NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

MAPPING MOBILE IPV6 PROVIDERS

by

Joseph Martineau

September 2020

Thesis Advisor: Robert Beverly
Co-Advisor: Justin P. Rohrer

THIS PAGE INTENTIONALLY LEFT BLANK

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

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

4. TITLE AND SUBTITLE
MAPPING MOBILE IPV6 PROVIDERS

5. FUNDING NUMBERS

6. AUTHOR(S) Joseph Martineau

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 exponential increase of IPv6 addresses and traffic on the Internet, relatively little is known about the topology of this space compared to the IPv4 Internet. However, recent discoveries in IPv6 probing have created rich datasets based on traceroutes across the IPv6 infrastructure. This thesis explores what constitutes a mobile autonomous system and analyzes this IPv6 topology data in order to classify which IP addresses belong to mobile providers. Topology maps have been created at the router-level of cellular ASes in order to visualize and define the characteristics of these IPv6 mobile providers. Mobile networks are of special interest due to the dominance of mobile devices in the IPv6 space. Understanding this topology is critical in numerous applications such as designing protocols, distributing content, and improving security, which benefits researchers, ISPs, and network administrators.

14. SUBJECT TERMS
networks, Internet measurement, cybersecurity, IPv6, topology, mobile, cellular

15. NUMBER OF PAGES
97

16. PRICE CODE

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

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

MAPPING MOBILE IPV6 PROVIDERS

Joseph Martineau
Civilian, Scholarship For Service
BS, California State University Monterey Bay, 2018

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

Justin P. Rohrer
Co-Advisor

Gurminder Singh
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Despite the exponential increase of IPv6 addresses and traffic on the Internet, relatively little is known about the topology of this space compared to the IPv4 Internet. However, recent discoveries in IPv6 probing have created rich datasets based on traceroutes across the IPv6 infrastructure. This thesis explores what constitutes a mobile autonomous system and analyzes this IPv6 topology data in order to classify which IP addresses belong to mobile providers. Topology maps have been created at the router-level of cellular ASes in order to visualize and define the characteristics of these IPv6 mobile providers. Mobile networks are of special interest due to the dominance of mobile devices in the IPv6 space. Understanding this topology is critical in numerous applications such as designing protocols, distributing content, and improving security, which benefits researchers, ISPs, and network administrators.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

1 Introduction **1**

1.1 Scope . . . 3

1.2 Research Questions . . . 3

1.3 Summary of Findings . . . 4

1.4 Thesis Structure . . . 5

2 Background **7**

2.1 Key Concepts . . . 7

2.2 Datasets . . . 12

2.3 Related Work . . . 12

3 Methodology **15**

3.1 Identifying Mobile IP Addresses . . . 16

3.2 Determining Mobile ASes . . . 18

3.3 Aggregating Traceroute Data . . . 19

3.4 Performing Longest Prefix Matching . . . 19

3.5 Creating Network Graphs . . . 20

3.6 Resolving Aliases . . . 22

3.7 Attributing MaxMind Prefixes and Domain Names . . . 22

3.8 Limitations . . . 23

4 Results **25**

4.1 Case Studies . . . 25

4.2 Other Mobile ASes . . . 52

4.3 User Agent Correlation . . . 52

5 Conclusions and Future Work **57**

5.1 Conclusions . . . 57

5.2 Future Work . . . 58

Appendix A Mobile Topology Graphs Ranked by Degree **61**

List of References **71**

Initial Distribution List **77**

List of Figures

Figure 1.1 Router-level and AS-level Topologies 2

Figure 2.1 Traceroute Example 8

Figure 3.1 Methodology Flowchart 15

Figure 3.2 Traceroute from Sprint's Looking Glass 16

Figure 3.3 IP to ASN Lookup of Sprint Router 17

Figure 3.4 Network Graph Traceroute Example 21

Figure 4.1 AS 1273 IPv4 Graph 27

Figure 4.2 AS 1273 IPv6 Graph 28

Figure 4.3 AS 1273 IPv4 Graph Degree Distribution 30

Figure 4.4 AS 1273 IPv6 Graph Degree Distribution 30

Figure 4.5 AS 1273 IPv4 Graph Ranked by Degree 31

Figure 4.6 AS 1273 IPv6 Graph Ranked by Degree 32

Figure 4.7 AS 4725 IPv4 Graph 33

Figure 4.8 AS 4725 IPv6 Graph 34

Figure 4.9 AS 4725 IPv4 Graph Degree Distribution 35

Figure 4.10 AS 4725 IPv6 Graph Degree Distribution 35

Figure 4.11 AS 5588 IPv4 Graph 36

Figure 4.12 AS 5588 IPv6 Graph 37

Figure 4.13 AS 5588 IPv6 Graph Size Distribution 38

Figure 4.14 AS 5588 IPv4 Graph Degree Distribution 39

Figure 4.15 AS 5588 IPv6 Graph Degree Distribution 39
Figure 4.16 AS 6167 IPv4 Graph 40
Figure 4.17 AS 6167 IPv6 Graph 41
Figure 4.18 AS 22394 IPv4 Graph 42
Figure 4.19 AS 22394 IPv6 Graph 42
Figure 4.20 AS 6167 IPv4 Graph Degree Distribution 43
Figure 4.21 AS 6167 IPv6 Graph Degree Distribution 44
Figure 4.22 AS 6167 IPv6 Graph Size Distribution 45
Figure 4.23 AS 9808 IPv6 Graph 46
Figure 4.24 AS 56040 IPv6 Graph 46
Figure 4.25 AS 56040 IPv6 Graph Degree Distribution 48
Figure 4.26 AS 55836 IPv4 Graph 49
Figure 4.27 AS 55836 IPv6 Graph 49
Figure 4.28 AS 55836 IPv4 Graph Degree Distribution 51
Figure 4.29 AS 55836 IPv6 Graph Degree Distribution 51
Figure 4.30 UA Distribution 54

Figure A.1 AS 4725 IPv4 Graph Ranked by Degree 61
Figure A.2 AS 4725 IPv6 Graph Ranked by Degree 62
Figure A.3 AS 5588 IPv4 Graph Ranked by Degree 62
Figure A.4 AS 5588 IPv6 Graph Ranked by Degree 63
Figure A.5 AS 6167 IPv4 Graph Ranked by Degree 64
Figure A.6 AS 6167 IPv6 Graph Ranked by Degree 65
Figure A.7 AS 9808 IPv6 Graph Ranked by Degree 66
Figure A.8 AS 56040 IPv6 Graph Ranked by Degree 67

Figure A.9 AS 55836 IPv4 Graph Ranked by Degree 68
Figure A.10 AS 55836 IPv6 Graph Ranked by Degree 69

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

Table 4.1 Case Study ASN Statistics 26

Table 4.2 AS 1273 Statistics 29

Table 4.3 AS 4725 Statistics 34

Table 4.4 AS 5588 Statistics 38

Table 4.5 AS 6167 Statistics 43

Table 4.6 AS 9808 and 56040 Statistics 47

Table 4.7 AS 55836 Statistics 50

Table 4.8 UA Prefix Information 54

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

AS Autonomous System

ASN Autonomous System Number

BGP Border Gateway Protocol

CAIDA Center for Applied Internet Data Analysis

CMAND Center for Measurement and Analysis of Network Data

CSV Comma-Separated Values

DNS Domain Name System

GDF GUESS Data Format

GUESS Graph Exploration System

HTTP Hypertext Transfer Protocol

ICMP Internet Control Message Protocol

IP Internet Protocol

IPv4 Internet Protocol version 4

IPv6 Internet Protocol version 6

IXP Internet Exchange Point

MIDAR Monotonic ID-based Alias Resolution

MRT Multi-Threaded Routing Toolkit

MTU Maximum Transmission Unit

NAT Network Address Translation

NPS Naval Postgraduate School

RIB Routing Information Base

UA User Agent

Acknowledgments

I would like to thank Dr. Robert Beverly and Dr. Justin P. Rohrer for their support, thoughtfulness, and great humor during this thesis process. I also want to thank my Scholarship for Service friends for the countless laughs we shared; my time at the Naval Postgraduate School would not have been anywhere near as enjoyable without them. Lastly, I would like to thank my partner, Alex, for her unwavering belief in my potential. When I feel lost or less than capable, she is always there to pull me back up.

This material is based upon work supported by the National Science Foundation under Grant Numbers 1565443 and 1855614. Any opinions, findings, and conclusions or recommendations expressed in this material 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

There are two versions of the Internet Protocol (IP) currently in use; the primary distinguishing feature of Internet Protocol version 6 (IPv6) is its larger 128-bit address space as compared to the 32-bit address space of Internet Protocol version 4 (IPv4) [1]. Despite the introduction of IPv6 in 1998 [2], the protocol languished for decades but has recently seen rapid and increased adoption due to the mounting exhaustion of IPv4 addresses [3], [4]. As the pool of available IPv4 addresses depletes, IPv6 has become an increasingly important area of interest.

