BEHAVE
Internet-Draft
Expires: August 16, 2005

N. Modadugu
Stanford University
February 15, 2005

NAT Behavioral Requirements for TCP
draft-modadugu-nat-tcp-00.txt

Status of this Memo

Copyright Notice

Abstract

This document specifies requirements for NAT devices when handling TCP traffic. In order to arrive at the requirements, basic terminology regarding NAT TCP handling is defined. The purpose of this document is to provide a specification of TCP handling by NAT devices so that TCP-using applications can work consistently.

Table of Contents

1. Scope 3
2. Introduction 3
3. Terminology 3
4. Network Address and Port Translation Behavior 3
5. Connection State and Timers 4
5.1 Timers 4
5.2 Sequence Number Adjustment 4
5.3 Connection Reset 5
6. Filtering Behavior 5
7. Requirements 5
7.1 Requirements Discussion 7
8. Security Considerations 7
9. IANA Considerations 7
10. IAB Considerations 7
11. Acknowledgments 8
12. References 8
12.1 Normative References 8
12.2 Informative References 8
Author's Address 9
Intellectual Property and Copyright Statements 10

1. Scope

This document is a counterpart to ``NAT Behavioral Requirements for Unicast UDP'' [2]. This document defines terminology related to NAT handling of TCP traffic and specifies requirements for NAT devices and implementations.

NAT behavior when handling TCP is also dependent on some transport protocol independent behavior (for instance, ICMP, and packet filtering behavior). Only (TCP) protocol dependent behavior is described in this document, and protocol independent behavior is referenced as necessary.

2. Introduction

Current NAT devices exhibit sufficiently differing behavior that some classes of network-using applications function either unpredictably or unreliably. Indeed, in some cases applications are simply incompatible with a network that employs NAT.

The most serious problem faced today by TCP-based applications that are deployed behind NAT devices is their inability to communicate directly with peers that are also behind NAT devices. A common, and unsatisfying, workaround to this problem is the use of a globally accessible proxy for data relay.

Along with specifying requirements for NAT handling of TCP traffic, this document also elucidates how meeting these requirements allows for direct peer-to-peer communication.

3. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [1].

Much of the NAT related terminology used in the document is identical to that used in [2].

4. Network Address and Port Translation Behavior

When handling TCP traffic, address and port translation behavior is the same as when handling UDP traffic. The reader is, hence, referred to [2] for a description of this type of behavior.

It should be noted, however, that Port Parity and Port Contiguity behaviors are not applicable when handling TCP traffic.

5. Connection State and Timers

When handling TCP traffic, it is necessary for a NAT device to keep track of the current TCP state for a connection in order to manage the corresponding binding.

Initiation of a TCP connection is marked by a TCP packet with the SYN flag set. Similarly, intent to close a connection is marked by a TCP packet with the FIN flag set.

A TCP connection that has yet to complete the ``three-way'' handshake is in the "Connecting" state, one that has completed the handshake is "Established". If either of the peers has initiated connection tear-down by sending a packet with the FIN flag set, then the connection is in the "Closing" state.

5.1 Timers

Connecting Timer: This timer starts when the initial SYN packet arrives on an internal interface of the NAT device. If the Connecting Timer expires before the connection state transitions to the Established state, then the binding for the connection is deleted.

Established Timer: Often connections will not been cleanly shutdown, perhaps due to system crash, and as a result leave bindings on any NAT devices that were used for data communication. The purpose of the Established timer is to detect and purge such defunct connections. An unfortunate side-effect of purging idle connections is that some genuinely idle connections get terminated (TCP connections may, but are not required to use the keep-alive option).

Closing Timer: This timer is started once a connection transitions into the Closing state. In case a TCP implementation, as an optimization, does not correctly implement connection closing, then when the close timer expires, the binding corresponding to the connection is marked closed.

