2020-09

ANALYSIS OF EUI-64'BASED ADDRESSING AND ASSOCIATED VULNERABILITIES

Thordarson, Kirstin E.

Monterey, CA; Naval Postgraduate School

https://hdl.handle.net/10945/66035

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

ANALYSIS OF EUI-64–BASED ADDRESSING AND ASSOCIATED VULNERABILITIES

by

Kirstin E. Thordarson

September 2020

Thesis Advisor:	Robert Beverly
Second Reader:	Erik Rye (CMAND)

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2020	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE ANALYSIS OF EUI-64–BASED ADDRESSING AND ASSOCIATED VULNERABILITIES	**5. FUNDING NUMBERS**
6. AUTHOR(S) Kirstin E. Thordarson	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Despite the adoption of security extensions in IPv6 addressing, usage of EUI-64–based addresses are known to be prevalent throughout the IPv6 address space. In particular, a high concentration of EUI-64 addresses are found on Customer Premises Equipment (CPE) infrastructure, including home gateways and routers. This thesis seeks to deepen current understanding of the IPv6 topology through an in-depth examination of EUI-64 addressing. Recent research has produced a series of rich IPv6 topology data sets that have yet to be fully leveraged for their insight into EUI-64 usage characteristics. Employing IPv6 topology data, this paper extracts and analyzes prefix assignment patterns and device identifications within EUI-64–rich networks. This thesis uses a combination of statistical and predictive analysis to execute the following research objectives: measure the distribution of device manufacturer/model on a given prefix; identify, describe, and predict EUI-64–based prefix rotation patterns; and apply findings within a cyber security context so as to further evaluate the security and privacy risks of EUI-64–based addresses in existing IPv6 addressing schemes.

14. SUBJECT TERMS cybersecurity, networking, internet measurement, internet security	**15. NUMBER OF PAGES** 91
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

THIS PAGE INTENTIONALLY LEFT BLANK

ANALYSIS OF EUI-64–BASED ADDRESSING AND ASSOCIATED VULNERABILITIES

Kirstin E. Thordarson
Civilian, Scholarship for Service
BA, University of San Francisco, 2017

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
September 2020

Approved by: Robert Beverly
Advisor

Erik Rye
Second Reader

Gurminder Singh
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Despite the adoption of security extensions in IPv6 addressing, usage of EUI-64–based addresses are known to be prevalent throughout the IPv6 address space. In particular, a high concentration of EUI-64 addresses are found on Customer Premises Equipment (CPE) infrastructure, including home gateways and routers. This thesis seeks to deepen current understanding of the IPv6 topology through an in-depth examination of EUI-64 addressing. Recent research has produced a series of rich IPv6 topology data sets that have yet to be fully leveraged for their insight into EUI-64 usage characteristics. Employing IPv6 topology data, this paper extracts and analyzes prefix assignment patterns and device identifications within EUI-64–rich networks. This thesis uses a combination of statistical and predictive analysis to execute the following research objectives: measure the distribution of device manufacturer/model on a given prefix; identify, describe, and predict EUI-64–based prefix rotation patterns; and apply findings within a cyber security context so as to further evaluate the security and privacy risks of EUI-64–based addresses in existing IPv6 addressing schemes.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

AFRINIC	African Network Information Centre
APNIC	Asia Pacific Network Information Centre
ARIN	American Registry for Internet Numbers
CGA	Cryptographically Generated Addresses
CGN	Carrier-Grade NAT
CPE	Customer Premises Equipment
DAD	Duplicate Address Detection
DHCPv6	Dynamic Host Configuration Protocol Version 6
DNS	Domain Name System
DoS	Denial of Service
DUID	DHCP Unique Identifier
EDAD	Enhanced Duplicate Address Detection
ESP	Encapsulating Security Payload
EUI-64	Extended Unique Identifier - 64 Bit
GPS	Global Positioning System
IANA	Internet Assigned Numbers Authority
IEEE	Institute of Electrical and Electronics Engineers
IID	Interface ID (Identifier)
IIJ	Internet Initiative Japan

IoT	Internet of Things
IP	Internet Protocol
IPSec	Internet Protocol Security
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
ISP	Internet Service Provider
LACNIC	Latin America and Caribbean Network Information Centre
LAN	Local Area Network
MAC	Media Access Control
NAT	Network Address Translation
NIC	Network Interface Controller
NPS	Naval Postgraduate School
OSI	Open Systems Interconnection
OUI	Organizationally Unique Identifier
PKI	Public Key Infrastructure
RFC	Request for Comment
RIPE NCC	Réseaux IP Européens Network Coordination Centre
RIR	Regional Internet Registry
SIIT	Stateless IP/ICMP Translation
SLAAC	Stateless Address Auto-Configuration
TTL	Time to Live
UDF	User Defined Function

ULA	Unique Local Address
VNPT	Vietnam Posts and Telecommunications

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

I would like to thank my advisors, Robert Beverly and Erik Rye, as well as Justin Rohrer and Thomas Krenc for their expert guidance and support in the research process. I also extend my deepest gratitude to the members of my cohort, who have allowed me to work (and sometimes struggle) alongside them for the last two years. Thank you to Alex Newhouse, for your unwavering belief in my abilities and for sharing with me your knowledge of and passion for data science.

Finally, this thesis, along with my studies at the Naval Postgraduate School, were made possible through the generous funding of National Science Foundation (NSF) and the Scholarship for Service (SFS) program. This material is based upon activities supported by the National Science Foundation under Agreement No 1565443. Any opinions, findings, and conclusions or recommendations expressed are those of the author and do not necessarily reflect the views of the National Science Foundation.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

Network topology maps at the IP, router, and provider level enable network and security researchers to better understand the operation of, and address problems in, the logical Internet. However, the enormity and relative sparseness of the Internet Protocol Version 6 (IPv6) address space obsoletes traditional and brute-force methods of address scanning and device discovery. The IPv6 address space is also characterized by ephemeral and dynamic addressing and a lack of address translation, which further convolute attempts to discover and map IPv6 infrastructure topology. Accordingly, current and future understanding of the Internet depends the development of active systems of IPv6 topology measurement capable of operating at speed and scale.

This thesis builds upon prior work which establishes one such system of active IPv6 topology discovery, focusing in particular on IPv6 Customer Premises Equipment (CPE). Using active measurement data derived from this study, we investigate addressing dynamics for a specific Internet Service Provider (ISP), facilitating not only the discovery of IPv6 CPE devices, but also the tracking of network prefix allocations over time. This work seeks to examine IPv6 addressing behaviors for CPE devices in both quantitative and qualitative terms, and to apply findings within a cyber security context so as to further evaluate potential security and privacy risks.

1.1 Motivation

The Internet has seen an exponential uptick in IPv6 adoption over the past 10 years. In 2020, Google reported approximately a third of its traffic arrives as IPv6 (Figure 2.1). One driver of the growth in IPv6 traffic has been residential deployments featuring CPE with IPv6 addresses.

As this rate of IPv6 adoption increases, a major area of concern is privacy and security in IPv6 addressing [1]–[4]. IPv6 addresses may be assigned via various methods, including Dynamic Host Configuration Protocol Version 6 (DHCPv6), Stateless Address Auto-Configuration (SLAAC), and manual assignment [5] [6]. While the network portion of an end-site IPv6 address may be up to 64 bits, the least significant 64 bits were historically assigned using a deterministic function of the host's Media Access Control (MAC) address called Extended Unique Identifier - 64 Bit (EUI-64). The use of the EUI-64 Interface ID (Identifier) (IID) format within SLAAC carries significant well-known security and privacy concerns [7]. Figure 1.1 depicts the creation of an EUI-64 IID, which is derived from an Institute of Electrical and Electronics Engineers (IEEE) MAC address, and makes up the lower 64 bits of a 128-bit IPv6 address. Specific mechanisms for the creation of an EUI-64 IID are discussed in more detail in section 2.1.

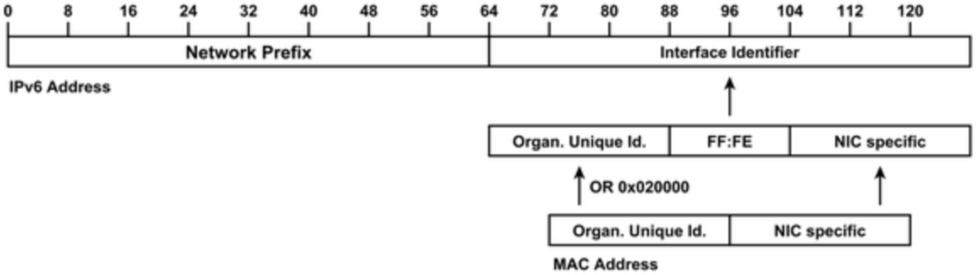


Figure 1.1. IPv6 Addresses containing the network prefix and the interface identifier. The modified EUI-64 format is derived from the MAC address by inserting a fixed pattern of two bytes. Source [3].

As EUI-64 addresses are derived from IEEE MAC addresses, they risk revealing sensitive information about the device to which they are assigned [7]. Information exposed through these addresses includes the manufacturer or model of a device, and in some cases its operating system [8]. Furthermore, the predictability and lack of privacy in this addressing scheme lends itself to the creation of an address space that is easier to scan or probe, and thereby more vulnerable to attack [9]. The weaknesses inherent to the EUI-64 addressing scheme are not new, and discussions of the format's deprecation have been ongoing. However, EUI-64 based addresses are still frequently used, particularly in mobile systems and Internet of Things (IoT) devices [3].

Despite known privacy and security vulnerabilities, previous studies have identified EUI-64 based addressing as commonplace in the IPv6 topology, estimating 651.4k or 45 percent of all IPv6 interface addresses to employ EUI-64 based addressing [10]. The publication of privacy extensions protocol (Request for Comment (RFC) 4941) in 2007 amplifies the interest of this statistic, as the document outlines viable mitigation for privacy concerns associated with the EUI-64 addressing scheme via an algorithm for ephemeral and randomized addresses assignment [11]. While the majority of end hosts currently employ the privacy extensions protocol, it appears that this is not necessarily the case elsewhere in the IPv6 topology. Many IPv6 router interface addresses are statically configured and do not necessarily employ either EUI-64 or privacy extension protocols. Of direct interest to this study ,however, are CPE that use EUI-64 and are, thus, trackable.

Previous work has created methods of tracking changes in CPE prefix assignment for a specific end node [7]. Cases in which service providers routinely change the assigned CPE prefix, but the CPE addresses is EUI-64, are of particular interest. In these cases, ostensible privacy benefits of ephemeral prefix assignment may be easily thwarted through implicit association to CPE's EUI-64 address. Furthermore, such behavior allows potential tracking of client devices, which in some cases is made even more trivial by predictable or repetitive prefix assignment patterns. The regular re-assignment of 64-bit network prefixes to EUI-64 IIDs is referred to in this work as "prefix rotation" and is described in more detail in section 2.1.9.

While precedent work alludes to the significance of understanding CPE prefix rotation, a more thorough quantification of service provider-induced network prefix changes is yet to be completed. Further lines of inquiry include how common the practice of CPE prefix rotation is within the IPv6 topology, as well as how frequently it is performed, and by which hardware or software mechanisms it is executed. The prevalence of EUI-64 addresses in current topology and their associated risks necessitates an in-depth examination of their use, as such an understanding can facilitate informed decision-making for both providers and end users as IPv6 adoption continues to increase. Furthermore, the contextualization of this information within an adversarial model reveals the gravity of privacy and security concerns in the rapidly developing IPv6 topology.

1.2 Research Questions

Hypothesis: Through statistical and predictive analysis of IPv6 topology data, this research may characterize the use of EUI-64 based addresses, facilitating the measurement of associated privacy and security risks.

Research questions being pursued in this work include:

- Given IPv6 prefixes containing many devices with EUI-64 addresses, what is the distribution of device manufacturers and/or models therein?
- What is the level of homogeneity in hardware for EUI-64 addressed devices, and to what extent can this facilitate cyber attacks?
- What is the IPv6 prefix rotation behavior in terms of frequency, locality, regularity, and entropy of prefix assignments?
- Given prior identification of prefix rotation, to what extent can the future prefix of an EUI-64 last hop be predicted?
- To what extent can particular rotation behaviors be mapped to inferred hardware or software implementations?

1.3 Scope

This thesis is an extension of recent work by Rye and Beverly [12], and explicitly relies on the use of data sets compiled through precedent efforts. The scope of this thesis is therefore limited according to the data available and the methods used in its collection.

Analyses pertaining to the temporal and spatial characterization of prefix assignments for EUI-64 addressed IPv6 devices are limited to a single ISP, Versatel. The selection of this ISP was based on results from Rye and Beverly's study, in which Versatel was found to host a large proportion of EUI-64 addressed CPE, and exhibited clearly defined prefix assignment behaviors.

As this thesis builds upon previous data collection efforts, the focus of the study will not be on accumulating more recent or higher quality data, but rather on furthering the analysis and understanding of extant, information-rich data. Validation of findings does not fall within the scope of this thesis, but may be the subject of future work.

1.4 Summary of Contributions

In this work, we contribute the following:

- Correlation EUI-64 last hop addresses with device manufacturers

- Quantitative analysis of temporal and spatial behaviors for prefix assignment among EUI-64 addressed devices
- Identification and analysis of anomalous prefix assignment schemes among EUI-64 addressed devices
- Qualitative analysis of potential privacy and security concerns related to prefix assignment behaviors

1.5 Thesis Structure

- Chapter 1 – Introduces the topic, motivations, and scope of this work; summarizes critical findings; and enumerates structure of thesis.
- Chapter 2 – Reviews key concepts and establishes scholastic context via discussion of related work.
- Chapter 3 – Describes employed methodologies.
- Chapter 4 – Examines results and critically analyzes work.
- Chapter 5 – Concludes thesis and offers suggestions for continued study.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Key Concepts and Related Work

This chapter provides an overview of key terms and concepts necessary to understand this thesis. It also situates our work within the context of previous related studies, which reveal the specific privacy and security vulnerabilities associated with EUI-64 based addressing and discuss potential methods for their mitigation.

2.1 Key Concepts

2.1.1 Internet Protocol Version 4 (IPv4) Exhaustion

Defined in RFC 791 [13], an IPv4 address is a distinct 32-bit unsigned integer assigned to network interfaces connected to the Internet. As of February 2011, the Internet Assigned Numbers Authority (IANA) IPv4 address space, whose purpose is to allocate blocks of IPv4 addresses to regional authorities, was exhausted [14]. In the years following, a series of Regional Internet Registry (RIR)s exhausted their IPv4 allocations: Asia Pacific Network Information Centre (APNIC) in April 2011, Latin America and Caribbean Network Information Centre (LACNIC) in June 2014, and American Registry for Internet Numbers (ARIN) in September 2015 [15]. The pattern continued with African Network Information Centre (AFRINIC) in April 2017 and Réseaux IP Européens Network Coordination Centre (RIPE NCC) in 2019, thus completely exhausting all RIR IPv4 address pools.

Total depletion of the approximately 4.3 billion addresses in the IPv4 architecture was anticipated since the late 1980s. In the following decades, the IPv6 architecture was developed to accommodate the rapid and continuous expansion of the Internet, and act as a remedy to the quickly dwindling pool of available IPv4 addresses. The most up-to-date information about IPv6 addressing architecture and protocol specifications may be found in RFC 4291 (February 2006) and RFC 8200 (July 2017), although their original counterparts date back to 1995 [16]–[19]. Detailed discussion of the IPv6 architecture may be found in Section 2.1.3.

Until IPv6 is fully deployed, each RIR has reserved blocks of IPv4 addresses [14]. ARIN and LACNIC both reserve the final /10, APNIC and RIPE NCC reserve the last /8, and AFRINIC reserves a /11 block [15], [20]–[24].

2.1.2 IPv6 Adoption and Address Space

RFC 4291 defines an IPv6 address as a 128-bit integer, represented as a string of 32 hexadecimal characters which are divided into eight colon-separated quartets. Specific elements of the IPv6 addressing format are discussed further in Section 2.1.3. This 128-bit addressing format provides a total address space that is approximately 37 orders of magnitude greater than that provided by IPv4's 32-bit addresses [25]. As such, IPv6 connectivity has increased exponentially since the top-level exhaustion of IPv4 in 2011 and subsequent exhaustion of RIR pools (Figure 2.1) [26]. As of July 2020, Google observes IPv6 connectivity among 32 percent of its users worldwide, and among 43 percent of users in the United States [26].

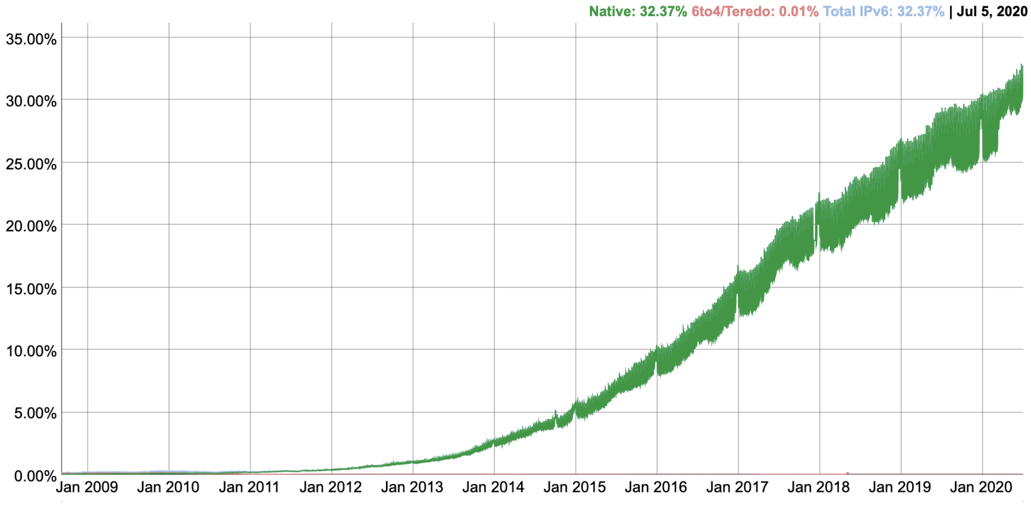


Figure 2.1. IPv6 connectivity among Google users. Source: [26].

Increased IPv6 is directly related to depletion of the IPv4 address space, and the comparatively massive address space in IPv6 [25], [26]. An IPv4 address, composed of 32-bits, allows for the creation of 2^{32} unique addresses, about 4.3×10^9 in total. In contrast, a 128-bit IPv6 address is capable of representing 2^{128} unique addresses, which equates to about 3.4×10^{38}.

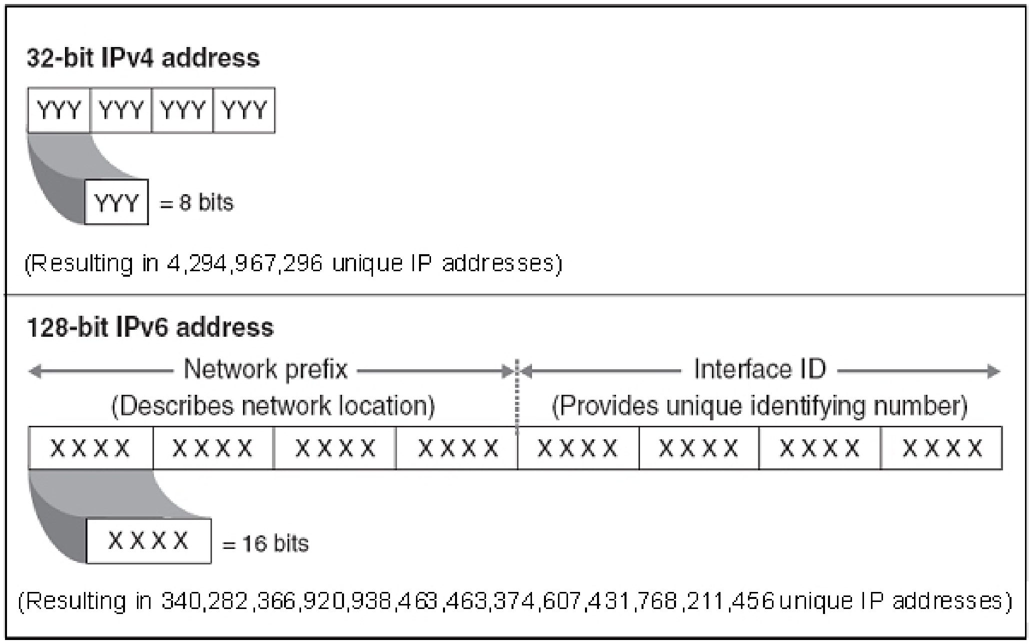

32-bit IPv4 address
YYY
YYY
YYY
YYY
YYY
= 8 bits
(Resulting in 4,294,967,296 unique IP addresses)
128-bit IPv6 address
Network prefix
(Describes network location)
Interface ID
(Provides unique identifying number)
XXXX
XXXX
XXXX
XXXX
XXXX
XXXX
XXXX
XXXX
XXXX
= 16 bits
(Resulting in 340,282,366,920,938,463,463,374,607,431,768,211,456 unique IP addresses)

Figure 2.2. Visualization of IPv4 and IPv6 address space calculations. Source: [27].

IPv6 thus yields a total address space that is significantly larger than that of IPv4 [25].

