Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 25, 2013

T. Tsou
Huawei Technologies (USA)
W. Liu
Huawei Technologies
S. Perreault
Viagenie
R. Penno
Cisco Systems, Inc.
M. Chen
FreeBit
October 22, 2012

Stateless IPv4 Network Address Translation
draft-tsou-stateless-nat44-02

Abstract

This memo describes a protocol for decentralizing IPv4 NAT to the customer-premises equipment (CPE) such that no state information is kept on the central NAT device. The CPE uses a restricted source port set that is encoded in its provisioned IPv4 WAN address. The NAT device performs only strictly stateless address (not port) translation.

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 April 25, 2013.

Table of Contents

1. Introduction 3
2. Terminology 3
3. Address Formats 4
4. Customer Provisioning 5
5. SLNAT44 Configuration 6
6. Port Set Computation 6
7. CPE Operation 7
7.1. ALG Handling 8
8. SLNAT44 Operation 8
8.1. Internal to External 8
8.2. External to Internal 8
8.3. Fragment Handling 9
9. Address Mapping Example 9
10. Security Considerations 10
11. Acknowledgements 10
12. References 10
12.1. Normative References 10
12.2. Informative References 10
Authors' Addresses 11

1. Introduction

IPv4 address exhaustion has become world-wide reality. NAT is one of the solutions to deal with the problem. The drawbacks of traditional NAT include statefulness and the need to track transport-layer sessions. This makes NAT complex, hard to scale up, and fragile.

This document describes a method of deploying stateless NAT as a backwards-compatible evolution of an IPv4-only network.

The assumed topology is illustrated in Figure 1.

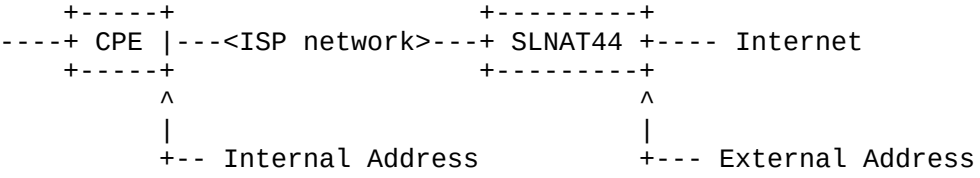

Figure 1: Stateless NAT44 topology

When CPE is configured working as a transparent bridge, internal addresses are directly assigned to the end hosts in the home network, as is shown in Figure 2.

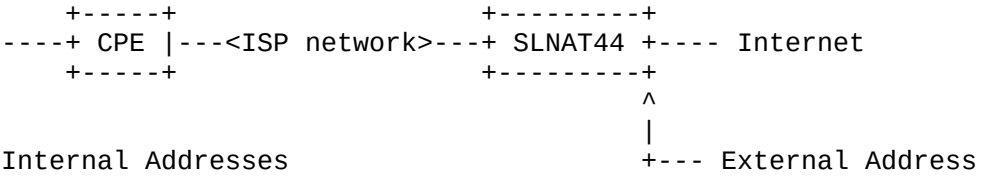

Figure 2: Stateless NAT44 topology: CPE as bridge

Note that SLNAT44 has no IPv6 component. Any deployment of IPv6 is unaffected by SLNAT44. Therefore, this document only describes IPv4 addresses and IPv4 packets. IPv6 is not discussed further.

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 following terms are used throughout this document:

Port set: Set of transport-layer ports that each CPE is assigned, to be used as source ports by packets sent by the CPE.

Port Set ID: A value from which a unique port set is algorithmically derived.

SLNAT44: Depending on the context, either the stateless NAT44 protocol or the stateless NAT44 device that translates between internal and external addresses. NAT44 in turn stands for "IPv4-to-IPv4 NAT".

Internal Address: The IPv4 address assignted to a CPE. It is used in the ISP network between the CPE and the SLNAT44.

External Address: The IPv4 address used on the Internet and routed to the SLNAT44.

Mapping rule: A set of parameters configured on the SLNAT44 (not on the CPE) describing the relationship between internal and external addresses.

3. Address Formats

Internal addresses have the format illustrated in Figure 3. The addresses are simply made of three parts concatenated together: the Internal Prefix, the External Suffix, and the Port Set ID.

