

TSVWG  
Internet-Draft  
Intended status: Best Current Practice  
Expires: February 15, 2016

R. Penno  
Cisco  
S. Perreault  
Jive Communications  
M. Boucadair  
France Telecom  
S. Sivakumar  
Cisco  
K. Naito  
NTT  
August 14, 2015

**Network Address Translation (NAT) Behavioral Requirements Updates**  
**draft-ietf-tsvwg-behave-requirements-update-04**

Abstract

This document clarifies and updates several requirements of [RFC4787](#), [RFC5382](#) and [RFC5508](#) based on operational and development experience. The focus of this document is NAT44.

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 February 15, 2016.

Copyright Notice

Copyright (c) 2015 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

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">1.1.</a>	Scope . . . . .	<a href="#">2</a>
<a href="#">1.2.</a>	Terminology . . . . .	<a href="#">3</a>
<a href="#">2.</a>	TCP Session Tracking . . . . .	<a href="#">3</a>
<a href="#">2.1.</a>	TCP Transitory Connection Idle-Timeout . . . . .	<a href="#">4</a>
<a href="#">2.2.</a>	TCP RST . . . . .	<a href="#">5</a>
<a href="#">3.</a>	Port Overlapping Behavior . . . . .	<a href="#">5</a>
<a href="#">4.</a>	Address Pooling Paired (APP) . . . . .	<a href="#">6</a>
<a href="#">5.</a>	EIF Protocol Independence . . . . .	<a href="#">6</a>
<a href="#">6.</a>	EIF Mapping Refresh . . . . .	<a href="#">7</a>
<a href="#">6.1.</a>	Outbound Mapping Refresh and Error Packets . . . . .	<a href="#">7</a>
<a href="#">7.</a>	EIM Protocol Independence . . . . .	<a href="#">7</a>
<a href="#">8.</a>	Port Parity . . . . .	<a href="#">7</a>
<a href="#">9.</a>	Port Randomization . . . . .	<a href="#">8</a>
<a href="#">10.</a>	IP Identification (IP ID) . . . . .	<a href="#">8</a>
<a href="#">11.</a>	ICMP Query Mappings Timeout . . . . .	<a href="#">8</a>
<a href="#">12.</a>	Hairpinning Support for ICMP Packets . . . . .	<a href="#">9</a>
<a href="#">13.</a>	IANA Considerations . . . . .	<a href="#">9</a>
<a href="#">14.</a>	Security Considerations . . . . .	<a href="#">9</a>
<a href="#">15.</a>	References . . . . .	<a href="#">10</a>
<a href="#">15.1.</a>	Normative References . . . . .	<a href="#">10</a>
<a href="#">15.2.</a>	Informative References . . . . .	<a href="#">11</a>
	Acknowledgements . . . . .	<a href="#">12</a>
	Contributors . . . . .	<a href="#">12</a>
	Authors' Addresses . . . . .	<a href="#">12</a>

## [1.](#) Introduction

[[RFC4787](#)], [[RFC5382](#)] and [[RFC5508](#)] greatly advanced NAT interoperability and conformance. But with widespread deployment and evolution of Network Address Translation (NAT) more development and operational experience was acquired some areas of the original documents need further clarification or updates. This document provides such clarifications and updates.

### [1.1.](#) Scope

The goal of this document is to clarify and update the set of requirements listed in [[RFC4787](#)], [[RFC5382](#)] and [[RFC5508](#)]. The document focuses exclusively on NAT44.



The scope of this document has been set so that it does not create new requirements beyond those specified in the documents cited above. Carrier-Grade NAT (CGN) related requirements are defined in [[RFC6888](#)].

## 1.2. 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 [\[RFC2119\]](#).

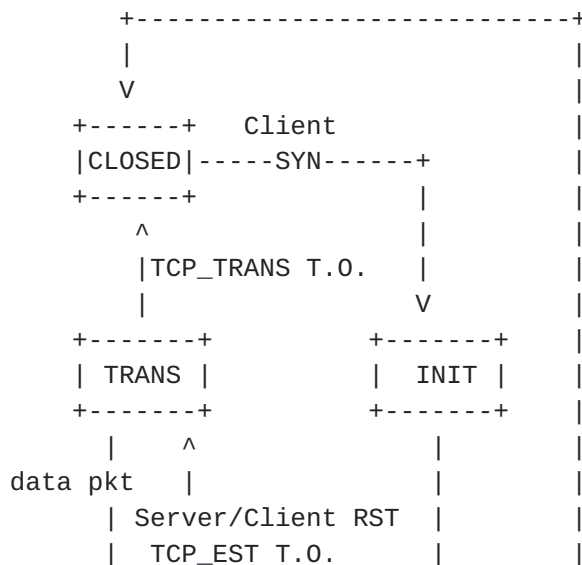
The reader is assumed to be familiar with the terminology defined in: [\[RFC2663\]](#), [\[RFC4787\]](#), [\[RFC5382\]](#), and [\[RFC5508\]](#).

