Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 13, 2015

A. Cooper
Cisco
F. Gont
Huawei Technologies
D. Thaler
Microsoft

October 10, 2014

Privacy Considerations for IPv6 Address Generation Mechanisms
draft-ietf-6man-ipv6-address-generation-privacy-02.txt

Abstract

This document discusses privacy and security considerations for several IPv6 address generation mechanisms, both standardized and non-standardized. It evaluates how different mechanisms mitigate different threats and the trade-offs that implementors, developers, and users face in choosing different addresses or address generation mechanisms.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 13, 2015.

Copyright Notice

Table of Contents

1. Introduction

IPv6 was designed to improve upon IPv4 in many respects, and mechanisms for address assignment were one such area for improvement. In addition to static address assignment and DHCP, stateless autoconfiguration was developed as a less intensive, fate-shared means of performing address assignment. With stateless autoconfiguration, routers advertise on-link prefixes and hosts generate their own interface identifiers (IIDs) to complete their addresses. Over the years, many interface identifier generation techniques have been defined, both standardized and non-standardized:

- Manual configuration

      * IPv4 address

      * Service port

      * Wordy

      * Low-byte

   o Stateless Address Auto-Cofiguration (SLAAC)

      * IEEE 802 48-bit MAC or IEEE EUI-64 identifier [RFC1972][RFC2464]

      * Cryptographically generated [RFC3972]

      * Temporary (also known as "privacy addresses") [RFC4941]

      * Constant, semantically opaque (also known as random) [Microsoft]

      * Stable, semantically opaque [RFC7217]

   o DHCPv6-based [RFC3315]

   o Specified by transition/co-existence technologies

      * IPv4 address and port [RFC4380]

   Deriving the IID from a globally unique IEEE identifier [RFC2462] was one of the earliest mechanisms developed. A number of privacy and security issues related to the interface IDs derived from IEEE identifiers were discovered after their standardization, and many of the mechanisms developed later aimed to mitigate some or all of these weaknesses. This document identifies four types of threats against IEEE-identifier-based IIDs, and discusses how other existing techniques for generating IIDs do or do not mitigate those threats.

2. Terminology

This section clarifies the terminology used throughout this document.

Public address:
: An address that has been published in a directory or other public location, such as the DNS, a SIP proxy, an application-specific DHT, or a publicly available URI. A host’s public addresses are intended to be discoverable by third parties.

Stable address:

An address that does not vary over time within the same network. Note that [RFC4941] refers to these as "public" addresses, but "stable" is used here for reasons explained in Section 4.

Temporary address:
An address that varies over time within the same network.

Constant IID:
An IPv6 Interface Identifier that is globally stable. That is, the Interface ID will remain constant even if the node moves from one network to another.

Stable IID:
An IPv6 Interface Identifier that is stable within some specified context. For example, an Interface ID can be globally stable (constant), or could be stable per network (meaning that the Interface ID will remain unchanged as long as a the node stays on the same network, but may change when the node moves from one network to another).

Temporary IID:
An IPv6 Interface Identifier that varies over time.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119]. These words take their normative meanings only when they are presented in ALL UPPERCASE.

3. Weaknesses in IEEE-identifier-based IIDs

There are a number of privacy and security implications that exist for hosts that use IEEE-identifier-based IIDs. This section discusses four generic attack types: correlation of activities over time, location tracking, address scanning, and device-specific vulnerability exploitation. The first three of these rely on the attacker first gaining knowledge of the target host’s IID. This can be achieved by a number of different attackers: the operator of a server to which the host connects, such as a web server or a peer-to-peer server; an entity that connects to the same network as the target (such as a conference network or any public network); or an entity that is on-path to the destinations with which the host communicates, such as a network operator.

3.1. Correlation of activities over time

As with other identifiers, an IPv6 address can be used to correlate the activities of a host for at least as long as the lifetime of the address. The correlation made possible by IEEE-identifier-based IIDs is of particular concern because MAC addresses are much more permanent than, say, DHCP leases. MAC addresses tend to last roughly the lifetime of a device's network interface, allowing correlation on the order of years, compared to days for DHCP.

