Addressing bit compression for stateless IPv6 prefix translation
draft-tremblay-pt66ac-00

Abstract

This document describes a method to extend IPv6 prefix translation [NAT66] (Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Address Translation (NAT66)," November 2008.) to statelessly translate between IPv6 prefixes of differing lengths. This technique, called PT66 addressing bit compression (PT66-AC), provides a way to map a shorter prefix (such as the fixed [6to4] prefix) to a longer global unicast prefix, while avoiding stateful translation. The difference between the prefix lengths is compensated for by removing an unused string of subnetting bits, as specified, from the addresses within the shorter prefix.

Requirements Language

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.

Status of this Memo

Table of Contents

---

1.   Introduction

Since IPv6 has become a standard, multiple transition mechanisms have
been proposed to stimulate the rate of IPv6 adoption. Some of these
mechanisms use specific addressing space and fixed lengths for
different types of prefixes. There is sometimes a need to translate a
protocol-specific addressing space into global unicast space to improve
the routing behavior. Of course, this improvement to routing comes at
the cost of lowered interoperability for some applications.

One issue with translating these prefixes is the challenge of obtaining a globally unique prefix of appropriate size. However, there are often a number of bits rarely used within the transition protocol specific prefixes. Using an address translator with addressing bit compression allows a deployment to drop these unused bits and make more efficient usage of the public addressing space.

For example, the IPv6 transition protocol [6to4] statelessly maps 32 bit IPv4 addresses into the 2002::/16 prefix, to create a /48 IPv6 prefix for every IPv4 address. Normally, return traffic destined to this prefix will be routed to the nearest 6to4 relay router. An ISP may wish to gain control over return routing of the 6to4 traffic by mapping their 6to4 address space into a globally unique IPv6 prefix. Implementing this without addressing bit compression would require an IPv6 /32 for every /16 of IPv4 global space assigned to an ISP. Acquiring the required IPv6 space from a Regional Internet Registry is therefore unrealistic in most cases. Using addressing bit compression, an ISP can make better use of the globally unique IPv6 space by reducing the size of the 6to4 subnet available to each user.

---

2. IPv6 Translator Model

The current stateless IPv6 prefix translation model is described in [NAT66] (Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Address Translation (NAT66)," November 2008.). In that model, IPv6 traffic is translated between an internal network and an external network.

An IPv6 translator model described in [NAT66] (Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Address Translation (NAT66)," November 2008.):

```
        External Network:  Prefix = 2001:0DB8:0001:/48
           ---------------------------------------
                            |
                            |
                      +---------+
                      |  PT66   |
                      | Device  |
                      +---------+
                            |
                            |
           --------------------------------------
         Internal Network:  Prefix = FD01:0203:0405:/48
```
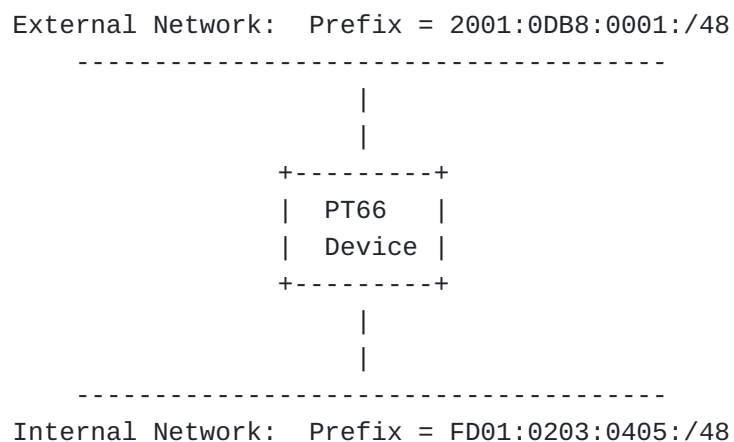
This translation model may be represented in a generic fashion by replacing the address characters with letters denoting the function of those address bits. The representation below describes a single translation mapping in a device that may implement several.