While intended to replace IPv4, the conversion to IPv6 is far from complete, and has been aided by a variety of transition mechanisms such as 6to4 (now deprecated), 6in4, Stateless IP/ICMP Translation (SIIT), and Dual Stack [28]–[30]. Furthermore, Carrier-Grade NAT (CGN) provides a viable means of extending extant IPv4 allocations on top of Network Address Translation (NAT) and RFC 1918 private address space [31], [32]. Recent research raises concerns over continued reliance on such transition mechanisms, indicating that they may act as crutch preventing a full IPv6 transition and concealing critical weaknesses inherent to the protocol design [33]. For example, the existence of protocol-specific dependencies for web resources creates a need for dual-stack environments that must be addressed before adoption of a fully IPv6 Internet is possible [34], [35].

2.1.3 IPv6 Address Architecture

IPv6 addresses consist of 128 bits, and are typically subdivided into three constituent parts: the routing prefix, the subnet ID, and the IID [16]. The lower 64 bits of the address always indicate the IID value, which is used to identify the host's network interface. The usage of the upper 64 bits is more flexible, and may be configured in various ways to represent different combinations of routing prefixes and subnet IDs, where the routing prefix is always at least 48-bits and the subnet ID is at most 16-bits in length. The subnet prefix acts as a network identifier, which serves to route traffic to a given Local Area Network (LAN). The IID ensures the uniqueness of a host's interface address within its local network [36].

```
|                          128 bits                              |
+----------------------------------------------------------------+
|                          node address                          |
+----------------------------------------------------------------+
```

Figure 2.3. 128-bit IPv6 address. Source [16].

```
|          n bits               |           128-n bits            |
+-------------------------------+---------------------------------+
|       subnet prefix           |           interface ID          |
+-------------------------------+---------------------------------+
```

Figure 2.4. IPv6 addresses are composed of a subnet prefix followed by an interface Identifier. Depending on the method of address assignment, the length of each field can vary. Source: [16].

Interface identifiers in IPv6 unicast addresses are used to identify interfaces on a link [37]. They therefore must be unique at least within the scope of that link, if not on a broader scope. The IID value may be created manually, assigned randomly and automatically, obtained from a DHCPv6 server, or generated via SLAAC using the EUI-64 addressing format, as described in Section 2.1.4 and 2.1.6, respectively. The IPv6 Address Architecture prevents assignment of duplicate Interface Identifiers through Duplicate Address Detection (DAD), also described in Section 2.1.4.

2.1.4 Stateless Address Auto-Configuration (SLAAC)

SLAAC is a commonly used method for automatically assigning 128-bit IPv6 addresses [6]. It is a stateless algorithm, and does not provide for control over exact address assignments. It only ensures that assigned addresses are unique and properly routable. The algorithm allows for autoconfiguration on a per-interface basis, for multicast-capable interfaces only.

SLAAC determines uniqueness of addresses via DAD, which sends a Neighbor Solicitation Message with the tentative address, which is replied to with a Neighbor Advertisement if the address is already taken [6], [38]. The original version of DAD was updated in 2015 by Enhanced Duplicate Address Detection (EDAD), which allows automated detection of looped back IPv6 Neighbor Discover messages used by DAD [39].

Global SLAAC addresses consist of an IID appended to a globally routable prefix of appropriate length, obtained from the Prefix Information options contained in Router Advertisements [6]. SLAAC addresses may be rendered local in scope if the router assigns a non-routable prefix, such as a Unique Local Address (ULA) prefix [40]. The IID may be assigned or selected according to a variety of methods, including those described in Sections 2.1.6 and 2.1.7.

There also exist stateful protocols for assigning IPv6 addresses, such as DHCPv6, which may be used in lieu of, or in conjunction with, SLAAC. It grants tighter control over exact address assignments than stateless protocols such as SLAAC, and it does not employ the EUI-64 method for creating interface identifiers [41]. IPv6 addresses may also be manually assigned, forgoing the various methods of autoconfiguration.

2.1.5 IEEE MAC Address

The IEEE MAC Address is a 48-bit Layer-2 hardware network identifier [42]. MAC addresses are static and persistent across network attachment points. A key attribute of an IEEE MAC address is their intended global uniqueness, wherein an individual MAC address is designed to correspond to a discrete device. Additionally, these addresses exist at a low level in the Open Systems Interconnection (OSI) stack and are therefore difficult to alter. The six composite octets of a MAC address are typically subdivided into the upper three and lower three bytes. The three most significant octets make up the Organizationally Unique Identifier (OUI), which advertises high-level information about the device, such as vendor, manufacturer, or organizational membership. The lower three octets specify the device Network Interface Controller (NIC). Recent work has developed predictive models for granular device fingerprinting based on the value of low-order octets, revealing further information about a given end device [43].

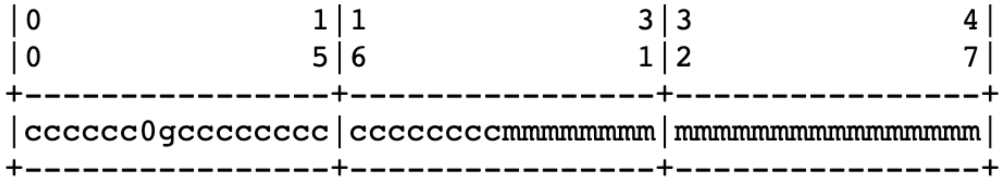
2.1.6 EUI-64 Addressing Mode

EUI-64 is a method of generating a 64-bit IID from a globally or locally unique token for use in a SLAAC address [37]. The IID composes the lower 64 bits of an IPv6 address, and may be either globally or locally unique depending on the type of token available.

The scope of the identifier is indicated by the universal/local bit (also called the “u” bit), while group membership is indicated by the individual/group bit (also called the “g” bit). The “g” and the “u” bit are the lowest and second lowest order bits of the highest order octet in the IID, respectively (Figure 2.5).

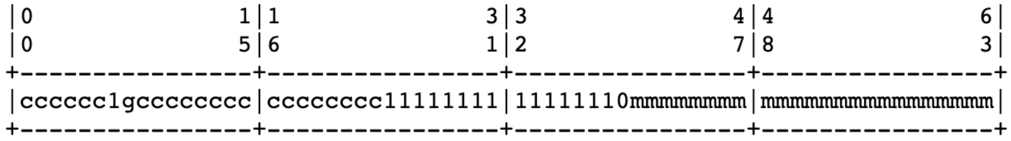
The majority of EUI-64 addresses are derived from IEEE MAC addresses; this variety of EUI-64 addresses constitutes the main focus of this thesis. The process of creating such IIDs is simple, and is defined in Appendix A of RFC 2373 and RFC 4921 [16], [37]. Despite the global scope of IEEE MAC addresses, the EUI-64 IIDs derived from them may be rendered local in scope by simply flipping the universal/local bit.

Take ,for example, the EUI-64 IID pictured in Figure 2.6 and the IEEE MAC address from which it is derived, shown in Figure 2.5.



```
|0              1|1              3|3              4|
|0              5|6              1|2              7|
+----------------+----------------+----------------+
|cccccc0gcccccccc|ccccccccmmmmmmmm|mmmmmmmmmmmmmmmm|
+----------------+----------------+----------------+
```

Figure 2.5. 48-bit IEEE MAC address. Source [37].



```
|0              1|1              3|3              4|4              6|
|0              5|6              1|2              7|8              3|
+----------------+----------------+----------------+----------------+
|cccccc1gcccccccc|cccccccc11111111|11111110mmmmmmmm|mmmmmmmmmmmmmmmm|
+----------------+----------------+----------------+----------------+
```

Figure 2.6. 64-bit EUI-64 IID created from the IEEE MAC address in Figure 2.5. Source: [37].

The EUI-64 IID is created by splitting the MAC Address in half, and inserting two hexadecimal octets, with the values 0xFF and 0xFE, in the center. This yields an IID composed of the IEEE MAC company ID, followed by the 0xFF and 0xFE octets, then the IEEE MAC vendor supplied ID, as pictured in Figure 2.6. Note that in the resulting EUI-64 identifier, the universal/local bit has been changed from "0" to "1", indicating a local scope, even though the IEEE MAC address from which the EUI-64 IID is derived is considered a globally unique identifier.

Recent research estimates that usage of EUI-64 based addressing is uncommon among IPv6 clients, with approximately 1 percent of clients employing this mode [36]. However, significant numbers of EUI-64 addresses still exist in IPv6 infrastructure, as approximately 45 percent of all IPv6 interface addresses employ EUI-64 based addressing [10], [44].

2.1.7 IPv6 Privacy Extensions

RFC 4941 describes an extension to IPv6 SLAAC for interfaces whose IID is derived from an IEEE identifier (usually this is an IEEE MAC Address) [11]. This extension allows for the creation of globally unique addresses using IIDs that change over time. Use of such ephemeral IIDs impede eavesdropping and device tracking, as they make it difficult to correlate different addresses with the same node.

There are two versions of the privacy extensions algorithm, depending on whether or not stable storage is available. When stable storage is available, the history value is obtained from the stored results of a previous iteration of the algorithm. This history value will be used to pseudo-randomly generate an interface identifier by appending its latter 64 bits to the SLAAC IID, taking the MD5 hash and setting the global/local bit to 0. After ensuring the generated identifier is not reserved or currently in use, the generated identifier will be used as the pseudo-random security extensions identifier, and the rightmost 64 bits of the MD5 digest will be saved as the next history value. In the case that there is no stable storage available, the history value is generated according to RFC 4086 [45].

The use of privacy extensions mitigates the privacy and security risks associated with SLAAC, as it generates a new IID value for each new prefix assignment. Consequently, a Privacy Extensions addressed device will be assigned an entirely new IPv6 address as it moves between networks. However, with this gain in privacy comes a loss in ease of use. The RFC authors recognize this trade-off, stating that "the desires of protecting individual privacy versus the desire to effectively maintain and debug a network can conflict with each other" [11]. Furthermore, it should be noted that the widespread use of temporary addresses could cause complications where global uniqueness is required. In addition, use of addresses generated via SLAAC with privacy extensions could cause problems in situations where servers refuse to grant access to clients for which no Domain Name System (DNS) name exists. While the continued use of EUI-64 based addressing may seem unintuitive given its privacy and security weaknesses, operational complications inherent to available solutions may provide explanation for this phenomenon.

2.1.8 IPv6 CPE Deployment and Discovery

CPE are devices which are physically present at a customer's home or business. Examples of CPE include routers, bridges, modems, or wireless access points. Such devices are often leased or sold to end users by their ISP along with internet service packages. In the realm of IPv6, the CPE is generally a router [12].

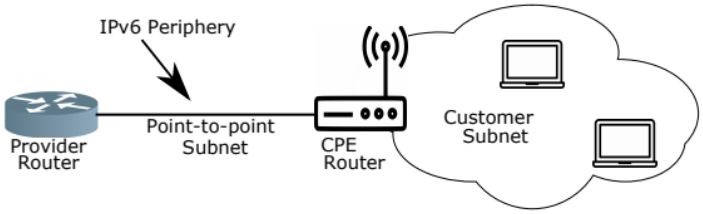


Figure 2.7. Common IPv6 architecture: an IPv6 subnet is assigned to the link between the provider and last hop CPE routers. There is no NAT or private addressing; a separate distinct routed IPv6 subnet is assigned to devices attached to the last hop CPE. Source: [12].

Figure 2.7 illustrates a typical residential deployment for IPv6 networks. The provider router connects to the CPE router by a point-to-point subnet assigned a public IPv6 prefix. The customer subnet, which connects the CPE router to end devices on the customer network, is also assigned a publicly routed prefix.

Many methods of network discovery exist, such as IPv6 scanning, passive collection, and hit lists, target end hosts [12]. However, this thesis surrounds the analysis of data pertaining specifically to the IPv6 periphery, which was collected via the methods described in Rye and Beverly's paper "Discovering the Network Periphery," called "edgy" [12]. Data was collected by sending hop-limited ICMPv6 packets using the high-speed randomized Yarrp topology prober [10]. Yarrp evades limitations of conventional active topology mapping via two key characteristics. First, Yarrp is stateless; by synthesizing information from ICMP replies as they arrive asynchronously, it increases parallelism and avoids the computational cost of per-trace state maintenance. Second, Yarrp complies with mandatory ICMPv6 rate limiting and avoids overloading routers and links with probe traffic by randomly permuting the *target IP address* x *Time to Live (TTL)* input domain via bijection provided by a symmetric RC5 block cipher with a 32-bit block size. The result is an efficient method of randomly probing <Target IP, TTL> pairs until the entire space is covered, which requires little communication overhead when distributed across various vantage points.

"Edgy" employs Yarrp with the specific goal of discovering devices on the IPv6 network periphery; often such devices are last-hop CPE routers connecting to end hosts as depicted in Figure 2.7. The "edgy" algorithm consists of an initialization stage followed by iterative rounds of probing. The initialization stage uses BGP and IPv6 hit list informed seed-traces to find candidate /48 prefixes based each trace's last responsive hop, effectively reducing the massiveness of the IPv6 address space and increasing the probability of finding responsive periphery devices. The probing stage then consists of randomly selecting IIDs within the candidate prefixes and sending Yarrp probes to each resulting target IP at progressively finer granularities until discovery threshold *n* is reached. The discovery threshold determines whether responsive prefixes will be subdivided for additional probing in the next round. Rounds of probing are conducted at /56, /60, /62, and /64 granularities. Using this method, "edgy" discovers over 37 million distinct IPv6 last hop addresses using BGP-informed seed, and over 29 million IPv6 last hop addresses using the hitlist informed seed.

Data collected using this algorithm revealed some concerning initial findings, including the discovery of approximately 16 million EUI-64 last hop addresses, accounting for approximately 42 percent of the total last hops discovered. However, these 16 million EUI-64 last hops were derived from only 5.4 million unique MAC addresses, meaning only 34 percent of them have unique IIDs. As a single MAC address generally correlates to a single device, the implication of this finding is that a corpus of 16 million EUI-64 IIDs somehow correspond to only 5.4 million discrete devices. Two root causes contribute to this phenomenon: provider-delegated prefix rotation and MAC address reuse. Each of these are examined in more detail in Chapter 4.

2.1.9 Prefix Rotation and Customer Prefixes

Much of this thesis deals with a specific phenomenon in EUI-64 based address assignment referred to colloquially as “prefix rotation”. The idea of prefix rotation builds on previous IPv6 security and privacy research, concerning the static nature of EUI-64 IIDs and the ability to track, and potentially predict, the address of CPE and client devices [7]. An EUI-64 IID will remain static even when a device is moved between networks, or when policy causes subnet prefixes to be reassigned. Certain operators implement policies which periodically assign new network prefixes, such as each time a device connects to the network. However, the EUI-64 IID remains static despite the assignment of a new prefix. The supposed security benefits of ephemeral addressing via prefix changes are thus trivially surpassed through enumeration of EUI-64 addresses which share the same IID over time [7].

In the case of CPE, the ISP is responsible for implementing such addressing policies. Previous research has revealed anecdotally that certain ISPs not only use EUI-64 based addressing, but also employ relatively predictable schemes for the reassignment of network prefixes. In such situations, the detriment to security and privacy is two-fold, as both halves of the IPv6 address (network prefix and IID) are now both trivially predictable rather than fully randomized. Recent work has shown that only a few ISPs perform regular prefix reassignments, including DTAG, Versatel, Netcologne, ANTEL, and Global Village [46]. Much of this thesis focuses specifically on identifying and quantifying prefix rotation behaviors in Versatel CPE.

2.2 Related Work

2.2.1 Evaluation of Vulnerabilities Across IPv6 Addressing Modes

Concerns related to privacy and security vulnerabilities in IPv6 addressing are not new, and have received increasing attention since the exhaustion of IANA’s IPv4 free pool in 2011. Although this thesis focuses on the continued use of EUI-64 based addressing in IPv6 infrastructure, it is important to contextualize this work within the greater dialogue of IPv6 privacy and security.

RFC 7721 [8] provides a comprehensive overview of vulnerabilities inherent to various forms of IPv6 address creation. The document details four categories of attack to which various forms of IPv6 address generation may be vulnerable: correlation of activities over time, location tracking, address scanning, and device-specific vulnerability exploitation. The relative privacy and security properties are evaluated across 8 forms of IPv6 address generation, including EUI-64 based IIDs (referred to in this document as IEEE-Identifier-Based IIDs); static, manually configured IIDs; constant, semantically opaque IIDs; cryptographically generated IIDs; stable, semantically opaque IIDs; temporary IIDs, DHCPv6 generation of IIDs; and transition and coexistence technologies.

Table 2.1. Privacy and security properties of IID generation mechanisms. Adapted from [8].

Privacy and Security Properties of IID Generation Mechanisms				
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 IPv6 link	No	No	No
Temporary	For temp address lifetime	No	No	No
DHCPv6	For lease lifetime	No	Depends on generation mechanism	No

Table 2.1 shows the susceptibility of each address generation method to each of the four attack categories. The degree of susceptibility is expressed in terms of the vulnerability "lifetime", or the period of time in which a given device addressed via a particular mode would be vulnerable to a specific category of attack. Note that the IEEE IID identifier mechanism is potentially vulnerable to all four categories of attack, which the RFC describes further.

2.2.2 Privacy and Security Vulnerabilities in EUI-64 Based Addressing

Correlation of activities over time enabled by EUI-64 IIDs last roughly for the lifetime of a device's network interface, thus allowing correlation on the order of years [8]. This concern is echoed in RFC 4941, [11] which recognizes that the use/re-use of a constant identifier over an extended period and across multiple independent activities always creates the possibility for activity correlation. Furthermore, as IPv6 addresses are a fundamental requirement of communication and must generally remain in the clear, the appearance of a constant identifier therein poses major challenges for attack mitigation [11]. Note that "frequently refreshing an IPv6 address may not mitigate correlation if an attacker has access to other longer-lived identifiers for a particular host", as is the case with network operators who regularly re-assign network prefixes for devices addressed with an EUI-64 IID [8].

Due to the static nature of EUI-64 IIDs, host-level location tracking is also possible. For example, a mobile host which repeatedly connects to the same server while employing an EUI-64 based addressing scheme reveals its movements via changes in the network prefix portion of its address. A more active variation of this attack is also possible. Suppose the attacker learns the target host's IID by connecting to the same IPv6 link, by running a server to which the target host connects, or by being on the path of the target host's communications. The attacker could then track the target host by sending a probe packet whose response would reveal the presence of the same IID on any given network [8].

Note also that the use of EUI-64 based identifiers can drastically reduce an attacker's target search space, nullifying many of the protections provided by the increased address space of IPv6. Several characteristics of EUI-64 IIDs contribute to the reduction of possible search space. First, the third and fourth bytes of EUI-64 SLAAC addresses always have a fixed value of 0xfffe, which reduces the total search space by 2^{16}. In addition, the first three bytes of the IID must correspond to a valid OUI, which indicates the device's NIC vendor. Since not all possible OUIs have been assigned, this further reduces the search space to only those addresses having valid, publicly assigned OUIs. Considerations pertaining to specific OUIs can narrow the pool of possible addresses even more. As some OUIs correspond to legacy devices, and therefore are unlikely to be used for IPv6 enabled systems, an attacker could safely omit such OUIs from a scanning attack. Additionally, the exact OUI in use by target network devices could be inferred via prior reconnaissance, drastically shrinking the possibility space of the attack. In the worst-case attack scenario, the original search space for a 64-bit IID is reduced from 2^{64} to n * 2^{24}, where n is the number of OUIs assigned to the target vendor [9]. Finally, since EUI-64 IIDs are derived from MAC addresses, it is not unlikely for addresses to be sequential within subnets, or for addresses to be nearby and/or sequential on other subnets in the same site [9]. Such simplistic addressing schemes are of particular interest to this study, as we seek to identify and quantify the predictability of prefix assignments among EUI-64 addressed Versatel CPE. Within Versatel networks, these devices are most often Fritz!Box devices [47].

The final attack category discussed in RFC 7721 is that of device-specific vulnerability exploitation. The use of IEEE MAC addresses in the generation of EUI-64 IIDs not only reduces the search space for scanning attacks, but also leaks information about target devices such as NIC vendor or operating system type. The OUI segment of an IEEE MAC address is publicly assigned, and coarsely identifies a device's manufacturer. Recent research has also developed methods for fine-grained device type and model predictions solely on the basis of their MAC address, using both the OUI segment and lower order bytes [43]. Such knowledge could be leveraged in a variety of ways by an attacker with knowledge of device/software-specific vulnerabilities. First, an adversary could obtain the IID for a particular target, then research and orchestrate an attack specific to the hardware or software type leaked by the address. Alternatively, the attacker could begin with knowledge of a device-specific attack, then scan the address space for addresses with EUI-64 IIDs whose composite OUIs reveal their vulnerability to the chosen attack [8].