As [RFC4941] explains,

> "[t]he use of a non-changing interface identifier to form addresses is a specific instance of the more general case where a constant identifier is reused over an extended period of time and in multiple independent activities. Anytime the same identifier is used in multiple contexts, it becomes possible for that identifier to be used to correlate seemingly unrelated activity. ... The use of a constant identifier within an address is of special concern because addresses are a fundamental requirement of communication and cannot easily be hidden from eavesdroppers and other parties. Even when higher layers encrypt their payloads, addresses in packet headers appear in the clear."

IP addresses are just one example of information that can be used to correlate activities over time. DNS names, cookies [RFC6265], browser fingerprints [Panopticlick], and application-layer usernames can all be used to link a host's activities together. Although IEEE-identifier-based IIDs are likely to last at least as long or longer than these other identifiers, IIDs generated in other ways may have shorter or longer lifetimes than these identifiers depending on how they are generated. Therefore, the extent to which a host's activities can be correlated depends on whether the host uses multiple identifiers together and the lifetimes of all of those identifiers. Frequently refreshing an IPv6 address may not mitigate correlation if an attacker has access to other longer lived identifiers for a particular host. This is an important caveat to keep in mind throughout the discussion of correlation in this document. For further discussion of correlation, see Section 5.2.1 of [RFC6973].

As noted in [RFC4941], in some cases correlation is just as feasible for a host using an IPv4 address as for a host using an IEEE identifier to generate its IID in its IPv6 address. Hosts that use static IPv4 addressing or who are consistently allocated the same address via DHCPv4 can be tracked as described above. However, the widespread use of both NAT and DHCPv4 implementations that assign the same host a different address upon lease expiration mitigates this

threat in the IPv4 case as compared to the IEEE identifier case in IPv6.

3.2. Location tracking

Because the IPv6 address structure is divided between a topological portion and an interface identifier portion, an interface identifier that remains constant when a host connects to different networks (as an IEEE-identifier-based IID does) provides a way for observers to track the movements of that host. In a passive attack on a mobile host, a server that receives connections from the same host over time would be able to determine the host's movements as its prefix changes.

Active attacks are also possible. An attacker that first learns the host's interface identifier by being connected to the same network segment, running a server that the host connects to, or being on-path to the host's communications could subsequently probe other networks for the presence of the same interface identifier by sending a probe packet (ICMPv6 Echo Request, or any other probe packet). Even if the host does not respond, the first hop router will usually respond with an ICMP Address Unreachable when the host is not present, and be silent when the host is present.

Location tracking based on IP address is generally not possible in IPv4 since hosts get assigned wholly new addresses when they change networks.

3.3. Address scanning

The structure of IEEE-based identifiers used for address generation can be leveraged by an attacker to reduce the target search space [I-D.ietf-opsec-ipv6-host-scanning]. The 24-bit Organizationally Unique Identifier (OUI) of MAC addresses, together with the fixed value (0xff, 0xfe) used to form a Modified EUI-64 Interface Identifier, greatly help to reduce the search space, making it easier for an attacker to scan for individual addresses using widely-known popular OUIs. This erases much of the protection against address scanning that the larger IPv6 address space was supposed to provide as compared to IPv4.

3.4. Device-specific vulnerability exploitation

IPv6 addresses that embed IEEE identifiers leak information about the device (Network Interface Card vendor, or even Operating System and/or software type), which could be leveraged by an attacker with knowledge of device/software-specific vulnerabilities to quickly find possible targets. Attackers can exploit vulnerabilities in hosts

whose IIDs they have previously obtained, or scan an address space to find potential targets.

4. Privacy and security properties of address generation mechanisms

Analysis of the extent to which a particular host is protected against the threats described in Section 3 depends on how each of a host's addresses is generated and used. In some scenarios, a host configures a single global address and uses it for all communications. In other scenarios, a host configures multiple addresses using different mechanisms and may use any or all of them.

[RFC3041] (later obsoleted by [RFC4941]) sought to address some of the problems described in Section 3 by defining "temporary addresses" for outbound connections. Temporary addresses are meant to supplement the other addresses that a device might use, not to replace them. They use IIDs that are randomly generated and change daily by default. The idea was for temporary addresses to be used for outgoing connections (e.g., web browsing) while maintaining the ability to use a stable address when more address stability is desired (e.g., in DNS advertisements).