L – Internal (or "Local") network IPv6 prefix

E – External network IPv6 prefix

N – Subnet bits

I – Interface Identifier (IID) bits

A generic IPv6 translator model

```
External Network Prefix = EEEE:EEEE:EEEE:NNNN:IIII:IIII:IIII:IIII
                   ---------------------------------------
                                     |
                                     |
                          +---------+
                          |  PT66   |
                          | Device  |
                          +---------+
                                     |
                                     |
                   ---------------------------------------
    Internal Network Prefix = LLLL:LLLL:LLLL:NNNN:IIII:IIII:IIII:IIII
```
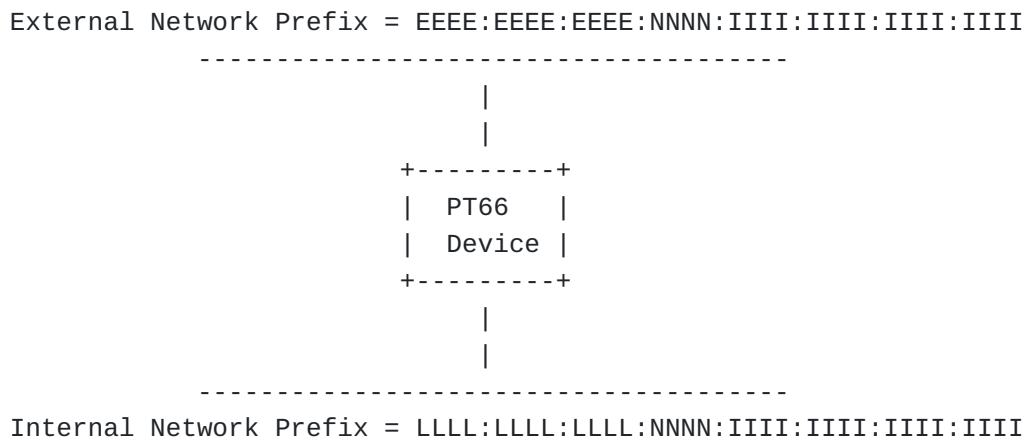
In this figure, L represents the Internal or Local prefix to be
translated to the External prefix bits E. The number of L bits is equal
to the number of E bits. Subnet bits N and Interface-ID bits I are not
changed by the translator.

---

3. IPv6 Prefix Translation with Addressing Compression

It is implicit that any basic [NAT66] (Wasserman, M. and F. Baker,
"IPv6-to-IPv6 Network Address Translation (NAT66)," November 2008.)
mapping must be between prefixes of the same size, as each prefix of is
overwritten by the other as traffic passes through. The section below
gives a description of how to map internal and external networks of
different lengths. In this example, the internal prefix used is a
larger /28, while the external prefix used is a smaller /40.

L - Local or Internal network IPv6 prefix

E – External network IPv6 prefix
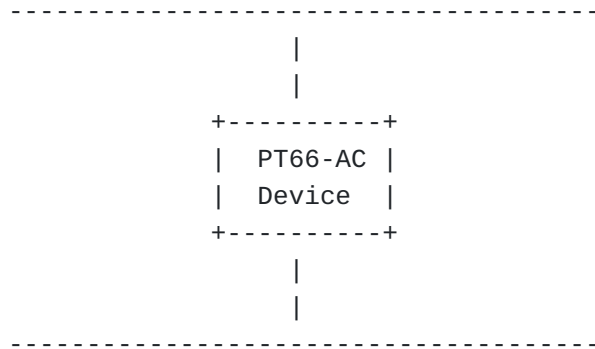
S - Shifted subnet bits

N - Unmodified subnet bits

C - Compressed (removed) bits

I – Interface Identifier (IID) bits

Generic Prefix Translator Model with Addressing Compression

```
External Network Prefix = EEEE:EEEE:EESS:SSSN:IIII:IIII:IIII:IIII
                 ---------------------------------------
                                    |
                                    |
                          +----------+
                          |  PT66-AC |
                          |  Device  |
                          +----------+
                                    |
                                    |
                 ---------------------------------------
     Internal Network Prefix = LLLL:LLLS:SSSS:CCCN:IIII:IIII:IIII:IIII
```