In this document, the term "NAT" refers to both "Basic NAT" and "Network Address/Port Translator (NAPT)" (see [Section 3 of \[RFC4787\]](#)). As a reminder, Basic NAT and NAPT are two variations of traditional NAT, in that translation in Basic NAT is limited to IP addresses alone, whereas translation in NAPT is extended to include IP address and Transport identifier (such as TCP/UDP port or ICMP query ID) (refer to [Section 2 of \[RFC3022\]](#)).

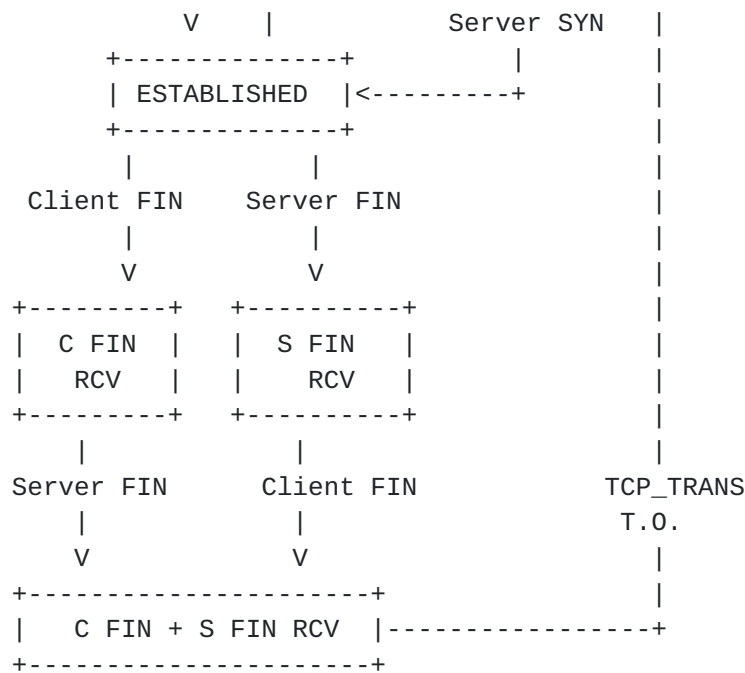
## 2. TCP Session Tracking

[RFC5382] specifies TCP timers associated with various connection states but does not specify the TCP state machine a NAT44 should follow as a basis to apply such timers.

Update: The TCP state machine depicted in Figure 1, adapted from [\[RFC6146\]](#), SHOULD be implemented by a NAT for TCP session tracking purposes.







## Legend:

- \* Messages sent to (resp. received from) the server are prefixed with "Server".
- \* Messages sent to (resp. received from) the client are prefixed with "Client".
- \* "C" means "Client-side"
- \* "S" means "Server-side".
- \* TCP\_EST T.O: refers to the established connection idle timeout as defined in [\[RFC5382\]](#).
- \* TCP\_TRANS T.O: refers to the transitory connection idle timeout as defined in [\[RFC5382\]](#).

Figure 1: State Machine

**2.1. TCP Transitory Connection Idle-Timeout**

The transitory connection idle-timeout is defined as the minimum time a TCP connection in the partially open or closing phases must remain idle before the NAT considers the associated session a candidate for removal (REQ-5 of [\[RFC5382\]](#)). But [\[RFC5382\]](#) does not clearly state whether these can be configured separately.

Clarification: This document clarifies that a NAT SHOULD provide different configurable parameters for configuring the open and closing idle timeouts.

To accommodate deployments that consider a partially open timeout of 4 minutes as being excessive from a security standpoint, a NAT



MAY allow to configure the timeout to be less than 4 minutes. Still, this specification recommends the default "transitory connection idle-timeout" minimum value to be set to 4 minutes.

## **2.2. TCP RST**

[RFC5382] leaves the handling of TCP RST packets unspecified.

