Network Working Group	C. Huitema
Internet-Draft	Microsoft Corporation
Obsoletes: 2765 (if approved)	C. Bao
Intended status: Standards Track	CERNET Center/Tsinghua University
Expires: April 29, 2010	M. Bagnulo
	UC3M
	M. Boucadair
	France Telecom
	X. Li
	CERNET Center/Tsinghua University
	October 26, 2009

TOC

IPv6 Addressing of IPv4/IPv6 Translators
draft-ietf-behave-address-format-01.txt

Status of this Memo

Copyright Notice

Abstract

This document discusses how an individual IPv6 address can be algorithmically translated to a corresponding IPv4 address, and vice versa, using only statically configured information. This technique is used in IPv4/IPv6 translators, as well as other types of proxies and gateways (e.g., for DNS) used in IPv4/IPv6 scenarios.

Table of Contents

1. Introduction

TOC

This document is part of a series of IPv4/IPv6 translation documents. A framework for IPv4/IPv6 translation is discussed in [I-D.ietf-behave-v6v4-framework] (Baker, F., Li, X., Bao, C., and K. Yin, “Framework for IPv4/IPv6 Translation,” October 2009.), including a taxonomy of scenarios that will be used in this document. Other documents specify the behavior of various types of translators and gateways, including mechanisms for translating between IP headers and other types of messages that include IP addresses. This 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. While specific types of devices are used herein as examples, it is the responsibility of the specification of such devices to reference this document for algorithmic mapping of the addresses themselves.

Section 2 of this document describes the format of "IPv4 Embedded IPv6 addresses", i.e. IPv6 addresses in which 32 bits contains an IPv4 address. These addresses can be used to represent IPv4 hosts to hosts in an IPv6 network. IPv6 addresses assigned to IPv6 hosts for use with stateless translation are referred to as "IPv4-translatable" IPv6 addresses; they are a variant of embedded addresses, and follow the format described in Section 2.

Section 3 discusses the choice of prefixes, the use of a well known prefix, and the use of embedded addresses with stateless and stateful translation.

1.1. Applicability Scope

TOC

This document is part of a series defining address translation services. We understand that the address format could also be used by other interconnection methods between IPv6 and IPv4, e.g. methods based on encapsulation. If the WG so decides, a future version of this document could also discuss the use of embedded addresses and prefixes for interconnection of IPv6 and IPv4 based on encapsulation.

1.2. Notations

TOC

In this document, an "IPv4/IPv6 translator" is an entity that translates IPv4 packets to IPv6 packets, and vice versa. It may do "stateless" translation, meaning that there is no per-flow state required, or "stateful" translation where per-flow state is created when the first packet in a flow is received.

In this document, an "address translator" is any entity that has to derive an IPv4 address from an IPv6 address or vice versa. This applies not only to devices that do IPv4/IPv6 packet translation, but also to other entities that manipulate addresses, such as name resolution proxies (e.g., DNS64 [I-D.bagnulo-behave-dns64] (Bagnulo, M., Sullivan, A., Matthews, P., Beijnum, I., and M. Endo, “DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers,” March 2009.)) and possibly other types of Application Layer Gateways (ALGs).

In this document, the "Well Known Prefix" is an IPv6 prefix assigned by IANA for use in an algorithmic mapping. Options for the actual allocation of the Well Known Prefix are discussed in Section 3.6.
In this document, a "Network Specific Prefix" is an IPv6 prefix assigned by an organization for use in algorithmic mapping. Options for the Network Specific Prefix are discussed in Section 3.3 and 3.4.

