Stateless IPv4 Network Address Translation
draft-tsou-stateless-nat44-00

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

Copyright Notice

Table of Contents

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.

```
             +-----+                    +---------+
 Home   -----+ CPE |---<ISP network>---+ SLNAT44 +---- Internet
network      +-----+                    +---------+
                ^                            ^
                |                            |
              +-- Internal Address       +--- External Address
```
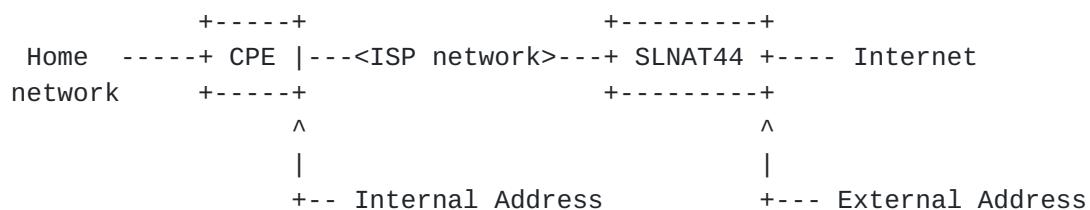
                   Figure 1: Stateless NAT44 topology

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 2. The
   addresses are simply made of three parts concatenated together: the
   Internal Prefix, the External Suffix, and the Port Set ID.

```
|0                   |a                                   |32
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    Internal Prefix   |    External Suffix    |   Port Set ID   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
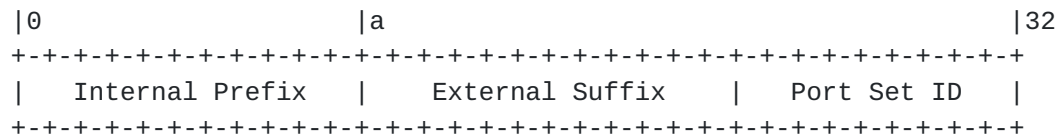
                    Figure 2: Internal Address format

   External Addresses have the format illustrated in Figure 3. It is
   made of two parts: the External Prefix and the External Suffix.

```
|0                                   |b                   |32
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|             External Prefix         |    External Suffix    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
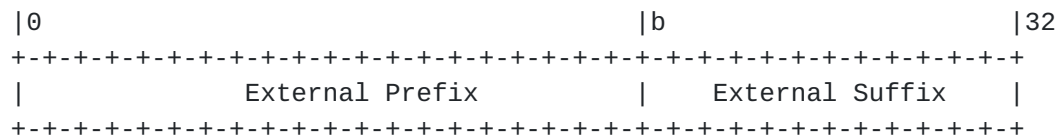
                    Figure 3: External Address format

   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:

   o Length of External Suffix: 32 - b

   o Length of Port Set ID: b - a

[4]. CPE Provisioning

   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.

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

   o IPv4 address (as usual)

   o Port Set ID

   o Port Set Mask (optional)

[5]. SLNAT44 Configuration

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

   o Internal Prefix

   o External Prefix

   o 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.

```
   |0          |5
   +-+-+-+-+-+-+
   |1 1 1 0 1|  Port Set ID = 29 (length n = 5 bits)
   +-+-+-+-+-+
    & & & & &
   +-+-+-+-+-+
   |1 1 1 1 1|  Port Set Mask
   +-+-+-+-+-+
    | | | | |
     V V V V V
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |1 1 1 0 1 x x x x x x x x x x x|  Port Set = 59392-61439
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |0                             |16
```

              Figure 4: Example Contiguous Port Set Computation


```
|0                |8
+-+-+-+-+-+-+-+-+-+
|0 0 1 0 1 1 1 1|  Port Set ID = 29 (length n = 8 bits)
+-+-+-+-+-+-+-+-+
 & & & & & & & &
+-+-+-+-+-+-+-+-+
|0 0 1 1 1 1 1 1|  Port Set Mask
+-+-+-+-+-+-+-+-+
 | | | | | | | |
  V V V V V V V V
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|x x 1 0 1 1 1 1 x x x x x x x x x x x| Port Set = 12032-12287, 28416-28671,
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                44800-45055, 61184-61439
|0                             |16
```

             Figure 5: 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

   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.

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:

o The SLNAT44 MUST limit the amount of resources devoted to the
   storage of fragmented packets in order to protect from DoS
   attacks.

o 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.

o 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. Security Considerations

10. Acknowledgements

Section 8.3 is adapted from [RFC6146].

11. References

11.1. Normative References

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

11.2. 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.

   [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.

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

   Phone:
   Email: Liushucheng@huawei.com

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

Email: simon.perreault@viagenie.ca