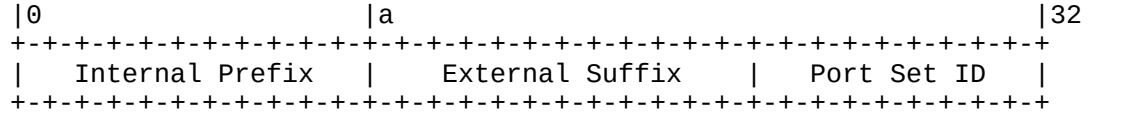

Figure 3: Internal Address format

External Addresses have the format illustrated in Figure 4. It is made of two parts: the External Prefix and the External Suffix.

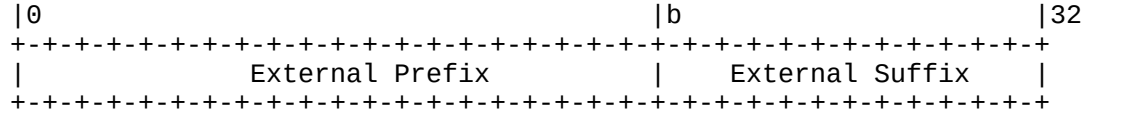

Figure 4: External Address format

The lengths of the Internal and External Prefixes, "a" and "b", are mandatory parameters of SLNAT44. They are determined by the ISP. They need not be communicated to the CPE. Other lengths can be

computed from them as follows:

- Length of External Suffix: 32 - b
- Length of Port Set ID: b - a

4. Customer Provisioning

Customer Provisioning is applied to the CPE when the CPE serves a gateway.

As part of its start up routine, the CPE is assigned an IPv4 address by the ISP using regular means (DHCP, PPP, etc.). This is the Internal Address.

In addition, using new provisioning options, the CPE is assigned a Port Set ID.

Optionally, a Port Set Mask is also provisioned to the CPE. This mask is of the same length as the Port Set ID (i.e., b-a bits). Its purpose is to allow discontinuous port ranges. If no mask is provided, a mask of all ones is assumed by default, which implies a continuous port range. System ports (0-1023) should not to be assigned to any CPE.

In summary, the CPE is provisioned with the following elements:

- IPv4 address (as usual)
- Port Set ID
- Port Set Mask (optional)

When the CPE is configured in the bridge mode, all the above features are provisioned directly to the end host behind the CPE.

Note: no matter in which mode the CPE is running, the customer provisioning could be either dynamic or static. Static provisioning implies an address planning for the private IPv4 address (i.e., RFC1918 addresses) inside in the domain. Static provisioning enables servers (passive daemons) at the home network being accessible within the domain. CPE running as bridge makes this feature easy to deploy while running as L3 gateway requires port redirection if an in-domain server at a host is demanded.

5. SLNAT44 Configuration

The SLNAT44 is configured with a set of mapping rules. Each rule contains:

- Internal Prefix
- External Prefix
- Port Set Mask (optional)

Prefixes include their length. For simplicity, rule prefixes MUST NOT overlap with other rules.

If it is absent, the Port Set Mask is assumed to be all ones by default.

6. Port Set Computation

Given a Port Set ID and a Port Set Mask, both n bits in length, the set of allowed ports is defined as the set of port numbers for which the higher-order n bits of their binary expression whose corresponding mask bits are 1 are equal to corresponding bits from the Port Set ID.