2. IPv4 Embedded IPv6 Address Format

TOC

IPv4 Embedded IPv6 Addresses are composed of a variable length prefix, the embedded IPv4 address, and a variable length suffix, as presented in the following diagram:

```
         +----+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
         |PLEN| 0-------------32--40--48--56--64--72--80--88--96--104-112-120-127-|
         +----+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
         |/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)      |
         +----+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

In these addresses, the prefix shall be either the "Well Known Prefix" defined in the addressing architecture to represent IPv4 mapped addresses, or a "Network Specific Prefix" unique to the organization deploying the address translators. (Options for the well known prefix are discussed in Section 3.6.)
Various deployments justify different prefix lengths. The tradeoff between different prefix lengths are discussed in Sections 3.3 and 3.4 of this document.
Bits 64 to 71 of the address are reserved for compatibility with the host identifier format defined in the IPv6 addressing architecture. These bits MUST be set to zero. The corresponding octet is noted "u" in the above diagram. When using a 96 prefix, the administrators MUST ensure that the bits 64 to 71 are compatible with the IPv6 addressing architecture.

The IPv4 address is encoded following the prefix, most significant bits first. Depending of the prefix length, the 4 octets of the address may be separated by the reserved octet "u". In particular:

- When the prefix is 32 bit long, the IPv4 address is encoded in positions 32 to 63.

- When the prefix is 40 bit long, 24 bits of the IPv4 address are encoded in positions 40 to 63, with the remaining 8 bits in position 72 to 79.

- When the prefix is 48 bit long, 16 bits of the IPv4 address are encoded in positions 48 to 63, with the remaining 8 bits in position 72 to 87.

- When the prefix is 56 bit long, 8 bits of the IPv4 address are encoded in positions 56 to 63, with the remaining 8 bits in position 72 to 95.

- When the prefix is 64 bit long, the IPv4 address is encoded in positions 72 to 103.

- When the prefix is 96 bit long, the IPv4 address is encoded in positions 96 to 127.

There are no remaining bits, and thus no suffix, if the prefix is 96 bit long. In the other cases, the remaining bits of the address constitute the suffix. These bits are reserved for future extensions, and should be set to a null value. (Different options for the suffix as discussed in Section 3.5.)

3. Deployment Guidelines and Choices

TOC

3.1. Deployment Using the Well Known Prefix

TOC

The Well Known Prefix MAY be used by organizations deploying translation services.
The Well Known Prefix SHOULD NOT be used to construct IPv4 translatable addresses. The host served by IPv4 translatable addresses should be able to receive IPv6 traffic bound to their IPv4 translatable address without incurring intermediate protocol translation. This is only possible if the specific prefix used to build the translatable addresses is advertized in inter-domain routing, and this kind of

specific prefix advertisement is not supported with the Well Known Prefix, as explained in Section 3.2.

The Well Known Prefix MUST NOT be used to represent non global IPv4 addresses, such as those defined in RFC 1918. Doing so would introduce ambiguous IPv6 address.

3.2. Impact on Inter-Domain Routing

TOC

The Well Known Prefix MAY appear in inter-domain routing tables, if service providers decide to provide IPv6-IPv4 interconnection services to peers. Advertisement of the Well Known Prefix SHOULD be controlled either by upstream and/or downstream service providers owing to inter-domain routing policies, e.g., through configuration of BGP. Organizations that advertize the Well Known Prefix in inter-domain routing MUST be able to provide address translation service.

When the translation relies on the Well Known Prefix, IPv4-mapped IPv6 prefixes longer than the Well Known Prefix MUST NOT be advertised in BGP (especially e-BGP) [rfc4271] because this imports IPv4 routing table into IPv6 one and therefore induces scalability issues to the global IPv6 routing table. Adjacent BGP speakers MUST ignore advertisements of IPv4-mapped IPv6 prefixes longer than the Well Known Prefix. BGP speakers SHOULD be able to be configured with the default Well Known Prefix.

When the translation service relies on Network Specific Prefixes, the global IPv6 routing policy guideline MUST be followed. In particular, if stateless translation is used, the IPv4-translatable addresses MUST be advertised with proper aggregation to the IPv6 Internet. Similarly, if translators are configured with multiple Network Specific Prefixes, these prefixes MUST be advertised to the IPv6 Internet with proper aggregation.

3.3. Choice of Prefix for Stateless Translation Deployments

TOC

Organization may deploy translation services using stateless translation. In these deployments, internal IPv6 hosts are addressed using "IPv4 translatable" IPv6 addresses, which enable them to be accessed by IPv4 hosts. The addresses of these external hosts are represented in "IPv4 Embedded" IPv6 addresses.