[RFC3484] originally specified that stable addresses be used for outbound connections unless an application explicitly prefers temporary addresses. The default preference for stable addresses was established to avoid applications potentially failing due to the short lifetime of temporary addresses or the possibility of a reverse look-up failure or error. However, [RFC3484] allowed that "implementations for which privacy considerations outweigh these application compatibility concerns MAY reverse the sense of this rule" and instead prefer by default temporary addresses rather than stable addresses. Indeed most implementations (notably including Windows) chose to default to temporary addresses for outbound connections since privacy was considered more important (and few applications supported IPv6 at the time, so application compatibility concerns were minimal). [RFC6724] then obsoleted [RFC3484] and changed the default to match what implementations actually did.

The envisioned relationship in [RFC3484] between stability of an address and its use in "public" can be misleading when conducting privacy analysis. The stability of an address and the extent to which it is linkable to some other public identifier are independent of one another. For example, there is nothing that prevents a host from publishing a temporary address in a public place, such as the DNS. Publishing both a stable address and a temporary address in the DNS or elsewhere where they can be linked together by a public identifier allows the host's activities when using either address to be correlated together.

Moreover, because temporary addresses were designed to supplement other addresses generated by a host, the host may still configure a more stable address even if it only ever intentionally uses temporary addresses (as source addresses) for communication to off-link destinations. An attacker can probe for the stable address even if it is never used as such a source address or advertised (e.g., in DNS or SIP) outside the link.

This section compares the privacy and security properties of a variety of IID generation mechanisms and their possible usage scenarios, including scenarios in which a single mechanism is used to generate all of a host's IIDs and those in which temporary addresses are used together with addresses generated using a different IID generation mechanism. The analysis of the exposure of each IID type to correlation assumes that IPv6 prefixes are shared by a reasonably large number of nodes. As [RFC4941] notes, if a very small number of nodes (say, only one) use a particular prefix for an extended period of time, the prefix itself can be used to correlate the host's activities regardless of how the IID is generated. For example, [RFC3314] recommends that prefixes be uniquely assigned to mobile handsets where IPv6 is used within GPRS. In cases where this advice is followed and prefixes persist for extended periods of time (or get reassigned to the same handsets whenever those handsets reconnect to the same network router), hosts' activities could be correlatable for longer periods than the analysis below would suggest.

The table below provides a summary of the whole analysis.

Mechanism(s)	Correlation	Location tracking	Address scanning	Device exploits
IEEE identifier	For device lifetime	For device lifetime	Possible	Possible
Static manual	For address lifetime	For address lifetime	Depends on generation mechanism	Depends on generation mechanism
Constant, semantically opaque	For address lifetime	For address lifetime	No	No
CGA	For lifetime of (modifier block + public key)	No	No	No
Stable, semantically opaque	Within single network	No	No	No
Temporary	For temp address lifetime	No	No	No
DHCPv6	For lease lifetime	No	Depends on generation mechanism	No

Table 1: Privacy and security properties of IID generation mechanisms

4.1. IEEE-identifier-based IIDs

As discussed in Section 3, addresses that use IIDs based on IEEE identifiers are vulnerable to all four threats. They allow correlation and location tracking for the lifetime of the device since IEEE identifiers last that long and their structure makes address scanning and device exploits possible.

4.2. Static, manually configured IIDs

Because static, manually configured IIDs are stable, both correlation and location tracking are possible for the life of the address.

The extent to which location tracking can be successfully performed depends, to a some extent, on the uniqueness of the employed Interface ID. For example, one would expect "low byte" Interface IDs to be more widely reused than, for example, Interface IDs where the whole 64-bits follow some pattern that is unique to a specific organization. Widely reused Interface IDs will typically lead to false positives when performing location tracking.

Whether manually configured addresses are vulnerable to address scanning and device exploits depends on the specifics of how the IIDs are generated.

4.3. Constant, semantically opaque IIDs

Although a mechanism to generate a constant, semantically opaque IID has not been standardized, it has been in wide use for many years on at least one platform (Windows). Windows uses the [RFC4941] random generation mechanism in lieu of generating an IEEE-identifier-based IID. This mitigates the device-specific exploitation and address scanning attacks, but still allows correlation and location tracking because the IID is constant across networks and time.