One primary driver of IPv6 is mobile networks due to the prevalence of cellular devices. In recent years, several mobile providers (i.e., businesses that provide cellular service to their customers) are now breaking away from IPv4 and switching to IPv6 as the backbone of their infrastructure. In fact, over eighty percent of the traffic originating from Verizon Wireless now operates over IPv6, and other mobile providers such as T-Mobile are in the process of phasing out IPv4 completely from their networks [5]. Undergoing this process greatly reduces the complexity and costs of managing IPv4 infrastructure. In Verizon's case, switching away from IPv4 eliminates the substantial overhead of managing private clouds with over 70 layers of Network Address Translation (NAT) [6]. However, some elements carry over from IPv4 to IPv6; there are numerous middleboxes in this mobile IPv4 infrastructure handling video and image transcoding that exist in the IPv6 space as well [7]. This thesis examines how the emerging mobile IPv6 infrastructure compares to the existing IPv4 infrastructure.

Despite the exponential growth of IPv6 addresses and traffic on the Internet, the topology of this space is not well-defined compared to the IPv4 Internet. This is primarily due to two reasons: the sparseness of the IPv6 address space and rate-limiting of Internet Control Message Protocol (ICMP) messages by IPv6 routers. Both of these issues massively negate the efficacy of IPv4 exhaustive probing techniques. However, recent breakthroughs in IPv6 probing techniques have made discovering larger portions of this topology feasible [8]. These new methods have led to an influx of data regarding the IPv6 topology which has not yet been thoroughly analyzed. Given the prevalence of mobile devices on the IPv6

Internet, mapping the topology of mobile IPv6 providers is of special interest due to the direct implications on concerns like security, content delivery, and privacy. For example, questions such as whether new attack vectors exist, how the infrastructure handles correlated failures (i.e., robustness), and how difficult it is for an adversary to monitor communications all relate to the topology of the network. While these questions could be asked of other network topologies, mobile is of greater importance due to the ubiquitous nature of cellular devices which often act as a user's primary method of accessing the Internet [9].

Lastly, there is a distinction to be made between router-level and autonomous system (AS)-level topologies. Router-level topologies can be deduced using datasets captured with Paris traceroutes and applying alias resolution, and AS-level topologies can be inferred through using Border Gateway Protocol (BGP) data sourced from public looking glass servers. Of note is that sending traceroute probes is an active measurement technique as packets that would otherwise not be sent enter the network, while BGP data can be measured passively as announcements go out periodically with no extra bandwidth used. Although both levels provide an idea of the structure of the Internet, AS-level topologies are much easier to create as the data is publicly available and represents a more coarse-grained view of the system. On the other hand, router-level topologies often rely on more private data and depict a fine-grained view of an individual system. Both topologies have their merits and distinctly inform concerns at macro- and micro-scales. An illustration containing these two types of topologies is shown in Figure 1.1.

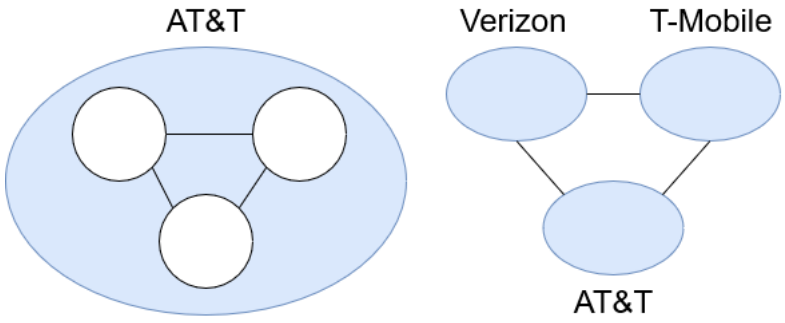

Figure 1.1. Router-level (left) and AS-level (right) Topologies.

1.1 Scope

This thesis limits itself to analyzing previously collected datasets. In other words, data collection is not in-scope for this thesis; the main focus is solely on finding and characterizing the mobile IPv6 topology using pre-existing data. This also means that measurements from personal devices (i.e., traceroutes from within mobile providers) have not been taken nor analyzed, although it would have been beneficial in gaining a deeper understanding of these topologies. Lastly, these mobile topologies have only been classified at the router-level as the desire is to gain insight into how individual cellular providers manage their infrastructure, and how that management differs from one another.

1.2 Research Questions

- How does the number of nodes and edges differ between IPv4 and IPv6 mobile topologies?
- How do other metrics such as average degree, average path length, and network diameter compare between IPv4 and IPv6 mobile topologies?
- How does the degree distribution compare between IPv4 and IPv6 mobile topologies?
- Does congruence exist between the IPv4 and IPv6 mobile infrastructure? In other words, do any of the same routers exist in both the IPv4 and IPv6 topology?

1.3 Summary of Findings

- Among the eight ASes analyzed, seven of them had a greater number of IPv4 nodes and edges compared to their IPv6 counterparts. An exception was found in Reliance Jio of India. India has the highest IPv6 adoption rate of all countries as measured by Akamai, which may be a possible explanation.
- There is a higher percentage of ingress nodes (i.e., routers that act as the first hop into an AS from the perspective of vantage points) in IPv6 topologies compared to IPv4. This was shown to be true in all case studies except in AS22394 as it was not applicable due to measurement issues. In two of the eight cases, the IPv6 mobile networks also displayed disjoint network behavior that resulted in fewer connected components.
- Network metrics such as average degree, average path length, and network diameter are not strongly correlated with determining whether a given mobile topology belongs to IPv4 or IPv6. In other words, these statistics do not prove useful in determining the type of mobile topology.
- The IPv4 degree distribution closely matches the expected network topology where most nodes are low degree and there is a consistent negative correlation where the number of nodes decreases as degree increases. In IPv6, the degree distribution is more erratic and does not exhibit this correlation.
- Congruence was shown to exist between IPv4 and IPv6 of a single mobile AS where the reverse domain name resolution of routers was mostly complete. In one other AS with partially resolved domain names, congruence was implied due to domain name similarities. This suggests that other mobile ASes could be running dual-stack IPv4/IPv6 infrastructure.
- Our devised method of user agent (UA)-based mobile network classification shows promise as an alternative way of finding mobile IP addresses. The method revealed clear thresholds in classifying cellular and non-cellular networks in our experiment with a web server access log.
- We determined that some providers use distinct cellular infrastructure for IPv6, while other providers run dual-stack IPv4/IPv6 infrastructure. Attacks on this shared infrastructure are a greater security concern as it affects both IPv4 and IPv6 traffic. The erratic degree distribution of IPv6 infrastructure may increase IPv6 network robustness as network availability is spread across more nodes.

1.4 Thesis Structure

- Chapter 1 introduces IPv6 and mobile as a large driver of the protocol and its adoption, specifies the scope of the thesis, presents research questions, summarizes the findings, and outlines the thesis structure.
- Chapter 2 provides background by discussing key concepts, the various datasets used, and related work.
- Chapter 3 defines the challenges faced in determining what constitutes a mobile AS, and the methodology used in inferring a mobile AS topology.
- Chapter 4 delves into case studies of several mobile ASes, which explores graph visualizations, metrics, and compares IPv4 and IPv6 topologies.
- Chapter 5 concludes the thesis and presents ideas for future work.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Background

This chapter discusses the background knowledge needed to understand the methodology and results of this thesis. This includes terminology and concepts related to network mapping and measurement. The datasets that this thesis analyzes are also described and related work is examined.

2.1 Key Concepts

This section provides background information on IPv6, traceroute, network topology, autonomous systems, cellular-specific infrastructure, and user agents. Current techniques for performing alias resolution on both IPv4 and IPv6 addresses are also elaborated upon.

2.1.1 IPv6

As mentioned in Chapter 1, IPv6 is the most up-to-date version of IP. Although IPv4 is still more widely used, IPv6 has gained traction in recent years due to the exhaustion of IPv4 addresses [3], [4]. One of the main motivations for IPv6 adoption is the 128-bit address space compared to the 32-bit address space of IPv4 [1]; IPv6 contains 2^{96} times more addresses than IPv4. The importance of the size of the IPv6 address space is amplified even further considering the increase in the number of cellular devices over the last decade [5]. For a cellular service provider, focusing on IPv6 and minimizing IPv4 infrastructure reduces both cost and complexity of networks [6].

This discussion also brings up the topic of Mobile IPv6 which allows nodes in transit (e.g., a cellphone in a car) to remain reachable while the device moves. The original, or home, address of a moving device is kept static which allows a connection to be maintained. Another driver of cellular providers using IPv6 is that Mobile IPv6 improved upon the Mobile IPv4 implementation and made it more efficient. For example, Mobile IPv4 requires special routers known as “foreign agents” while Mobile IPv6 does not, route optimization support is built into IPv6 while it requires nonstandard extensions in IPv4, and Mobile IPv4 requires IP encapsulation to reach a device not at its own address (i.e., the device has

changed cells) which is greater overhead [10]. As a cellular service provider, Mobile IPv6 having only standard routers, more efficient routing, and less overhead leads to lower costs and reduced complexity of the network infrastructure.

2.1.2 Tracerouting

Traceroute is a tool that infers a sequence of IP interfaces from a source machine to a destination machine. This is done by the packet expiring at each hop along the way, and each router reporting back to the source when this occurs. Although tracerouting is often used to diagnose network connectivity issues, it can also be employed as a tool for active network measurement. However, the standard utility has faults when utilized for this task. At the core of these faults is load balancing, which routers use to distribute packets across a network. This creates the problem where packets sent via standard traceroute might take paths across different routers. When the packets expire, the messages displayed might imply that two routers are linked when they are actually unrelated [11].