5.2 Sequence Number Adjustment

A number of NAT devices support Application Level Gateways (ALGs) that can modify application layer data. As a result of modification, packets may contain fewer, or more bytes than were contained in the original packet.

NAT devices that support such ALGs also keep track of TCP sequence numbers, and correct future packets as they traverse the NAT. NAT devices that support Sequence number adjustment are "Support Sequence Number Adjustment", and those incapable of making these adjustments are "Do not Support Sequence Number Adjustment".

5.3 Connection Reset

TCP supports connection abortion through the Reset flag: upon receiving a TCP packet with the RST (Reset) flag set, the client immediately tears down the connection and does not send or receive further data on the connection.

A NAT device that recognizes packets that have the RST flag set are "Reset aware", otherwise they are "Reset unaware".

6. Filtering Behavior

Filtering behavior for TCP traffic is same as the filtering behavior for UDP traffic. Briefly: incoming packets are only allowed to pass through if a binding is present for the address and port pair(s) presented by the packet. Further filtering behavior is dependent on whether the NAT device is endpoint address or port independent.

For our discussion, it is worthwhile noting that some NAT devices do not allow incoming TCP packets with the SYN flag set, regardless of the presence of a binding for the connection.

If there is no binding for an incoming SYN packet, then some NAT devices respond with a RST (TCP Reset) packet, and some silently discard the SYN. A NAT device that responds with a TCP Reset packet exhibits "Reset on No Binding" behavior. Similarly, a NAT device that remains silent exhibits "Silence on No Binding".

If a NAT allows to pass through, after applying all other filtering rules, a SYN packet, then it is "Allow Incoming SYN". Otherwise it is "Disallow Incoming SYN".

7. Requirements

Many of the requirements that apply to TCP traffic also apply to UDP traffic, and hence a significant portion of this section is lifted from [2].

In order to avoid ambiguity and confusion, the complete set of requirements for NAT TCP behavior are listed in this section. However, only the discussion of additional, or differing requirements is provided in this document.

REQ-1 A NAT MUST have an "External NAT mapping is endpoint independent" behavior.

REQ-2 It is RECOMMENDED that a NAT have an "IP address pooling" behavior of "Paired". Note that this requirement is not applicable to NATs that do not support IP address pooling.

REQ-3 It is RECOMMENDED that a NAT have a "Port assignment" behavior of "Port preservation".

a) A NAT MUST NOT have a "Port assignment" behavior of "Port overloading".

b) If the host's source port was in the range 1-1023, it is RECOMMENDED the NAT's source port also be in the same range. If the host's source port was in the range 1024-65535, it is RECOMMENDED that the NAT's source port also be in that range.

REQ-4 The NAT mapping Refresh Direction MUST have a "NAT Outbound refresh behavior" of "True".

a) The NAT mapping Refresh Direction MAY have a "NAT Inbound refresh behavior" of "True".

b) The NAT mapping Refresh Direction MUST have a "NAT refresh method behavior" of "Per mapping" (i.e. refresh all sessions active on a particular mapping).

REQ-5 It is RECOMMENDED that a NAT have an "External filtering is endpoint address dependent" behavior.

REQ-6 It is RECOMMENDED that a NAT have "Silence on No Binding" behavior.

REQ-7 It is RECOMMENDED that a NAT have an "Allow Incoming SYN" behavior.

REQ-8 A NAT MUST support "Hairpinning".

a) A NAT Hairpinning behavior MUST be "External source IP address and port".

REQ-9 It is RECOMMENDED that a NAT is "Supports Sequence Number Adjustment".

a) If a NAT supports ALGs that are capable of inserting or removing bytes from TCP packets, then the NAT MUST have "Supports Sequence Number Adjustment" behavior.

b) If a NAT that has "Supports Sequence Number Adjustment" behavior, it is RECOMMENDED the NAT also have "RST Aware" behavior.