In this translation model, the N bits (4 bits represented) from the
internal subnet are left untouched. The C bits (12 bits represented)
are the compressed bits, not present in the external network prefix.
The S bits (20 bits represented) are bitwise shifted to the right by
the length of the C bits when translating from the internal network to
the external network. This compression of the subnet allows for the
adaptation from the larger prefix to the smaller prefix to be made. The
L bits (28 bits represented) are replaced by the E bits (40 bits
represented) when translating from the internal network to the external
network.
In this document, traffic received on the internal network interface
and forwarded by the external traffic interface is called egress
traffic. Inversely, any traffic received on the external network
interface and forwarded to the internal network interface is called
ingress traffic.

3.1. Configuration Elements

The IPv6 prefix translation with addressing compression model (PT66-AC)
makes use of the following variables:

L Length of the internal network prefix

E Length of the external network prefix

S Number of shifted bits

C Number of compressed bits

N Number of unchanged subnet bits

In order to avoid confusion and increase readability, both the full
name and variable notation of the above parameters are used throughout
this document.

In this translation model, the sum of the length of the internal
network prefix (L), of the number of shifted bits (S), of the number of
compressed bits (C) and of the number of unchanged bits (N) always
equals 64 (L + S + C + N = 64). As well, the sum of the length of the
external network (E), of the number of shifted bits (S) and of the
number of unchanged bits (N) always equals 64 (E + S + N = 64).
Hence the number of compressed bits (C) is always equal to the length
of the external prefix minus the length of the local prefix (C = E –
L). Only the length of each prefix need be defined, and the length of
the compressed bits (C) can be derived from those values. Similarly, in
order to know the number of shifted bits (S) and unchanged bits (N),
one of these two must be defined in the configuration of the
translation mapping. Therefore, the elements that must be defined in a
translation mapping with compression are the internal network prefix
and its length (L), the external network prefix and its length (E) and
either the number of shifted bits (S) or unchanged bits (N).

With these configuration elements defined for each translation mapping,
an IPv6 translator can statelessly translate ingress and egress traffic
between prefixes of differing length.

3.2. Egress Translation Behavior for Compressed Bits

The value of the bits compressed is hereto referred to as the
'compressed bit pattern'. By default, any egress traffic MUST be
dropped if the compressed bit pattern is not equal to zero. Optionally,
a compressed bit pattern different from all zeros may be defined by
configuration and used to filter incoming traffic. The compressed bit
pattern length MUST be equal to the difference between the length of
the external prefix and the length of the internal prefix (C must equal
E – L).

The translation device SHOULD issue an ICMP type 1 (Destination
Unreachable) message with code 5 (Source address failed ingress/egress
policy) to the source address when dropping traffic for the first time
for each source address/destination address pair.

3.3. Ingress Translation Behavior for Compressed Bits

For traffic entering the external network interface of the translator,
the compressed bits MUST be restored with zeros by default, unless a
different compressed bit pattern has been configured for the mapping.
In such a case, the user-defined compressed bit pattern MUST be used to
restore the compressed bits.

3.4. Translator Operation

For traffic entering the external network interface of the translator,
the compressed bits MUST be restored with zeros by default, unless a
different compressed bit pattern has been configured for the mapping.
In such a case, the user-defined compressed bit pattern MUST be used to
restore the compressed bits.

3.4.1. Egress Traffic Operation

For each egress packet, the prefix translator MUST use the appropriate
translation mapping with the longest matching internal prefix. The S
bits of the IPv6 source address are shifted bitwise to the right by the
number of compressed bits (C). The leftmost bits of the source address
(L) are replaced by the bits of the external network prefix (E).
Unchanged bits (N) of the source address MUST NOT be modified.