A significant amount of research is dedicated to the discovery and mitigation of IPv6-specific vulnerabilities. In addition to the weaknesses in privacy and subsequent security risks discussed in RFC 7721, there also exist known cyber attacks which exploit SLAAC directly [48]. Choudhary, Barker, and Ullrich et al. all present work which pinpoint weaknesses in the trust model around which SLAAC operates [2], [49], [50]. The model allows any node to construct a globally routable address by obtaining a global prefix and a link local address through a combination of SLAAC and Neighbor Discovery protocols, with no approval or control. This presents the possibility of man in the middle and several types of Denial of Service (DoS) attacks. For example, a malicious node could decide to serve as a router on the link by sending router advertisements and responding to node solicitations until it is selected as an on-link router by another node, thus achieving a man in the middle position. This can be extended to a form of DoS attack, in which the router advertises invalid prefixes, causing new nodes to auto-configure with a bad address, rendering themselves unreachable [2], [49]. Other forms of DoS attack include preventing other nodes from acquiring a link-local address by falsely replying to DAD requests, taking legitimate routers offline by spoofing the target router's address and issuing a zero lifetime advertisement, and overwhelming a legitimate router's resources by flooding it with requests for addresses with invalid IIDs [49], [50]. Ullrich et al. [50] also systematically enumerate and evaluate the effectiveness of countermeasures for each of the known IPv6 vulnerabilities.

2.2.3 Tracking EUI-64 Addressed Devices

Recall that using a host's MAC address in the generation of an IPv6 IID causes the resulting address to be global in scope, and thus a party could potentially track the geographic location and monitor traffic of end user through their network enabled devices. Dunlop et al. [4] conducted geotemporal tracking of an EUI-64 addressed Android mobile device. Using a production network with different subnets covering different geographic areas, they created a tracking method which uses a script to continuously send echo requests to the different subnets on campus, and store the time and location of each echo reply. Correlating these replies over time provided a reliable method of geotemporal tracking. They also successfully isolated and collected all network traffic for a particular node by placing a sensor at the network border to collect all outbound IPv6 traffic, and using a packet sniffer to filter the traffic related to the node in question. As such, Dunlop et al. provide proof of geotemporal tracking and networking monitoring for target nodes with EUI-64 addresses.

They also contextualize these methods within a series of use cases, both malicious and otherwise. For example, geotemporal tracking and traffic monitoring could have legitimate applications in forensics and hostage rescue situations, but the same methods could be used for stalking and terrorist or adversarial reconnaissance. Furthermore, correlating geolocation and network traffic data could allow law enforcement to determine the location and identity of criminal perpetrators, but it could also allow criminal or adversarial groups to analyze social networks in order to isolate key individuals to target. Dunlop et al. also quantify the relative puissance of IID-based tracking against other methods, such as Global Positioning System (GPS). They note that IID-based tracking is often more resilient, as it does not rely on a single device (usually a mobile phone) running with specific settings enabled, but rather allows tracking of all Internet-capable, EUI-64 addressed devices carried by an individual. They also consider the philosophical question as to whether individuals are entitled to baseline levels of privacy, arguing that privacy is a basic right, regardless of an individual's intention or culpability. This prioritization of individual privacy is widely corroborated, both culturally and academically [51]–[53].

Dunlop et al. conclude their study by presenting alternative methods of IPv6 addressing, and evaluating their relative effectiveness in privacy protection. Methods discussed include IID obfuscation via Cryptographically Generated Addresses (CGAs) or SLAAC Privacy Extensions, the use DHCPv6, and the use of Encapsulating Security Payload (ESP) in Internet Protocol Security (IPSec). While each method does improve privacy protections over EUI-64 addresses, none accomplish this without caveat. The use of CGAs increases privacy and prevents specific types of DoS attacks, but incurs significant computational cost [4], [54]. Privacy extensions avoid this increased cost, and the lack of management overhead makes it a scalable solution, but the use of default parameters can result in addresses that remain static for up to a week, effectively nullifying the protections provided by an obfuscated IID [11]. DHCPv6 provides added convenience and automation, as a server would theoretically assign a new address upon each new client connection. However, this often does not occur in practice, and RFCs exist which discourage the use of temporary addresses [5]. Furthermore, local tracking is still possible via the DHCP Unique Identifier (DUID) which is communicated between the client and server [5]. Finally, the use of ESP in IPSec would provide strong protection against tracking as the target node's entire packet would be encrypted. Furthermore, cryptographic burden would be offloaded to tunnel endpoints. However, tracking would still be possible within subnets where tunnel endpoints reside, and the encryption method depends on a stable and secure Public Key Infrastructure (Public Key Infrastructure (PKI)) [49].

2.2.4 Defense and Mitigation for IPv6 Privacy

Many options exist for the generation of IPv6 addresses. Accordingly, much of the defensive cyber security dialogue in IPv6 is dedicated to identifying existing modes of address generation as defensive alternatives to other, more vulnerable methods. Carpene and Woodward [1] evaluate relative impact on privacy for multiple means of IPv6 address assignment. In addition to corroborating the concerns with EUI-64 based addressing discussed in 2.2.2 and 2.2.3, they also discuss potential pitfalls pertaining to SLAAC with privacy extensions, DHCPv6, and manual assignment. They argue that while the use of privacy extensions in SLAAC addressing prevents persistent, device-level traceability, the addressing method is still easily discernible amongst network traffic, as each address will effectively appear as a random hexadecimal string. Furthermore, they maintain that the use of privacy extension addresses can still reveal information about the network segment despite their randomized and ephemeral nature. For example, by monitoring communications for an extended period of time, calculating the life cycle of an IID, and cross-referencing with unique IIDs from a given network, an attacker could determine the number of unique hosts on a network segment [1]. The use of DHCPv6, Carpene and Woodward argue, can also be easily identifiable based on characteristic patterns of employment. They posit that a network administrator is likely to choose to allocate a range of addresses in the pursuit of job simplification. This would then result in an easily recognizable pattern in which a series of addresses with a shared network prefix also display incremental numbers. They also expect that users of manually configured addresses will similarly choose a simple representation, thereby facilitating simple identification and tracking of devices despite the massive scale of the IPv6 address space.

Despite the weaknesses inherent to other methods of IPv6 address generation, it is clear that vulnerabilities associated with EUI-64 SLAAC addressing are the gravest and most urgent. This idea is echoed in Ullrich's 2017 publication for RIPE Labs, which states that "more bits for addressing are a curse and a savior at the same time: on the one hand, more address space provides more flexible addressing, but could also lead to information leakage. The IPv6 address format chosen by a host or a network operator has significant impact on how well a node is protected against malicious actions by adversaries" [3]. The article updates the work of Carpene and Woodward by discussing newer methods of address generation such as the use of Semantically Opaque IIDs [55], which allows SLAAC addressing to be used without loss of security or privacy. RFC 7217 defines an alternative algorithm for automated IID generation, which produces IPv6 addresses which are stable within each subnet, but, unlike SLAAC with EUI-64 IID generation, will change the IID assignment as the host moves between different subnets. Ullrich also corroborates the concerns expressed in previous publications surrounding EUI-64 IIDs, noting that despite continuing discussions around its deprecation, the addressing mode is still widely used, especially among mobile systems and IoT devices [3].

Carpene, Woodward, and Ullrich suggest that although certain methods of IPv6 address generation are inherently weaker than others, mitigation of privacy and security concerns in IPv6 is not necessarily as simple as merely substituting one addressing mode for another. Plonka et al. [56] present a method of detecting and locating the remote monitoring of IPv6 nodes, regardless of the addressing mode employed, via nonce-based inverse surveillance. The strategy consists of dissemination and propagation phases, in which researchers actively spread 64-bit nonces embedded in IPv6 source addresses via a yarrp-based traceroute survey then passively listen for the propagation of those nonces through activities which indicate remote surveillance. Activities of interest include receipt of unsolicited packets in response to nonced source addresses, entries in DNS Database (DNSDB), or the execution of reverse DNS queries. Plonka et al. captured this activity by configuring an open-source DNS server to act as the authoritative reverse DNS name server for the /36 IPv6 address block used to make nonced source addresses in the dissemination stage. Results of this study indicate that this method is capable of detecting suspicious reverse DNS queries and TCP connection attempts, improving reachability measurements, detecting unauthorized sharing of passive DNS data, and detecting eavesdropping.

2.2.5 Measuring Prefix Rotation in RIPE Atlas Data

Padmanabhan et al. [46] pursue adjacent goals to this work, characterizing the behavior of EUI-64 addressed IPv6 nodes through analysis of RIPE Atlas IP echo data [57]. Using EUI-64 addressed CPE, they regularly probed the RIPE Atlas IP echo server and collected the returned IPv6 addresses from August 2014 to December 2019. In analyzing this data, they found that IPv6 address lifetimes tended to be on the order of months, with only a few AS's performing reassignments, including DTAG, Versatel, Netcologne, ANTEL, and Global Village. Pandamanabhan et al. also examined the frequency and distance of EUI-64 prefix rotations by analyzing common prefix lengths between successive addresses. In doing so, they found that common prefix lengths vary according to the particular CPE and ISP in question, suggesting that the combination of these properties determines prefix assignments. Furthermore, they suggest that this knowledge could potentially be leveraged to track or predict prefix assignments for EUI-64 addressed nodes.

2.2.6 Contributions

Previous work has established the privacy and security risks associated with SLAAC based addressing and the use of EUI-64 IIDs, and suggests viable methods of mitigation. While EUI-64 addressing is increasingly uncommon among IPv6 clients, recent work shows it to be prevalent in IPv6 infrastructure, particularly among CPE devices. As of 2018, approximately 45 percent of all IPv6 interface addresses employed EUI-64 based addressing [10]. As such, this thesis seeks to further analyze patterns of EUI-64 addressing in IPv6 architecture, with particular focus on the Versatel ISP.

Contributions of this work to the larger field of IPv6 research include an in-depth assessment of Versatel prefix assignment behaviors as well as methods of analysis that can be expanded in the examination of other ISPs, or the IPv6 infrastructure at large. The work also provides quantitative characterizations of EUI-64 addressed devices in terms of device manufacturer. Finally, it suggests potential methods of prefix pattern identification and prediction of future prefix assignments.

CHAPTER 3:
Methodology

3.1 Input Data

3.1.1 Data from the IPv6 Network Periphery

This thesis is motivated by results from a recent novel IPv6 network measurement technique that discovered areas of the IPv6 topology about which little was previously known [12]. One observation noted from this prior work was the phenomenon where particular IIDs are seen in multiple prefixes of a provider over time due to the provider employing mechanisms to "rotate" the prefix. A subsequent focused measurement campaign was performed by Rye (the same author of [12]) to specifically help shed light on this rotation phenomenon.

To study prefix rotation, Rye focused on a single service provider, Versatel, that was empirically observed to exhibit rotation. Versatel (AS8881), which also goes by the brand names "1&1" and "1&1 Versatel" [58], is a German access provider [59] that provides Fiber to the Premises to business and residential customers. As Versatel presented the opportunity for a detailed examination of EUI-64 prefix changes, the methodology discussed in this chapter is applied specifically to last hop data collected from Versatel networks using the methods described as follows.

Versatel prefixes were probed during a 126 day span, between November 7, 2019, and March 12, 2020. Rye selected 700 /48 prefixes of Versatel to be exhaustively probed. Each day, a single vantage point uses Yarrp to probe a random address within all constituent /64s of the 700 /48 Versatel prefixes. Target destination addresses were created by concatenating each /64 prefix with a different randomly generated IID. Each /48 therefore produced 2^{64-48} or 65,536 target destination addresses for a total of 45,875,200 target destination addresses across all 700 /48s. These targets are fed to Yarrp, which was seeded with the same random seed value each run, meaning that each (destination, TTL) pair was probed in the same order, and at approximately the same time, each day. Probing was initiated by a cron job to begin at the same time each day and Yarrp used the same random seed in each run to maintain a deterministic random probing order. Probing was performed using ICMPv6 packets, as they are generally considered less abusive than TCP or UDP packets, and ICMP is meant for diagnostic and error reporting. Yarrp was configured to probe at 10,000 packets per second such that all of the probing would complete in a single day, without overwhelming infrastructure close to the vantage point.

Rye coordinated with the operator of a well-connected server in Lausanne, Switzerland in order to use it as the vantage point for Yarrp probing. Since Yarrp randomizes the order in which it probes target addresses, it distributes the load across probed networks. He also set the neighborhood TTL setting to 5 in order to minimize impact on nearby infrastructure. During the course of data collection, Rye maintained a web page with opt-out instructions at the source address of probe packets, but nonetheless received no opt-put requests.

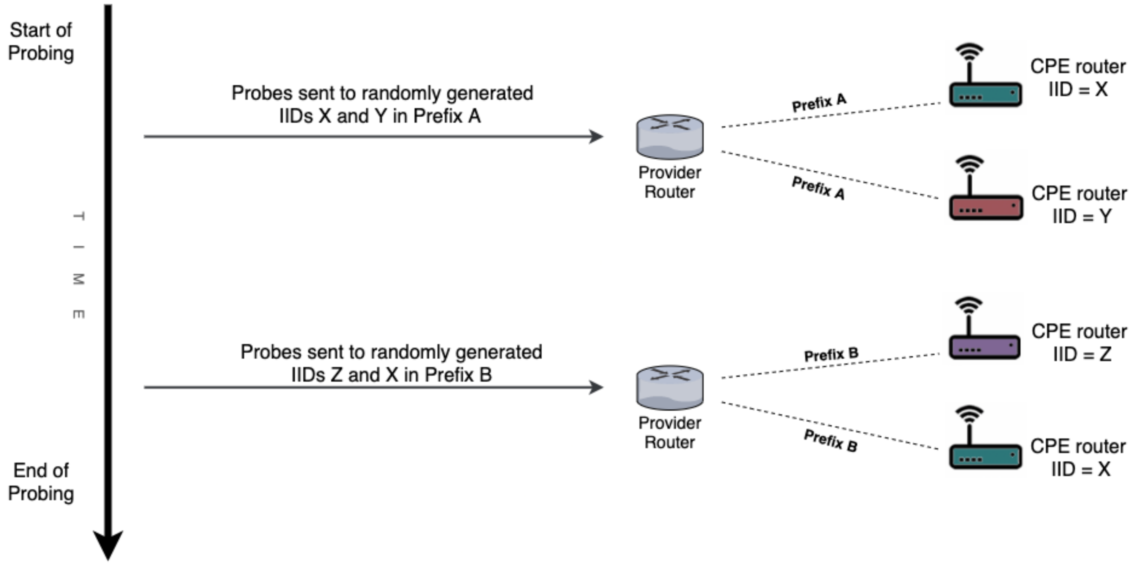


Figure 3.1. Illustration of probing a single CPE twice on the same day due to prefix movements

It is important to note for discussions our analysis of results in Chapter 4 that the probing method used may sometimes reach the same CPE both before and after a prefix change, resulting in multiple observations of the same IID on the same day. Figure 3.1 illustrates this phenomenon. Conversely, an IID could evade observation if it happens to change its prefix to one that has already been probed before its previous prefix is reached.

3.1.2 Open Source Data

In addition to the primary corpus of data provided through the processes described in Sections 2.1.8 and 3.1.1, certain analyses involved the use of the IEEE's open-source database of MAC address assignments [60]. Use of this database allowed for the reverse lookup of device manufacturers based on the three-byte OUI portion of the MAC address from which EUI-64 addresses are derived, as detailed in Sections 2.1.5 through 2.1.6. The data is available for download as a text or csv file, and contains two columns – the 3-byte hexadecimal OUI and the name and address of the organization to which the identifier is assigned.

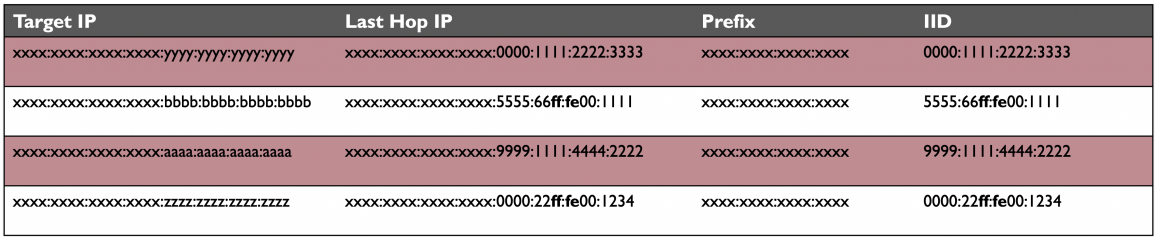
3.2 Analytical Tools and Workflow

The input data set consists of approximately 519 Gigabytes of IPv6 last hop records. The amount of data and the computational intensity of analyses in question were sufficient to justify the use of distributed computing tools. The preliminary stages of this work therefore involved transferring data to the Hadoop Distributed File System which is native to the Grace node of Hamming, the Naval Postgraduate School (NPS) high-powered computing cluster [61]. Subsequent analyses could then be performed using PySpark, the Python API for Apache Spark, a distributed cluster computing framework for programming with implicit data parallelism. Code written in PySpark is deployed to data for in-place processing across multiple distributed compute nodes on Grace. This allows for large amounts of data to be efficiently processed and manipulated as required by the research question at hand. As such, efforts upfront are rewarded with speed and flexibility later on. Communication of results necessitated visualizations, which were created using a combination of the plotly package in Python [62] and the ggplot package in R [63].

3.2.1 Pre-processing

A number of considerations came into play during initial processing of the data. The first and most evident of these was to filter the data such that only records related to EUI-64 last hops are retained. Filtering was performed on a per-traceroute basis. The last hop record correlates to the response received from a CPE device connecting provider routers with the end hosts to which the trace route probes are sent, as described in Section 2.1.8.

We searched for the characteristic ff:fe at the fourth and fifth bytes of an IPv6 IID, and filtered out any row in which these bytes are not found. Figure 3.2 illustrates the filtering process for the input data frame, where rows colored in red would be removed while rows in white would be retained.



Target IP	Last Hop IP	Prefix	IID
xxxx:xxxx:xxxx:xxxx:yyyy:yyyy:yyyy:yyyy	xxxx:xxxx:xxxx:xxxx:0000:1111:2222:3333	xxxx:xxxx:xxxx:xxxx	0000:1111:2222:3333
xxxx:xxxx:xxxx:xxxx:bbbb:bbbb:bbbb:bbbb	xxxx:xxxx:xxxx:xxxx:5555:66**ff:fe**00:1111	xxxx:xxxx:xxxx:xxxx	5555:66**ff:fe**00:1111
xxxx:xxxx:xxxx:xxxx:aaaa:aaaa:aaaa:aaaa	xxxx:xxxx:xxxx:xxxx:9999:1111:4444:2222	xxxx:xxxx:xxxx:xxxx	9999:1111:4444:2222
xxxx:xxxx:xxxx:xxxx:zzzz:zzzz:zzzz:zzzz	xxxx:xxxx:xxxx:xxxx:0000:22**ff:fe**00:1234	xxxx:xxxx:xxxx:xxxx	0000:22**ff:fe**00:1234

Figure 3.2. Example data frame illustrating filtering for EUI-64 IIDs

Note that this method does allow for a slight possibility of false positives, given the possibility that a randomly assigned IID could technically match this specific pattern at the byte positions in question. However, the frequency of such addresses can be considered statistically insignificant; it is estimated that the possibility of misidentifying an EUI-64 address is about 1 in 2^{16} [36].

A significant focus of this study involves tracking of individual MACs and IIDs, so this information needed to be extracted in bulk from the EUI-64 last hop addresses observed in the input data. This was achieved via a two-step step process. Obtaining the IID for each last hop address was a simple matter of splitting the 128 bit hexadecimal last hop address in half, such that the latter four hextets are retained. Conversely, a column containing the 64-bit network prefix for each EUI-64 last hop address was also created using the same "split point", this time retaining the four most significant hextets.

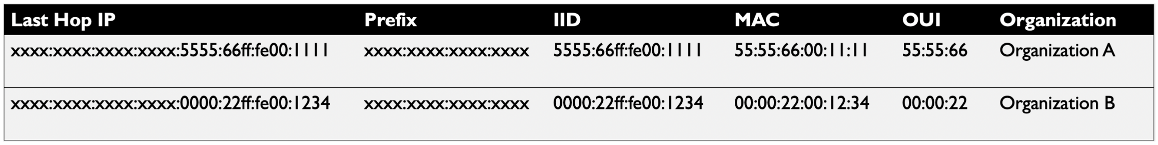
Next, to derive the MAC, we split the IID into three chunks – the 3-byte OUI, the 2-byte ff:fe separator, and the 3-byte NIC. The separator was then discarded and we concatenated the OUI and NIC to form a MAC address. Finally, recall that some EUI-64 addresses are given a local scope by flipping the local/global bit, as described in Section 2.1.6. In order to reconstruct the correct MAC address for locally-scoped MAC addresses, we converted each MAC address to its binary representation, checked the value of the global/local bit, and flipped its value to zero if it had been set to one.

Performing simple manipulations in advance across the full data set facilitated later analyses, as the basic blocks of information had been neatly placed into a columnar representation. The “what” and the “who” are encapsulated in the IID and MAC address columns, while the “where” and the “when” are indicated by the SLAAC prefix and date/timestamp columns.

3.2.2 Manufacturer Classification

The input data records the addresses of IPv6 last hop CPE devices, namely routers. Many end users perform little customization when configuring their home network, therefore default addressing behaviors are likely to be expressed en masse within the input data set. Further, using the methods described in this section, such default behaviors can potentially be mapped to the manufacturer of the device in question.

Creating this mapping involved a multi-step process using both data sets discussed in Section 3.1.1 and 3.1.2. First, the IEEE OUI data set was stripped for relevant information such that only the assigned OUI and organization name were retained, and the associated address and contact information were dropped. We added a new column to the input data frame by splitting each MAC address in half, and retaining the first three hexadecimal octets. Next, the open-source IEEE OUI assignment data was read in and cleaned such that OUIs in each of the respective data frames followed a standard format using colon separation and lowercase alphanumeric characters. Finally, the two dataframes were merged by performing a join on the OUI column. The resulting dataframe encompassed all information contained in the original input data set, with the addition of the device manufacturer associated with each EUI-64 IID. Figure 3.3 shows an example of the processed data frame. Information synthesized from this data frame is presented in Section 4.1.1.