Update: This document adopts a similar default behavior as in [RFC6146]. Concretely, when the NAT receives a TCP RST matching an existing mapping, it MUST translate the packet according the NAT mapping entry. Moreover, the NAT SHOULD wait for 4 minutes before deleting the session and removing any state associate with it if no packets are received during that 4 minutes timeout.

Admittedly, the NAT has to verify whether received TCP RST packets belong to a connection. These verification checks are required to avoid off-path attacks.

If the NAT removes immediately the NAT mapping upon receipt of a TCP RST message, stale connections may be maintained by endpoints if the first RST message is lost between the NAT and the recipient.

## **3. Port Overlapping Behavior**

REQ-1 from [RFC4787] and REQ-1 from [RFC5382] specify a specific port overlapping behavior; that is the external IP address and port can be reused for connections originating from the same internal source IP address and port irrespective of the destination. This is known as endpoint-independent mapping (EIM).

Update: This document clarifies that this port overlapping behavior may be extended to connections originating from different internal source IP addresses and ports as long as their destinations are different.

The following mechanism MAY be implemented by a NAT:

If destination addresses and ports are different for outgoing connections started by local clients, a NAT MAY assign the same external port as the source ports for the connections. The port overlapping mechanism manages mappings between external packets and internal packets by looking at and storing their 5-tuple (protocol, source address, source port, destination address, destination port).





This enables concurrent use of a single NAT external port for multiple transport sessions, which allows a NAT to successfully process packets in an IP address resource limited network (e.g., deployment with high address space multiplicative factor (refer to [Appendix B. of \[RFC6269\]](#))).

#### **4. Address Pooling Paired (APP)**

The Address Pooling Paired (APP) behavior for a NAT was recommended in REQ-2 from [\[RFC4787\]](#), but the behavior when a public IPv4 runs out of ports was left undefined.

Clarification: This document clarifies that if APP is enabled, new sessions from a host that already has a mapping associated with an external IP that ran out of ports SHOULD be dropped.

The administrator MAY provide a configurable parameter that allows a NAT to starting using ports from another external IP address when the one that anchored the APP mapping ran out of ports. This is a trade-off between service continuity and APP strict enforcement. (Note, this behavior is sometimes referred as 'soft-APP'.)

Update: This behavior SHOULD apply also for TCP.

#### **5. EIF Protocol Independence**

REQ-8 from [\[RFC4787\]](#) and REQ-3 from [\[RFC5382\]](#) do not specify whether EIF mappings are protocol-independent. In other words, if an inbound TCP SYN creates a mapping, it is left undefined whether inbound UDP packets destined to that mapping should be forwarded.

Update: This document specifies that EIF mappings SHOULD be protocol-independent in order allow inbound packets for protocols that multiplex TCP and UDP over the same IP address and port through the NAT and also maintain compatibility with stateful NAT64 . The administrator MAY provide a configuration parameter to make it protocol-dependent. The default value of this configuration parameter is to allow for protocol-independent EIF.

Applications that can be transported over a variety of transport protocols and/or support transport fall back schemes won't experience connectivity failures as a function of the underlying transport protocol or the filtering mode enabled at the NAT.



## **6. EIF Mapping Refresh**

The NAT mapping Refresh direction may have a "NAT Inbound refresh behavior" of "True" according to REQ-6 from [\[RFC4787\]](#), but [\[RFC4787\]](#) does not clarify how this behavior applies to EIF mappings. The issue in question is whether inbound packets that match an EIF mapping but do not create a new session due to a security policy should refresh the mapping timer.

Clarification: This document clarifies that even when a NAT has an inbound refresh behavior set to 'TRUE', such packets SHOULD NOT refresh the mapping. Otherwise a simple attack of a packet every 2 minutes can keep the mapping indefinitely.

Update: This behavior SHOULD apply also for TCP.

### **6.1. Outbound Mapping Refresh and Error Packets**

Update: In the case of NAT outbound refresh behavior there are certain types of packets that should not refresh the mapping even if their direction is outbound. For example, if the mapping is kept alive by ICMP Errors or TCP RST outbound packets sent as response to inbound packets, these SHOULD NOT refresh the mapping.

## **7. EIM Protocol Independence**

REQ-1 from [\[RFC4787\]](#) and REQ-1 from [\[RFC5382\]](#) do not specify whether EIM are protocol-independent. In other words, if a outbound TCP SYN creates a mapping it is left undefined whether outbound UDP can reuse such mapping and create session. On the other hand, stateful NAT64 [\[RFC6146\]](#) clearly specifies three binding information bases (TCP, UDP, ICMP).

