periodically re-probe the path to determine whether path MTU reductions, e.g, due to route fluctuations, have occurred.

4.10. Decapsulator MRU Discovery

The MTU assurance scheme must provide a means for an encapsulator to discover the maximum receive unit (MRU) for each decapsulator.

5. IANA Considerations

This document does not introduce any IANA considerations.

6. Security Considerations

This document does not introduce any security considerations.

7. Acknowledgments

This document represents the mindshare of many contributors.

8. Informative References

[FRAG] Mogul, J. and C. Kent, "Fragmentation Considered Harmful, In Proc. SIGCOMM '87 Workshop on Frontiers in Computer Communications Technology.", August 1987.

[I-D.gont-tcpm-icmp-attacks] Gont, F., "ICMP attacks against TCP", draft-gont-tcpm-icmp-attacks-05 (work in progress), October 2005.

[I-D.heffner-frag-harmful] Heffner, J., "Fragmentation Considered Very Harmful", draft-heffner-frag-harmful-02 (work in progress), June 2006.

[I-D.ietf-pmtud-method] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", draft-ietf-pmtud-method-10 (work in progress), September 2006.

[RFC2529] Carpenter, B. and C. Jung, "Transmission of IPv6 over IPv4 Domains without Explicit Tunnels", RFC 2529, March 1999.

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

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[RFC3931] Lau, J., Townsley, M., and I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, March 2005.

[RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.

[RFC4214] Templin, F., Gleeson, T., Talwar, M., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol

(ISATAP)", RFC 4214, October 2005.

[RFC4380] Huitema, C., "Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs)", RFC 4380, February 2006.

[RFC4459] Savola, P., "MTU and Fragmentation Issues with In-the-Network Tunneling", RFC 4459, April 2006.

Author's Address

Fred L. Templin (editor)
Boeing Phantom Works
P.O. Box 3707
Seattle, WA 98124
USA

Email: fred.l.templin@boeing.com

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).