For example, imagine five routers: A, B, C, D, and E. Router A is a load balancer and is linked to routers B and C. Router C is linked to router D. Routers B and D are linked to router E. The first hop reports router A, the second hop reports router C, and the third hop reports router E. The second hop took the path from A to C, but the third hop took the path from A to B to E. This would lead one to conclude that router C and E are directly connected, which is not the case. This example is shown in Figure 2.1.

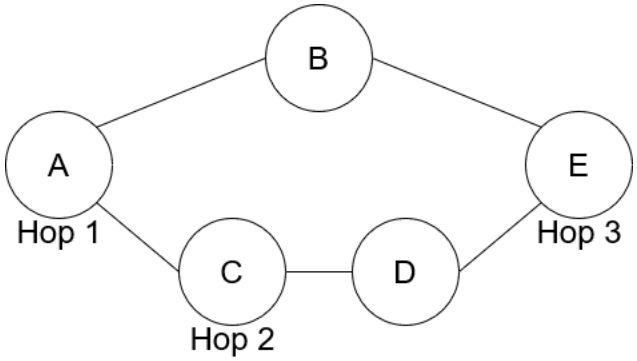

Figure 2.1. Traceroute Example.

In order to address these issues, Paris traceroute was created as an improvement of the traceroute tool. By controlling the content of packet headers, Paris traceroute returns a more precise path of which routers are actually linked. This is because load balancing is based on various packet header fields (e.g., protocol, checksum, port numbers), and handling these fields in a certain manner can control and/or detect load balancing [11]. The traceroute datasets utilized for this thesis and described in a later section employ Paris traceroute which ensures that network topologies are created with a high level of accuracy.

2.1.3 Autonomous Systems

An autonomous system is a collection of routing prefixes[1] typically under the control of a single entity and its defined policy [12]. These ASes are assigned a unique Autonomous System Number (ASN) for routing that allows them to be identified. In this thesis, the ASes examined are under the control of cellular providers (e.g., AT&T, Sprint, T-Mobile, Verizon) and are identified to be mobile based on prefix data from MaxMind, which is a company that specializes in IP address data. One point to clarify is that mapping an AS is equivalent to mapping its router-level infrastructure. Although based on the definition of an AS one might include the end-user devices, it is not feasible to map clients as those devices do not typically respond to traceroutes and thus do not appear in the data. As such, the network topology of cellular ASes will be defined as the routers that reside within them and how they connect to one another.

2.1.4 Topology

Introduced in Chapter 1, the topology of a network refers to the composition of its infrastructure. When talking about network topology, three levels of granularity are commonly used: interface level, router level, and AS level. As the most fine-grained view, interface-level topology concerns itself with how the individual interfaces of routers connect to one another. Providing a less fine-grained view is the router-level topology, which focuses only on how routers are linked to each other. This is the topology the thesis maps and is created using datasets from vantage points around the world sending out Paris traceroute probes. The router-level topologies mapped each pertain to a specific AS, although this type of

[1]Network prefixes define the collection of IP addresses within a network (e.g., 192.168.0.0/16 defines IP addresses from 192.168.0.0 – 192.168.255.255).

topology could technically span multiple ASes. While interface-level topologies could also be created from this dataset, such a fine-grained view might be too detailed for the analyses this thesis performs. AS-level topologies are the most coarse-grained of the three types and instead examine how ASes themselves connect to one another. These types of topologies can be created using BGP data gathered from looking glass servers (e.g., RouteViews) that employ passive measurement techniques. Although this thesis did not examine the AS-level topology of mobile providers, it may be something to consider in future work.

2.1.5 Alias Resolution

Applying alias resolution is key in transforming interface-level measurements into accurate router-level topologies. Once the IP addresses of routers within a mobile AS have been aggregated from the data, alias resolution must be performed on the set of collected addresses. A single router can have multiple interfaces and thus multiple IP addresses that appear in traceroute data. As these aliases all identify the same node, it would be redundant and an inaccurate representation of the topology to include a given router more than once. Not only that, but any edges coming from a node with aliases further exacerbates the flawed topological representation. For example, a router could be a critical point in the infrastructure but could have a dozen aliases with a dozen links each; if these aliases are not resolved, the router would seem far less important than its actual status as a node of high degree. In other words, such a router could be a single point of failure in the network but would not appear as such without alias resolution.

The three techniques used for performing alias resolution in this thesis are Monotonic ID-based Alias Resolution (MIDAR), *speedtrap*, and Domain Name System (DNS) analysis. MIDAR is used for IPv4 alias resolution and utilizes the IP ID value of response packets returned from probes sent to a given router. This IP ID value is implemented as an incrementing counter within a router. If probes are sent to suspected aliases at the same time, then the IP ID values returned should be nearly identical. MIDAR also uses multiple vantage points, various probe techniques, and a sliding-window probe scheduling algorithm in order to enable Internet-scale alias resolution [13].

The next technique, *speedtrap*, allows alias resolution on a set of IPv6 addresses. In contrast to IPv4 packets, IPv6 packet headers have no IP ID field innately. *Speedtrap* instead in-

duces fragmentation by sending routers ICMP "packet too big" messages with a Maximum Transmission Unit (MTU) field that is smaller than the packet size. Fragmented responses may then be sent to target hosts if the router follows IPv6 protocol; however, some routers choose to ignore these ICMP "packet too big" messages. Unlike MIDAR, the IP ID value of IPv6 packets is similar across multiple routers and has no natural velocity (i.e., the IP ID value only increments in response to these fragmented packets). Due to this limitation, a different approach has to be taken. *Speedtrap* instead utilizes router responses to fragmented IPv6 packets in "a particular temporal pattern that produces distinguishing per-router fingerprints" [14]. Although these techniques are not perfect (e.g., they require non-random IP IDs to be returned and for routers to respond which may not be the case), they allow network topologies to be created much more accurately.

The last technique, DNS analysis, examines the domain name of routers to determine if two IP addresses are aliases [15]. Unlike the previous two techniques, this is a process that is done manually. To perform DNS analysis, the domain names associated with two or more IP addresses are compared to check if they are identical or at least similar. When it comes to similarity, the domain names must have some discernible naming convention for this technique to work. As DNS analysis is a labor-intensive way of performing alias resolution, it is not used in the creation of topology maps. Instead, this technique comes into play during the topological analysis in Chapter 4 to determine the congruence of IPv4 and IPv6 cellular networks.

2.1.6 Cellular-specific Infrastructure

Like traditional wireless networks (e.g., home Wi-Fi), the last hop of a cellular network is wireless. However, unlike a wireless network at home or at work, cellular networks are composed of geographic regions called cells that contain cellular towers to facilitate wireless communication. Mobile devices such as cellphones contain portable transceivers that allow communication with these towers using radio frequencies. Once a signal reaches the cellular tower, the data enters a unified IP core for handling either voice or data services [16]. The cellular-specific infrastructure is not elucidated in the topology maps, but it is good to have an understanding of how mobile networks differ from traditional networks.

2.1.7 User Agents

A User Agent (UA) is a piece of software that represents a user. In Hypertext Transfer Protocol (HTTP), the UA is often the web browser being used. These UAs include a User-Agent string that identifies the device that is making the HTTP request. At a minimum, this string typically includes the browser name, browser version, and operating system of the client [17]. These UA strings can be logged by web servers and are used in Chapter 4 to show a method of determining correlation between mobile UAs and mobile IP addresses.

2.2 Datasets

Most of the data used for this thesis was sourced from the Center for Applied Internet Data Analysis (CAIDA). The primary datasets used to create the network graphs were CAIDA's IPv4 and IPv6 traceroute data from the Archipelago (ARK) measurement infrastructure [18], [19]. Datasets from ARK to perform reverse DNS name lookup of routers were also used [20], [21]. The final CAIDA dataset used was RouteViews Prefix-to-AS mappings, which provided a correlation for MaxMind cellular prefixes and mobile ASes [22]. The MaxMind GeoIP2 Connection Type database defined cellular prefixes to determine what constitutes a mobile AS [23]. BGP Routing Information Base (RIB) data from RouteViews allowed the creation of radix trees in order to map IP addresses to ASNs [24], [25]. Lastly, private web access log data from the Center for Measurement and Analysis of Network Data (CMAND) lab was used to perform UA string correlation as an alternative method of finding cellular prefixes [26]. All the CAIDA/ARK and RouteViews datasets pertained to January 27, 2019, the MaxMind data was sourced from May 5, 2020, and the access log contained data from September 5, 2015 to July 8, 2020.

2.3 Related Work

There are numerous related works that have contributed to informing and enabling mapping the topologies of mobile providers. This section details these key pieces of the literature in chronological order of release. Huffaker et al. defined best practices in data selection for creating topology maps and compared topological graphs at the interface-, router-, and AS-levels [27]. This paper demonstrated the types of topological metrics one might focus on when doing comparisons such as the number of nodes and edges, average degree, and degree distribution which this thesis uses for comparing mobile topologies. While not

strictly related to mapping, Czyz et al. explored global datasets and found an exponential increase in native IPv6 traffic and adoption rate [28]. This increase in IPv6 traffic could be correlated with the growth of cellular devices in the last decade which also implies the growth of mobile networks. Giotsas et al. observed the convergence of structure and routing paths in IPv4 and IPv6 topologies [29]. Convergence is related to the idea of congruence between IPv4 and IPv6 topologies that this thesis investigates, which is when identical nodes exist between the two mappings (i.e., the same router is used in IPv4 and IPv6).

