12-3-2012

Exposing Potential Privacy Issues with IPv6 Address Construction

Clinton Carpene
Edith Cowan University

Andrew Woodward
Edith Cowan University

DOI: 10.4225/75/57b556f5cd8d7
10th Australian Information Security Management Conference, Novotel Langley Hotel, Perth, Western Australia, 3rd-5th December, 2012

EXPOSING POTENTIAL PRIVACY ISSUES WITH IPV6 ADDRESS CONSTRUCTION

Clinton Carpene and Andrew Woodward
SRI - Security Research Institute, Edith Cowan University, Perth, Western Australia
c.carpene@ecu.edu.au, a.woodward@ecu.edu.au

Abstract

The usage of 128 bit addresses with hexadecimal representation in IPv6 poses significant potential privacy issues. This paper discusses the means of allocating IPv6 addresses, along with the implications each method may have upon privacy in different usage scenarios. The division of address space amongst the global registries in a hierarchal fashion can provide geographical information about the location of an address, and its originating device. Many IPv6 address configuration methods are available, including DHCPv6, SLAAC (with or without privacy extensions), and Manual assignment. These assignment techniques are dissected to expose the identifying characteristics of each technique. It is seen that use of the modified EUI-64 in SLAAC can allow agents to simply decipher an interface's MAC address over layer 3 communications, whilst discernable patterns can be used to identify the presence of DHCPv6 or manual address assignment. Additionally, the frequency and lifetime of unique addresses originating from a single network prefix may allude to privacy addresses in use within the network. Together these issues pose a risk to the privacy of IPv6 users, as it may allow for tracking of users of portable network devices.

Keywords
Internet security, IPv6, privacy, network security.

INTRODUCTION

The Internet Protocol version 6, commonly referred to as IPv6, was proposed to address the exhaustion of the currently used IPv4 address space, among other issues, and is being increasingly implemented as the successor Layer 3 network communications protocol. Since its ratification in 1998, IPv6 adoption rates have slowly increased. Recent 'World IPv6 Day' campaigns have assisted in increasing the exposure of the protocol and number of services with IPv6 offerings (Arkko & Keranen, 2012). Currently, Google reports that approximately 0.75% of its traffic is IPv6 based, with close to 99% of this being native IPv6 traffic (Google, 2012).

In addition to social media which now holds significant private information about its users, there is increasing attention being made to the volumes of internet and World Wide Web usage data. This data is used by organisations such as Google and Yahoo to more specifically target advertising to those browsing the Web, but data can also be mined to determine other information about a person. For example it has been claimed that the data stored by the Target chain of department stores about its shoppers can be used to determine whether a customer is pregnant based on their buying patterns (Golgowski, 2012). Users of twitter who post tweets from devices which have geolocation enabled are unwittingly creating a pattern of data which can be used to track their movements over time (Hannay & Baatard, 2011). There are frequent reports of issues with Facebook privacy settings leading to users having their personal information made available to the wider internet. Reports of these types of privacy issues have raised concerns in the wider community about technology in general, and more specifically whether the exploitation of data made available by users of these services is creating security problems for them.

In addition to the privacy concerns about IPv6, there are also some security issues in relation to its implementation. One benefit of IPv6 is the inclusion of native support for IPSec tunnelling, increasing the security of end to end user communications. Another aspect of the implementation of IPv6 is that it negates the need for network address translation (NAT), as management of IP addresses becomes irrelevant with IPv6. This has been labelled as a security negative, but in reality, should have no impact on security and should make network boundary security easier to configure.

With more services becoming IPv6 enabled, and the eventual transition to the protocol moving closer, security aspects of the protocol become increasingly important. One aspect of protocol security that has garnered a small degree of concern is the makeup of the IPv6 addresses themselves, and there has already been preliminary investigation of this issue (Dunlop *et al.*, 2011a; Dunlop *et al.*, 2011b). Since the address space is so vast, it is possible that addresses could be used to fingerprint users or devices. Using certain characteristics of an IPv6 address, an agent may be able to make informed assumptions pertaining to the address' construction. With services like SLAAC and DHCPv6 making the allocation of IPv6 addresses automated, simplistic and efficient; and a huge address space reducing the requirement for address recycling; addresses themselves may become tools for tracking users. This paper explores the composition of the IPv6 addressing protocol, and examines the implications of its implementation on user privacy.