Last Hop IP	Prefix	IID	MAC	OUI	Organization
xxxx:xxxx:xxxx:xxxx:5555:66ff:fe00:1111	xxxx:xxxx:xxxx:xxxx	5555:66ff:fe00:1111	55:55:66:00:11:11	55:55:66	Organization A
xxxx:xxxx:xxxx:xxxx:0000:22ff:fe00:1234	xxxx:xxxx:xxxx:xxxx	0000:22ff:fe00:1234	00:00:22:00:12:34	00:00:22	Organization B

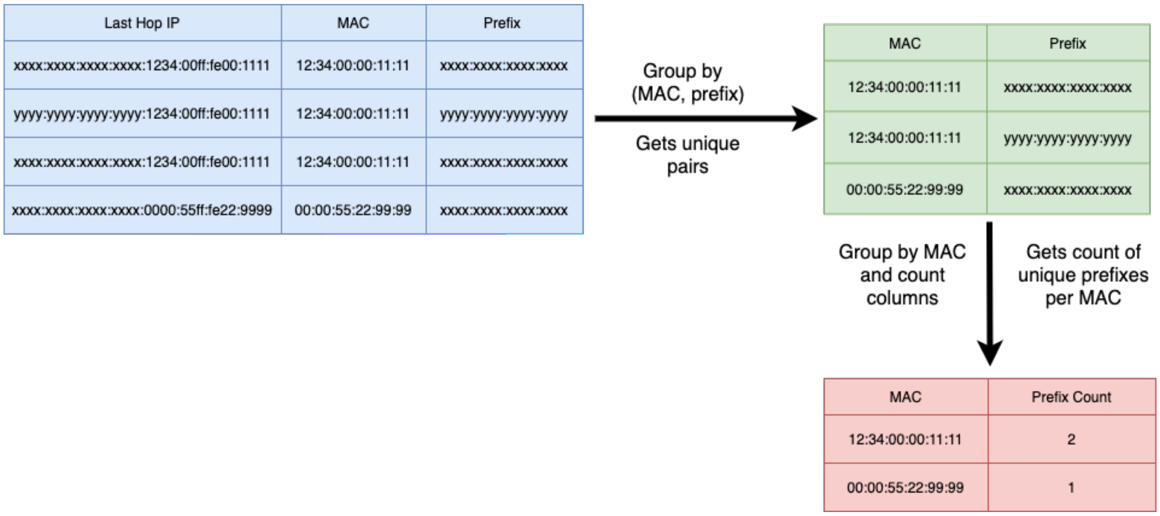
Figure 3.3. Sample data illustrating pre-processing outcome data frame

3.3 Measurement of Prefix Rotation

Given the difference between the number of unique last hop addresses and unique MAC addresses observed in the “edgy” data, one area of interest for this study was creating a detailed characterization of prefix rotation behaviors, as described in Section 2.1.9. As such, the following sections explain the methods used to perform quantitative analysis of prefix changes on a per-IID basis. This discussion is subdivided based on two investigative themes: temporal and spatial.

3.3.1 Temporal Characterization

To understand how prefixes change for a given IID over time, we determined the total unique prefixes per IID over the 126 days represented in the input data set. We begin by reading in a data frame containing a MAC and prefix column for each last hop, we then double group the data frame by MAC, followed by prefix. The result is a data frame filled with rows of distinct <MAC, prefix> pairs. We then perform a subsequent group operation on the MAC column of this intermediate data frame, and count all rows for each MAC group. The output of this group and count provides the total unique prefixes per distinct MAC address. Figure 3.4 illustrates this process.



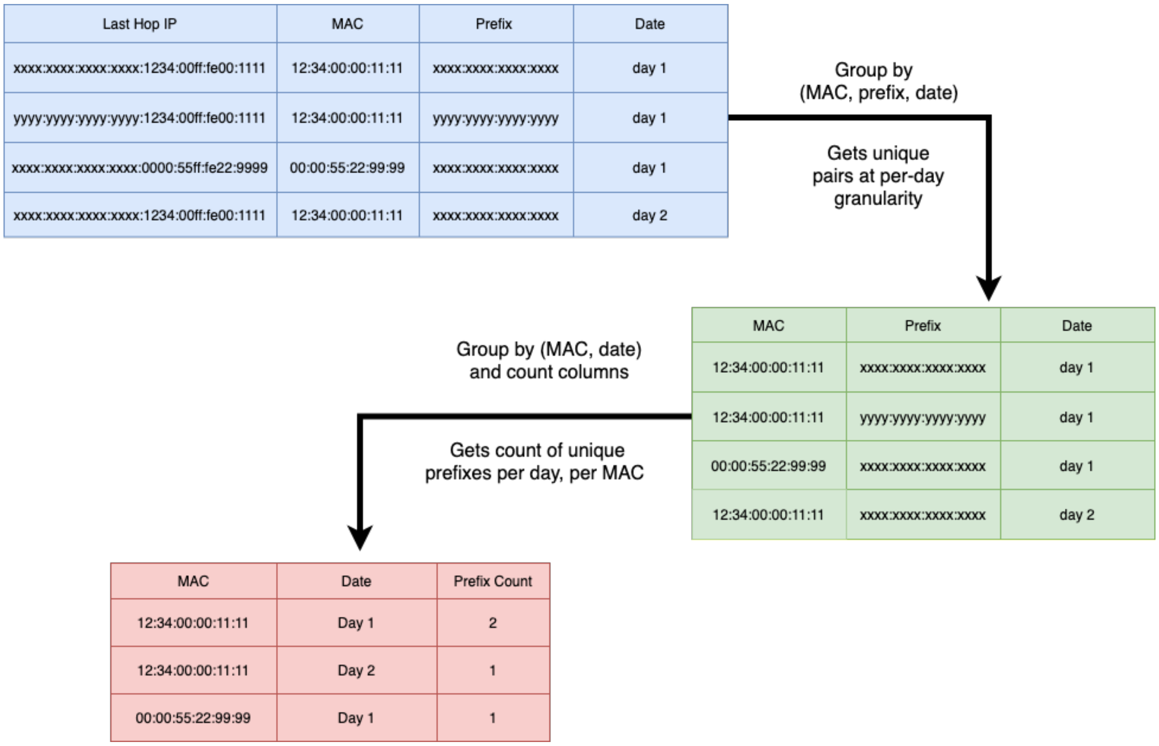
Last Hop IP	MAC	Prefix
xxxx:xxxx:xxxx:xxxx:1234:00ff:fe00:1111	12:34:00:00:11:11	xxxx:xxxx:xxxx:xxxx
yyyy:yyyy:yyyy:yyyy:1234:00ff:fe00:1111	12:34:00:00:11:11	yyyy:yyyy:yyyy:yyyy
xxxx:xxxx:xxxx:xxxx:1234:00ff:fe00:1111	12:34:00:00:11:11	xxxx:xxxx:xxxx:xxxx
xxxx:xxxx:xxxx:xxxx:0000:55ff:fe22:9999	00:00:55:22:99:99	xxxx:xxxx:xxxx:xxxx

MAC	Prefix
12:34:00:00:11:11	xxxx:xxxx:xxxx:xxxx
12:34:00:00:11:11	yyyy:yyyy:yyyy:yyyy
00:00:55:22:99:99	xxxx:xxxx:xxxx:xxxx

MAC	Prefix Count
12:34:00:00:11:11	2
00:00:55:22:99:99	1

Figure 3.4. Visualization for calculation of total prefixes per unique IID

To further refine the findings from the previous analysis, we calculated the number of unique prefixes observed per MAC, per day. We begin by reading in a data frame with MAC, prefix, and date columns for each last hop collected as described in 3.1.1. We then triple group the data successively by MAC, prefix, and date. This produces an intermediary data frame composed of rows with unique <MAC, prefix, date> combinations. This data frame is then double grouped by MAC then date, allowing prefixes observed per MAC to be distinguished by their date of observation. Finally, we count the rows for each resulting double-group, outputting the total unique prefixes observed per day for a given MAC. Figure 3.5 illustrates this process.



Last Hop IP	MAC	Prefix	Date
xxxx:xxxx:xxxx:xxxx:1234:00ff:fe00:1111	12:34:00:00:11:11	xxxx:xxxx:xxxx:xxxx	day 1
yyyy:yyyy:yyyy:yyyy:1234:00ff:fe00:1111	12:34:00:00:11:11	yyyy:yyyy:yyyy:yyyy	day 1
xxxx:xxxx:xxxx:xxxx:0000:55ff:fe22:9999	00:00:55:22:99:99	xxxx:xxxx:xxxx:xxxx	day 1
xxxx:xxxx:xxxx:xxxx:1234:00ff:fe00:1111	12:34:00:00:11:11	xxxx:xxxx:xxxx:xxxx	day 2

MAC	Prefix	Date
12:34:00:00:11:11	xxxx:xxxx:xxxx:xxxx	day 1
12:34:00:00:11:11	yyyy:yyyy:yyyy:yyyy	day 1
00:00:55:22:99:99	xxxx:xxxx:xxxx:xxxx	day 1
12:34:00:00:11:11	xxxx:xxxx:xxxx:xxxx	day 2

MAC	Date	Prefix Count
12:34:00:00:11:11	Day 1	2
12:34:00:00:11:11	Day 2	1
00:00:55:22:99:99	Day 1	1

Figure 3.5. Visualization for calculation of prefixes per day, per unique IID

For the purpose of this study, we define one day as the 24-hour period between two consecutive dates during the data collection period. As described in Section 3.1.1, each day of data collection began at the same time, but end times may have varied within a 2-hour window depending on the performance of data collection tools for a given day. Recall also that probing was run toward a given /64 prefix once per day.

The two analyses described above present a macroscopic view of overall prefix rotation frequency. We then performed further investigations to better understand the cycling behaviors of EUI-64 IID prefix changes. The first of these measured the time passed between repeated observations of a given prefix per unique IID. Figure 3.6 shows a simplified example of the measurement performed in this analysis.

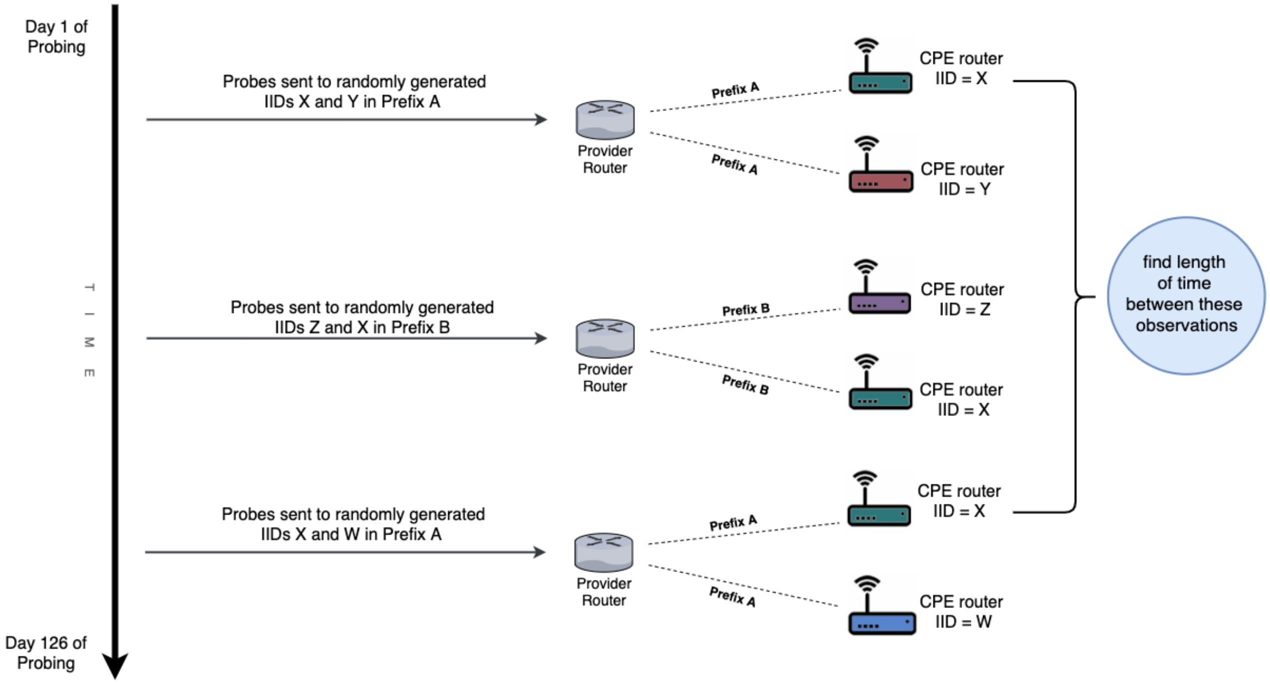


Figure 3.6. Illustration for measurement of intervals between repeated prefix observations for a given IID

To perform this analysis, we began by reading in a data frame composed of prefix, MAC, and date columns for each last hop, and then converting the date column from a string to a Unix timestamp. Next, we added on a column containing the concatenation of the prefix and MAC fields in each row, which acted as a unique ID for each <MAC, prefix> tuple. We then group on this new ID column and sort by date to create a "date lag" column, which features the value in the date column from the row immediately above it. The resulting data frame consisted of rows of chronologically ordered observations of a given MAC in a specific prefix. Next, we created a new column by taking the difference between the date and "date lag" columns. The values of the resulting column represent the time elapsed between prefix repetitions per MAC, and therefore per EUI-64 IID. Note that this method records every interval between any prefix repetition for each unique IID; which could potentially return up to 126 records for a single IID. In order to make this data more easily digestible, we took the mean value for all inter-repetition periods associated with a particular IID, and rounded this value to the nearest integer. This results in a single record for each unique MAC observed repeating its prefix at least one time.

The next series of experiments examine the same phenomenon of alternating prefixes from a different angle, focusing instead on the number of consecutive days spent in a single /64 per unique EUI-64 IID. The first of these experiments measures the amount of time in which a specific prefix remains "empty", or devoid of EUI-64 IIDs, after a given IID moves to another prefix. For this experiment we start off with a data frame composed of MAC, date, and prefix columns. We then group on the prefix and date columns to find distinct prefix observations for each date captured within the input data set. Next, we converted the date column from a string to a timestamp, and created a "date lag" column using a window function that partitions on prefix and orders by date. We use our previously described method of obtaining a "date diff" column from the date and "date lag" columns, and convert the resulting integer values from seconds to days. We then subtract 1 from all non-zero values in the "date diff" column, such that the resulting value represents the number of full 24 hour periods in which no EUI-64 IID is observed for a given prefix. Finally, by grouping on the prefix column and taking the mean of the "date diff" column for each group, we find the mean length of time a given prefix is consecutively empty.

In addition, we analyzed the reverse question, measuring instead the consecutive lengths of time in which a given prefix is occupied by some EUI-64 IID. We began by reading in an intermediary data frame from the previous analysis featuring prefix, date, "date lag", and days empty columns. We then created a temporary ID column by concatenating the contents of the days empty and prefix columns at each row. Next, we created a new column composed of a running count of consecutive observations of <days empty, prefix> pairs. This column indicates the consecutive lengths of time in which a given IID is observed empty for zero days, or stated otherwise, it measures the length of periods of consecutive occupation. We then filtered the data frame such that all rows in which the days empty column showed a non-zero value were discarded. This step effectively removes all records of "gaps" in EUI-64 occupation for a given prefix, while retaining the periods of consecutive occupation as well as their unique identifiers produced by the running count operation. Next, we create an "occupied duration" column by performing a count on the ID column. This returns the number of rows in which an ID value is observed, therefore counting the consecutive days of EUI-64 occupation per prefix. Finally, to find the mean period of EUI-64 occupation for each prefix, we group on the prefix column and find the mean value within each distinct group. The results of this analysis allowed us to determine the mean lengths of consecutive

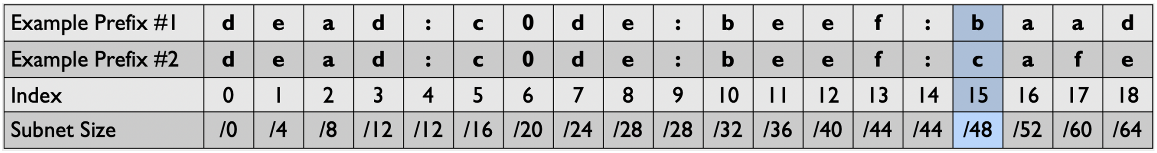
EUI-64 IID occupation on a per-prefix basis.

In order to further refine the results of the precedent analysis, we followed a similar workflow to calculate the duration periods in which a prefix is occupied by a specific IID. In contrast to the previous iteration, this experiment distinguishes between unique EUI-64 IIDs, so as to determine whether periods of consecutive occupation are composed of several successive IIDs, or a single unique IID. We began with data featuring MAC, date, and prefix rows. We then grouped the data frame by prefix, MAC, and date to obtain chronologically distinct observations of unique prefix and MAC pairs. We then converted the date column from a string format to a Unix timestamp. By partitioning on prefix and MAC and ordering by date, we created a "date lag" column featuring the value of the date row immediately precedent. Next, we added a new column containing the difference between the date and "date lag" column in seconds. We convert the seconds in this column to days by dividing each value by the number of seconds in a day. We subtract one from all non-zero values in this column, similar to the process described in the previous experimental iteration. The resulting column represents the number of consecutive 24 hour periods in which a given prefix and MAC combination goes unobserved. We then perform a series of transformations which are almost identical to those used in our previous iteration. The only modification was to change every grouping operation to include the MAC column in addition to the prefix column. The resulting output is a data frame featuring the mean duration in days per unique prefix in which consecutive occupations by a single unique MAC have occurred.

3.3.2 Spatial Characterization

The contents of this subsection discuss the methods used to measure the distance traveled by a given EUI-64 IID within the IPv6 address space if and when its prefix changes. For the purpose of this analysis, distance is defined as a function of the similarity between two last hop addresses, which is correlated to a smallest common subnet which encompasses the two addresses. Spatial analyses build upon insights provided by temporal analyses, by improving our understanding of not only when and how often prefixes change for a given IID, but also how similar one prefix is to the next. This ultimately allows patterns and anomalous behaviors in prefix assignments to be more easily identified.

First we take a high-level view by calculating the total range of prefix travel per IID. We begin by reading in a data frame with MAC, last hop, date, and prefix columns. We then create a new column by converting the contents of the prefix column from a hexadecimal string to an long integer representation of the same value. Next we group the data by MAC, and create columns featuring the maximum and minimum long integer prefix values for each group. We then make two more columns in which the maximum and minimum long integer prefixes are converted back to their hexadecimal string representations via a PySpark User Defined Function (UDF). Finally, we define the encompassing prefix for each IID's maximum and minimum prefix. We created a function which takes as input the hexadecimal string representations of each maximum and minimum prefix. It locates and saves the index of their least significant common character, then uses a dictionary lookup to associate this index with an IPv6 subnet size. Figure 3.7 shows an example prefix with array indices and their associated subnet sizes, as would be used for each dictionary lookup. Moving from most to least significant digits, the two sample prefixes first differ at the 15th index of the two character arrays. This index would be associated with a /48 common subnet size in the dictionary lookup.



Example Prefix #1	d	e	a	d	:	c	0	d	e	:	b	e	e	f	:	b	a	a	d
Example Prefix #2	d	e	a	d	:	c	0	d	e	:	b	e	e	f	:	c	a	f	e
Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Subnet Size	/0	/4	/8	/12	/12	/16	/20	/24	/28	/28	/32	/36	/40	/44	/44	/48	/52	/60	/64

Figure 3.7. Illustration of prefix index and subnet size association

The resulting column expresses the total range of prefix travel per IID in terms of common subnet size. Note that this method will "round up" to the nearest nibble boundary in cases where the two prefixes in question align outside the nibble boundary. We consider such instances to be edge cases, as RIPE NCC, prescribes that delegated IPv6 prefixes always be a multiple of 4 [64].

We enhance the specificity of this analysis by calculating the common subnet on a per-change basis for each IID. We begin with the data set described in 3.1.1, which features MAC, last hop, date, and prefix columns, and then sort the data frame by MAC then date. Using a window function which groups by MAC and sorts by date, we create a prefix lag column to contain the prefix value from the precedent row. Then, we use the same function for finding the common subnet of two IPv6 prefixes as described in the previous analysis, this time using the prefix and prefix lag columns as input. The resulting column resulting column reflects the common subnet between two consecutive rows, and consequently between two consecutive observations of a given IID traveling through prefixes.

CHAPTER 4: Analysis of Results

This chapter discusses quantitative and qualitative analysis of EUI-64 addressing behaviors. All findings discussed in the following sections are unified by the high-level goal of characterizing EUI-64-based addressing behaviors in IPv6 networks within the context of security and privacy. Pursuing this goal involved an iterative process in which the breadth of investigation became increasingly specific as research progressed. As such, the information presented below results from a series of analyses performed with varying scopes in terms of input data. The scope of the input data used will be noted with each discussion, as it influences the conclusions that may be drawn from resulting outputs.