Almeida et al. characterized load balancing as a common feature of the IPv6 Internet, similar to the IPv4 Internet [30]. This impacts the topologies inferred from traceroute datasets; the presence of load balancing can create an incomplete view of a topology as some routers and links may never be iterated through and thus not appear at all. Beverly et al. developed state-of-the-art IPv6 topology discovery techniques and provided background for IPv4 topology discovery and how methods differ [8]. This thesis utilizes the datasets this work enabled by creating topology graphs and performing comparative analyses between the IPv4 and IPv6 topologies of cellular providers. Carisimo et al. studied the evolution of the Internet toward a multimedia network as large content providers have gained a foothold in the core of the AS ecosystem [31]. This transition has topological implications as the shifting infrastructure can alter the layout of networking devices within ASes. Jia et al. tracked IPv6 deployment and found that most core Internet providers have deployed IPv6, but that adoption in Europe and Asia is greater than in North America [32]. During the analyses of mobile topologies, this thesis considers the IPv6 adoption rate as a possible explanation for topological differences.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3:
Methodology

This chapter discusses the challenges of classifying mobile ASes and the steps taken in creating a mobile AS topology. Broadly, these steps include aggregating the traceroute data from CAIDA's Archipelago (ARK) [18], [19], using a radix tree to perform longest prefix matching of traceroute hops to determine mappings between IP addresses and ASes, creating network graphs, resolving aliases for nodes within each AS, and attributing a MaxMind prefix and domain name to each node. Figure 3.1 provides a high-level summary of the process. Of note is that classifying the mobile topology is a hard problem; these are initial steps toward making progress on the problem but is by no means a comprehensive solution.

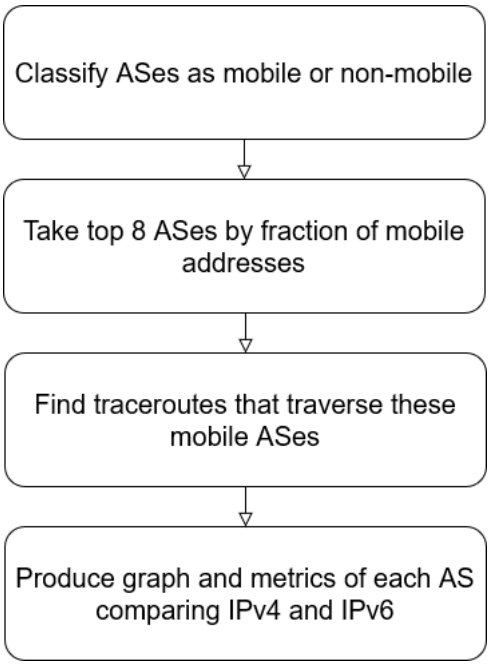

Figure 3.1. Methodology Flowchart.

3.1 Identifying Mobile IP Addresses

At present, no reliable method exists in the network measurement community to identify mobile (cellular) IP addresses: i.e., whether an IP address is used for either mobile infrastructure or for mobile hosts. As a first intuition, one might take a naive approach and perform a WHOIS registry lookup for ASes that have the word "mobile", "cellular", "cell" or other related terms in their name, or search for well-known cellular providers. Regardless of how one searches through the registry for these keywords, this technique is error prone and by no means exhaustive. For example, one might search on CAIDA's ASRank (which is a tool that pulls from WHOIS registry info) for "Sprint", which would lead them to AS1239 [33]. Based on this, one might conclude that AS1239 is cellular infrastructure. However, this is not the case as AS1239 is actually a wired network. This can be seen by running a traceroute from Sprint's looking glass which resides on wired infrastructure [34], as shown in Figure 3.2:

Figure 3.2. Traceroute from Sprint's Looking Glass.

From here, longest prefix matching of the first hop (2600:0:2:1239:144:232:22:71) against RIB data can be performed to determine the ASN of the IP address. This technique is elaborated upon in a later section. For now, Team Cymru's IP to ASN mapping service can be used to verify the result:

Figure 3.3. IP to ASN Lookup of Sprint Router.

As shown in Figure 3.3, AS1239 contains routers that are part of Sprint's wired network. This does not necessarily preclude AS1239 from having a mixed infrastructure (i.e., a network that has both fixed and mobile IP addresses within the same AS), although later analysis showed AS1239 to only be wired. However, it does show that the assumption of using the organization name registered to an ASN to conclude that an AS is mobile is an unreliable method.

To actually identify mobile ASes, the MaxMind GeoIP2 Connection Type commercial database [23] provides a comma-separated values (CSV) file with prefixes and connection types. These connection types are not limited to cellular; dial-up, cable, DSL, and corporate connection types are also included. Based on the size of the prefixes, 79 percent of the IPv4 address space is covered, excluding private addresses. The cellular prefixes encompass eight percent of the IPv4 address space. Since much of the IPv6 address space is unused, it is difficult to conclude IPv6 coverage. Despite this, the IPv6 cellular prefixes account for seven percent of the IPv6 address space represented by the MaxMind database. This leads one to believe MaxMind's IPv6 coverage is approximately equal to its IPv4 coverage. Lastly, MaxMind specifically claims that this database is "about 95 percent accurate for identifying cellular connections" [23].

Despite this claim, there is little information about how the data is obtained and how this proprietary database constructed. Although the database is used in this research, the extent of its validity is unverified. This is important to note as this database is used as a point of comparison when using the CMAND access log to perform UA string correlation as an alternative way of finding mobile IP addresses. Assuming both UA string correlation and the MaxMind data is sound, prefixes that MaxMind claims to be cellular should be reflected in the UA string correlation results.

3.2 Determining Mobile ASes

Now that a set of known mobile prefixes has been obtained, there must be a way to correlate prefixes to ASNs. In order to achieve this goal, the RouteViews Prefix-to-AS mappings dataset is used. This file contains one column with prefixes and another column with ASNs. To determine what constitutes a mobile AS, Python is used to test each prefix in the RouteViews Prefix-to-AS dataset against the MaxMind cellular prefixes. If a RouteViews prefix is a subnet of any MaxMind cellular prefix (i.e., all of the IP addresses in the RouteViews prefix are contained within the MaxMind prefix), then the ASN listed is considered to contain mobile addresses and have cellular infrastructure. The other case where a MaxMind prefix is a subnet of a RouteViews prefix was not considered as the initial case alone returned numerous ASes to examine. However, future work may find this investigation worth pursuing. Once the set of mobile ASes is returned, the ASes are then sorted by the total number of mobile IP addresses from largest to smallest. This is done by iterating through the Prefix-to-AS mappings and summing the number of mobile addresses associated with each AS. This step is taken as approximately 300 ASes are returned; a method to prioritize which ASes get examined provides a useful starting point for the mapping process.

There is a consideration to be made as to whether each prefix in the RouteViews .csv file associated with a given ASN should have to be a subnet of a MaxMind cellular prefix in order for an AS to be classified as mobile. Ultimately, this was decided against as both datasets are missing some coverage of the IP address space (i.e., the MaxMind data covers 79 percent of the total IPv4 space and the RouteViews data covers 93 percent of the total IPv4 space). If such a requirement were enforced, the potential for false negatives would increase. Besides, this step mainly serves as a filtering function for determining which ASes are worth mapping. When the traceroute data is processed, the addresses are ensured to be within the MaxMind cellular prefixes. This does bring up the question of what the definition of a mobile AS is. Is it an AS that has at least one prefix, all prefixes, or some number of prefixes contained within a known cellular prefix? In this thesis, the definition of a mobile AS is that it must contain at least one cellular prefix. However, other definitions certainly have merit and would be worth exploring in future research.

3.3 Aggregating Traceroute Data

After creating a list of mobile ASes, the next step is to aggregate traceroute data which is used to correlate IP addresses with ASNs. The traceroute data comes from the two CAIDA ARK topology datasets and was collected with the scamper package [35]. This data comes from 254 IPv4 vantage points and 92 IPv6 vantage points. Each vantage point outputs an individual file in scamper's output file format called warts. This is where the aggregation part comes in, which is done by using a Bash script that converts each warts file to .csv and concatenates the results. This leads to two data files: a 25 gigabyte IPv4 file and an 18 gigabyte IPv6 file.

All the traceroute data comes from January 27, 2019, as that was the most recent publicly available data when this work began. Looking at multiple time periods could mitigate potential anomalies in the data, but a single 24-hour period is sufficient for this exploration into mobile ASes. Future work could examine and compare mobile ASes using different days of traceroute data, or even reduce the number of vantage points used in order to craft a particular view of the topology.

3.4 Performing Longest Prefix Matching

A radix tree is a data structure typically used for storing and retrieving IPv4 and IPv6 prefixes and allows for fast lookups. In this thesis, the py-radix library handles the radix tree implementation [36]. RouteViews RIB files[2] store BGP information about network prefixes, each with an associated AS path. This AS path can be used to determine which AS originated a given prefix, which creates a mapping from prefixes to ASes. The RIB file for the date matching the traceroute data is processed and each prefix is stored as a node in the radix tree. The origin AS is also associated with each node as an additional data member. Once the radix tree is built, best-match search of a traceroute IP address returns the node with the longest matching prefix. The origin AS is then extracted from the returned node and the mapping between IP address and AS is saved. This process is repeated for every IP address in the traceroute. In summary, the radix tree creates a mapping between a prefix and an AS, and then searching through the tree using an IP address defines a mapping between an IP address and an AS. Once an IP-address-to-AS mapping has been defined, the AS of

[2]These files are in the Multi-Threaded Routing Toolkit (MRT) format, and were converted to a text-based format using mrt2bgpdump.py of the mrtparse library [37].