IPV6 ORIGINS AND CONCEPTS

IPv6 is a protocol operating at layer 3 of the Open Systems Interconnection (OSI) model, and layer 2 of the TCP/IP model. The development of IPv6 commenced in 1992 when members of the Internet Engineering Task Force (IETF) foresaw a future requirement for a larger address space. The protocol, known simply at the time as IPNG (IP Next Generation), was the foundation of what is now commonly referred to as IPv6. The IPv6 protocol was introduced by the IETF in 1998 through RFC 2460 (Deering & Hinden, 1998).

Since the primary constraint of IPv4 is address space, IPv6 has ensured address space availability for the foreseeable future. Many comparisons have been made between IPv4 and IPv6's address space. However, the reality is that there are simply many orders of magnitude more addresses available. Where IPv4 has 32 bits of address space, IPv6 defines 128 bits of address space. On a functional level this distinction means that IPv6 has 2^{128} (i.e. 340, 282, 366, 920, 938, 463, 463, 374, 607, 431, 768, 211, 456 or 340 undecillion) possible unique addresses, compared to IPv4's 2^{32} (i.e. 4, 294, 967, 296, or 4.29 billion) possible unique addresses. To put this into context, IPv6 provides an IP address for every grain of sand on the planet, or 16000 trillion addresses for every human on the planet. With such a large amount of addresses available, the IETF have been able to shift their focus from the conservation of addresses, to simplicity and efficiency of allocation. The result is that, at least for the foreseeable future, each device on the Internet can be provided with a globally unique IPv6 address.

Network address translation (NAT) and port address translation (PAT/NAPT) have increased the longevity of IPv4 by providing a public/private abstraction to Internet access and communication. This system helped to preserve and efficiently distribute the scarce IPv4 resources. The system, however, is no longer required. Although useful for its intentions, NAT introduces many issues when attempting to provide services that require end-to-end connectivity (such as peer-to-peer file services and VoIP). The decreased reliance on NAT/NAPT however introduces an element of privacy that was previously a nonissue: an IP address being used as a device identifier. Previously NAT obscured the activities of a device to a degree, by providing a translation between a publicly routable global IP, and a locally routable private IP. This abstraction means that from a global perspective, the actions of a device are partially concealed, as many devices may exist behind a single public IP address. With IPv6 there are enough address to provide globally routable, unique addresses for every device in the world, at least for the foreseeable future.

Under current specifications, IPv6 addresses can be broken into 3 types including unicast, multicast and anycast addresses. Each scheme has a specific purpose and application for usage. These types each have separate scopes, of which the unicast types includes global unicast addresses, which are the subject of the research. Under the global unicast IPv6 format, an address is divided into three major sections; the global routing prefix portion (highest order); the subnet ID (middle order); and the interface ID portion (lowest order). The address space is divided as per Figure 1.

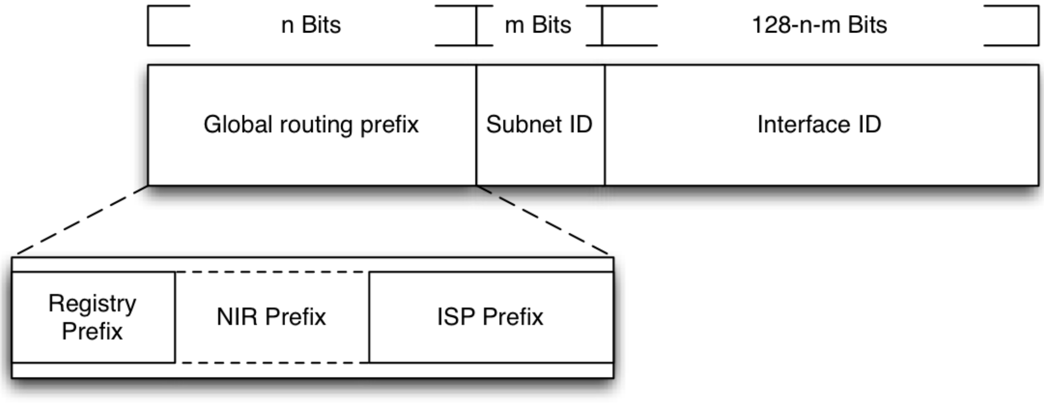

Figure 1 - A representation of the division of global unicast IPv6 addresses, highlighting the hierarchal prefix allocation.

RFC4291 defines that unless the first 3 bits of an address are 000, the interface ID should be 64 bits in length (Hinden & Deering, 2006).