4.4. Cryptographically generated IIDs

Cryptographically generated addresses (CGAs) [RFC3972] bind a hash of the host's public key to an IPv6 address in the SEcure Neighbor Discovery (SEND) [RFC3971] protocol. CGAs may be regenerated for each subnet prefix, but this is not required given that they are computationally expensive to generate. A host using a CGA can be correlated for as long as the lifetime of the combination of the public key and the chosen modifier block, since it is possible to rotate modifier blocks without generating new public keys. Because the cryptographic hash of the host's public key uses the subnet prefix as an input, even if the host does not generate a new public key or modifier block when it moves to a different network, its location cannot be tracked via the IID. CGAs do not allow device-specific exploitation or address scanning attacks.

4.5. Stable, semantically opaque IIDs

[RFC7217] specifies a mechanism that generates a unique random IID for each network. A host that stays connected to the same network could therefore be tracked at length, whereas a mobile host's

activities could only be correlated for the duration of each network connection. Location tracking is not possible with these addresses. They also do not allow device-specific exploitation or address scanning attacks.

4.6. Temporary IIDs

A host that uses only a temporary address mitigates all four threats. Its activities may only be correlated for the lifetime a single temporary address.

A host that configures both an IEEE-identifier-based IID and temporary addresses makes the host vulnerable to the same attacks as if temporary addresses were not in use, although the viability of some of them depends on how the host uses each address. An attacker can correlate all of the host's activities for which it uses its IEEE-identifier-based IID. Once an attacker has obtained the IEEE-identifier-based IID, location tracking becomes possible on other networks even if the host only makes use of temporary addresses on those other networks; the attacker can actively probe the other networks for the presence of the IEEE-identifier-based IID. Device-specific vulnerabilities can still be exploited. Address scanning is also still possible because the IEEE-identifier-based address can be probed.

If the host instead generates a constant, semantically opaque IID to use in a stable address for server-like connections together with temporary addresses for outbound connections (as is the default in Windows), it sees some improvements over the previous scenario. The address scanning and device-specific exploitation attacks are no longer possible because the OUI is no longer embedded in any of the host's addresses. However, correlation of some activities across time and location tracking are both still possible because the semantically opaque IID is constant. And once an attacker has obtained the host's semantically opaque IID, location tracking is possible on any network by probing for that IID, even if the host only uses temporary addresses on those networks. However, if the host generates but never uses a constant, semantically opaque IID, it mitigates all four threats.

When used together with temporary addresses, the stable, semantically opaque IID generation mechanism [RFC7217] improves upon the previous scenario by limiting the potential for correlation to the lifetime of the stable address (which may still be lengthy for hosts that are not mobile) and by eliminating the possibility for location tracking (since a different IID is generated for each subnet prefix). As in the previous scenario, a host that configures but does not use a stable, semantically opaque address mitigates all four threats.

4.7. DHCPv6 generation of IIDs

The security/privacy implications of DHCPv6-based addresses will typically depend on the specific DHCPv6 server software being employed. We note that recent releases of most popular DHCPv6 server software typically lease random addresses with a similar lease time as that of IPv4. Thus, these addresses can be considered to be "stable, semantically opaque."

On the other hand, some DHCPv6 software leases sequential addresses (typically low-byte addresses). These addresses can be considered to be stable addresses. The drawback of this address generation scheme compared to "stable, semantically opaque" addresses is that, since they follow specific patterns, they enable IPv6 address scans.

4.8. Transition/co-existence technologies

Addresses specified based on transition/co-existence technologies that embed an IPv4 address within an IPv6 address are not included in Table 1 because their privacy and security properties are inherited from the embedded address. For example, Teredo [RFC4380] specifies a means to generate an IPv6 address from the underlying IPv4 address and port, leaving many other bits set to zero. This makes it relatively easy for an attacker to scan for IPv6 addresses by guessing the Teredo client’s IPv4 address and port (which for many NATs is not randomized). For this reason, popular implementations (e.g., Windows), began deviating from the standard by including 12 random bits in place of zero bits. This modification was later standardized in [RFC5991].

5. Miscellaneous Issues with IPv6 addressing

5.1. Geographic Location

Since IPv6 subnets have unique prefixes, they reveal some information about the location of the subnet, just as IPv4 addresses do. Hiding this information is one motivation for using NAT in IPv6 (see RFC 5902 section 2.4).