4.1 Quantitative Analysis of EUI-64 Devices

The following section concerns the results of quantitative analyses, which are guided by the following research questions:

1. Given IPv6 prefixes with many EUI-64 addresses, what is the distribution of device manufacturers and/or models therein?
2. What is the IPv6 prefix rotation behavior in terms of frequency, locality, regularity, and entropy of prefix assignments?

Synthesis of results are organized roughly by the scope of the question at hand and the input data used to answer it.

4.1.1 Manufacturers of EUI-64 Addressed Devices

Our examination of OUIs among EUI-64 addressed Versatel CPE revealed high levels of homogeneity for device manufacturers. The results of this analysis are shown in Table 4.1.

Table 4.1. Manufacturers for EUI-64 addressed CPE in Versatel networks

Manufacturers for EUI-64 addressed CPE in Versatel networks		
CPE Manufacturer	Percent of Versatel EUI-64 Addresses	Count of Unique MACs
AVM GmbH	84.4	1,243,699
Huawei	15.6	229,291
Other Vendors	.03	535

Table 4.1 shows that near 84 percent of Versatel CPE examined in this study were produced by the same manufacturer, suggesting that AVM GmbH CPE implement SLAAC with EUI-64 as a default addressing scheme. The second most common manufacturer was Huawei, accounting for only 15 percent of EUI-64 addressed CPE in Versatel networks. The "Other Vendors" category includes a set of 32 different CPE vendors, whose collective presence in Versatel subnets account for approximately .03 percent of the total devices observed in the input data. AVM GmbH produce the Fritz!Box line of residential gateway devices, the sales of which have captured a significant portion of the overall market share for CPE in Germany. The rightmost column of Table 4.1 shows the count of unique CPE MACs which responded to edgy probes.

4.1.2 Frequency of Prefix Rotations

This subsection reviews the results of quantitative analyses surrounding temporal characterization of prefix rotation behaviors for EUI-64 IIDs. Recall from Section 2.1.9 that prefix rotation is not universal to all ISPs. In addition, prefix rotation schemes may vary from provider to provider, making generalized quantification difficult in both experimental design and verification. In the interest of scoping and accuracy, the results discussed in this subsection therefore pertain specifically to last hop addresses residing within Versatel networks.

The first analysis shows a broad view of the overall frequency of prefix changes through examination of the total number of unique prefixes observed per IID among Versatel EUI-64 last hops.

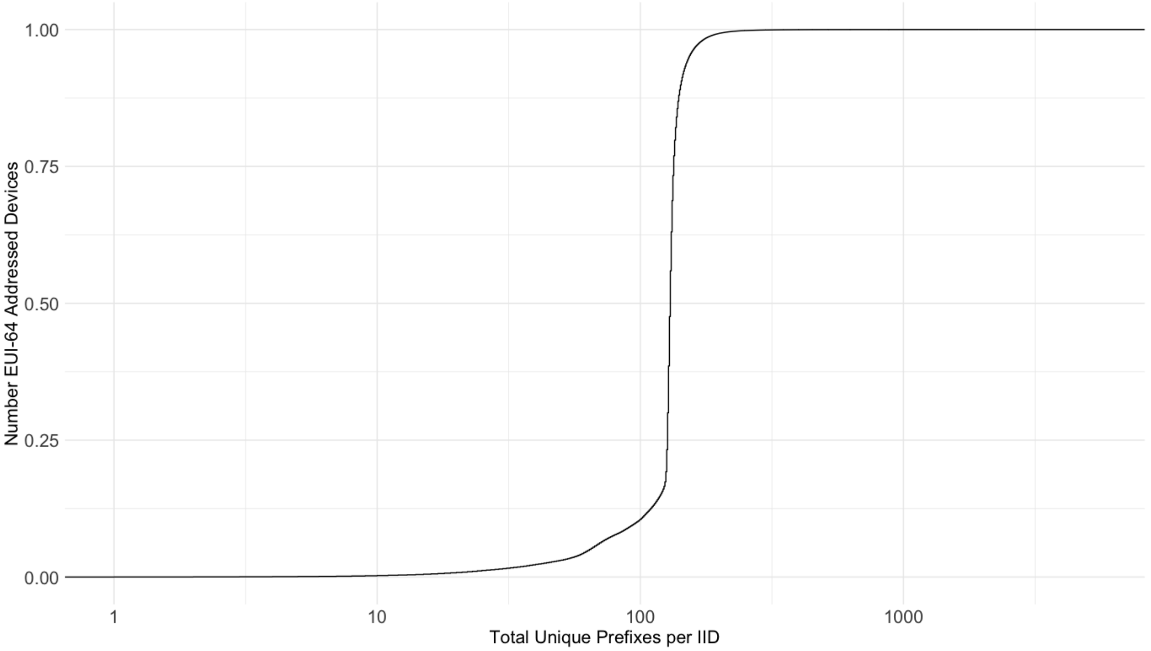


Figure 4.1. CDF – total unique prefixes per IID

Recall that the input data was collected over a period of 126 days, which accounts for the spike just past 100 on the x-axis of Figure 4.1. Note also that less than one-eighth of Versatel IIDs were observed in fewer than 100 unique prefixes over 126 days. This indicates not only a general trend of daily prefix reassignment, but also one of new, unique prefix assignments. The distribution of unique prefixes features outliers on the opposite end of the graph.

The mean and the mode value for unique prefixes veer toward the right of the graph, which has a mean value of 125.87, and a mode of 126. The range of unique prefix counts per IID is large, with a minimum value of 1 and a maximum value of 5376. This suggests that while some IIDs remain static in a single prefix, others are present in numerous prefixes per day. The existence of this latter group implies that such IIDs must either churn through many prefixes per day, or be derived from devices with non-unique MAC addresses.

Furthering this line of inquiry, the next analysis determines the number of unique prefixes observed daily per IID. Note that due to the probing methods described in Section 3.1.1, we may see the same (prefix, IID) pair multiple times in a day. For example, a customer have a /56 allocation, within which we would probe each /64 but potentially receive responses from the same last hop CPE each time.

Figure 4.2 and Table 4.2 both reference the average number of unique prefixes observed per day for each unique EUI-64 IID, measured over a 126 day period. Figure 4.2 confirms that the data is distinctly biased toward values between 1 and 2, as the graph features a long tail extending to the right and a spike to the right side of the 1 mark on the x-axis. It can therefore be concluded that the majority of Versatel EUI-64 IIDs appear to occupy more than one prefix per day. However, the methods of data collection used allow for the possibility that any given CPE may be probed both before and after a daily prefix change, as described in Section 3.1.1 and illustrated in Figure 3.1. Recall also from Section 3.1.1 that the opposite can occur, and a given IID may be missed completely were it to change prefixes before its previous prefix is probed. Given this additional consideration, this indicates that majority of EUI-64 addressed Versatel CPE change prefixes approximately once a day.

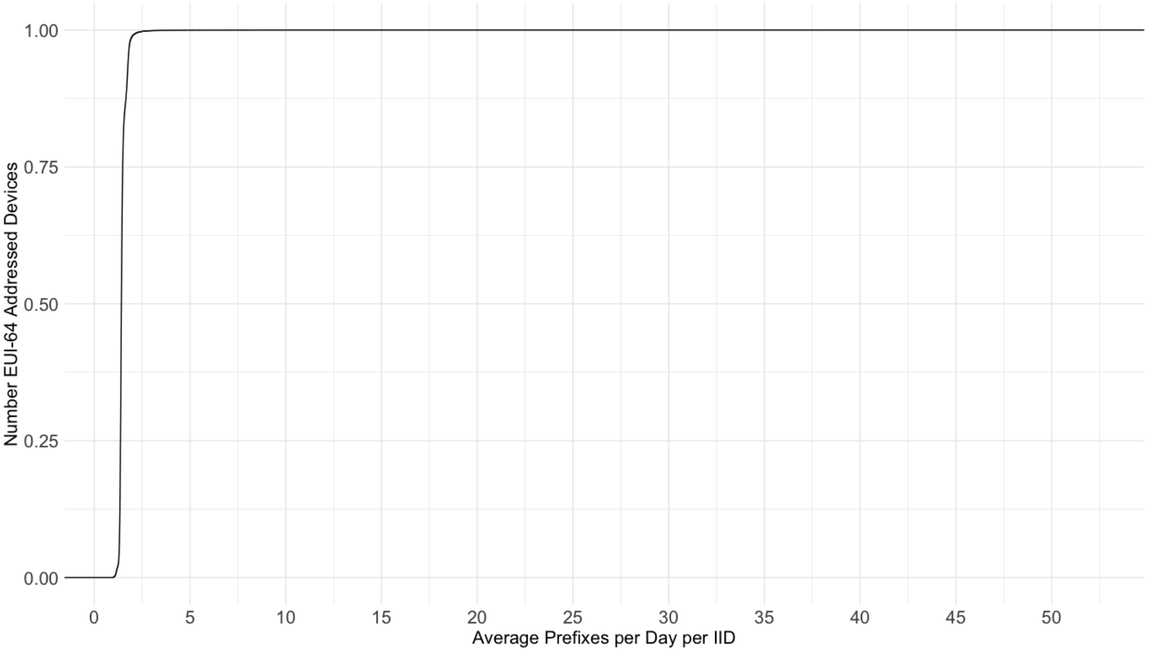


Figure 4.2. CDF – total unique prefixes per IID, per day

Table 4.2 corroborates this inference, as it shows a median prefix per day value of 1.41, and a mean value of 1.46. However, there are also a small number of IIDs observed in exactly one prefix per day. It is inconclusive whether such IIDs change prefixes exactly once per day or if they remain in the same prefix over time. However, synthesizing the results of this experiment with those previous indicate the latter is more likely to be true, as both analysis showed a minimum value of 1. Furthermore, given the nature of data collection methods previously discussed, IIDs observed between an average of 1 and 2 prefixes per day are more likely to be actually rotating their prefixes, as opposed to remaining entirely static.

Table 4.2. Summary statistics – total unique prefixes per IID, per day

Summary statistics – total unique prefixes per IID, per day				
Mean	Median	Std. Dev.	Min	Max
1.46	1.41	0.26	1	52.28

Furthermore, we see in Table 4.2 a maximum value of approximately 52. While we suspect that many IIDs do change prefixes slightly more frequently than once per day, outlier values represent an improbable frequency of prefix change given the nature of data collection described in Section 3.1.1. Instead, we offer a few possible explanations. First, extremely high numbers of daily prefix observations for a given MAC are could be attributable to MAC address re-use among EUI-64 addressed CPE. Second, power outages or malfunctioning devices could also contribute to this phenomenon, as a given device may request a new prefix each time it comes back online. A more detailed example of this behavior may be found in Section 4.1.4, as well as 4.2.1.

The preceding sections examined experimental results through an IID-centric lens, tracking the assignment of prefixes to individual IIDs. However, the following discussion reverses the analytical perspective, measuring instead the occupation of individual prefixes by EUI-64 IIDs. Figure 4.3 and Figure 4.4 depict the periods of time in which no EUI-64 IID is observed in a given prefix.

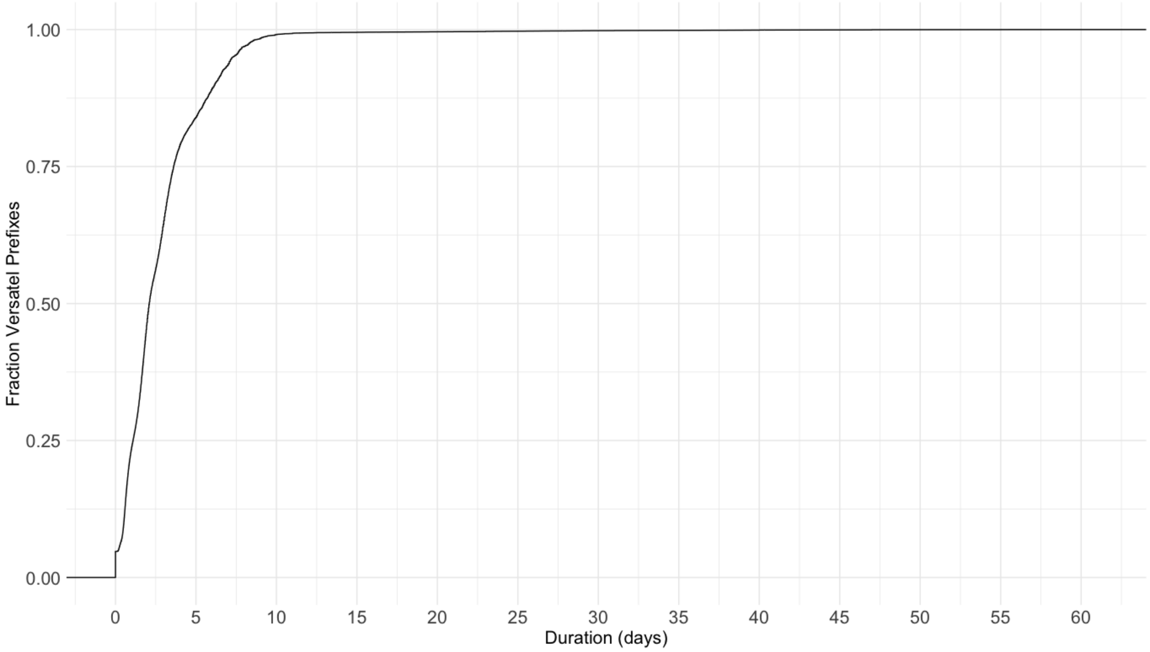


Figure 4.3. CDF – mean days in which no EUI-64 IID is observed in a given prefix

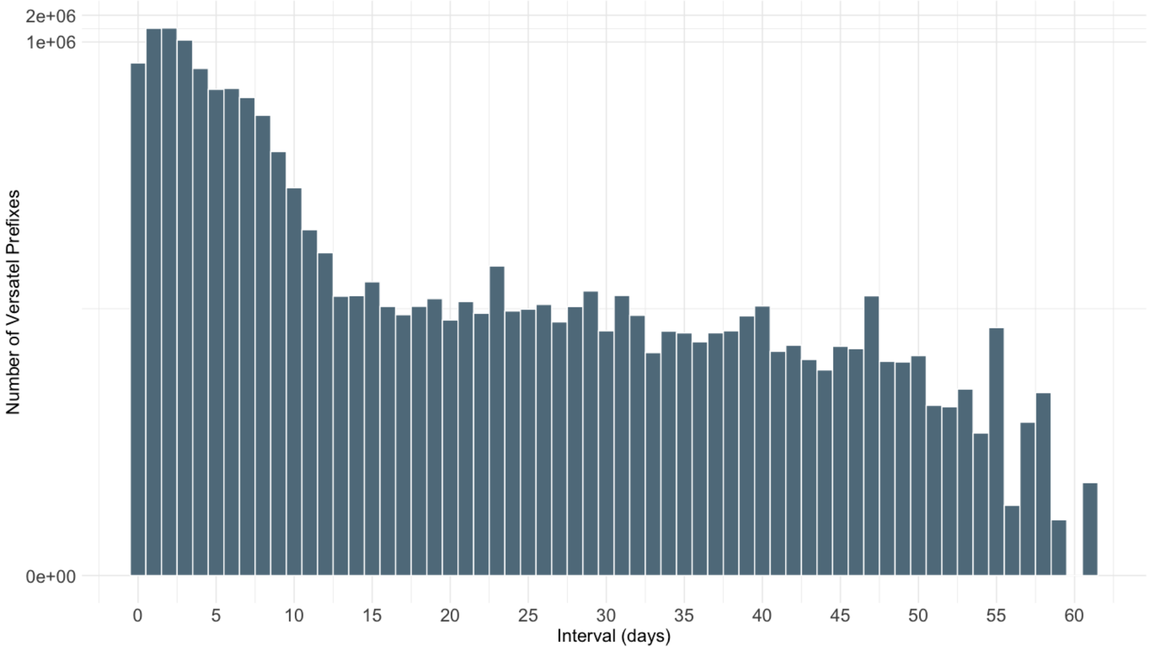


Figure 4.4. Histogram – mean days in which no EUI-64 IID is observed in a given prefix

The distribution of data is right-skewed with some prefixes experiencing significantly long periods with no EUI-64 IID observations. However, we see that the majority of prefixes are devoid of EUI-64 IIDs for no longer than two days on average, indicating a fairly rapid turn-over. These findings are enriched further when discussed in conjunction with Figure 4.5 and Figure 4.6.

Figure 4.5 and Figure 4.6 depict the mean number of days elapsed in which a given prefix is continuously occupied by some EUI-64 IID. Note that this analysis does not take into consideration the exact IID in question, but only measures periods of consecutive occupation by any EUI-64 IID. Note that over half of Versatel prefixes had a mean consecutive occupation period of 1 day. This indicates that a majority of prefixes experience short periods of emptiness interspersed with short periods of occupation. In other words, there is a high level of IID churn, as new IIDs rapidly leave and enter Versatel prefixes.

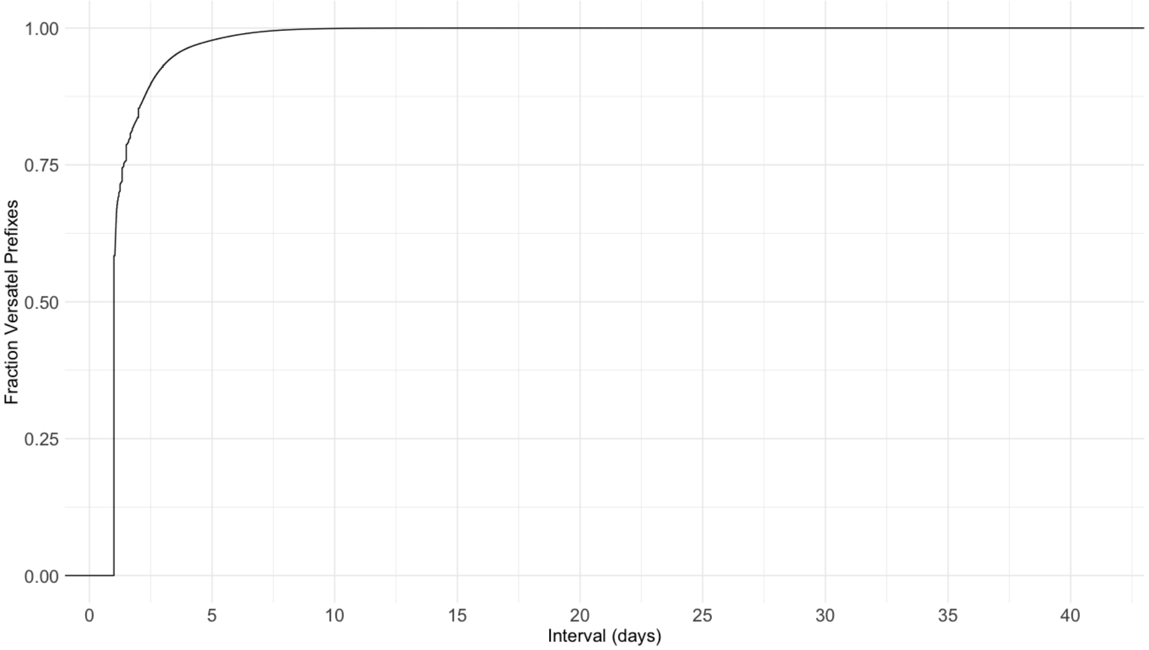


Figure 4.5. CDF – mean days in which prefix is consecutively occupied by any EUI-64 IID

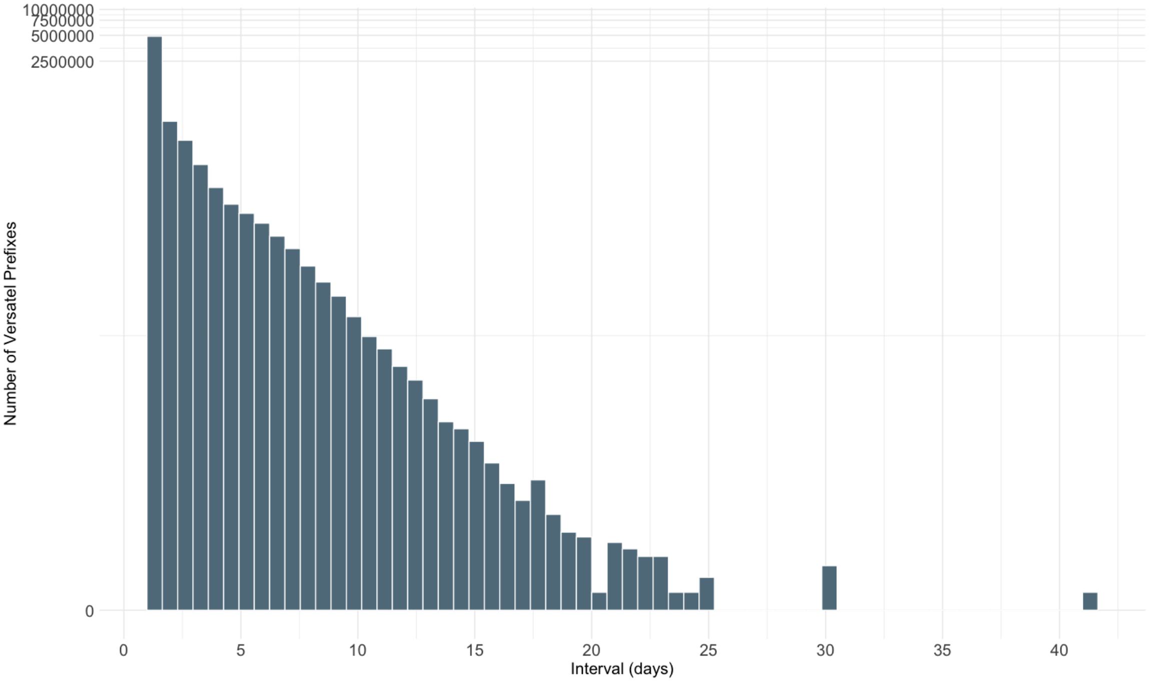


Figure 4.6. Histogram – mean time periods in which a prefix is consecutively occupied by any EUI-64 IID (pseudo-log scale)

Similar to the graphs showing periods of EUI-64 emptiness, Figure 4.5 and Figure 4.6 are right skewed. However, the overall distribution of data is less uniform, and the slope of the histogram in 4.8 is significantly steeper than that of Figure 4.6. This indicates that periods of EUI-64 occupation are generally shorter than periods which are EUI-64 devoid.

