C. Bao, Ed.
X. Li
CERNET Center/Tsinghua University
October 24, 2011

Extended IPv6 Addressing for Encoding Port Range
draft-bcx-behave-address-fmt-extension-00

Abstract

This document discusses an extension of the algorithmic translation between IPv4 and IPv4-translatable IPv6 addresses. The extended address format contains transport-layer port range information which allows several IPv6 nodes to share a single IPv4 address with each node managing a different range of ports. This address format extension can be used for IPv4/IPv6 translation, as well as IPv4 over IPv6 tunneling.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

This document discusses an extension of the address format defined in [I-D.ietf-behave-address-format]. The original address format document specifies how an individual IPv6 address is translated to a corresponding IPv4 address, and vice versa, in cases where an algorithmic mapping is used. To face the IPv4 public address exhaustion, it is desirable to assign fractional IPv4 addresses to IPv6 nodes which can share a single IPv4 address with each node managing a different range of ports.

In [I-D.ietf-behave-address-format] Section 3.5, it states:

> "There have been proposals to complement stateless translation with a port-range feature. Instead of mapping an IPv4 address to exactly one IPv6 prefix, the options would allow several IPv6 nodes to share an IPv4 address, with each node managing a different range of ports. If a port range extension is needed, it could be defined later, using bits currently reserved as null in the suffix."

This document defines one of such a suffix encoding scheme and the corresponding port-range algorithm.

1.1. Applicability Scope

The address format extension can be used for both IPv4/IPv6 translation and IPv4 over IPv6 tunneling. However, in this document we will use IPv4-translatable addresses in the stateless translation to discuss this specific address format extension. The descriptions of the other algorithms for their specific use case can be defined later.

1.2. Conventions

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].

2. Extended IPv4-tranlatable IPv6 Address

In Section 2.2 of [I-D.ietf-behave-address-format], IPv4-embedded IPv6 address format is defined which composed of a variable length prefix, the embedded IPv4 address, and a variable length suffix, as presented in the following diagram, in which PL designates the prefix length:

```
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    |PL| 0-------------32--40--48--56--64--72--80--88--96--104-112-120-|
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    |32|     prefix    |v4(32)         | u | suffix                    |
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    |40|     prefix        |v4(24)     | u |(8)| suffix                |
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    |48|     prefix            |v4(16) | u | (16)  | suffix            |
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    |56|     prefix                |(8)| u |  v4(24)   | suffix        |
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    |64|     prefix                    | u |   v4(32)      | suffix    |
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
    |96|     prefix                                    |    v4(32)     |
    +--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

Figure 1: Address Format

It is clear that only Prefix lengths (PL) with 32, 40, 48, 56 and 64 have suffix portions, with maximum suffix lengths of 56, 48, 40, 32, 24, respectively. In order to use different Prefix length and unify the port range encoding method, the suffix should be 24 bits or less. Furthermore, we choose the port-range coding suffix which is directly following the embedded IPv4 address and padding zeros after the suffix.

2.1. Port-range Coding Algorithm

The address-sharing scheme is shown in the following figure.

```
                                   ---------------------------------
                             -----|IPv4-translatable-0|port-range-0 |
                           /       ---------------------------------
                          /        ---------------------------------
                         |--------|IPv4-translatable-1|port-range-1 |
                         |         ---------------------------------
     --------------------|         ---------------------------------
    | IPv4-addr|any ports|-----------|IPv4-translatable-2|port-range-2 |
     --------------------|         ---------------------------------
                         |         ---------------------------------
                         |--------|IPv4-translatable-3|port-range-3 |
                         |         ---------------------------------
                          \                       ...
                           \       ---------------------------------
                             -----|IPv4-translatable-K|port-range-K |
                                   ---------------------------------
                                                  ...
```

Figure 2: Address Sharing Scheme

In the above figure, the IPv4-translatable-0, IPv4-translatable-1, IPv4-translatable-2, ..., IPv4-translatable-K are sharing the same IPv4 address IPv4-addr, but port number range for different IPv4-translatable addresses (i.e. port-range-0, port-range-1, port-range-2 , port-range-K) are not overlapped. When an IPv6 node using IPv4-translatable addresses communicates with the IPv4 Internet via a translator, it looks like a single host using IPv4 address (IPv4-addr) communicating with the IPv4 Internet.

There exist many port-range coding schemes and each one may have its advantages and disadvantages, as well as has its best application scenario. In this document, we will introduce a simple one, which we believe is suitable for the IPv4/IPv6 stateless translation. This encoding scheme uses the modulus operator to define the port number range.

If the sharing ratio is N, then:

- o Given K (K=0, 1, ..., N-1), the allowed port number (P) are P=j*N + K-1, where j=0, 1, ..., (65536-N)/N.

- o Given P, the IPv6 node index (K) is K=(P%N) (% is the Modulus Operator).

For example, If N=128, then IPv6 node K=5 is only allowed to use port numbers 5, 133, 261, 389, 517, 645, 773, 901, ... 65,413 as the source port, while the packets with these port numbers as the destination port number will be send to IPv6 node K=5.

The modulus operator has several features:

1. The port ranges are not overlapped for different IPv6 nodes.

2. The individual ports for each IPv6 node are not continues and the whole 65536 port range is equally shared by IPv6 nodes.

3. The port number range can be uniquely determined by given sharing ratio (N) and the IPv6 node index (K).

Based on the modulus operator, We need to encode N and K in the suffix as an extended IPv4-translatable IPv6 address.

2.2. Extended Address Format

Since the transport port number is a 16 bit integer, the sharing ratio (N) and the IPv6 node index (K) can both have the value from 0

to 65535. In theory, 32 bits (16 bits for sharing ratio and 16 bits for IPv6 node index) are required for encoding the port range based on the modulus operation. In order to fit into 24 bit or less suffix range, we need to do compression.

First, we can use number of bits to represent the sharing ratio when the sharing ratio is bit-wise, hence 4 bits is enough for N.

Secondly, if sharing ratio N is very high, each IPv6 node can only use a small number of concurrent sessions. For example, if N=4096, each IPv6 node will have 16 concurrent sessions, which may be too small for most of the applications. Therefore, if we set the maximum sharing ratio N=4096, then 12 bits are enough for the IPv6 node index. In this case, we can design suffix which consists of 16 bits for encoding the port range.

Based on the above discussion, the extended address format is shown in the following figure.