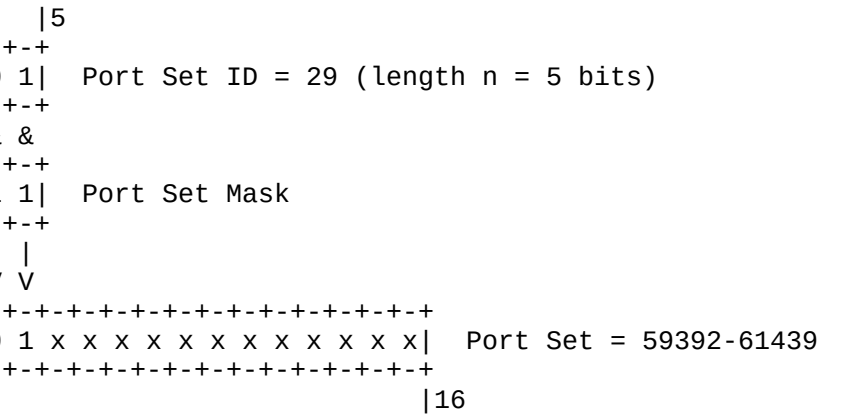

Figure 5: Example Contiguous Port Set Computation

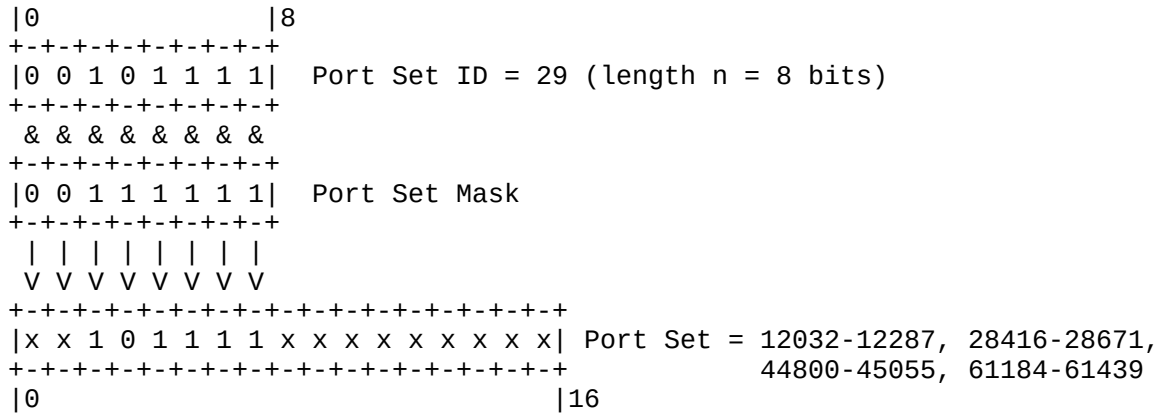

Figure 6: Example Non-Contiguous Port Set Computation

It follows that the number of ports in the set is 2^(16-x), where x is the number of ones in the Port Set Mask.

This computation is performed by the CPE as part of its provisioning routine as well as by the SLNAT44 for dropping packets with ports outside the allowed range.

For the purposes of SLNAT44, a "source port" corresponds to either a TCP source port, a UDP source port, or an ICMPv4 identifier, while a "destination port" corresponds to either a TCP destination port, a UDP destination port, or an ICMPv4 identifier. Note that an ICMPv4 identifier plays the role of both source and destination port.

Transport protocols other than TCP and UDP, as well as ICMPv4 types without an identifier field, are not supported.

7. CPE Operation

CPE can be configured as either a gateway or transparent bridge.

In the gateway mode, packets sent from the CPE MUST have the provisioned IPv4 address as source and MUST have a source port that is within the allowed set. This is usually accomplished by having the CPE run a NAT44 configured with the provisioned address and allowed port set and having it process all packets sent out the WAN interface.

Packets received by the CPE on its WAN interface with a destination port outside the allowed range MUST be dropped.

In the bridge mode, however, CPE only transfers packets and therefore the service of stateless NAT44 is performed by the SLNAT44 directly towards end hosts that possibly running as in-domain servers.

Regardless of any mode of the CPE, the operation involves injecting private addresses (or prefixes) into the intra-domain backbone routing infrastructure. It is necessary to operationally ensure that there are no private addresses/prefixes are leaking into the backbone route tables unless they are assigned by the SLNAT44 to CPEs or directly to hosts.

7.1. ALG Handling

If the CPE implements application level gateways (ALGs) such as FTP, RSTP or PPTP, it must ensure that ports present in the payload when tanslated fall within the allowed range.

8. SLNAT44 Operation

8.1. Internal to External

When it receives a packet on an internal interface, the SLNAT44 finds the rule whose Internal Prefix matches the packet’s source address. It extracts the Port Set ID from the packet’s source address. It then checks if the packet’s source port is within the allowed set, using the rule’s Port Set Mask. If it is not, the packet MUST be dropped.