3.4.2. Ingress Traffic Operation

For each ingress packet, the prefix translator MUST use the appropriate
translation mapping. Selection of the appropriate mapping is outside of
the scope of this document. The behavior of the translator when no
mapping is matched is also out of scope for this document. Asynchronous
mappings between internal and external prefixes, where an ingress
source address translated does not in turn translate back to the same
destination address on egress traffic, SHOULD be prevented.
Once a suitable mapping is found, the leftmost bits of the destination
address are replaced by the corresponding bits from the local network
prefix (L). Starting at the bit position equal to the length of the
external prefix (E), the Shifted bits are bitwise shifted to the left
by the number of compressed bits (C). The C bits are restored with
zeros, or the compressed bit pattern if set in the mapping

configuration. Unchanged bits (N) of the destination address MUST NOT
be modified.
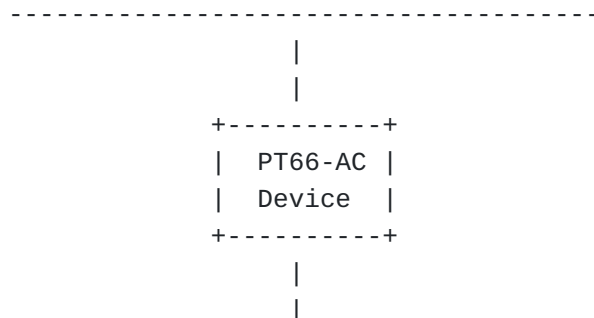
---

4. Examples

4.1. 6to4 Subnet Compression

When an IPv6 prefix translator is used with a 6to4 prefix as the
internal network, it is possible to compress seldom-used subnet bits in
order to reduce the size of the external global unicast prefix. As
specified in [6to4], a 6to4 router will define a /48 for each IPv4
global unicast address. This provides 16 bits of subnet addressing
available to the user. The 6to4 protocol being often implemented by
residential gateways, only one LAN subnet (/64) is utilized in most
implementations. It is therefore possible to reduce the number of
available /64 subnets available from 65536 (16 bits) to 16 (4 bits) or
even a single one (0 subnet bits). The following example compresses the
16 subnet bits to 4 bits, leaving room for 16 /64 subnets.
The prefix translation model for 6to4 is similar to the generic model,
with the internal network bits (L) being defined as 2002::/16. The 32
bits of the 6to4 router IPv4 address are the shifted bits (S).
This example uses 192.0.2.1 as the 6to4 router IPv4 address and
2001:DB0::/28 as the provider global unicast prefix. A single subnet
number 0001 is used. A single host exchanges traffic from Interface ID
::A.

```
      External Network Prefix = EEEE:EEES:SSSS:SSSN:IIII:IIII:IIII:IIII
                                2001:0DBC:0000:2011:0000:0000:0000:000A
                        ---------------------------------------
                                        |
                                        |
                                 +----------+
                                 |  PT66-AC |
                                 |  Device  |
                                 +----------+
                                        |
                                        |
                        ---------------------------------------
          Internal 6to4 Prefix = LLLL:SSSS:SSSS:CCCN:IIII:IIII:IIII:IIII
                                 2002:C000:0201:0001:0000:0000:0000:000A
```

In this example, the ISP configures the PT66-AC with the following
mapping: 2002::/16, 2001:DB0::/28, C=12 (or N=4)
During operation, egress traffic from address 2002:C000:0201:1::A is
translated as follow:

```
    2002:C000:0201:0001::A  Egress packets source address
    2002:C00C:0000:2011::A  Shift S bits (32 bits) to the right
                            by C bits (12 bits)
    2001:0DBC:0000:2011::A  Replace L bits by E bits. This is the new
                            egress source address.
```

The inverse operations takes place for ingress traffic with
destinations within 2001:db0::/28:

```
    2001:0DBC:0000:2011::A  Incoming destination address
    2002:0DBC:0000:2011::A  Leftmost bits are replaced by the L bits
                            (16 bits)
    2002:C000:0201:2011::A  The 32 S bits are shifted to the left
                            by C bits (12 bits)
    2002:C000:0201:0001::A  The C bits are restored. This is the new
                            destination address on the local network
```

This example demonstrated how to compress the 12 first bits of the 6to4
subnet bits in order for an ISP to use a longer global unicast prefix
for the external network.