Organizations deploying stateless translation SHOULD assign a Network Specific Prefix to their translation service. Both "IPv4 translatable" and "IPv4 Embedded" MUST be constructed as specified in section 2. "IPv4 translatable" addresses MUST use the selected Network Specific Prefix. Both types of addresses SHOULD use the same prefix. Using the same prefix ensures that internal IPv6 hosts will use the most efficient paths to reach the hosts served by "IPv4 translatable" addresses.

The intra-domain routing protocol must be able to deliver packets to the hosts served by "IPv4 translatable" addresses. This may require routing on some or all of the embedded IPv4 address bits. Security considerations detailed in the security section requires that routers check the validity of the "IPv4 translatable" source addresses, using some form of reverse path check.

Forwarding, and reverse path checks, should be performed on the combination of the "prefix" and the IPv4 address. In theory, routers should be able to route on prefixes of any length. However, there is some suspicion that routing on prefixes larger than 64 bit may be slower, or possibly not supported by some router. But routing efficiency is not the only consideration in the choice of a prefix length. Organization also need to consider the availability of prefixes, and the potential impact of null identifiers.

If a /32 prefix is used, all the routing bits are contained in the top 64 bits of the IPv6 address, leading to excellent routing properties. These prefixes may however be hard to obtain, and allocation of a /32 to a small set of IPv4 translatable addresses may be seen as wasteful. In addition, the /32 prefix and a null suffix leads to a null identifier, an issue that we discuss in section 3.5.

Intermediate prefixes like /40, /48 or /56 appear as compromise. Only some of the IPv4 bits are part of the /64 addresses. Reverse checks, in particular, may have a limited efficiency. Reverse checks limited the most significant bits of the IPv4 address will reduce the possibility of spoofing external address, but would allow internal hosts to spoof internal addresses.

We propose here a compromise, based on using no more than 1/256th of an organization's allocation of IPv6 addresses for the translation service. For example, if the organization is an ISP, with an allocated prefix /32 or shorter, the ISP could dedicate a /40 prefix to the translation service. An end site with a /48 allocation could dedicate a /56 prefix to the translation service.

The recommended prefix length is also a function of the deployment scenario. The stateless translation can be used for Scenario 1, Scenario 2, Scenario 5 and Scenario 6 defined in [I-D.ietf-behave-v6v4-framework] (Baker, F., Li, X., Bao, C., and K. Yin, “Framework for IPv4/IPv6 Translation,” October 2009.). For different scenarios, the prefix length recommendations are:

*For scenario 1 (an IPv6 network to the IPv4 Internet) and scenario 2 (the IPv4 Internet to an IPv6 network), we recommend using a /40 prefix for an ISP holding a /32 allocation, and a /56 prefix for a site holding a /40 allocation.

*For scenario 5 (an IPv6 network to an IPv4 network) and scenario 6 (an IPv4 network to an IPv6 network), we recommend using a /64 prefix.

3.4. Choice of Prefix for Stateful Translation Deployments

TOC

Organizations MAY deploy translation services based on stateful translation technology. The organizations may decide to use either a Network Specific Prefix or the Well Known Prefix. The Well Known Prefix SHOULD be used when no Network Specific Prefix is available.

When these services are used, internal hosts are addressed through standard IPv6 addresses, while IPv4 hosts are represented by IPv4 embedded addresses, as specified in section 2.

The stateful nature of the translation creates potential stability issue when the organization deploys multiple translators. If several translators use the same prefix, there is a risk that packet belonging to the same connection may be routed to different translators as the internal routing state changes. This issue can be mitigated either by assigning different prefixes to different translators, or by ensuring that all translators using same prefix coordinate their state.

The stateful translation can be used in the scenarios defined in [I-D.ietf-behave-v6v4-framework] (Baker, F., Li, X., Bao, C., and K. Yin, “Framework for IPv4/IPv6 Translation,” October 2009.). The general recommendation is to use the Well Known Prefix, with two exceptions:

- *In all scenarios, the translation MAY use a Network Specific Prefix, if deemed appropriate for management reasons.

- *The Well Known Prefix MUST NOT be used for scenario 3 (the IPv6 Internet to an IPv4 network), as this would lead to using the Well Known Prefix with non global IPv4 addresses. That means a Network Specific Prefix MUST be used in that scenario.

3.5. Choice of Suffix

TOC

The address format described in Section 2 recommends a null suffix. Before making this recommendation, we considered different options: checksum neutrality; the encoding of a port range; and a value different than 0.