If the packet’s source port is within the allowed set, the SLNAT44 builds the External Address by concatenating the rule’s External Prefix with the External Suffix extracted from the packet’s source address. It then replaces the packet’s source address with this External Address. The IPv4 and transport-layer checksums are updated as necessary. The packet is then forwarded as usual.

8.2. External to Internal

When it receives a packet on an external interface, the SLNAT44 finds the rule whose External Prefix matches the packet’s destination address. It then builds the Internal Address by concatenating the rule’s Internal Prefix, the External Suffix extracted from the packet’s destination address, and the Port Set ID computed by applying the rule’s Port Set Mask to the packet’s destination port’s higher-order bits. It then replaces the packet’s destination address with this Internal Address. The IPv4 and transport-layer checksums are updated as necessary. The packet is then forwarded as usual.

8.3. Fragment Handling

If the incoming IP packet contains a fragment, then more processing may be needed. This specification leaves open the exact details of how a SLNAT44 handles incoming IP packets containing fragments, and simply requires that the external behavior of the SLNAT444 be compliant with the following conditions.

The SLNAT44 MUST handle fragments. In particular, SLNAT44 MUST handle fragments arriving out of order, conditional on the following:

- The SLNAT44 MUST limit the amount of resources devoted to the storage of fragmented packets in order to protect from DoS attacks.
- As long as the SLNAT44 has available resources, the SLNAT44 MUST allow the fragments to arrive over a time interval. The time interval SHOULD be configurable and the default value MUST be of at least 2 seconds.
- The SLNAT44 MAY require that the UDP, TCP, or ICMPv4 header be completely contained within the fragment that contains fragment offset equal to zero.

For incoming packets carrying TCP or UDP fragments with a non-zero checksum, SLNAT44 MAY elect to queue the fragments as they arrive and translate all fragments at the same time. In this case, the incoming tuple is determined as documented above to the un-fragmented packets. Alternatively, a SLNAT44 MAY translate the fragments as they arrive, by storing information that allows it to compute the necessary port number for fragments other than the first. In the latter case, subsequent fragments may arrive before the first, and the rules (in the bulleted list above) about how the SLNAT44 handles (out-of-order) fragments apply.

Implementers of SLNAT44 should be aware that there are a number of well-known attacks against IP fragmentation; see [RFC1858] and [RFC3128]. Implementers should also be aware of additional issues with reassembling packets at high rates, described in [RFC4963].

9. Address Mapping Example

An operator has two public ranges of size /18 and /19 called foo and bar respectively. It plans to use 10/8 as its internal address prefix and PSID (port range) of length 5. Two prefixes of the internal network

The internal prefixes lengths are:

- 32 - 18 - 5 = 13 (derived from foo)
- 32 - 19 - 5 = 14 (derived from bar)

This will give the following possible mappings:

- foo/18 <--> 10.0.0.0/13
- bar/19 <--> 10.128.0.0/14

Author note: Discuss the where internal prefixes are overlapping

10. Security Considerations

The security considerations related to IP address sharing documented in RFC 6269 [RFC6269] and RFC 6056 [RFC6056] apply to SLNAT44.

11. Acknowledgements

Section 8.3 is adapted from [RFC6146].

12. References

12.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

12.2. Informative References

[RFC1858] Ziemba, G., Reed, D., and P. Traina, "Security Considerations for IP Fragment Filtering", RFC 1858, October 1995.

[RFC3128] Miller, I., "Protection Against a Variant of the Tiny Fragment Attack (RFC 1858)", RFC 3128, June 2001.

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

[RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", BCP 156, RFC 6056, January 2011.

[RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.

[RFC6269] Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", RFC 6269, June 2011.

Authors' Addresses

Tina Tsou
Huawei Technologies (USA)
2330 Central Expressway
Santa Clara, CA 95050
USA

Phone: +1 408 330 4424
Email: tina.tsou.zouting@huawei.com

Will Liu
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Email: liushucheng@huawei.com

Simon Perreault
Viagenie
246 Aberdeen
Quebec, QC G1R 2E1
Canada

Email: simon.perreault@viagenie.ca

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

Email: repenno@cisco.com

Maoke Chen
FreeBit
3-6 Maruyama-cho
Shibuya-ku, Tokyo 150-0044
Japan

Email: fibrib@gmail.com