NETWORK PREFIX ALLOCATION

The prefix can be dissected into many possible descriptors; identifying the type of address (e.g. multicast, link-local, node-local or global unicast), region, nation, service provider and subnet (as evidenced in Figure 1).

This dissection is made possible by the strict allocation guidelines enforced by the address space owners: the Internet Assigned Numbers Authority (IANA). As with IPv4, allocation of numbers to Internet registries (IR) is performed from a top-down approach; whereby a superscope is allocated at the highest order, then divided up amongst the lower orders (Figure 2).

The possible Internet registries include (in hierarchal order):

- Regional internet registries (RIR)
- National internet registries (NIR)
- Local internet registries (LIR)

Currently there exists 5 RIRs globally (IANA, 2010):

- AfriNIC - African Network Information Center
- APNIC - Asia/Pacific Network Information Centre
- ARIN - American Registry for Internet Numbers
- LACNIC - Latin America and Caribbean Network Information Centre
- RIPE NCC - Reseaux IP Europèens Network Coordination Centre

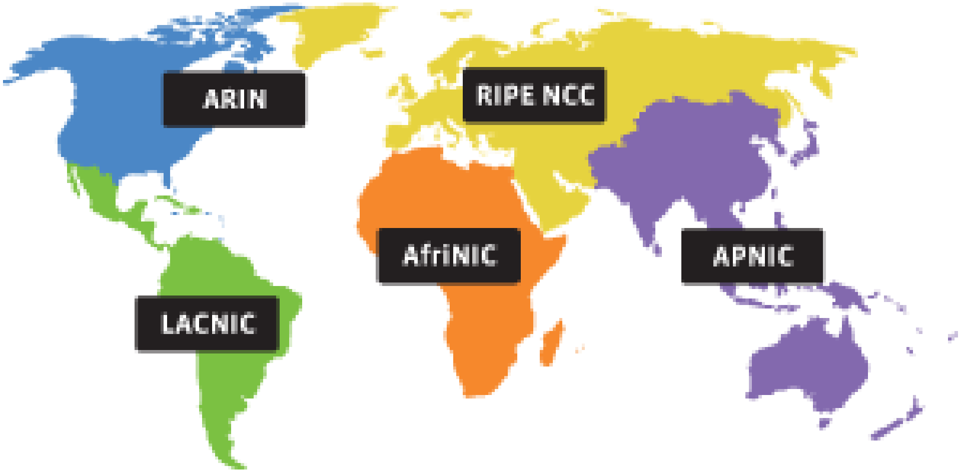

Figure 17 - World map highlighting the geographical coverage of each RIR (Internet Assigned Numbers Authority, n.d.)

The IANA dictates which IPv6 block or range each RIR is allocated (Hinden & Deering, 2006; IANA, 2008), with most RIRs distributing their ranges directly to LIRs, who in turn service end users (and potentially other ISPs). A notable exception to this policy is APNIC whose realm covers the entire Asia/Pacific region. Consequently, APNIC assigns ranges to NIRs for further distribution to LIRs.

The LIRs (which includes Internet Service Providers) then allocate network ranges to customers. The length of range allocations varies depending on the requirements of the region, ISP, and customer. Current RIR policies dictate that customers should receive between a /56 and /48 network prefix (American Registry for Internet Numbers, 2012; The JPNIC IPv6 policy drafting team, 2011), giving each user between 8 and 16 bits of subnet space. This means that at a minimum each customer is able to allocate 256 unique networks of 64 bits. Under some circumstances LIRs may wish to provide customers with a /64 (1 network of 18, 446, 744, 073, 709, 551, 616 addresses) or /128 (1 network of 1 address) subnet. An example of these allocations would be in usage with point-to-point protocol (PPP) connections (which represent many residential grade services).

Under the addressing scheme it is possible to derive many points of information from an IPv6 prefix:

1. The IANA prefix identifies the type of address;
2. The registry prefix denotes which RIR the address belongs to;
3. The NIR prefix (if applicable) can identify which nation the address belongs to;
4. The ISP prefix can identify which Internet service provider the address belongs to;
5. The global routing or site prefix can identify the entity that owns the address;
6. Finally the subnet prefix can identify which network a device is on.

HOST ADDRESS CONFIGURATION

The host portion of the address, known as the Interface Identifier (IID), is determined by one of many methods; stateless address auto configuration (SLAAC), dynamic host configuration protocol version 6 (DHCPv6), or manual assignment. The IID will be arguably the most important aspect of any future research in this area, since it can be used to identify an interface on a device, and therefore a device itself.