The "neutrality checksum" option would give a chosen value to 16 of the suffix bits to ensure that the "IPv4 embedded" IPv6 address has the same 16 bit complement to 1 checksum as the embedded IPv4 address. There have been discussion of this checksum in the working group mailing list, and some push to standardize a checksum format. However, we observed that the neutrality checksum alone does eliminate checksums computation during stateful translation, as only one of the two addresses would be checksum neutral. In the case of stateless translation, translators may want to recomputed the checksum anyhow, to verify the validity of the translated datagrams. In doubt, we picked the simplest alternative, to not specify a neutrality checksum.

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 hosts to share an IPv4 address, with each host managing a different range of ports. But these schemes are not yet specified in work group documents. If a port range extension is needed, it could be defined later, using bits currently reserved as null in the suffix.

When a /32 prefix is used, the null suffix results in a null identifier. We understand the conflict with Section 2.6.1 of RFC4291, which specifies that all zeroes are used for the subnet-router anycast address. However, in our specification, there would be only one IPv4 translatable host in the /64 subnet, and the anycast semantic would not create confusion. We thus decided to keep the null suffix for now. (Different authors of this document have different opinions.)

3.6. Choice of the Well Known Prefix

TOC

We are faced with three choices for the Well Known Prefix:

- *reuse the IPv4-mapped prefix, ::FFFF:0:0/96, as specified in RFC 2765 Section 2.1;
- *request allocation of a new /96 prefix;
- *or request IANA to allocate a /32 prefix.

Each of these choices has pros and cons. We expect this issue to be debated and resolved by the BEHAVE working group, and present here our analysis of the options.

The main advantage of the existing IPv4-mapped prefix is that it is already defined. Reusing that prefix will require minimal standardization efforts. However, being already defined is not just and advantage, as there may be side effects of current implementations. When presented with the IPv4-mapped prefix, several versions of Windows and MAcOS may generate IPv4 packets, but will not send IPv6 packets. If we used the IPv4-mapped prefix, these hosts would not be able to support translation without modification. This will defeat the main purpose of the translation techniques.

Allocating a new prefix would diminish the risk of undesirable side effects in current implementations. The main cost will be the registration cost with IANA. We will also need to update the recommendation for textual representations of IPv6 addresses, if we want to ensure the dotted decimal representation of the IPv4 component in the IPv4 embedded IPv6 addresses.

If we allocate a new prefix, choosing a /32 prefix would allow the embedded IPv4 address to fit within the top 64 bits of the IPv6 address. This would facilitate routing and load balancing when an organization deploys several translators. However, such destination-address based load balancing may not be desirable, as it is not compatible with STUN in the deployments involving multiple stateful translators, each one having a different pool of IPv4 addresses. STUN compatibility would only be achieved if the translators managed the same pool of IPv4 addresses and were able to coordinate their translation state.

We should also note that according to Section 2.2 of RFC 4291, in the legal textual representations of IPv6 addresses, dotted decimal can only appear at the end. We could simply forego the dotted decimal notation, but that would make the address format harder to use, and log files harder to read. We could also update RFC4291 to allow textual representation of addresses using the assigned WKP and having the interface identifier set to all zeros. We could also embed the IPv4 address both in the last 32 bits of the interface identifier and the last 32 bits of the 64 bit prefix, allowing to use the textual representation as defined in RFC4291 and also have the possibility of including the IPv4 address in the prefix part. Moreover, we could request for IANA to assign a /32 for the WKP and then operators could simply decide whether to use it as a /32 or pad it with zeros and use it as a /96.

Allocating a new /96 prefix would not enable the same routing and load balancing options as a /32 prefix, but would allow for decimal notation of IPv4 addresses without requiring an update to RFC 4291.

4. Security Considerations

TOC

4.1. Protection Against Spoofing

TOC

By and large, address translators can be modeled as special routers, are subject to the same risks, and can implement the same mitigation. There is however a particular risk that directly derived from the practice of embedding IPv4 addresses in IPv6: address spoofing.

An attacker could use an IPv4 embedded address as the source address of malicious packets. After translation, the packets will appear as IPv4 packets from the specified source, and the attacker may be hard to track. If left without mitigation, the attack would allow malicious IPv6 nodes to spoof arbitrary IPv4 addresses.