IP addresses that appear in the traceroute data can be identified. Cellular IP addresses that belong to mobile ASes are added as nodes of the router-level topological graph.

3.5 Creating Network Graphs

Now that IP-address-to-AS mappings and mobile ASes are known, the next step can be done. By iterating through the traceroute data and tracking hops between IP addresses in mobile ASes of interest, router-level topology graphs can be created. More specifically, mobile IP addresses are found by iterating through the IP-address-to-AS mappings and filtering on a given AS of interest. From there, a GUESS (Graph Exploration System) Data Format (GDF) file can be made. The GDF file format is one used with the graphing software Gephi. This is essentially a more robust .csv file that allows one to specify nodes and edges along with some attributes (e.g., weight, color) [38].

These extra attributes are important, as it is desirable to track the ingress (first hop into the network) and egress (last hop before exiting the network) nodes of a mobile topology. For clarification, these ingress and egress nodes are with respect to the vantage point sending the traceroute probes rather than from a mobile client's perspective. This tracking is done by assigning colors to nodes: green nodes are ingress only, red nodes are egress only, blue nodes are both ingress and egress, and black nodes are neither ingress nor egress (i.e., intermediate nodes). When the traceroute data is iterated through, this attribution is achieved by tracking the previous router visited belonging to the AS of interest. If there was no previous router visited, then the node is an ingress node. If the next hop is not in the AS, then the router is an egress node. If there was a previous router visited and the next hop is in the AS, then the router is an intermediate node. If a given node has already been flagged as an ingress or egress node and the other condition is met, then the node is removed from the set of nodes and added again with an updated color.

The reason such importance is given to these nodes is due to the implications they have on the network topology. If an AS only had a single or very few ingress nodes in its IPv6 topology, those nodes would be of special importance in terms of securing or monitoring the infrastructure. Likewise, having a similar percent of ingress nodes between the IPv4 and IPv6 topologies of the same AS would also be noteworthy as this could be a potential indicator of running dual-stack IPv4/IPv6 infrastructure. The goal of labelling these nodes

is to gain greater insight into how the IPv4 and IPv6 topologies compare, which would otherwise remain unexplored if no coloring occurred.

As an example of labelling these nodes, imagine there are five routers within an AS. These routers are labeled A through E. The first traceroute is from an arbitrary source to an arbitrary destination. The traceroute reaches its first hop into the AS, router A, which is the ingress node. In the first traceroute shown in Figure 3.4, router A is colored green as it served as an entry into the AS. The next hops of this traceroute are from A to B, B to C, C to D, then from D to a router outside the AS. Routers B and C are colored black as they acted as intermediate hops into the AS (i.e., the prior hop and the following hop were both in the AS). Router D is colored red because it was the last router reached before leaving the AS, which is an egress node.

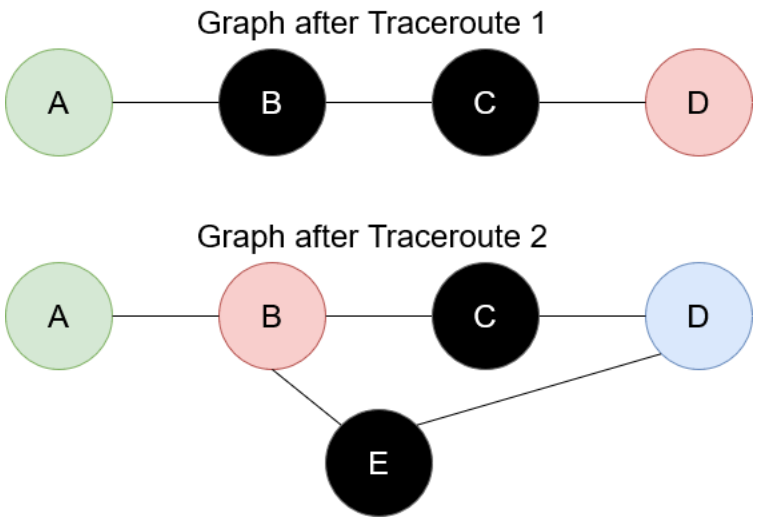

Figure 3.4. Network Graph Traceroute Example.

Building upon this example, a second traceroute is run from another vantage point to another destination. This traceroute enters the AS at router D. Router D was seen as an egress node in the first traceroute and as an ingress node in the second traceroute. This changes its color from red to blue to represent its dual usage. The next hops in the AS are from D to E, E to B, then to a destination not within the AS. As router E had a previous and a next hop within the AS, it is colored black as an intermediate node. In this traceroute, router B acted

as the last hop so it is now an egress node colored red. Even though in the first traceroute it was an intermediate node, ingress and egress nodes take priority in the coloring. For every traceroute in the dataset that traverses the AS, the graph of the AS may change.

During this process, hops between routers within the same AS are being added to the set of edges. Once the entire traceroute has been iterated through, the node and edge definitions can be saved to GDF format. With the implemented algorithm, intermediate nodes are only inferred from this set of edges by Gephi and are not explicitly added to the set of nodes. This is also why these nodes are black as it is the default color. This is initially fine, until other attributes such as MaxMind prefixes and DNS names are added to all the nodes, including intermediate nodes. Despite this minor issue, it is straightforward to edit the GDF files programmatically and explicitly define the inferred nodes in the same way Gephi might. This is done by keeping track of the defined nodes and the edge nodes. If an edge node already existed as a defined node, then it is left alone. This is because ingress and egress nodes take precedence over intermediate nodes. If the edge node was not already defined, then it is added to the node definitions as an intermediate node.

3.6 Resolving Aliases

As mentioned previously, alias resolution is an important step in ensuring the accuracy of a network topology. To reiterate, IPv4 alias resolution is performed using MIDAR while the technique for IPv6 alias resolution is *speedtrap*. While the ideas behind these techniques themselves might be complicated, making use of them is simple enough. MIDAR can be used as a web service [39] where the set of IPv4 addresses within a mobile AS can be passed as a text file and sets of aliases are returned as a text file. *Speedtrap* comes included with the scamper library used for warts files earlier, and takes and returns the same parameters but in IPv6 instead. Occurrences of these aliases are then overwritten in the GDF files, using the first alias returned as the primary IP address. During this process of overwriting aliases, node colors are also tracked and ingress/egress nodes are combined as necessary.

3.7 Attributing MaxMind Prefixes and Domain Names

Another important aspect of managing the data is to attribute MaxMind prefixes and domain names to each node. These domain names are defined by the DNS PTR records for the

IP address of each router interface (represented as a node) that appears in a topology. As the MaxMind prefixes are already known, checking each IP address for the prefix it resides in and appending it as an attribute to its node definition is all that is needed to be done. Next, the domain names must be attributed to each node definition. This is key in determining the congruence between networks, which provides insight as to how the IPv4 and IPv6 topologies are correlated. CAIDA provides a DNS names dataset, which provides IP addresses and the domain name associated with them. As the topology data contains IP addresses, the domain names can be matched from this dataset and added as an attribute. This action was taken; however, there were many unknown domain names as a number of IP addresses appeared in the traceroute data but the domain name was not resolved in the DNS names data. In order to resolve this, dig was used to perform reverse DNS lookups which map an IP address to a domain name. These domain names were then added as an attribute to the existing data. With this, the GDF files are now complete and ready for analysis.

3.8 Limitations

Mapping mobile topologies poses a challenging problem and this methodology is by no means perfect. This research makes a best effort attempt toward classifying mobile infrastructure, but there are a number of limitations and possible sources of error. To begin with, the MaxMind Connection Type Database [23] may be incorrect. Due to the proprietary nature of the database, it is difficult to ascertain the soundness of its data collection methodology and the data itself. Errors in what this dataset considers cellular addresses can lead to errors in mapping mobile providers. Additionally, the network boundaries may be incorrect. Traceroutes return IP addresses, and a customer's router could have an interface numbered with its provider's IP address range. Border mapping is a separate area of research [40], and the simplistic manner in which this work handles borders may have errors.

The next source of error is wrong and/or unresolved aliases. Alias resolution is not a perfect technique, and both incorrect and unresolved aliases create topological errors. The sets of traceroutes in the CAIDA ARK data [18], [19] are also incomplete; there is not complete coverage through every mobile AS as routers may be missed or simply do not respond to probes. There is also the issue of having a limited number of vantage points which determine how much of the network and which parts of the network are seen. For example, there were no vantage points originating from cellular providers, so it is highly probable that much

provider-specific infrastructure was never mapped. The number of vantage points between IPv4 and IPv6 also vary which could lead to crafting topologies of differing specificities.

Lastly, the DNS PTR records are not complete or even necessarily correct. When looking for congruence using this information, it is possible for congruent infrastructure to be mistaken as incongruent if the DNS names are incorrect. The reverse where incongruent infrastructure is mistaken as congruent is also possible for the same reason. With both DNS PTR records and the classification of mobile IP addresses, we have no ground truth. All these limitations create some amount of potential error in the methodology and should be considered when interpreting the results.