Due to IPv6's large address space the IETF introduced methods of creating IPv6 addresses in standard and meaningful ways. Where IPv4's smaller address space resulted in a conservation requirement, IPv6 has been designed to ensure the address allocation is simple, and consistent.

SLAAC

Stateless address auto-configuration (SLAAC) is an address assignment policy introduced into IPv6 by RFC 4862 (Thomson et al., 2007). The policy defines a method for hosts to self-configure a unique IPv6 address, without the requirement for manual intervention or DHCP. To accomplish this, the host is provided with the network prefix portion of the IPv6 address (first 64 bits). The host then derives the Interface Identifier (IID) portion of the address using a method known as modified EUI-64. The static nature of IIDs generated by SLAAC can also lend itself to unwanted device tracking. A security research group operating out of Virginia tech in the USA proved this notion by monitoring network communications over the Virginia Tech University campus' IPv6 network. From the monitored traffic, Dunlop *et al.* were able to track devices moving between network segments by the static nature of their IID (2011a, 2011b). They achieved this by mapping the physical locations of the logical, layer 3 network segments, onto a campus map. They then analysed the movement of a single IID traversing the campus network.

MODIFIED EUI-64 FOR IID GENERATION

Layer 2 Ethernet protocols implement hardware identifiers to enable communication over a physical medium. These identifiers, commonly known as Ethernet addresses or Media Access Control (MAC) addresses, are written into the firmware of the Ethernet network interface controller (NIC). Each NIC possesses a practically unique MAC address by its manufacturer, with blocks of addresses assigned to the manufacturer by the IEEE (Institute of Electrical and Electronics Engineers). A MAC address commonly uses the IEEE MAC-48 standard for formation. IEEE MAC-48 defines a 24-bit Organisationally Unique ID (OUI) and a 24-bit extension identifier, allocated by the IEEE Standards Association, and the assigned company or organisation respectively (as illustrated in Figure 3) (IEEE Standards Association, 2012, n.d.).

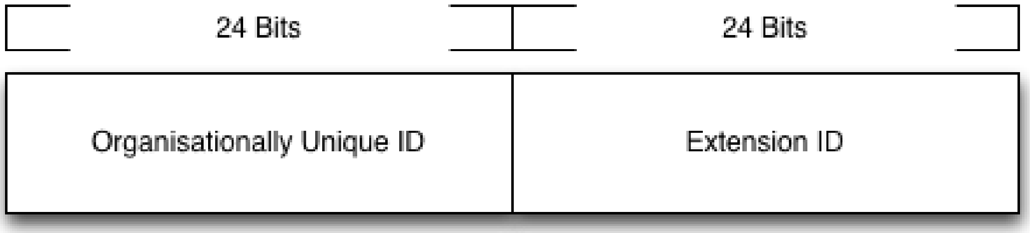

Figure 18 - Construction of IEEE MAC-48 address highlighting OUI and Extension ID division (IEEE Standards Association, 2012).

The OUI can be used to identify what device manufacturer supplied the NIC or even in some cases, the manufacturer of the device itself. Previously, this information would only be ascertainable by inspecting the Ethernet frames passing over a local network (since frames are repackaged at each layer 3 hop, stripping and replacing the source and destination Ethernet addresses from the frames). With the information being used to produce an IID however, an agent could reverse engineer the originating MAC address, and subsequent device/vendor information.

The IEEE defines a set of standard formats for the creation of unique 64 bit identifiers known as EUI-64 (Extended Unique Identifier - 64 bit) (IEEE Standards Association, n.d.). The IETF has adopted a modified version of the IEEE's EUI-64 standard for usage in SLAAC Interface Identifier (IID) generation, which is defined in RFC 4291 (Narten, Draves. & Krishnan, 2007). Henceforth the usage of the term EUI-64 will refer to the modified EUI-64 specification detailed in RFC 4291.

EUI-64 defines a standard approach to converting EUI numbers to IIDs. Many EUIs exist of varying lengths, however the most common are; IEEE MAC-48 which are used commonly as MAC addresses; EUI-48 which are used for identifying some software products; and EUI-64 which are used as Ethernet addresses on some IEEE 802.15.4 devices (IEEE Standards Association, n.d.)