```
+--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|PL| 0-------------32--40--48--56--64--72--80--88--96--104-112-120-|
+--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|32|     prefix    |v4(32)         | u | suffix|      zero         |
+--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|40|     prefix        |v4(24)     | u |(8)| suffix|     zero      |
+--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|48|     prefix            |v4(16) | u | (16)  | suffix|    zero   |
+--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|56|     prefix                |(8)| u |  v4(24)   | suffix| zero  |
+--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|64|     prefix                    | u |   v4(32)      | suffix|zer|
+--+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

Figure 3: Extended Address Format

Since different suffixes are more specifics in the original address format defined in [I-D.ietf-behave-address-format], the routing considerations in that document are also applied here. Furthermore, the port range is embedded in the extended IPv4-translatable IPv6 addresses and bound to the IPv6 node index (K), therefore the packets containing extended IPv4-translatable IPv6 addresses as the destination can be routed to different IPv6 nodes.

2.3. Suffix for Port Range Encoding

The most significant 4 bits define the multiplexing ratio and the least significant 12 bits define the IPv6 node index. The multiplexing ratio, the suffix range and the number of corresponding

concurrent ports are as shown in the following figure.

ratio	suffix range	# of Ports
1	0000 - 0000	65,536
2	1000 - 1001	32,768
4	2000 - 2003	16,384
8	3000 - 3007	8,192
16	4000 - 400f	4,096
32	5000 - 501f	2,048
64	6000 - 603f	1,024
128	7000 - 707f	512
256	8000 - 80ff	256
512	9000 - 91ff	128
1,024	a000 - a3ff	64
2,048	b000 - b7ff	32
4,096	c000 - cfff	16

Figure 4: Suffix for Port Range Encoding

2.4. Stateless Suffix Translation Algorithm

For the stateless translation, the IPv6 nodes are required to follow the port number range defined by the extended IPv4- translatable address format when communicating with the IPv4 Internet. The port number handling algorithm is:

- o If the packets are from IPv4 to IPv6, the IPv4 source addresses are translated to the IPv4-converted addresses and the source port numbers are unchanged before and after translation; the IPv4 destination addresses are translated to the extended IPv4-translatable addresses based on the destination port number and the destination port numbers are unchanged before and after translation. Note that this means that only a specific IPv6 node can receive the packets for a specific port number. When it is port 80, that specific IPv6 node can setup http redirect service for other IPv6 nodes which also provide web services with non-standard port numbers (e.g. 81, 82, etc.).

- o If the packets are from IPv6 to IPv4, the IPv6 source addresses and the source port numbers are checked, if the source port number matches the port number range defined by the extended IPv4-translatable address format, the IPv6 source addresses (which are the IPv4-translatable addresses) are translated to the IPv4 addresses and the source port numbers are unchanged before and after translation; the destination IPv6 addresses (which are the IPv4-converted addresses) are translated to the IPv4 destination

addresses and the destination port numbers are unchanged before and after translation. However, if the source port number does not match the port number range defined by the extended IPv4-translatable address format, the packets will be dropped.

2.5. Partial-state Suffix Translation Algorithm

Stateless translation requires that IPv6 nodes generate source port number in the range defined by the extended IPv4-translatable address. If this condition does not hold, the partial-state suffix translation algorithm can be used.

The reason we call this partial-state is that:

- o The address mapping is fully algorithm based, as defined in section 3.4. The states are used for port number mapping only.
- o There will be no port mapping table created if the the source port number from IPv6 to IPv4 is in the range defined by the extended IPv4-translatable address.
- o For the destination port number of the packet from the IPv4 to IPv6, there will be no port mapping table created.

The partial-state suffix translation algorithm can be defined later.

2.6. ICMP Packet Handling

The ICMP errors should be translatable using the same algorithm (that is, an error such as Destination Unreachable includes the original TCP (or UDP) header in the ICMP payload, and the that TCP (or UDP) port number can be used to translate the ICMP packet into an IPv6-encoded packet and back again.

Since ICMP echo-request/echo-reply packets only contain identification field, not the transport port numbers, similar to NAT, special actions can be taken for translating the ICMP echo-request/echo-reply packets, which can be defined later.

3. IANA Considerations

This memo adds no new IANA considerations.

4. Security Considerations

There is no special security consideration.

5. Acknowledgements

6. References

6.1. Normative References

[I-D.ietf-behave-address-format]
Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", draft-ietf-behave-address-format-10 (work in progress), August 2010.

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

6.2. Informative References

[I-D.ietf-behave-v6v4-framework]
Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", draft-ietf-behave-v6v4-framework-10 (work in progress), August 2010.

Authors' Addresses

Congxiao Bao (editor)
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing, 100084
China

Phone: +86 10-62785983
Email: congxiao@cernet.edu.cn

Xing Li
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing, 100084
China

Phone: +86 10-62785983
Email: xing@cernet.edu.cn