We further refine these results by examining the number of days in which a given prefix is occupied by a specific IID. In contrast to the previous analysis, this iteration takes into account the specific IID occupying the prefix in question.

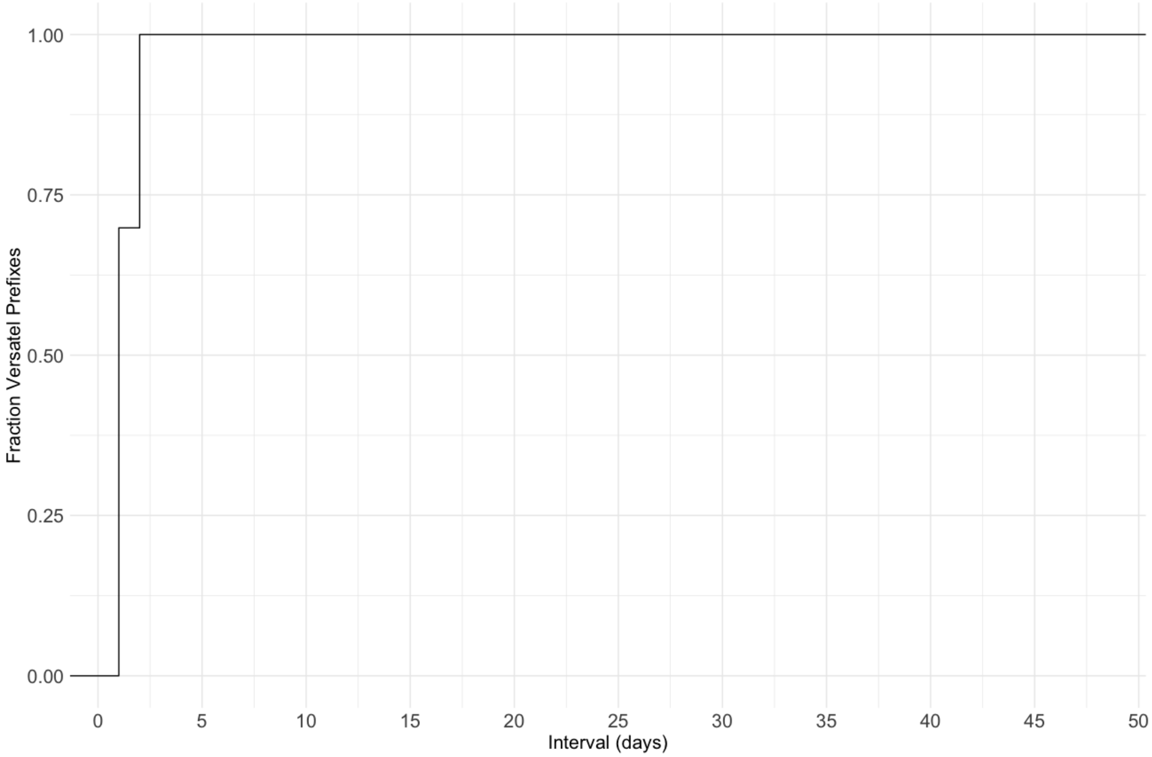


Figure 4.7. CDF – days in which prefix is consecutively occupied by a unique EUI-64 IID

The majority of data is still clustered around 1 day. This is fairly unsurprising, as a day-long occupation by definition implies an occupation by a single IID. However, while the slope of the CDF in Figure 4.5 is fairly gradual after the uptick at 1, Figure 4.7 displays a sharp uptick at 1, followed by another sharp uptick around the mean value of 2.8. This pattern of deeply asymmetrical data is perpetuated in Figure 4.8, indicating that periods of consecutive occupation by a unique IID are generally shorter than those which do not take into account IID specificity.

Interestingly, the outlier values to the far right of Figure 4.6 are echoed in Figure 4.8, indicating that the longest continuous occupations for a given prefix are not composed of a continuous series of varying IIDs, but rather correlate with long periods of occupation by a single, static IID.

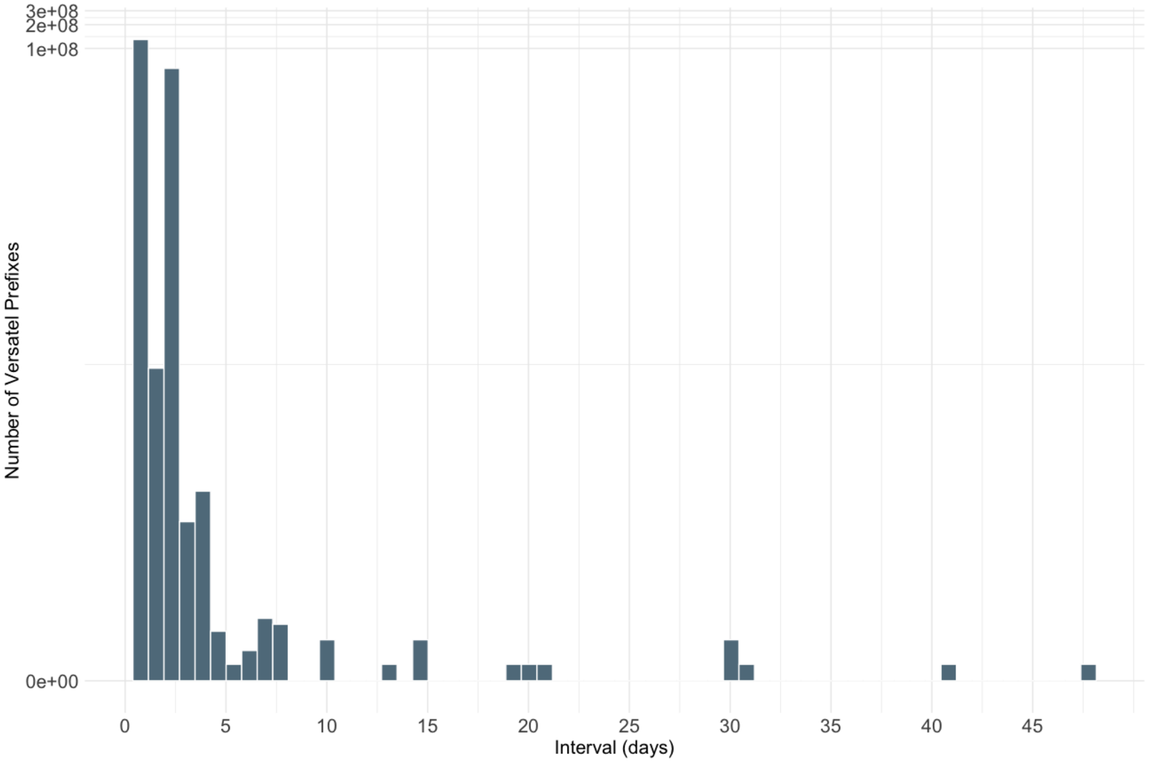


Figure 4.8. Histogram – days in which prefix is consecutively occupied by a unique EUI-64 IID (pseudo-log scale)

4.1.3 Locality of Prefix Rotation

The following subsection takes into consideration prefix rotation behaviors in terms of the distance traveled within the IPv6 address space. As with the previous subsection, the following results represent an additional facet of analysis pertaining to Versatel EUI-64 last hops. We build upon previous analyses to understand not only when and how often prefixes change, but also the level of spatial divergence between prefixes observed for a given IID.

First, we examine the total range of prefix travel for each unique IID observed in Versatel networks within the input data set. Figure 4.10 depicts the full range of prefix travel for each IID in terms of the number of EUI-64 addressed devices whose largest and smallest prefix values fall into a given IPv6 prefix subnet size. Figure 4.9 presents the same information in terms of a cumulative distribution of total EUI-64 IIDs observed in the input data. Due to the uneven distribution of subnet sizes, a pseudo-log scale has been applied to the y-axis of

Figure 4.10 such that outlier values remain visible in the graph.

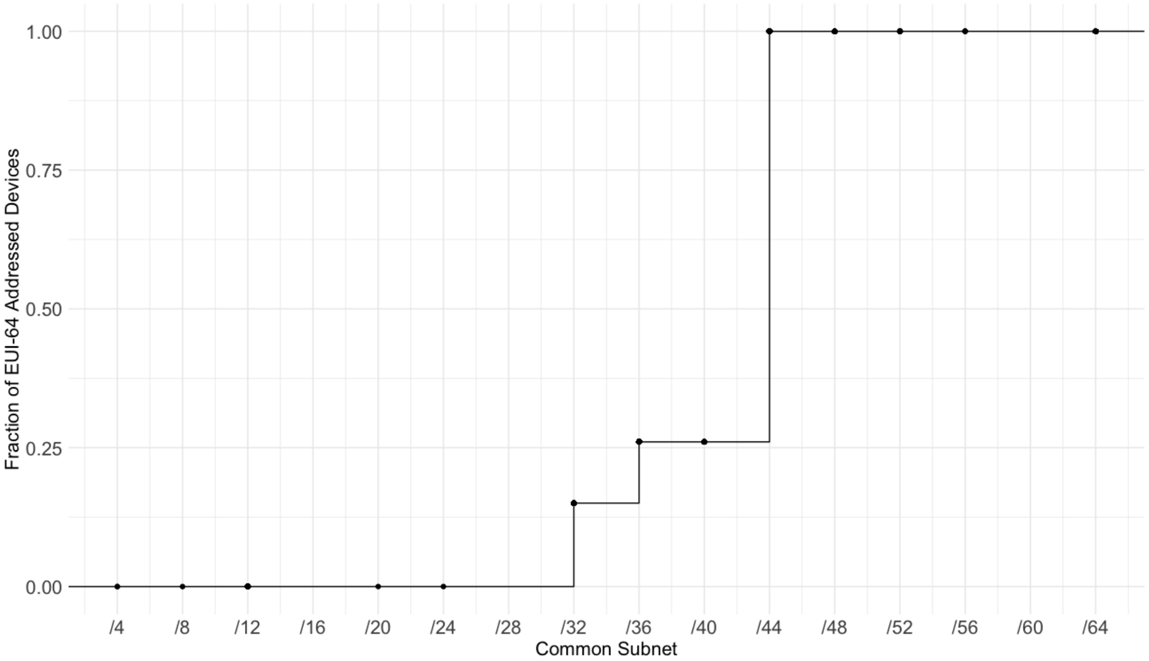


Figure 4.9. CDF - common subnet for maximum and minimum prefix per IID

Table 4.3 illustrates mock IPv6 prefix strings for the top three common subnet sizes shown in Figure 4.11. Take for example the most frequently observed common subnet size of /44. We conclude for any given IID in the /44 category that, in the event of any observed prefix change, the new prefix will have at least the three most significant hextets in common with the prefix immediately antecedent. As such, we may also derive the maximum network size within which this IID resides, reflected in its subnet categorization on the x-axis of Figure 4.10 and Figure 4.11.

Table 4.3. Example prefix strings for 3 most frequent common prefix lengths. Matching hexadecimal digits between two network prefix addresses are denoted by "m". Most significant non-matching digits are denoted by "x". Arbitrary digits are denoted by "a".

Example prefix strings for common prefix lengths	
Common Subnet Size	Example Prefix String
/44	mmmm:mmmm:mmmm:xaaa
/36	mmmm:mmmm:mmxa:aaaa
/32	mmmm:mmmm:mxaa:aaaa

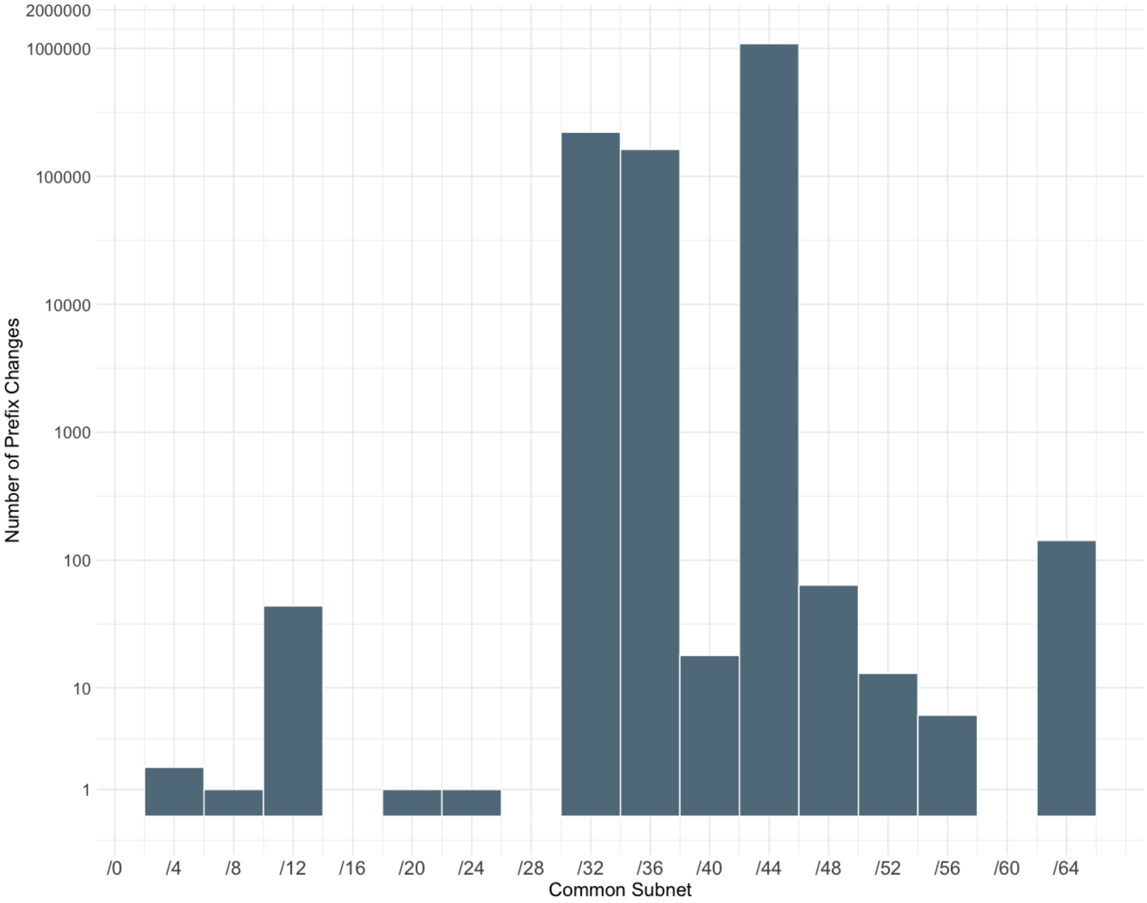


Figure 4.10. Histogram - common subnet for maximum and minimum prefix per IID (pseudo-log scale)

The results of this analysis suggest that the vast majority of IIDs eventually change prefixes, and do so within a relatively specific range of addresses. Furthermore, a non-trivial amount of addresses remain static across the full 126 days of data collection. This strongly correlates EUI-64 based SLAAC addressing with the significant reduction of the total possible address space for a given IPv6 device. Note also that very few IIDs travel within a common subnet larger than a /32. However, among these devices Figure 4.8 features a spike at /12. The cause for this anomalous uptick is discussed further in Section 4.1.4.

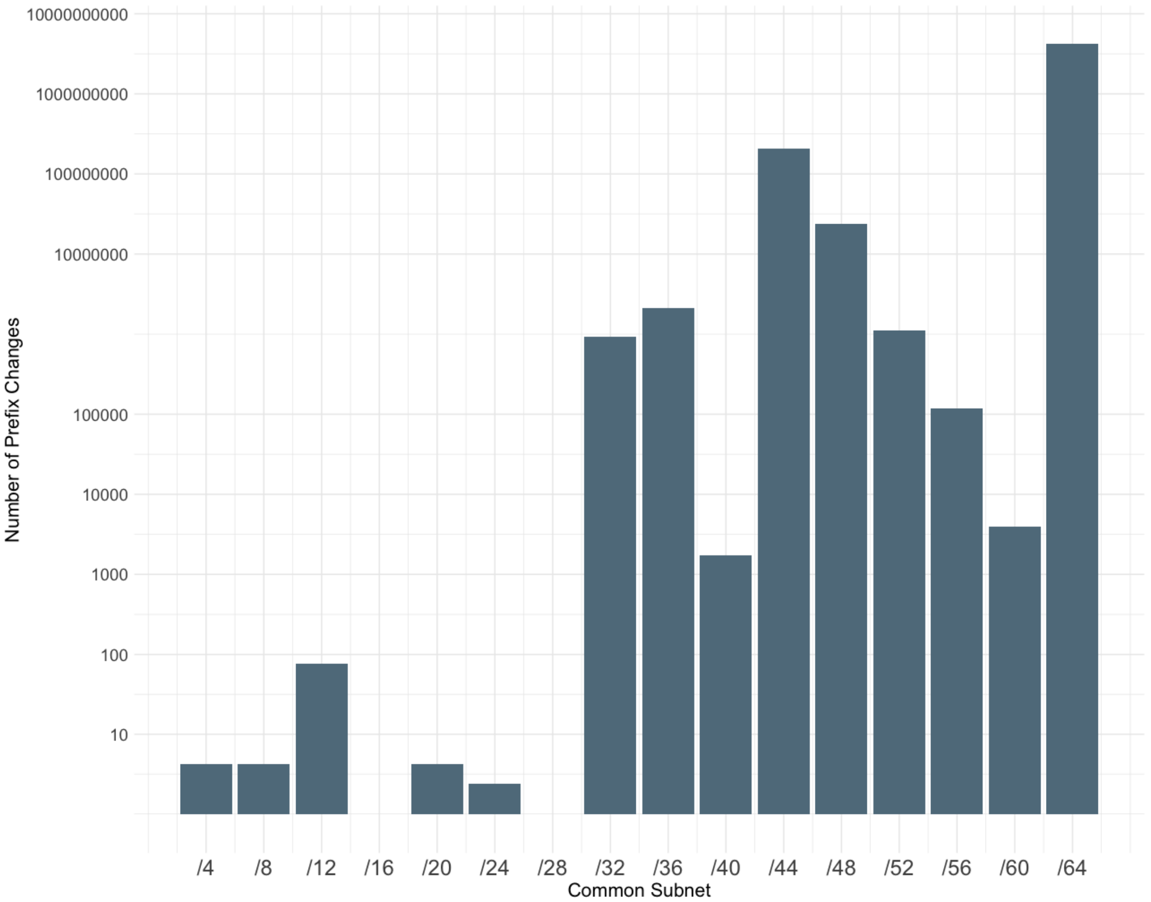


Figure 4.11. Histogram – common subnet per prefix rotation, per unique IID (pseudo-log scale)

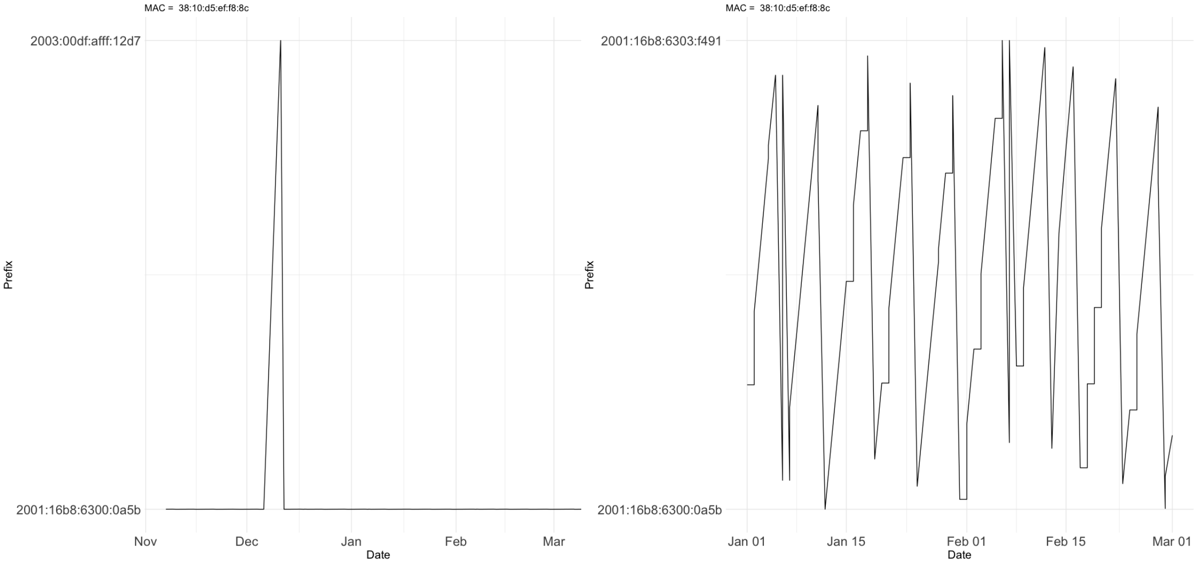
Figure 4.11 illustrates the common subnet on a per-observation basis for each EUI-64 IID in Versatel networks. In other words, each value on the x-axis represents the result of a comparison of the 64-bit network prefixes associated with two consecutive observations of a given IID. Similar to Figure 4.10, the uneven distribution of data necessitated the use of a pseudo-log scale on the y-axis of Figure 4.11. The pseudo-log scale maps numbers to a signed logarithmic scale with a smooth transition to linear scale around 0, which makes smaller values more visible when the distribution of data is skewed toward significantly larger values.