4.2. 6to4 IPv4 and Subnet Compression

As an improvement over the previous example, the ISP also has the
opportunity to compress the IPv4 prefix part of the IPv4 address in the
6to4 prefix. That information is redundant, because it can be recovered
from the mapping when translating ingress traffic. Omitting this allows
the ISP to use an even smaller prefix on the external interface.
This example is similar the previous one, with the addition that the
IPv4 prefix 192.0.2.0/24 will also be compressed. The external prefix
used by the ISP is now 2001:db8:FFFF:F000::/52. The 6to4 subnet remains
compressed from 16 to 4 bits.
A translation mapping is configured as follow: 2002:C000:0200::/40,
2001:0DB8:FFFF:F000::/52, C=12 (or N=4)

```
     External Network Prefix = EEEE:EEEE:EEEE:ESSN:IIII:IIII:IIII:IIII
                               2001:0DB8:FFFF:F011:0000:0000:0000:000A
                    ---------------------------------------
                                        |
                                        |
                               +----------+
                               |  PT66-AC |
                               |  Device  |
                               +----------+
                                        |
                                        |
                    ---------------------------------------
       Internal 6to4 Prefix = LLLL:LLLL:LLSS:CCCN:IIII:IIII:IIII:IIII
                               2002:C000:0201:0001:0000:0000:0000:000A
```

Egress operation:

```
    2002:C000:0201:0001::A  Egress packets source address
    2002:C000:0201:0011::A  Shift S bits (8 bits) to the
                            right by C bits (12 bits)
    2001:0DB8:FFFF:F011::A  Replace L bits by E bits. This
                            is the new egress source address.
```

Ingress operation for destinations within 2001:db8:FFFF:F000::/52:

```
    2001:0DB8:FFFF:F011::A  Incoming destination address
    2002:C000:02FF:F011::A  Leftmost bits are replaced
                            by the L bits (40 bits)
    2002:C000:0201:F011::A  The S bits (8 bits) are shifted to
                            the left by C bits (12 bits)
    2002:C000:0201:0001::A  The C bits are restored. This is the
                            new destination address on local network
```

This example demonstrates that both the IPv4 addressing and the subnet
parts of a 6to4 prefix can be compressed. It is therefore possible to
translate a provider's own 6to4 addressing subset (from its own global
IPv4 unicast space) to a set of global IPv6 unicast prefixes that are
small enough not to require an extremely large IPv6 global unicast
assignment.
The resulting IPv6 unicast prefixes are also completely decoupled from
IPv4 and avoid the need for injection of IPv4 routing information in
the IPv6 routing table in order to achieve better routing.

5. IANA Considerations

This document makes no request of IANA.
Note to RFC Editor: this section may be removed on publication as an RFC.

---

6. Security Considerations

All of the security considerations of [NAT66] are still applicable when address bit compression is applied. The stateless address mapping of NAT66 allows for direct inbound connections to internal nodes, something not present in NAT44.
In addition, an implementation of NAT66 with addressing bit compression must be aware of the potential for a DOS attack against the device by an internal node intentionally sending packets that do not match the compressed bit pattern. This would force the generation of ICMP unreachable messages to the source address, which could be spoofed, and could potentially impact the operation of both the NAT66 device and the legitimate user of the spoofed source address. For this reason, it is recommended that the rate of ICMP unreachable messages generated is limited.
Another possibility of resource exhaustion is to receive packets on an interface that doesn't have any mapping for the source or destination. For example, traffic with an internal destination coming on the outside interface. Such traffic SHOULD be silently dropped.

---

7. Acknowledgements

Thanks to the following people (in alphabetical order) for their participation and review:

Victor Kuarsingh, Rogers Communications

Yiu Lee, Comcast

---

8. Informative References

| [NAT66] | Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Address Translation (NAT66)," November 2008. |

Appendix A. An Appendix

Authors' Addresses

	Jean-Francois Tremblay
	Videotron
	150 Beaubien Ouest
	Montreal, Quebec H2V 1C4
	Canada
Email:	jf@jftremblay.com
	Scott Beuker
	Shaw Communications
	200 - 3636 23rd St NE
	Calgary, Alberta T2E 8Z5
	Canada
Email:	Scott.Beuker@sjrb.ca