When taken with a standard MAC-48 address, the format requires the IID to be determined by first taking the MAC-48 address of the interface, flipping the universal/local bit and padding 0xfffe between bits 24 and 40. A typical IID in IPv6 will resemble the format above, in base 16 hexadecimal format. An example of how IPv6 uses EUI-64 to generate an IID is as follows:

- 00:aa:12:34:56:fe - Standard MAC-48 address
- 02:aa:12:34:56:fe - Flip the universal/local bit of MAC-48 (7th bit of the highest order byte).
- 02aa:12<-->34:56fe - Split address at 24th bit.
- 02aa:12ff:fe34:56fe - Pad 16 bits using 1's', with the least significant bit set to 0' (fffe) between bits 24 and 40.

The process results in an Interface Identifier, which can be used to simply derive the initial MAC address (by reverse flipping the universal/local bit and then removing the 0xfffe padding).

The process for generating a modified EUI-64 address out of a standard IEEE EUI-64 address is simpler still:

- 00aa:1234:56fe:aa56 - Standard EUI-64 address
- 02aa:1234:56fe:aa56 - Flip the universal/local bit of EUI-64 (7th bit of the highest order byte).

Dunlop *et al*. (2011b) have already determined that an IID constructed using EUI-64 can be reverse engineered to determine the origin MAC address.

PRIVACY EXTENSIONS

Originally the IID of a device was intended to be persistent, to the point of almost being permanent (in a similar vein to a device's MAC address). The engineers, who originally developed SLAAC, and the notion of automatic addressing in IPv6, didn't foresee a reason to change the IID since theoretically there is almost zero chance of duplicate addresses being encountered on a single network segment. Also, IPv6 had provisions to handle address duplication should the situation arise. It soon became realised that linking an identifier to a device's MAC address may not be a desirable situation, since the address can be reconstructed outside of a local Ethernet segment. Additionally, maintaining a persistent IID between network sessions was also considered undesirable. As a result, the privacy extensions field was introduced into the IPv6 header specification in 2007 under the RFC 4941 (Narten, Draves & Krishnan, 2007). This document defines a method of cryptographically generating a temporary IID to prevent persistent traceability on a device. This cryptographic process requires performing an MD5 hash on the IID as well as an incremental number. The temporary IIDs created during the process have a lifespan for the duration of a network session, at which point regeneration of a new IID takes place.

Privacy extensions may be easily discernable amongst network traffic. This distinction is possible given that the addresses will be effectively random hexadecimal strings, lacking any discernable pattern (such as collapsed notation, or the EUI-64 padding). Additionally, given the temporary nature of IIDs with privacy extensions, an agent monitoring communications for extended periods of time could monitor the lifecycle of an IID, in conjunction with any overlaps with unique IIDs originating from a network, to determine the number of unique hosts on the network segment.

At the time of writing, privacy extensions are enabled by default on Microsoft Windows operating systems. Linux and Unix systems (including Mac OSX) currently offer privacy extensions as a configurable option.

DHCPV6

Dynamic Host Configuration Protocol version 6 (DHCPv6) is a complementary service used for the automatic configuration of host computers in an enterprise environment. The technology allows a device to issue addresses to hosts, as well as track address assignments. Similar to DHCPv4 an administrator can configure a DHCPv6 server to supply connected devices with IPv6 addresses whilst they are connected to the network. Part of this requirement is that the administrator must specify valid ranges of addresses that will be allocated to connecting clients.

The interest in DHCPv6 from this research's perspective comes not from the method of IP assignment, but the choice of particular ranges. It is predicted that administrators opting to use DHCPv6 within their network would do so under the expectation that it makes their jobs simpler. It is therefore expected that the ranges of addresses used for allocation will be easily discernible when compared with an address configured with SLAAC. An administrator, for example, may allocate the range of addresses "::0' – '::FF' for distribution in a particular subnet. This range would allow 256 devices to connect to the network with valid IP addresses. Witnessing a number of unique IPv6 addresses originating from a single network prefix, with incremental numbers (such as '::1', '::2', etc.) could indicate the network is using DHCPv6 to assign client addresses.

As an example of this scenario at the time of writing, the search engine provider Google Inc.'s IPv6 webpage's DNS AAAA host record (ipv6.google.com) resolves to the IPv6 address 2404:6800:4006:803::1012. The IID of this address exhibits collapsed notation for the 3 higher order quartets. Such a situation indicates that SLAAC is not being used to configure the IPv6 address, and that DHCPv6 or manual assignment is more probable.