Update: EIM mappings SHOULD be protocol-dependent. A configuration parameter MAY be provided in order allow protocols that multiplex TCP and UDP over the same source IP address and port number to use a single mapping.

## **8. Port Parity**

Update: A NAT MAY disable port parity preservation for all dynamic mappings. Nevertheless, A NAT SHOULD support means to explicitly request to preserve port parity (e.g., [\[I-D.ietf-pcp-port-set\]](#)).

Note: According to [\[RFC6887\]](#), dynamic mappings are said to be dynamic in the sense that they are created on demand, either implicitly or explicitly:



1. Implicit dynamic mappings refer to mappings that are created as a side effect of traffic such as an outgoing TCP SYN or outgoing UDP packet. Implicit dynamic mappings usually have a finite lifetime, though this lifetime is generally not known to the client using them.
2. Explicit dynamic mappings refer to mappings that are created as a result, for example, of explicit PCP MAP and PEER requests. Explicit dynamic mappings have a finite lifetime, and this lifetime is communicated to the client.

## **9. Port Randomization**

Update: A NAT SHOULD follow the recommendations specified in [Section 4 of \[RFC6056\]](#), especially:

"A NAT that does not implement port preservation [\[RFC4787\]](#) [\[RFC5382\]](#) SHOULD obfuscate selection of the ephemeral port of a packet when it is changed during translation of that packet. A NAT that does implement port preservation SHOULD obfuscate the ephemeral port of a packet only if the port must be changed as a result of the port being already in use for some other session. A NAT that performs parity preservation and that must change the ephemeral port during translation of a packet SHOULD obfuscate the ephemeral ports. The algorithms described in this document could be easily adapted such that the parity is preserved (i.e., force the lowest order bit of the resulting port number to 0 or 1 according to whether even or odd parity is desired)."

## **10. IP Identification (IP ID)**

Update: A NAT SHOULD handle the Identification field of translated IPv4 packets as specified in [Section 5.3.1 of \[RFC6864\]](#).

## **11. ICMP Query Mappings Timeout**

[Section 3.1 of \[RFC5508\]](#) precises that ICMP Query Mappings are to be maintained by a NAT. However, the specification doesn't discuss Query Mapping timeout values. [Section 3.2 of \[RFC5508\]](#) only discusses ICMP Query Session Timeouts.

Update: ICMP Query Mappings MAY be deleted once the last the session using the mapping is deleted.



## **12. Hairpinning Support for ICMP Packets**

REQ-7 from [[RFC5508](#)] specifies that a NAT enforcing 'Basic NAT' must support traversal of hairpinned ICMP Query sessions.

Clarification: This implicitly means that address mappings from external address to internal address (similar to Endpoint Independent Filters) must be maintained to allow inbound ICMP Query sessions. If an ICMP Query is received on an external address, a NAT can then translate to an internal IP.

REQ-7 from [[RFC5508](#)] specifies that all NATs must support the traversal of hairpinned ICMP Error messages.

Clarification: This behavior requires a NAT to maintain address mappings from external IP address to internal IP address in addition to the ICMP Query Mappings described in [Section 3.1 of \[RFC5508\]](#).

## **13. IANA Considerations**

This document does not require any IANA action.

## **14. Security Considerations**

NAT behavioral considerations are discussed in [[RFC4787](#)], [[RFC5382](#)], and [[RFC5508](#)].

Because some of the clarifications and updates (e.g., [Section 2](#)) are inspired from NAT64, the security considerations discussed in [Section 5 of \[RFC6146\]](#) apply also for this specification.

The update in [Section 3](#) allows for an optimized NAT resource usage. In order to avoid service disruption, the NAT MUST invoke this functionality only if packets are to be sent to distinct destination addresses.

Some of the updates (e.g., [Section 6](#), [Section 9](#), and [Section 11](#)) allow for an increased security compared to [[RFC4787](#)], [[RFC5382](#)], and [[RFC5508](#)]. Particularly:

- o The updates in [Section 6](#) and [Section 11](#) prevent an illegitimate node to maintain mappings activated in the NAT while these mappings should be cleared.
- o Port randomization ([Section 9](#)) complicates tracking hosts located behind a NAT.





[Section 4](#) and [Section 12](#) propose updates that increase the serviceability of a host located behind a NAT. These updates do not introduce any additional security concerns to [\[RFC4787\]](#), [\[RFC5382\]](#), and [\[RFC5508\]](#).