CHAPTER 4:
Results

This chapter delves into the analysis of mobile IPv4 and IPv6 topologies and their topological similarities and differences. Case studies of eight different, geographically dispersed ASes that contain both IPv4 and IPv6 mobile addresses constitute the results of this analysis. These eight ASes were specifically chosen because they represent large networks in multiple geographic regions in order to provide a more representative picture of the differences between IPv4 and IPv6 infrastructure.

An overall evaluation of these case studies and any trends observed are elaborated upon in the next chapter. Furthermore, there are more mobile ASes than the ones analyzed here; we discuss limitations in examining these other ASes. Lastly, we confirm that using MaxMind's Connection Type database [23] is a reasonable approach for our methodology by validating its mobile network prefix labels against a different inference technique based on HTTP user agent correlation.

4.1 Case Studies

These case studies analyze the mobile topology of eight ASes across six different providers: Vodafone Group in Europe, SoftBank Mobile Corp. in Japan, T-Mobile in the Czech Republic, Verizon Wireless in the United States, China Mobile in China, and Reliance Jio in India. The Verizon Wireless and China Mobile case studies each contain two ASes for examination. These analyses are focused on the following questions:

- How do the number of nodes and edges differ between IPv4 and IPv6?
- How do other metrics such as average degree, average path length, and network diameter compare between IPv4 and IPv6 topologies?
- How does the degree distribution compare between IPv4 and IPv6?
- Does congruence exist between the IPv4 and IPv6 infrastructure? In other words, do any of the same routers exist in both the IPv4 and IPv6 topology?

As an additional note about congruence, the main method for determining such is using

domain names. This is why reverse DNS lookup was performed previously; if these domain names match between IPv4 and IPv6 nodes, then congruence likely exists. As a caveat, this method is not a conclusive determination that two IP addresses are the same router. We do not have ground truth that allows us to prove with certainty that identical domain names map to the same router (i.e., the DNS PTR records could be incorrect), but it is a promising sign. The percent of congruence is calculated by taking the number of nodes that match divided by the total number of nodes in the IPv6 topology.

Table 4.1. Case Study ASN Statistics.

Name	**ASN**	**IPv4 Mobile Prefix Count**	**IPv6 Mobile Prefix Count**	**IPv4 Trace-route Count**	**IPv6 Trace-route Count**	**IPv4 Vantage Point Count**	**IPv6 Vantage Point Count**
Vodafone	1273	55	1	290,278	38,117	254	87
SoftBank	4725	67	2	10,918	1,880	254	85
T-Mobile	5588	108	12	17,240	14,142	254	85
Verizon	6167	10	4	100,902	55,008	254	85
Verizon	22394	9	2	4,075	2,092	254	85
China Mobile	9808	99	3	358,729	115,416	254	85
China Mobile	56040	18	1	11,652	7,360	254	83
Reliance Jio	55836	14	10	550	139,810	203	86

Table 4.1 provides statistics for the case study ASes regarding the number of MaxMind cellular prefixes [23] seen and the amount of CAIDA ARK traceroutes and vantage points [18], [19] used to create each IPv4 and IPv6 topology. To provide points of reference, there were 16,233 IPv4 cellular prefixes and 3,234 IPv6 cellular prefixes within the MaxMind dataset. The ARK traceroute datasets spanned 254 IPv4 vantage points that sent 20,582,417 traceroutes, and 92 IPv6 vantage points that sent 12,477,586 traceroutes. Despite the number of traceroutes analyzed, all the limitations discussed in Chapter 3 still apply (e.g., the traceroutes may not have complete coverage of any given AS, and unequal numbers of vantage points skews coverage).

4.1.1 Vodafone Group PLC (AS 1273)

The most interesting case study is that of AS 1273 which belongs to the Vodafone Group in Europe [41]. This is because the set of DNS names for both IPv4 and IPv6 successfully resolved for the majority of nodes. In every other case, the DNS names fail to resolve for IPv4 and/or IPv6. Due to this limitation in the data, AS 1273 is the only AS where congruence can be properly evaluated.

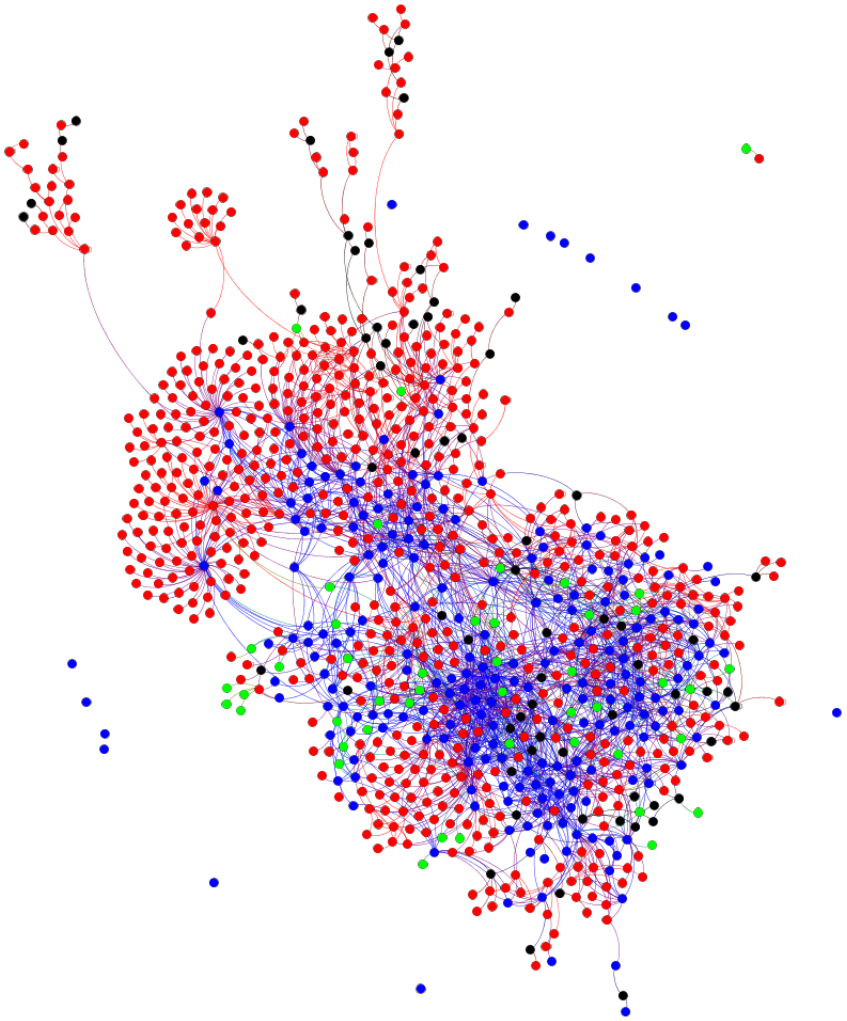

Figure 4.1. AS 1273 IPv4 Graph.

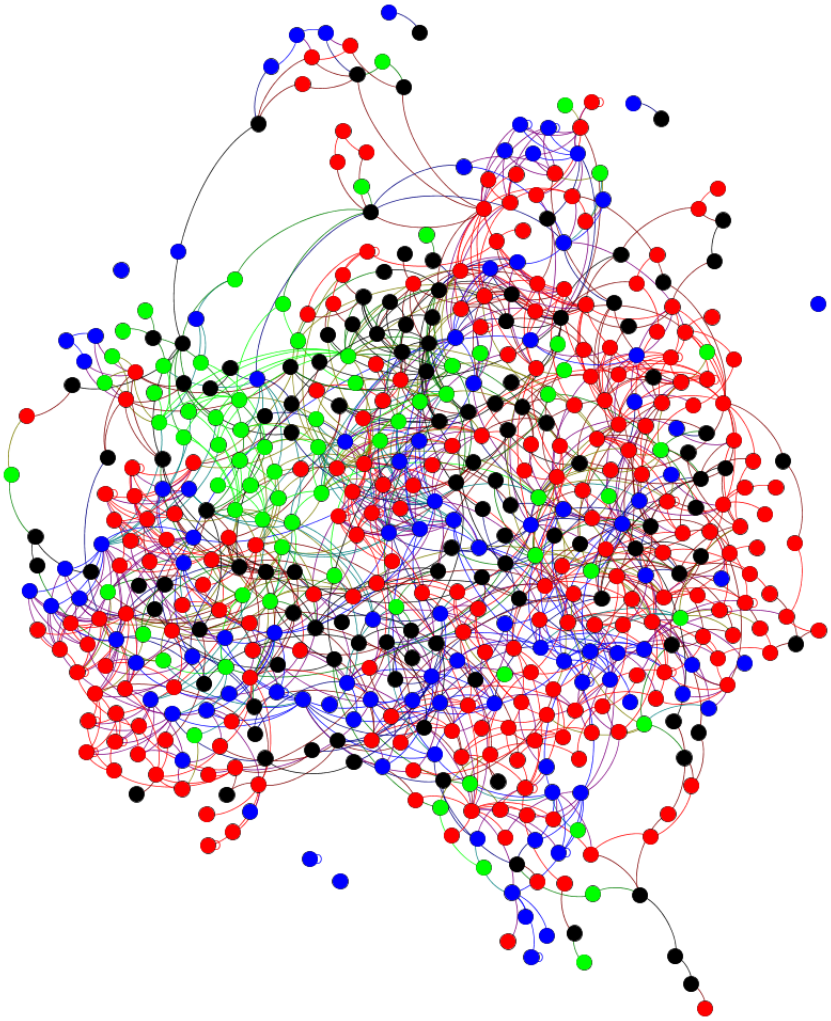

Figure 4.2. AS 1273 IPv6 Graph.