MANUAL ASSIGNMENT

As is possible with IPv4, IPv6 allows users to manually configure their IPv6 address. As long as the interface identifier is unique on the network, the IPv6 address is valid (as the network prefix is allocated by the router and should be globally unique). Similarly to DHCPv6 it is expected that users wishing to manually configure their IPv6 address will likely use a simple representation. The reasons one may opt for manual configuration are varied; a user may have devices on the network for which they wish to maintain static, easily memorable addresses (such as web severs, media servers, file servers, home routers, etc.); a user may not have a local DNS infrastructure, and opt to use simple IPv6 addresses amongst the network to make administration of devices simpler; a user may wish to spoof their IPv6 address in an attempt to bypass IPv6 based system access restrictions (e.g. in the case of a user being banned from an online game, based on their IPv6 address).

Whatever the reason for the manual configuration, it remains a potentially identifiable portion of the address. As mentioned previously, it is predicted that when posed with the option to select addresses for systems, a user will likely choose an address that is simply represented (such as <prefix>::1).

CONCLUSION

The implementation of IPv6 may lead to an increase in user security through the inclusion of IPSec in the protocol, but its main benefit is the large address space making exhaustion a distant, if not irrelevant problem. Unfortunately it also means a step back for user privacy in some respects. Whilst NAT cannot be considered a formidable security solution by any description, it does provide an abstraction layer between network communications and the originating device. This translation certainly increases user privacy since an outside agent cannot distinguish without further investigation what device made a particular request beyond a NAT gateway. Although IPv4 can operate in a NAT-free fashion, the reality is that it is unable to, given the sheer number of devices requiring access to network communications. IPv6 introduces a paradigm shift whereby address space conservation is no longer a valid concern. It is seen, however, that doing so can impose privacy risks to users in many different ways. Due to the range allocation methods, the network prefix is shown to provide some basic geolocational risk, as is present with IPv4. Techniques of host address configuration also provide some characteristics that potentially expose their construction. Future research in this area will attempt to simulate a portion of the global IPv6 address allocation and to determine whether a device can be tracked across boundaries based on tis IP address. It can be concluded that based on the discussion of the protocol in this paper, that the privacy concerns in relation to IPv6 addressing are not trivial.

REFERENCES

American Registry for Internet Numbers. (2012). Number Resource Policy Manual: ARIN.

Arkko, J. & Keranen, A. (2012). Some Measurements on World IPv6 Day from End-User Perspective.

Deering, S. E. & Hinden, R. M. (1998). Internet protocol, version 6 (IPv6) specification: IETF.

Dunlop, M., Groat, S., Marchany, R. & Tront, J. (2011a, 2-5 May 2011). *The Good, the Bad, the IPv6.* Paper presented at the 2011 Ninth Annual Communication Networks and Services Research Conference (CNSR).

Dunlop, M., Groat, S., Marchany, R. & Tront, J. (2011b). *IPv6: Nowhere to Run, Nowhere to Hide*. Paper presented at the 44th Hawaii International Conference on System Sciences (HICSS), 2011, Hawaii.

Golgowski, N. (2012). How Target knows when its shoppers are pregnant - and figured out a teen was before her father did. Retrieved from http://www.dailymail.co.uk/news/article-2102859/How-Target-knows-shoppers-pregnant--figured-teen-father-did.html

Google. (2012). IPv6 Adoption Retrieved from http://www.google.com/ipv6/statistics.html

Hannay, P. & Baatard, G. (2011). GeoIntelligence: Data Mining Locational Social Media Content for Profiling and Information Gathering.

Hinden, R. & Deering, S. (2006). IP Version 6 Addressing Architecture: Network Working Group.

IANA. (2008). IPv6 Global Unicast Address Assignments.

IANA. (2010). IPv6 Address Space registry, 2012, from http://www.iana.org/assignments/ipv6-address-space/ipv6-address-space.xml

IEEE Standards Association. (2012). Registration Authority: ORGANIZATIONALLY UNIQUE IDENTIFIER (OUI) OR 'COMPANY_ID': IEEE Standards Association.

IEEE Standards Association. (n.d.). Guidelines for 64-bit Global Identifier (EUI-64) Registration Authority: IEEE.

Internet Assigned Numbers Authority. (n.d.). rir-map-small: Image.

Narten, T., Draves, R. & Krishnan, S. (2007). Privacy Extensions for Stateless Address Autoconfiguration in IPv6: IETF.

S. Thomson, Narten, T. and Jinmei, T. (2007). IPv6 Stateless Address Autoconfiguration: Network Working Group.

The JPNIC IPv6 policy drafting team. (2011). IPv6 address allocation and assignment policy: APNIC.