In comparing Figure 4.10 and Figure 4.11, we most notably see a drastic increase in the /64 category. This increased size of the /64 bar may be attributed to the increased likelihood of observing a given IID in the same prefix between two consecutive observations. This suggests that many IIDs experience periods of relative stasis throughout their lifetime, but still change prefixes on either a periodic or one-time basis. Furthermore, we may reasonably conclude that the common subnet for an IID's full range of movement, as shown in Figure 4.10, is not necessarily indicative of the distance for every movement made by the same IID. Note also that the subnet categories to the right of Figure 4.11 see a more even distribution than those in Figure 4.10. From this, we can conclude that prefix rotation behaviors display greater variance on a per-change level, suggesting that IIDs do not always travel the exact same distance with every jump.

4.1.4 Anomalous Behaviors

During the course of quantitative experiments shared above, we discovered a number of anomalous behaviors among EUI-64 addressed devices in Versatel networks. The first is alluded to in our discussion of Figure 4.10, and pertains to the spike at the /12 category. We found 44 unique MAC addresses which reside primarily within Versatel network prefixes, but move for a relatively short period to a Deutsche Telekom prefix. Figure 4.12 illustrates this behavior, depicting a MAC which jumps to a Deutsche Telekom prefix for one day, then returns to a limited set of Versatel prefixes.

We found every IID represented in the /12 category in Figure 4.10 exhibits similar behavior. This accounts for the fact that the /12 bar is roughly the same size in Figure 4.11, as IID exhibiting a full range of movement with a /12 common prefix size do so because of a single jump outside of Versatel prefixes. We performed an exhaustive lookup for each of the OUIs in this set of 44 unique MACs, and found that they are all registered to AVM GmbH.



a)Full range – 126 days

b) Detailed view - January through March

Figure 4.12. Example of a unique MAC address observed briefly in Deutsche Telekom prefixes

Figure 4.13 shows the dates of occupation by aforementioned MACs for each of the observed Deutsche Telekom prefixes. While a few MACs appear to reside in Deutsche Telekom prefixes for relatively long periods, the majority remain for between 1 and 3 days, then return to Versatel prefixes. Although the exact origin for this phenomenon is unknown, we suggest some possible root causes. First, we suggest that the MACs identified in Figures 4.12 and 4.13 could be dual-homed devices, in which each respective network is associated with a different ISP. MACs shown with a turquoise dot in Figure 4.13 were observed in both Deutsche Telekom and Versatel prefixes on the same date, suggesting a greater likelihood of falling under this dual-homed category.

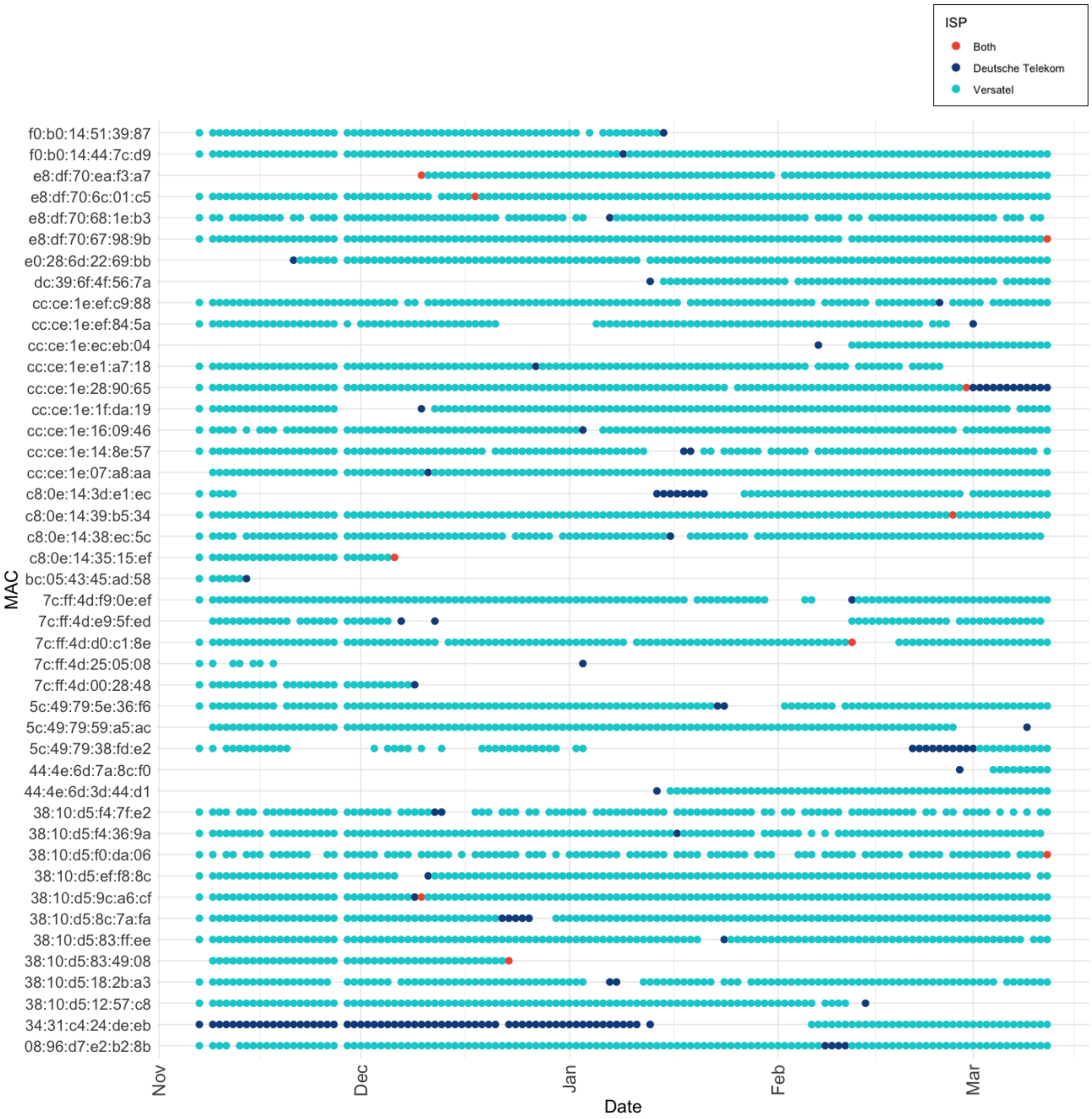
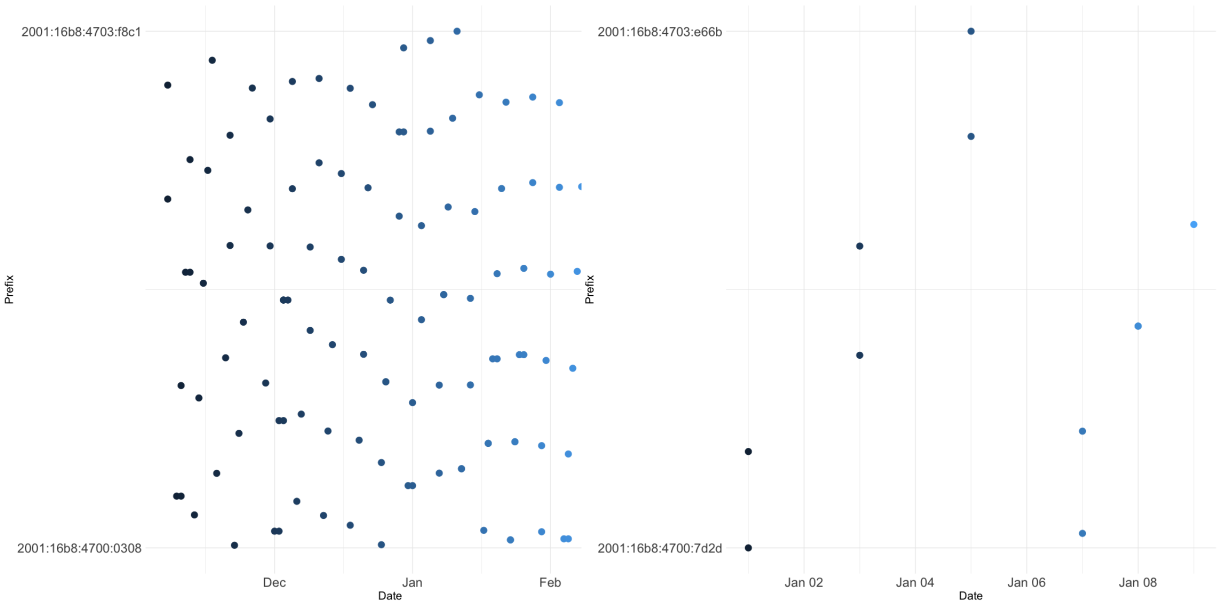


Figure 4.13. Dates in which MACs jump from Versatel to Deutsche Telekom

However, certain MACs shown in Figure 4.13 do not cleanly fit into this explanation. For example, the device toward the bottom of the figure, with MAC address 34:31:c4:24:de:eb, resides in a Deutsche Telekom prefix for about half of the data collection period, then appears to go offline for an extended interval, finally coming back online and remaining within a Versatel prefix. This pattern of behavior more probably represents that of a customer changing internet providers, rather than that of a dual-homed CPE. A final, but less conclusive, explanation could be MAC re-use. However, the device manufacturers and prefixes observed within the aforementioned group of MACs are respectively associated with German telecommunications companies and ISPs, suggesting that this phenomenon is more likely and artifact of the two phenomena previously discussed.

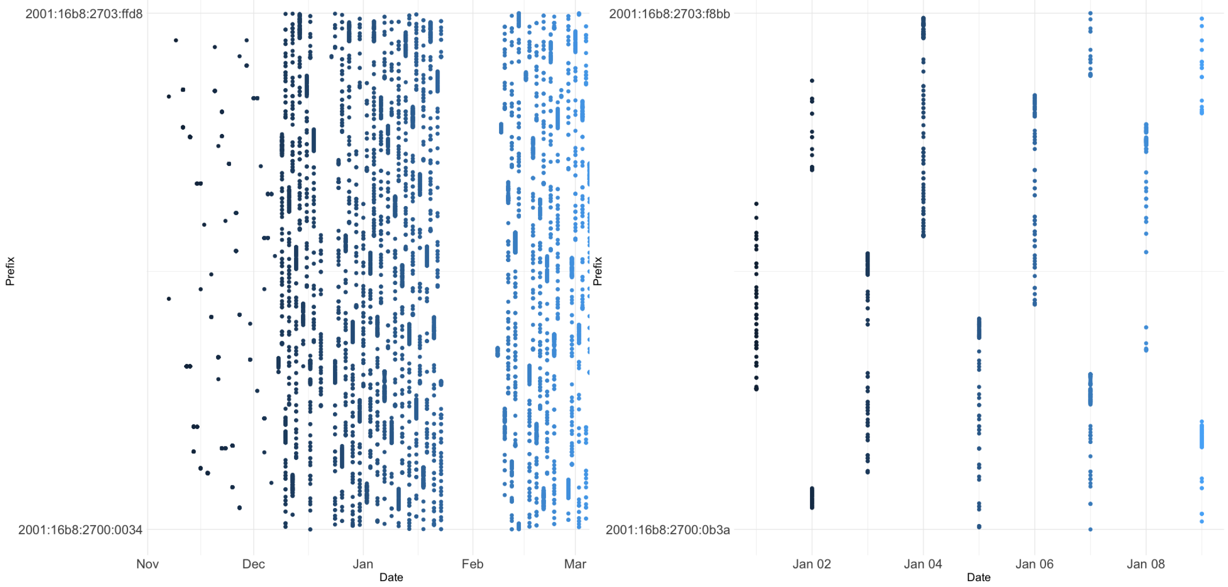
Although we have not found conclusive evidence of MAC address re-use in the IIDs shown above, other analyses strongly indicate its ubiquity among Versatel CPE. Recall from previous discussion of Figure 4.1 that Versatel CPE are observed in an average of 1.46 unique prefixes per day. Furthermore, while over 80 percent of IIDs were observed in fewer than two unique prefixes per day, less than 10 percent are seen in exactly one. This suggests that a typical IID will change its prefix about once a day, while relatively few remain completely static. This more common behavior is punctuated by outlier IIDs observed in thousands of unique prefixes over 126 days. Each of these cases is illustrated in Figure 4.14 and Figure 4.15, respectively. Both MAC addresses illustrated below are registered to AVM GmbH.



a) 126 days

b) One week

Figure 4.14. Prefixes over time for MAC e8:df:70:d9:b1:9f - observed in a total of 100 prefixes over 126 days.



a) 126 days

b) One week

Figure 4.15. Prefixes over time for MAC f0:b0:14:5a:59:03 - observed in a total of 5376 prefixes over 126 days.

Figure 4.14 provides a good example of typical behavior among EUI-64 Versatel CPE. Based solely on gross unique prefix counts, one may not expect this MAC to appear in multiple prefixes per day, as it was observed in a total of 100 prefixes over 126 days. While we occasionally observe this MAC in a single prefix per day, Figure 4.14 indicates that observation in two distinct prefixes per day is more common. This is an artifact of the probing method described in Section 3.1.1, which was performed over 126 24-hour periods, during which a given IID could have been probed before and after a given prefix change. While each individual MAC displays relatively unique behavior, global analyses suggest minor variations on the patterns displayed in Figure 4.14 are the most common.

Figure 4.15 depicts the MAC which appears in the greatest number of unique prefixes throughout the 126 day data collection period. Possibly, it represents an extreme case of MAC re-use. In the right panel, we see one week of prefix observations for this MAC, which illustrates a clear and consistent pattern of numerous unique prefix observations per day. Interestingly, the left panel indicates that this level of consistency does not extend over the full 126 days of data. Most notably, we see a gap with zero observations centered around February, as well as a region of relative sparseness from November to December in which fewer unique prefixes are observed per day. Taking this into consideration, the behavior shown in these plots may also correlate to that of a malfunctioning device. Note that assigned prefixes monotonically increase within discrete chunks of time. This could indicate a given device which frequently power cycles, receiving the next closest prefix in the free pool upon each power-up. This scenario could also account for periods of relative sparseness and gaps in observations, since the device may have been taken offline due to erratic behavior. Finally, MAC re-use would likely show a more random distribution of prefix observations, making the latter explanation more probable. Ultimately, the cause of these discrepancies is unclear, and may be the subject of future study.

4.2 Qualitative Analysis of EUI-64 Devices

This section provides a qualitative discussion based on the quantitative analyses reviewed in the previous section. In particular, it is guided by the following research questions:

1. To what extent can particular rotation behaviors be mapped to inferred hardware or software implementations?
2. What is the level of homogeneity in hardware for EUI-64 addressed devices, and to what extent can this facilitate cyber attacks?
3. Given prior identification of prefix rotation, to what extent can the future prefix of an EUI-64 last hop be predicted?

Accordingly, subsections are organized by themes correlating to the above questions.

4.2.1 Prefix rotation as an artifact of hardware or software implementations

Previous sections establish that the count of unique MAC addresses represented in EUI-64 IIDs in the input data set is outnumbered by the count of unique last hops. In this section, we suggest two main causes for this disparity, and evaluate their relative weight in terms of the quantitative results shared in Section 4.1.

First, we suggest ISP-imposed prefix rotation, in which the 64-bit SLAAC network prefix periodically changes for an EUI-64 IID. If the majority of EUI-64 IIDs are subject to prefix rotation, this would result in the observation of a limited number of discrete devices within multiplicatively many network prefixes. We suggest that the practice of ISP enforced prefix rotation, along with the probing methods used in the collection of input data account for the bulk of this discrepancy. We have found no evidence that prefixes are intentionally re-assigned more frequently than once per day. In fact, analysis of frequency and entropy of prefix assignments suggest the opposite is true, given the observation that most IIDs remain in a given prefix for longer than one day, evidenced by Figures 4.6 - 4.8.

We see in Figure 4.1 and Figure 4.2, along with their associated tables, substantially large outliers observed in as many as 52 prefixes per day. Such outlier values account for the remainder of the disparity between unique IIDs (and accordingly unique devices), and unique last hop addresses in our input data. Non-unique MACs in EUI-64 addressed CPE could be one potential explanation for outlier IIDs observed in many unique prefixes per day. Furthermore, we discovered a number of anomalous behaviors as described in 4.1.4. The conjectured cause of such behaviors varies from case to case, but Figure 4.15 illustrates a commonly observed pattern of behavior among aforementioned outlier IIDs, suggesting malfunctioning devices as another potential cause for IIDs which appear in an unusually high number of prefixes per day. However, it should be noted that our interpretation of anomalous behavior is the product of a case-by-case analysis of anomalous EUI-64 IIDs, and therefore cannot be extrapolated to make generalized statements.

Of the IIDs observed in three or more unique prefixes per day, two registered to LANCOM, and were observed in an average of 3.5 and 5.8 prefixes per day, respectively. As such, they do not represent the most extreme example of redundant IID observations. All others were registered to AVM GmbH, the maker of Fritz!Box devices. Accordingly, all of the most egregious examples of frequent IID observation across prefixes correspond to Fritz!Box devices. However, we cannot definitively attribute these anomalous behaviors to a given hardware implementation due to the disproportionate representation of AVM GmbH CPE among Versatel networks, illustrated in 4.1. Furthermore, given the scope of this research we cannot extend any of the above claims to ISPs other than Versatel. Recall that prefix rotation behaviors can be unique to a given ISP, so the significance of these findings should be contextualized within their appropriate scope.

4.2.2 Significance of Findings for IPv6 Security and Privacy

This section evaluates the extent to which individual behaviors discussed in Section 4.1 contribute to the overall homogeneity or predictability of EUI-64 based addressing and associated prefix rotations. Furthermore, it contextualizes this analysis in terms of its relative impact to IPv6 security and privacy.

Quantitative analyses revealed the specific instances in which EUI-64 based addressing is generally homogeneous. First, we see in Table 4.1 that about 99 percent of EUI-64 addressed CPE are made by only 2 major vendors, with a single vendor accounting for 84 percent of this portion. From this we may conclude that hardware associated with EUI-64 CPE in Versatel networks is strikingly homogeneous. This finding bears significant weight in an adversarial context, as it substantially increases the probability of success as well as the size of the attack surface in the event of device-specific exploitation. Homogeneity in OUIs also implies a drastically reduced target search space, as 99 percent of EUI-64 addressed devices will share a pool of only 2 unique sequences from which the most significant 40 bits of the 64-bit IID may be constructed. This effectively reduces a portion of the IPv6 address search space from 2^{40} potential combinations to 2 potential combinations. Furthermore, we see in Table 4.1 that an attacker wishing to exploit EUI-64 addressed CPE within Versatel networks has about an 84 percent chance of correctly guessing the hardware manufacturer of the target device if they assume it is an AVM GmbH/Fritz!Box device.

We also see in Figure 4.10 and Figure 4.11 that the total distance of prefix movement for a given IID is disproportionately likely to fall into one of three subnet sizes: /44, /36, or /32. In addition to the significant reduction in the search space for IIDs by homogeneous OUIs, this suggests that the search space for 64-bit network prefixes is also drastically reduced in the majority of EUI-64 based addressing schemes. The most common of these subnets is a /44; as illustrated in Table 4.3, rotating prefixes which remain within this subnet size will always have the same 3 most significant hextets. As such, a potential prefix search space of 2^{64} is reduced to a space of 2^{20}. For a common subnet of /36 the space is reduced to 2^{28}, and for a /32 it is reduced to 2^{32}. The search space is further reduced when examining a given IID at the per-change level. As shown in Figure 4.11, the most common encompassing subnet prefix sizes on a per-rotation basis are /64, /44, and /48. An encompassing prefix with a /64 size simply indicates that the prefix remained the same, and the search space is therefore 2^{0} or 1. Finally, for a /48 the search space is reduced to 2^{16}. This not only facilitates scanning for device discovery and fingerprinting, but also trivializes the tracking of individual devices despite frequently rotating network prefixes.

It should be noted that prefix assignments are more difficult to predict on a per-change level as opposed to the more generalized findings discussed above. For example, per-change distance between prefixes showed a relatively wide variance, as depicted in Figure 4.11. However, looking at overall range of change as shown in Figure 4.10 reduces this variance, rendering per-IID movement more predictable, albeit at a coarser granularity. As such, predicting a given jump is easily generalized to a total range of travel, since we can say with sufficient confidence that a given IID will always remain within a subnet of a given size. However, we cannot as easily say whether a prefix jump will differ by the same distance every time. We see also that temporal frequency of individual prefix rotations can be intermittent and vary on a per-change basis. Temporal analyses such as predicting the time of the next prefix movement are complicated further by non-unique MAC addresses. Therefore, while we may be reasonably certain about the mean frequency of prefix reassignment in a qualitative and probabilistic sense, methods of quantitative and granular temporal prediction may be the subject of future work.