Figures 4.1 and 4.2 illustrate that the created topology of an AS can be quite complex. However, even at a glance it is apparent that there is not a one-to-one correspondence between the IPv4 and IPv6 topologies of AS 1273. The force-directed graph algorithm, ForceAtlas2, dictated the layout of these graphs and every other graph in this thesis. This is a generic, intuitive way of network graph visualization that focuses on maintaining consistent edge lengths and uniform node distribution [42]. In other words, the location of a node has no particular meaning other than for readability purposes. To better gain insight into these

topologies, Table 4.2 presents graph statistics.

Table 4.2. AS 1273 Statistics.

Metric	**IPv4**	**IPv6**
Total Nodes	962	526
Total Edges	2572	1417
Ingress (Green) Nodes	46	71
Egress (Red) Nodes	626	229
Ingress/Egress (Blue) Nodes	224	104
Intermediate (Black) Nodes	66	122
Average Degree	5.35	5.39
Average Path Length	4.54	4.55
Network Diameter	14	12

In terms of how the IPv4 and IPv6 graphs appear visually and based on the different number of nodes and edges, these topologies are dissimilar at the surface level. Yet by digging deeper into the statistics underlying these topologies, similarities can be found. Both IPv4 and IPv6 have an almost identical average degree and average path length, and network diameters that are quite close. The ratio of nodes to edges is also nearly identical with 0.374 (962 / 2572) nodes-to-edges in IPv4 and 0.371 (526 / 1417) nodes-to-edges in IPv6. The next piece up for examination is the degree distribution of these topologies.

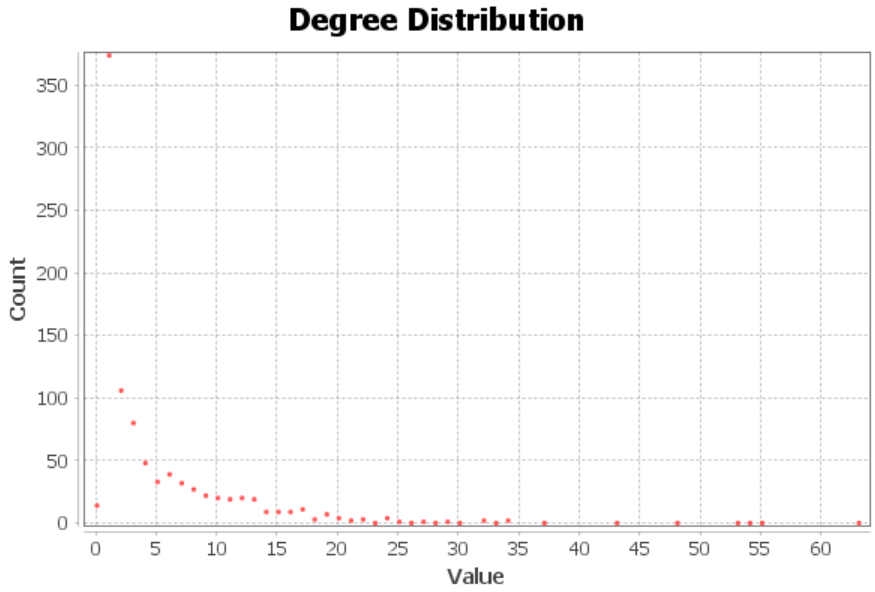

Figure 4.3. AS 1273 IPv4 Graph Degree Distribution.

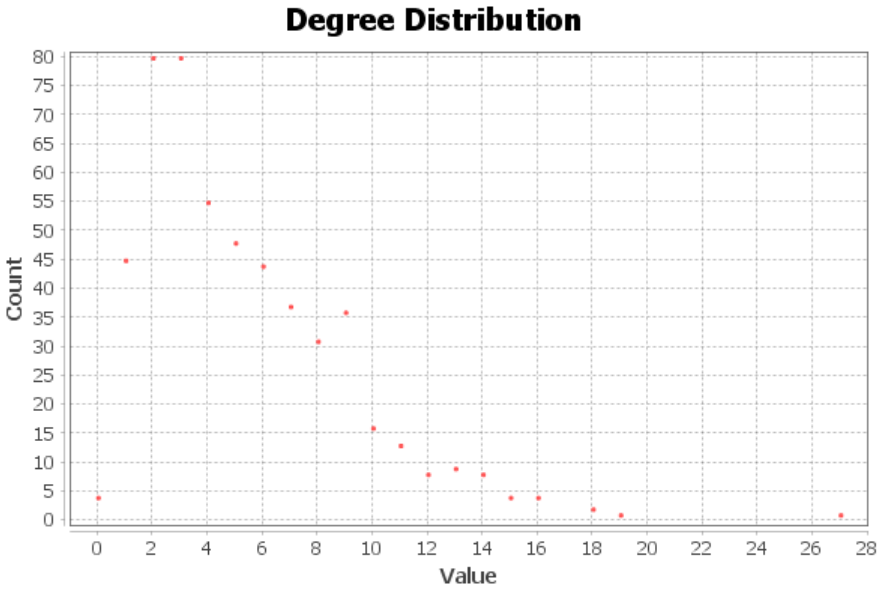

Figure 4.4. AS 1273 IPv6 Graph Degree Distribution.

As shown in Figures 4.3 and 4.4, the degree distribution of these topologies is fairly typical of what one would expect to see in a network. Most of the nodes are of a relatively low degree between two and ten, with the count of nodes decreasing as the degree value increases. Although certainly not exact and with more variance in the IPv6 distribution possibly due to the smaller sample size, the degree distribution trend seems shared between IPv4 and IPv6. All of these similar statistics are indicators of congruence, but to further examine congruence in the infrastructure requires the use of DNS names.

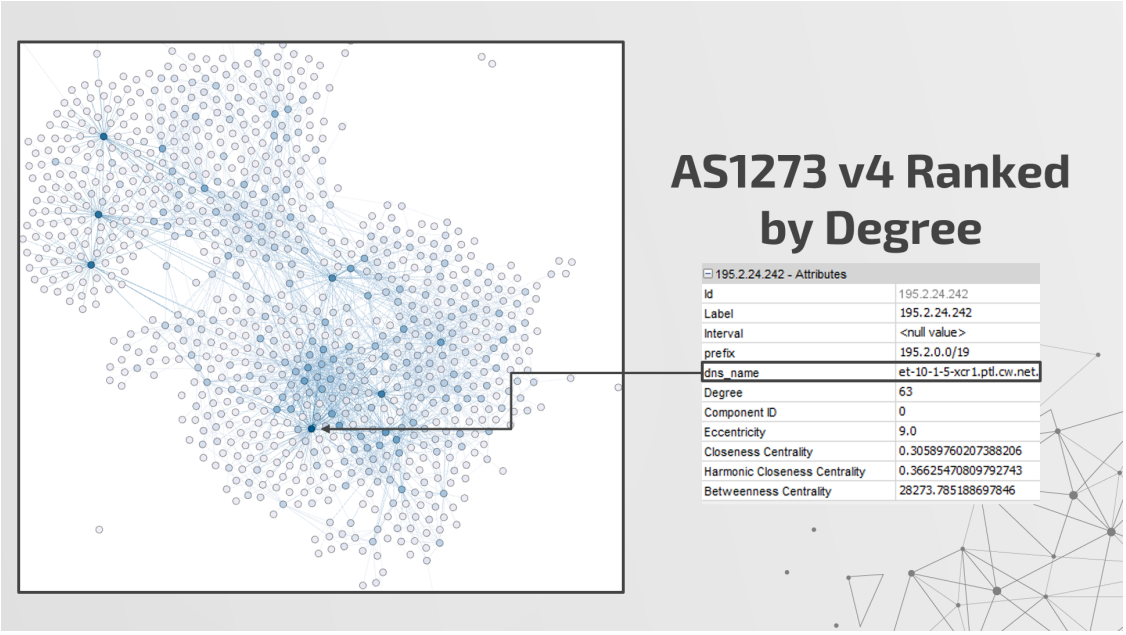

Figure 4.5. AS 1273 IPv4 Graph Ranked by Degree.