REQ-10 If a NAT includes ALGs, it is RECOMMENDED that all of those ALGs be disabled by default.

a) If a NAT includes ALGs, it is RECOMMENDED that the NAT allow the user to enable or disable each ALG separately.

REQ-11 A NAT MUST have deterministic behavior, i.e., it MUST NOT change the NAT mapping or the External External Filtering Behavior at any point in time or under any particular conditions.

REQ-12 It is RECOMMENDED that a NAT support ICMP Destination Unreachable.

a) The ICMP timeout SHOULD be greater than 2 seconds.

REQ-13 A NAT MUST support fragmentation of packets larger than link MTU.

REQ-14 A NAT MUST support receiving in order fragments, so it MUST be "Received Fragment Ordered" or "Received Fragment Out of Order".

a) A NAT MAY support receiving fragmented packets that are out of order and be of type "Received Fragment Out of Order".

7.1 Requirements Discussion

This section only discusses the requirements that are additional, or differ from the requirements listed for NAT handling of UDP traffic in [2].

REQ-3 The recommendation that "Port Assignment" be "Port Preservation" differs from the recommendation in [2] for handling UDP traffic. We recommend that NAT devices preserve ports as preservation helps with (1) peer-to-peer connection establishment, since port numbers are more predictable; and (2) the implementation is not more complicated than other port assignment strategies (any assignment scheme needs a secondary port assignment scheme in case the preferred port is in use).

REQ-6 While this recommendation is a deviation from standard TCP behavior, the recommendation is a good compromise since it helps clients make direct peer-to-peer connections. It is worthwhile noting that firewalls generally do not respond to an incoming connection request with a TCP Reset.

REQ-7 This is one of the recommendations that, when met, enables clients to make direct peer-to-peer connections.

REQ-9 NAT devices typically support one or more ALGs, in which case supporting sequence number adjustment is not much of an additional burden. For a discussion on "Reset Unaware" see Section 5.3.

8. Security Considerations

NAT devices that do not keep track of per-connection sequence numbers and are also "Reset Aware", can be easily fooled by forged RST packets that contain arbitrary sequence numbers. Such an attack results in a Denial-of-Service, since an attacker can cause some bindings to be deleted.

9. IANA Considerations

There are no IANA considerations.

10. IAB Considerations

Refer to [2].

11. Acknowledgments

The author would like to acknowledge Francois Audet and Cullen Jennings for their draft [2] from which much material has been borrowed.

Thanks to Francois Audet, Kaushik Biswas, Dan Boneh, Bryan Ford, Cullen Jennings, Eric Rescorla, Hovav Shacham, and Senthil Sivakumar for useful comments and discussion.

12. References

12.1 Normative References

[1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119, March 1997.

[2] Audet, F., "NAT Behavioral Requirements for Unicast UDP", Internet-Draft 2005, January 2005.

[3] Daigle, L., "IAB Considerations for UNilateral Self-Address Fixing (UNSAF) Across Network Address Translation", RFC 3424, November 2002.

12.2 Informative References

[4] Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations", RFC 2663, August 1999.

[5] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, January 2001.

[6] Holdrege, M. and P. Srisuresh, "Protocol Complications with the IP Network Address Translator", RFC 3027, January 2001.

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

[8] Ziemba, G., Reed, D. and P. Traina, "Security Considerations for IP Fragment Filtering", RFC 1858, October 1995.

[9] Miller, I., "Protection Against a Variant of the Tiny Fragment Attack", RFC 3128, June 2001.

[10] Ford, B., Srisuresh, P. and D. Kegel, "State of Peer-to-Peer(P2P) communication across Network Address Translators(NATs)", draft-srisuresh-behave-p2p-state-00 (work in progress), December 2004.

Author's Address

Nagendra Modadugu
Stanford University
Computer Science Department
353 Serra Mall
Stanford, CA 94305
USA

EMail: nagendra@cs.stanford.edu

Intellectual Property Statement

Disclaimer of Validity

Copyright Statement

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.