The mitigation is to implement reverse path checks, and to verify throughout the network that packets are coming from an authorized location.

4.2. Secure Configuration

TOC

The prefix and format need to be the same among multiple devices in the same network (e.g., hosts that need to prefer native over translated addresses, DNS gateways, and IPv4/IPv6 translators). As such, the means by which they are learned/configured must be secure. Specifying a default prefix and/or format in implementations provides one way to configure them securely. Any alternative means of configuration is responsible for specifying how to do so securely.

5. IANA Considerations

TOC

A future version of this memo will request an IPv6 prefix assignment as a Well-Known Mapped Prefix, that is used to represent IPv4 hosts, and which must start with binary 000.
[EDITOR'S NOTE: 0/8 is reserved by the IETF (and not allocated by IANA), so all that is needed is to specify the prefix herein since it is an allocation from IETF not from IANA.]
OPEN ISSUE: The prefix length of this block has not yet been determined. Some possibilities are /16, /32, /48 or /96.

6. Acknowledgements

TOC

Many people in the Behave WG have contributed to the discussion that led to this document, including Andrew Sullivan, Andrew Yourtchenko, Brian Carpenter, Congxiao Bao, Dan Wing, Ed Jankiewicz, Fred Baker, Hiroshi Miyata, Iljitsch van Beijnum, John Schnizlein, Keith Moore, Kevin Yin, Magnus Westerlund, Marcelo Bagnulo Braun, Margaret Wasserman, Masahito Endo, Phil Roberts, Philip Matthews, Remi Denis-Courmont, Remi Despres, William Waites and Xing Li.

7. Contributors

TOC

The following individuals co-authored drafts from which text has been incorporated, and are listed in alphabetical order.

Dave Thaler
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
USA

Phone: +1 425 703 8835
Email: dthaler@microsoft.com

Congxiao Bao
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing,   100084
China
Phone: +86 62785983
Email: congxiao@cernet.edu.cn

Fred Baker
Cisco Systems
Santa Barbara, California 93117
USA
Phone: +1-408-526-4257
Fax: +1-413-473-2403
Email: fred@cisco.com

Hiroshi Miyata
Yokogawa Electric Corporation
2-9-32 Nakacho
Musashino-shi, Tokyo 180-8750
JAPAN
Email: h.miyata@jp.yokogawa.com

Marcelo Bagnulo
Universidad Carlos III de Madrid
Av. Universidad 30
Leganes, Madrid 28911
ESPANA
Email: marcelo@it.uc3m.es

Xing Li
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing,   100084
China
Phone: +86 62785983
Email: xing@cernet.edu.cn

8. References

TOC

8.1. Normative References

TOC

[RFC2026]	Bradner, S., “The Internet Standards Process -- Revision 3,” BCP 9, RFC 2026, October 1996 (TXT).
[RFC4291]	Hinden, R. and S. Deering, “IP Version 6 Addressing Architecture,” RFC 4291, February 2006 (TXT).

8.2. Informative References

TOC

[I-D.bagnulo-behave-dns64]	Bagnulo, M., Sullivan, A., Matthews, P., Beijnum, I., and M. Endo, “DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers,” draft-bagnulo-behave-dns64-02 (work in progress), March 2009 (TXT).
[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-03 (work in progress), October 2009 (TXT).
[I-D.wing-behave-nat64-referrals]	Wing, D., “Referrals Across an IPv6/IPv4 Translator,” draft-wing-behave-nat64-referrals-01 (work in progress), October 2009 (TXT).
[RFC1918]	Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, “Address Allocation for Private Internets,” BCP 5, RFC 1918, February 1996 (TXT).
[RFC2765]	Nordmark, E., “Stateless IP/ICMP Translation Algorithm (SIIT),” RFC 2765, February 2000 (TXT).
[RFC2766]	Tsirtsis, G. and P. Srisuresh, “Network Address Translation - Protocol Translation (NAT-PT),” RFC 2766, February 2000 (TXT).
[RFC3484]	Draves, R., “Default Address Selection for Internet Protocol version 6 (IPv6),” RFC 3484, February 2003 (TXT).
[RFC3493]	Gilligan, R., Thomson, S., Bound, J., McCann, J., and W. Stevens, “Basic Socket Interface Extensions for IPv6,” RFC 3493, February 2003 (TXT).
[RFC4271]	Rekhter, Y., Li, T., and S. Hares, “A Border Gateway Protocol 4 (BGP-4),” RFC 4271, January 2006 (TXT).
[RFC4380]	

	Huitema, C., “Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs),” RFC 4380, February 2006 (TXT).
[RFC4862]	Thomson, S., Narten, T., and T. Jinmei, “IPv6 Stateless Address Autoconfiguration,” RFC 4862, September 2007 (TXT).
[RFC5214]	Templin, F., Gleeson, T., and D. Thaler, “Intra-Site Automatic Tunnel Addressing Protocol (ISATAP),” RFC 5214, March 2008 (TXT).
[RFC5389]	Rosenberg, J., Mahy, R., Matthews, P., and D. Wing, “Session Traversal Utilities for NAT (STUN),” RFC 5389, October 2008 (TXT).

Authors' Addresses

TOC

Christian Huitema
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
U.S.A.
Email: huitema@microsoft.com

Congxiao Bao
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing, 100084
China
Phone: +86 10-62785983
Email: congxiao@cernet.edu.cn

Marcelo Bagnulo
UC3M
Av. Universidad 30
Leganes, Madrid 28911
Spain
Phone: +34-91-6249500
Fax:
Email: marcelo@it.uc3m.es
URI: http://www.it.uc3m.es/marcelo

Mohamed Boucadair
France Telecom
3, Av Francois Chateaux
Rennes 350000
France
Email: mohamed.boucadair@orange-ftgroup.com

	Xing Li
	CERNET Center/Tsinghua University
	Room 225, Main Building, Tsinghua University
	Beijing, 100084
	China
Phone:	+86 10-62785983
Email:	xing@cernet.edu.cn