5.2. Network Operation

It is generally agreed that IPv6 addresses that vary over time in a specific network tend to increase the complexity of event logging, trouble-shooting, enforcement of access controls and quality of service, etc. As a result, some organizations disable the use of temporary addresses [RFC4941] even at the expense of reduced privacy [Broersma].

5.3. Compliance

Some IPv6 compliance testing suites required (and might still require) implementations to support MAC-derived suffixes in order to be approved as compliant. This document recommends that compliance testing suites be relaxed to allow other forms of address generation that are more amenable to privacy.

5.4. Intellectual Property Rights (IPRs)

Some IPv6 addressing techniques might be covered by Intellectual Property rights, which might limit their implementation in different Operating Systems. [CGA-IPR] and [KAME-CGA] discuss the IPRs on CGAs.

6. Security Considerations

This whole document concerns the privacy and security properties of different IPv6 address generation mechanisms.

7. IANA Considerations

This document does not require actions by IANA.

8. Acknowledgements

The authors would like to thank Bernard Aboba, Tim Chown, Rich Draves, Robert Moskowitz, Erik Nordmark, and James Woodyatt for providing valuable comments on earlier versions of this document.

9. Informative References

[Broersma] Broersma, R., "IPv6 Everywhere: Living with a Fully IPv6-enabled environment", Australian IPv6 Summit 2010, Melbourne, VIC Australia, October 2010, October 2010, <http://www.ipv6.org.au/10ipv6summit/talks/Ron_Broersma.pdf>.

[CGA-IPR] IETF, "Intellectual Property Rights on RFC 3972", 2005.

[I-D.ietf-opsec-ipv6-host-scanning] Gont, F. and T. Chown, "Network Reconnaissance in IPv6 Networks", draft-ietf-opsec-ipv6-host-scanning-04 (work in progress), June 2014.

[KAME-CGA]
KAME, "The KAME IPR policy and concerns of some technologies which have IPR claims", 2005.

[Microsoft]
Microsoft, "IPv6 interface identifiers", 2013.

[Panopticlick]
Electronic Frontier Foundation, "Panopticlick", 2011.

[RFC1972] Crawford, M., "A Method for the Transmission of IPv6 Packets over Ethernet Networks", RFC 1972, August 1996.

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

[RFC2462] Thomson, S. and T. Narten, "IPv6 Stateless Address Autoconfiguration", RFC 2462, December 1998.

[RFC2464] Crawford, M., "Transmission of IPv6 Packets over Ethernet Networks", RFC 2464, December 1998.

[RFC3041] Narten, T. and R. Draves, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 3041, January 2001.

[RFC3314] Wasserman, M., "Recommendations for IPv6 in Third Generation Partnership Project (3GPP) Standards", RFC 3314, September 2002.

[RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.

[RFC3484] Draves, R., "Default Address Selection for Internet Protocol version 6 (IPv6)", RFC 3484, February 2003.

[RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, March 2005.

[RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, March 2005.

[RFC4380] Huitema, C., "Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs)", RFC 4380, February 2006.

[RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.

[RFC5991] Thaler, D., Krishnan, S., and J. Hoagland, "Teredo Security Updates", RFC 5991, September 2010.

[RFC6265] Barth, A., "HTTP State Management Mechanism", RFC 6265, April 2011.

[RFC6724] Thaler, D., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, September 2012.

[RFC6973] Cooper, A., Tschofenig, H., Aboba, B., Peterson, J., Morris, J., Hansen, M., and R. Smith, "Privacy Considerations for Internet Protocols", RFC 6973, July 2013.

[RFC7217] Gont, F., "A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)", RFC 7217, April 2014.

Authors' Addresses

Alissa Cooper
Cisco
707 Tasman Drive
Milpitas, CA 95035
US

Phone: +1-408-902-3950
Email: alcoop@cisco.com
URI: https://www.cisco.com/

Fernando Gont
Huawei Technologies
Evaristo Carriego 2644
Haedo, Provincia de Buenos Aires 1706
Argentina

Phone: +54 11 4650 8472
Email: fgont@si6networks.com
URI: http://www.si6networks.com

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

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