The updates in [Section 5](#) and [Section 7](#) allow for a better NAT transparency from an application standpoint. Hosts which require a restricted filtering behavior should enable security-dedicated features (e.g., ACL) either locally or by soliciting a dedicated security device (e.g., firewall).

The update in [Section 8](#) induces security concerns that are specific to the protocol used to interact with the NAT. For example, if PCP is used to explicitly request parity preservation for a given mapping, the security considerations discussed in [\[RFC6887\]](#) should be taken into account.

The update in [Section 10](#) may have undesired effects on the performance of the NAT in environments in which fragmentation is massively experienced. Such issue may be used as an attack vector against NATs.

## **[15. References](#)**

### **[15.1. Normative References](#)**

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC4787] Audet, F., Ed. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", [BCP 127](#), [RFC 4787](#), DOI 10.17487/RFC4787, January 2007, <<http://www.rfc-editor.org/info/rfc4787>>.
- [RFC5382] Guha, S., Ed., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", [BCP 142](#), [RFC 5382](#), DOI 10.17487/RFC5382, October 2008, <<http://www.rfc-editor.org/info/rfc5382>>.
- [RFC5508] Srisuresh, P., Ford, B., Sivakumar, S., and S. Guha, "NAT Behavioral Requirements for ICMP", [BCP 148](#), [RFC 5508](#), DOI 10.17487/RFC5508, April 2009, <<http://www.rfc-editor.org/info/rfc5508>>.



- [RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", [BCP 156](#), [RFC 6056](#), DOI 10.17487/RFC6056, January 2011, <<http://www.rfc-editor.org/info/rfc6056>>.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", [RFC 6146](#), DOI 10.17487/RFC6146, April 2011, <<http://www.rfc-editor.org/info/rfc6146>>.
- [RFC6864] Touch, J., "Updated Specification of the IPv4 ID Field", [RFC 6864](#), DOI 10.17487/RFC6864, February 2013, <<http://www.rfc-editor.org/info/rfc6864>>.
- [RFC6888] Perreault, S., Ed., Yamagata, I., Miyakawa, S., Nakagawa, A., and H. Ashida, "Common Requirements for Carrier-Grade NATs (CGNs)", [BCP 127](#), [RFC 6888](#), DOI 10.17487/RFC6888, April 2013, <<http://www.rfc-editor.org/info/rfc6888>>.

## **15.2. Informative References**

- [I-D.ietf-pcp-port-set] Qiong, Q., Boucadair, M., Sivakumar, S., Zhou, C., Tsou, T., and S. Perreault, "Port Control Protocol (PCP) Extension for Port Set Allocation", [draft-ietf-pcp-port-set-09](#) (work in progress), May 2015.
- [RFC2663] Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations", [RFC 2663](#), DOI 10.17487/RFC2663, August 1999, <<http://www.rfc-editor.org/info/rfc2663>>.
- [RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", [RFC 3022](#), DOI 10.17487/RFC3022, January 2001, <<http://www.rfc-editor.org/info/rfc3022>>.
- [RFC6269] Ford, M., Ed., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", [RFC 6269](#), DOI 10.17487/RFC6269, June 2011, <<http://www.rfc-editor.org/info/rfc6269>>.
- [RFC6887] Wing, D., Ed., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", [RFC 6887](#), DOI 10.17487/RFC6887, April 2013, <<http://www.rfc-editor.org/info/rfc6887>>.



## Acknowledgements

Thanks to Dan Wing, Suresh Kumar, Mayuresh Bakshi, Rajesh Mohan, Lars Eggert, and Gorrry Fairhurst for their review and discussion.

## Contributors

The following individual contributed text to the document:

Sarat Kamiset, Insieme Networks, United States

## Authors' Addresses

Reinaldo Penno  
Cisco Systems, Inc.  
170 West Tasman Drive  
San Jose, California 95134  
USA

Email: [repenno@cisco.com](mailto:repenno@cisco.com)

Simon Perreault  
Jive Communications  
Canada

Email: [sperreault@jive.com](mailto:sperreault@jive.com)

Mohamed Boucadair  
France Telecom  
Rennes 35000  
France

Email: [mohamed.boucadair@orange.com](mailto:mohamed.boucadair@orange.com)

Senthil Sivakumar  
Cisco Systems, Inc.  
United States

Email: [ssenthil@cisco.com](mailto:ssenthil@cisco.com)



Kengo Naito  
NTT  
Tokyo  
Japan

Email: [k.naito@nttv6.jp](mailto:k.naito@nttv6.jp)