CHAPTER 5: Conclusions and Future Work

The main goal of this thesis was to better understand the specific behavior of EUI-64 addressed CPE within Versatel networks. In particular, we chose to measure the temporal and spatial properties of prefix rotation by leveraging the inherent persistence of EUI-64 IIDs. The overarching concern throughout this work is that every CPE represented in our input data fails to provide the full capacity of security and privacy benefits offered by other forms of IPv6 addressing. The main findings presented by this study therefore quantify and contextualize the extent to which a failure to provide security and privacy to end users has occurred.

5.1 Conclusions

This work has established several core findings which are summarized in this section. First, we found high levels of homogeneity in the distribution of manufacturers for EUI-64 addressed CPE within networks served by the German ISP, Versatel. Most strikingly, over 84 percent of EUI-64 CPE in Versatel networks belonged to the Fritz!Box line of residential gateway devices, produced by German manufacturer AVM GmbH.

We also contributed quantitative and qualitative characterizations of prefix assignment behaviors among EUI-64 addressed CPE in Versatel Networks. The majority of Versatel IIDs occupy on average between 1 and 2 unique prefixes per day. However, we also discovered IIDs which are observed in unusually large numbers of unique prefixes per day; the most pronounced example of these is observed in approximately 52 prefixes per day on average. We suggest possible rationale for this discovery, including the existence of non-unique MACs among EUI-64 CPE, as well as malfunctioning devices or power outages which lead to frequent power cycling and subsequent re-assignment of prefixes. We found the latter of these explanations to be more likely, and present an illustrative case study.

In addition to these IIDs which are frequently observed across large swathes of prefixes per day, we discussed a second anomalous behavior discovered among our input data. Our analysis of prefix rotation behaviors unearthed 44 unique MACs which are observed in both Versatel and Deutsche Telekom prefixes over the course of our 126 day data collection period. We conducted macroscopic analysis which suggests this phenomenon may be attributed to dual-homed CPE or to end users who switch their subscription to a new ISP.

Our temporal analysis of prefix rotation behaviors provides insight as to when and how often a new prefix is assigned to a given EUI-64 IID. First, we found that mostEUI-64 IIDs tend to move between Versatel prefixes at least once per day. This is evidenced by the fact that the majority of prefixes in question are both occupied by and devoid of EUI-64 IIDs for short periods of time. However, it should be noted that the periods in which Versatel prefixes remain unoccupied by EUI-64 IIDs have a more even distribution than periods of occupation, indicating an unevenness within this overall trend of IID churn. This unevenness is driven by the fact that there are a greater number of possible prefixes within which an EUI-64 IID may be observed than there are unique EUI-64 IIDs. We elaborated upon this finding by showing that that periods of consecutive occupation become shorter and more frequent when taking into consideration the uniqueness of EUI-64 IIDs.

We conducted spatial analysis of prefix assignments on both a per-IID and per-change basis. We found that the majority of IIDs stay within the same /44 prefix throughout the full 126 day observation period, with the next most common subnet sizes being /36 and /32. The fourth most common was a /64 subnet, indicating that statically assigned prefixes are somewhat common among EUI-64 CPE in Versatel networks. Upon changing the scope of analysis to a per-change, per-IID granularity, we found that individual jumps from prefix to prefix generally take place over smaller distances than that of the full range of travel for a given IID. As such, the common subnet for individual changes is more specific, with the most common size being /64 followed by /44 and /48. This indicates that an individual prefix change which covers the maximum distance observed for an IID is less common than one which remains closer to the initially assigned prefix. We consider this finding significant, as it implies the feasibility of predicting individual prefix movements for a given IID.

Finally, we contextualized our findings within an adversarial context so as to elucidate potential privacy and security concerns. Areas of concern include the homogeneity of hardware implementations for EUI-64 addressed CPE, as well as the predictability of prefix assignment frequency and locality. Noting the homogeneity of hardware, we assert the increased risk and expanded attack surface for device-specific exploits. Furthermore, the tendency for most IIDs to remain within a fixed, fairly specific, subnet of assigned prefixes reduces the search space for device discovery and exploitation. Despite findings which indicate nominal increases in security and privacy, such as frequently rotating prefixes, the broader concern remains that every CPE represented in this study employs the EUI-64 mode of IID generation, and therefore exhibits characteristics which are not commensurate with the full potential for privacy and security in IPv6 addressing.

5.2 Future Work

Our research raised additional questions and lines of inquiry which fall outside the scope of this work, but could serve to substantially augment understanding of EUI-64 addressing among IPv6 CPE.

First, our analysis primarily concerns a single ISP, Versatel. Recall that we selected Versatel due to its high concentration of EUI-64 addressed CPE, as well as its exhibition of prefix rotation behaviors prior to this study. However, other ISPs may also contain EUI-64 addressed devices or engage in similar addressing behaviors. Examination of other ISPs would contribute to a more comprehensive measurement of IPv6 addressing behaviors, allowing for the isolation of ISP specific policies and the identification of more universal trends. For example, an evaluation of relative frequency of EUI-64 addressing across multiple ISPs would give a global view of the state of security and privacy practices in current CPE deployments. Since many residential end users tend to use default configurations, such findings could potentially be used to identify models for best practices at the provider level. An inter-ISP study would also aid in creating a macroscopic view of the mechanisms driving prefix assignment and rotation behaviors among EUI-64 addressed devices, potentially allowing for mappings to be determined between observed behaviors and specific ISP policies or hardware implementations. However, as many addressing behaviors can be ISP specific, it should be noted that the exact techniques used in our examination of Versatel may not translate directly to the study of other ISPs. Further, validation of our own results would help quantify the effectiveness of our methods used, and possibly their applicability in adjacent studies.

Some of our experiments allowed us to draw initial conclusions, but were qualified by gaps in understanding. Recall that our discussion surrounding Figure 4.3 and Figure 4.4 dealt only with periods of emptiness in regard to EUI-64 IIDs. As such, we could only make assertions within this relatively narrow scope, and could not differentiate whether the prefix in question was truly devoid of any IPv6 IID, or whether this was only true of EUI-64 IIDs. Making this differentiation would require an expansion in the scope of input data, and therefore remains a subject for future work. The results of this study could potentially build a more detailed understanding of prefix assignment in relation to the type of IID in question. For example, it would allow us to understand whether certain prefixes are reserved for certain device types (such as residential CPE), and are consequently occupied by specific types of IIDs.

Throughout the course of this work, we discovered a number of anomalous behaviors described in Section 4.1.4. Discussions included two key phenomena, the first being IIDs which are observed in both Versatel and Deutsche Telekom prefixes, and the second being IIDs which are seen in a large number of unique Versatel prefixes per day. While we were able to make educated hypotheses as to their root cause, we lack the conclusive evidence and ground truth necessary to validate these claims. A more in depth analysis of these anomalous prefix assignment patterns among EUI-64 addressed CPE in Versatel networks could test our initial hypotheses and offer new insights.

Furthermore, the results of our experiments lacked the depth necessary to identify the exact implementation responsible for individual address assignments. Instead, our findings represent a more general view of prefix assignments. The development of methods for differentiating between addresses which are assigned manually and those assigned via automated mechanisms such as DHCPv6 may be the subject of future study. We also suggest further quantitative analysis and measurement of predictability in prefix assignment schemes for EUI-64 addressed CPE. While this study contributes general temporal and spatial characterization of prefix rotation among Versatel CPE, further study of predictability for prefix assignments at the individual device level would enrich this understanding. For example the use of pattern mining and/or machine learning algorithms could potentially trivialize the identification and forecasting of simple prefix assignment patterns for a given IID.

Finally, the measurements and findings presented in this study are the result of rapid, active probing techniques. As such, accuracy in predictions for the exact location of individual CPE device movements may be reduced. However, the ability to make generalized measurements and predictions of addressing behaviors via data collected from fast probers reduces the massiveness of the IPv6 search space, and thus increases the tractability of the core problem at hand.

THIS PAGE INTENTIONALLY LEFT BLANK

List of References

[1] C. Carpene and A. Woodward, "Exposing potential privacy issues with IPv6 address construction," in *Australian Information Security Management Conference*, 2012.

[2] K. Barker, "The security implications of IPv6," *Network Security*, vol. 2013, pp. 5–9, 06 2013.

[3] J. Ullrich. (2017, 12). IPv6 Addresses, Security and Privacy. *RIPE Labs*. [Online]. Available: https://labs.ripe.net/Members/johanna_ullrich/ipv6-addresses-security-and-privacy

[4] M. Dunlop, S. Groat, R. Marchany, and J. Tront, "The good, the bad, the IPv6," in *2011 Ninth Annual Communication Networks and Services Research Conference*, 2011, pp. 77–84.

[5] J. Bound, B. Volz, T. Lemon, C. E. Perkins, and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)," IETF, RFC 3315, July 2003. [Online]. Available: https://tools.ietf.org/rfc/rfc3315.txt

[6] S. Thomson, T. Narten, and T. Jinmei, "IPv6 Stateless Address Autoconfiguration," September 2007. [Online]. Available: https://www.rfc-editor.org/rfc/rfc4862.txt

[7] E. C. Rye, J. Martin, and R. Beverly, "EUI-64 considered harmful," *CoRR*, vol. abs/1902.08968, 2019, [Online]. Available: http://arxiv.org/abs/1902.08968

[8] A. Cooper, F. Gont, and D. Thaler, "Security and Privacy Considerations for IPv6 Address Generation Mechanisms," IETF, RFC 7721, March 2016. [Online]. Available: https://tools.ietf.org/html/rfc7721

[9] F. Gont and T. Chown, "Network Reconnaissance in IPv6 Networks," IETF, RFC 7707, March 2016. [Online]. Available: https://www.ietf.org/rfc/rfc7707.txt

[10] R. Beverly, R. Durairajan, D. Plonka, and J. P. Rohrer, "In the IP of the Beholder: strategies for active IPv6 topology discovery," in *Proceedings of the Internet Measurement Conference 2018* (IMC '18). New York, NY, USA: Association for Computing Machinery, 2018, p. 308–321. [Online]. Available: https://doi.org/10.1145/3278532.3278559

[11] T. Narten, R. Draves, and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6," IETF, RFC 4941, September 2007. [Online]. Available: https://www.rfc-editor.org/rfc/rfc4941.txt

[12] E. C. Rye and R. Beverly, "Discovering the IPv6 network periphery," in *Proceedings of the Passive and Active Network Measurement Conference (PAM)*, Mar. 2020.

[13] I. S. Institute, "Internet Protocol," IETF, RFC 791, September 1981. [Online]. Available: https://tools.ietf.org/rfc/rfc791.txt

[14] Internet Corporation for Assigned Names and Numbers (ICANN), "Available pool of unallocated IPv4 internet addresses now completely emptied," *ICANN*, February 2011, [Online]. Available: https://www.icann.org/en/system/files/press-materials/release-03feb11-en.pdf

[15] G. Huston. (2020). IPv4 Address Report. Accessed July 15 2020. [Online]. Available: https://www.potaroo.net/tools/ipv4/index.html

[16] R. Hinden and S. Deering, "IP version 6 addressing architecture," IETF, RFC 4291, February 2006. [Online]. Available: https://www.rfc-editor.org/rfc/rfc4291.txt

[17] S. E. Deering and R. M. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," IETF, RFC 8200, July 2017. [Online]. Available: https://tools.ietf.org/rfc/rfc8200.txt

[18] S. E. Deering and R. M. Hinden, "IP Version 6 Addressing Architecture," IETF, RFC 18840, December 1995. [Online]. Available: https://tools.ietf.org/rfc/rfc1884.txt

[19] S. E. Deering and R. M. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," IETF, RFC 1883, December 1995. [Online]. Available: https://tools.ietf.org/rfc/rfc1884.txt

[20] ARIN. IPv4 Addressing Options. Accessed July 15 2020. [Online]. Available: https://www.arin.net/resources/guide/ipv4/

[21] LACNIC. (2014, June). No more IPv4 addresses in Latin America and the Caribbean. Accessed July 15 2020. [Online]. Available: https://www.lacnic.net/1532/2/lacnic/no-more-ipv4-addresses-in-latin-america-and-the-caribbean

[22] APCNIC. Resource Ranges Allocated by APNIC. Accessed July 15 2020. [Online]. Available: https://www.apnic.net/manage-ip/manage-resources/address-status/apnic-resource-range/

[23] RIPE NCC. IPv4 Address Allocation and Assignment Policies for the RIPE NCC Service Region. Accessed July 15 2020. [Online]. Available: ripe.net/publications/docs/ripe-680

[24] AFRNIC. AFRINIC IPv4 Exhaustion. Accessed July 15 2020. [Online]. Available: https://afrinic.net/exhaustion

[25] K. Claffy, “Tracking IPv6 evolution: Data we have and data we need,” *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 3, p. 43–48, July 2011. Available: https://doi.org/10.1145/2002250.2002258

[26] Google. (2020). IPv6 Adoption Statistics. [Online]. Available: https://www.google.com/intl/en/ipv6/statistics.html

[27] Federal Communications Commission. (2020, January). Internet Protocol Version 6: IPv6 for Consumers. [Online]. Available: https://www.fcc.gov/consumers/guides/internet-protocol-version-6-ipv6-consumers

[28] O. Troan and B. Capenter, “Deprecating the Anycast Prefix for 6to4 Relay Routers,” IETF, RFC 7526, May 2015. [Online]. Available: https://tools.ietf.org/rfc/rfc7526.txt

[29] F. Baker, X. Li, C. Bao, and K. Yin, “Framework for IPv4/IPv6 Translation,” IETF, RFC 6144, April 2011. [Online]. Available: https://tools.ietf.org/rfc/rfc6144.txt

[30] E. Nordmark and R. E. Gilligan, “Basic Transition Mechanisms for IPv6 Hosts and Routers,” IETF, RFC 4213, October 2005. [Online]. Available: https://tools.ietf.org/rfc/rfc4213.txt

[31] Y. Rekhter, R. G. Moskowitz, D. Karrenberg, G. J. de Groot, and E. Lear, “Address Allocation for Private Internets,” IETF, RFC 1918, February 1996. [Online]. Available: https://tools.ietf.org/rfc/rfc1918.txt

[32] S. Jiang, D. Guo, and B. Carpenter, “An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition,” IETF, RFC 6264, June 2011. [Online]. Available: https://tools.ietf.org/rfc/rfc6264.txt

[33] E. B. Davies, S. Krishnan, and P. Savola, “IPv6 Transition/Coexistence Security Considerations,” IETF, RFC 4942, September 2007. [Online]. Available: https://tools.ietf.org/rfc/rfc4942.txt

[34] C. Deccio, “Measuring the Dual-Stack IPv6/IPv4 Experience,” in *NPS/CAIDA 2020 Virtual IPv6 Workshop*, June 2020, [Online]. Available: https://www.cmand.org/workshops/202006-v6/slides/deccio.pdf

[35] G. Huston. (2020, June). Measuring IPv6. [Online]. Available: https://www.potaroo.net/ispcol/2020-06/m6w.html

[36] D. Plonka and A. Berger, “Temporal and spatial classification of active ipv6 addresses,” in *Proceedings of the 2015 Internet Measurement Conference* (IMC ’15). New York, NY, USA: ACM, 2015, pp. 509–522. [Online]. Available: http://doi.acm.org/10.1145/2815675.2815678

[37] R. M. Hinden and S. E. Deering, "IP version 6 addressing architecture," IETF, RFC 2373, July 1998. [Online]. Available: https://tools.ietf.org/html/rfc2373

[38] T. Narten, E. Nordmark, W. Simpson, and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)," IETF, RFC 4861, September 2007. [Online]. Available: https://www.rfc-editor.org/rfc/rfc4861.txt

[39] R. Asati, H. Singh, W. Beebee, C. Pignatar, E. Dart, and W. George, "Enhanced Duplicate Address Detection," IETF, RFC 7527, April 2015. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7527.txt

[40] B. H. Robert Hinden, "Unique Local IPv6 Unicast Addresses," IETF, RFC 4193, October 2005. [Online]. Available: https://tools.ietf.org/rfc/rfc4193.txt

[41] T. Mrugalski, B. Volz, A. Yourtchenko, M. C. Richardson, S. Jiang, T. Lemon, and T. Winters, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)," IETF, RFC 4815, November 2018. [Online]. Available: https://tools.ietf.org/rfc/rfc8415.txt

[42] IEEE, "I.E.E.E. Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications," IEEE, Tech. Rep., Dec 2005.

[43] J. Martin, E. Rye, and R. Beverly, "Decomposition of MAC address structure for granular device inference," in *Proceedings of the 32nd Annual Conference on Computer Security Applications* (ACSAC '16). New York, NY, USA: Association for Computing Machinery, 2016, p. 78–88. [Online]. Available: https://doi.org/10.1145/2991079.2991098

[44] O. Gasser, Q. Scheitle, P. Foremski, Q. Lone, M. Korczyński, S. D. Strowes, L. Hendriks, and G. Carle, "Clusters in the expanse: Understanding and unbiasing ipv6 hitlists," in *Proceedings of the Internet Measurement Conference 2018* (IMC '18). New York, NY, USA: Association for Computing Machinery, 2018, p. 364–378. [Online]. Available: https://doi.org/10.1145/3278532.3278564

[45] D. E. 3rd and J. Schiller, "Randomness Requirements for Security," IETF, RFC 4086, June 2005. [Online]. Available: https://tools.ietf.org/html/rfc4086

[46] R. Padmanabhan, J. Rula, P. Richter, S. Strowes, and A. Dainotti, "Analyzing IPv6 address assignment practices," in *NPS/CAIDA 2020 Virtual IPv6 Workshop*, June 2020, [Online]. Available: https://www.cmand.org/workshops/202006-v6/slides/padmanabhan.pdf

[47] 1und1. Fritzbox bei 1und1. [Online]. Available: https://dsl.1und1.de/fritzbox

[48] Alec Waters. (2011, April). SLAAC Attack – 0day Windows Network Interception Configuration Vulnerability. [Online]. Available: https://resources.infosecinstitute.com/slaac-attack/

[49] A. R. Choudhary, "In-depth analysis of IPv6 security posture," in *2009 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing*, 2009, pp. 1–7.

[50] J. Ullrich, K. Krombholz, H. Hobel, A. Dabrowski, and E. Weippl, "IPv6 security: Attacks and countermeasures in a nutshell," in *8th USENIX Workshop on Offensive Technologies (WOOT 14)*. San Diego, CA: USENIX Association, Aug. 2014, [Online]. Available: https://www.usenix.org/conference/woot14/workshop-program/presentation/ullrich

[51] D. J. Solove, "'I've got nothing to hide' and other misunderstandings of privacy," *GWU Law School Public Law Research Paper*, vol. 44, no. 289, p. 745, July 2007, [Online]. Available: https://ssrn.com/abstract=998565

[52] R. Clarke, "Information technology and dataveillance," *Commun. ACM*, vol. 31, no. 5, p. 498–512, May 1988, [Online]. Available: https://doi.org/10.1145/42411.42413

[53] S. Farrell and H. Tschofenig, "Pervasive Monitoring is an Attack," IETF, RFC 7258, May 2014. [Online]. Available: https://tools.ietf.org/rfc/rfc7258.txt

[54] T. Aura, "Cryptographically Generated Addresses," IETF, RFC 3972, March 2005. [Online]. Available: https://tools.ietf.org/rfc/rfc3972.txt

[55] F. Gont, "A Method for Generating Semantically Opaque Interface Identifiers with IPv6 Stateless Address Autoconfiguration (SLAAC)," IETF, RFC 7217, April 2014. [Online]. Available: https://tools.ietf.org/rfc/rfc7217.txt

[56] L. M. Roberts and D. Plonka, "Watching the watchers: Nonce-based inverse surveillance to remotely detect monitoring," *arXiv preprint arXiv:2005.07641*, 2020.

[57] RIPE NCC. About RIPE Atlas. [Online]. Available: https://atlas.ripe.net/landing/about/

[58] 1&1 Versatel. [Online]. Available: https://www.1und1.net/

[59] CAIDA, "The CAIDA UCSD AS Classification Dataset," 2019. [Online]. http://www.caida.org/data/as-classification.

[60] IEEE. IEEE Registration Authority: Assignments. [Online]. Available: https://regauth.standards.ieee.org/standards-ra-web/pub/view.html#registries

[61] NPS HPC. Available HPC Resources. [Online]. Available: https://hamming.uc.nps.edu/Resources.html

[62] Plotly. Plotly Open Source Graphing Libraries. [Online]. Available: https://plotly.com/graphing-libraries/

[63] ggplot. Ggplot 2. [Online]. Available: https://ggplot2.tidyverse.org/index.html

[64] J. Žorž, S. Steffann, P. Dražumerič, M. Townsley, A. Alston, G. Doering, J. Palet, J. Linkova, L. Balbinot, K. Meynell, and L. Howard, "Best Current Operational Practice for Operators: IPv6 prefix assignment for end-users - persistent vs non-persistent, and what size to choose," RIPE, Tech. Rep. 690, October 2017. [Online]. Available: https://www.ripe.net/publications/docs/ripe-690

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California