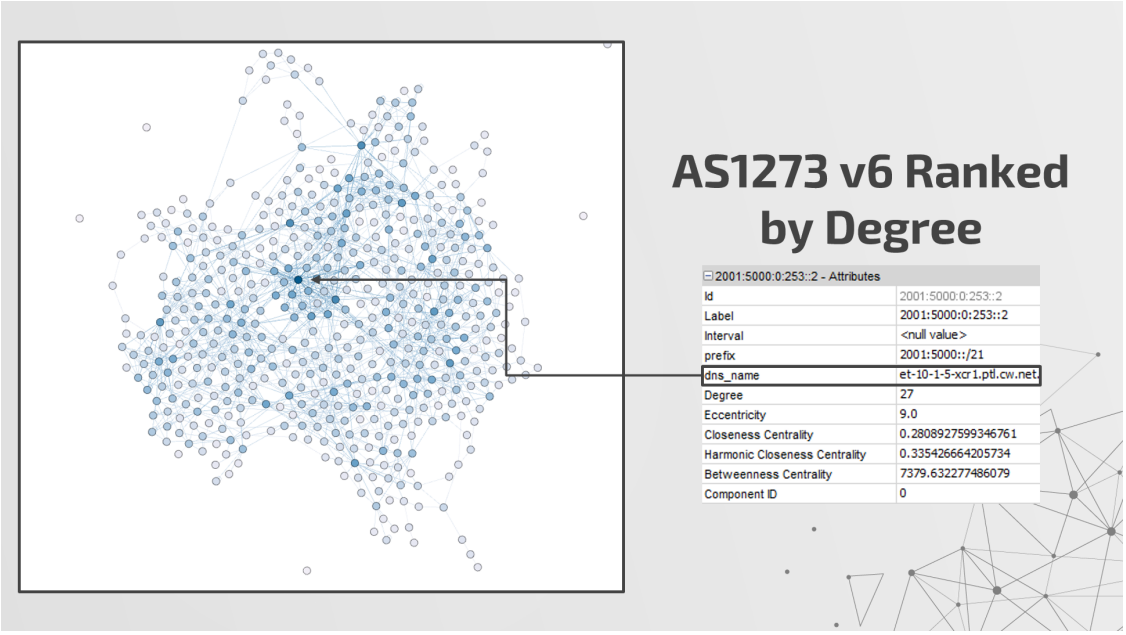

Figure 4.6. AS 1273 IPv6 Graph Ranked by Degree.

In Figures 4.5 and 4.6, the graphs have a degree ranking overlay applied to them which means that nodes of a higher degree are given a color of a darker shade. Both of these figures show the highest degree node in their respective graph selected. These nodes are 195.2.24.242 and 2001:5000:0:253::2, and both of them resolve to the same DNS name et-10-1-5-xcr1.ptl.cw.net. This is a clear sign that these IPv4 and IPv6 topologies have some amount of congruence. Interestingly enough, this node acts as both an ingress and egress node in the IPv4 topology, but only acts as an egress node in the IPv6 topology. The trend does not continue where the second-highest degree nodes resolve to the same DNS name. However, this is logical; the IPv4 topology has nearly twice as many nodes as the IPv6 topology and it would be anomalous if the DNS mapping were completely consistent between the two.

In the end, there were 263 shared nodes between IPv4 and IPv6 which means that 50 percent of IPv6 nodes shared a DNS name with an IPv4 node. The congruence of these shared nodes plays a greater role in determining the resilience and robustness of the network as a whole compared to their non-shared counterparts. With regard to security, these congruent nodes would be the ones that are of higher concern as they have the potential to impact the greatest number of users. As a caveat, this assumes the non-congruent nodes truly are not shared

between IPv4 and IPv6. If the issue is a lack of coverage in the traceroute data, then the non-congruent nodes may actually be congruent and therefore have a similar level of importance in the network.

4.1.2 SoftBank Mobile Corp. (AS 4725)

The next topology to examine is that of SoftBank Mobile Corp. in Japan, which is registered to AS 4725 [43]. Unfortunately this is one of those mobile ASes where most of the IPv4 DNS names resolved, but none of the IPv6 DNS names resolved. This increases the challenge of determining congruence; however, the IPv6 infrastructure of AS 4725 poses this issue even without DNS names due to its size.

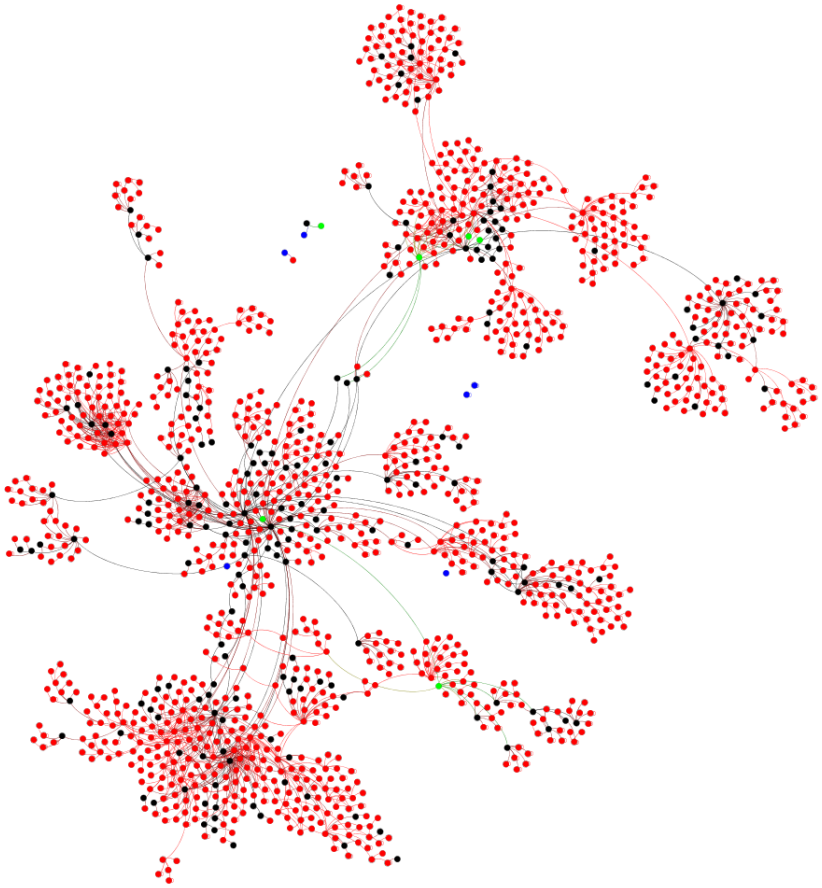

Figure 4.7. AS 4725 IPv4 Graph.

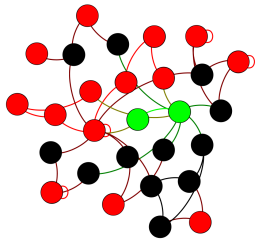

Figure 4.8. AS 4725 IPv6 Graph.

Based on Figures 4.7 and 4.8, the IPv4 infrastructure of AS 4725 massively eclipses the IPv6 topology. This result is actually unexpected; according to Akamai's *State of the Internet IPv6 Adoption Visualization*, Japan has a 35.8 percent IPv6 adoption rate [44]. Especially when one considers that mobile is a large driver of IPv6, the lack of IPv6 infrastructure here is surprising. Potential explanations for this are enumerated in a later section.

Table 4.3. AS 4725 Statistics.

Metric	**IPv4**	**IPv6**
Total Nodes	1408	28
Total Edges	2293	41
Ingress (Green) Nodes	6	2
Egress (Red) Nodes	1203	14
Ingress/Egress (Blue) Nodes	6	0
Intermediate (Black) Nodes	193	12
Average Degree	3.26	2.93
Average Path Length	5.97	2.91
Network Diameter	13	5

The two AS 4725 statistics that immediately stick out compared to AS 1273 are the number of ingress and egress nodes. In AS 4725, the number of ingress nodes is minuscule and the number of egress nodes is massive compared to the total number of nodes. While the same trend exists in AS 1273 where the ingress nodes make up the minority and the egress nodes constitute the majority, the scale and magnitude are on different levels for each AS. From

a network resilience standpoint, attacking the few ingress nodes could potentially cause a service disruption.

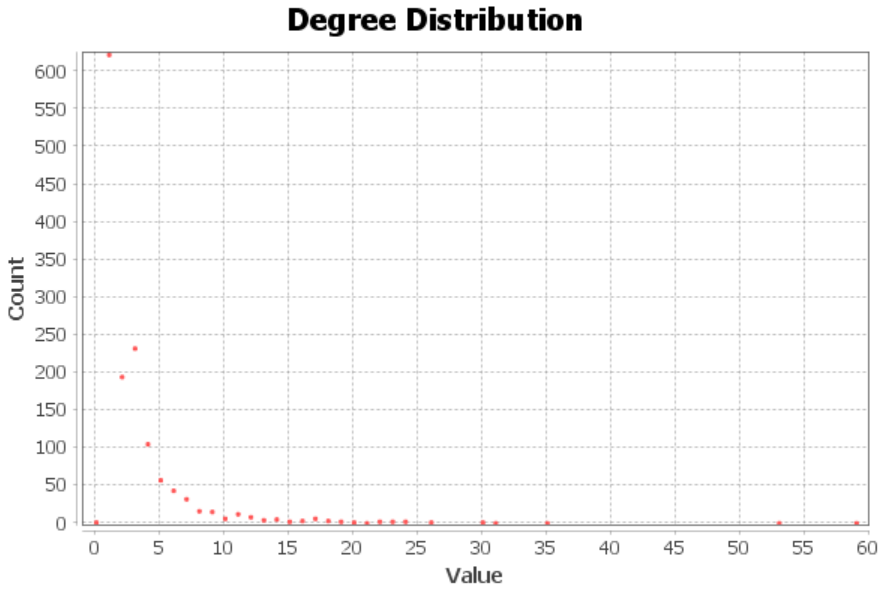

Figure 4.9. AS 4725 IPv4 Graph Degree Distribution.

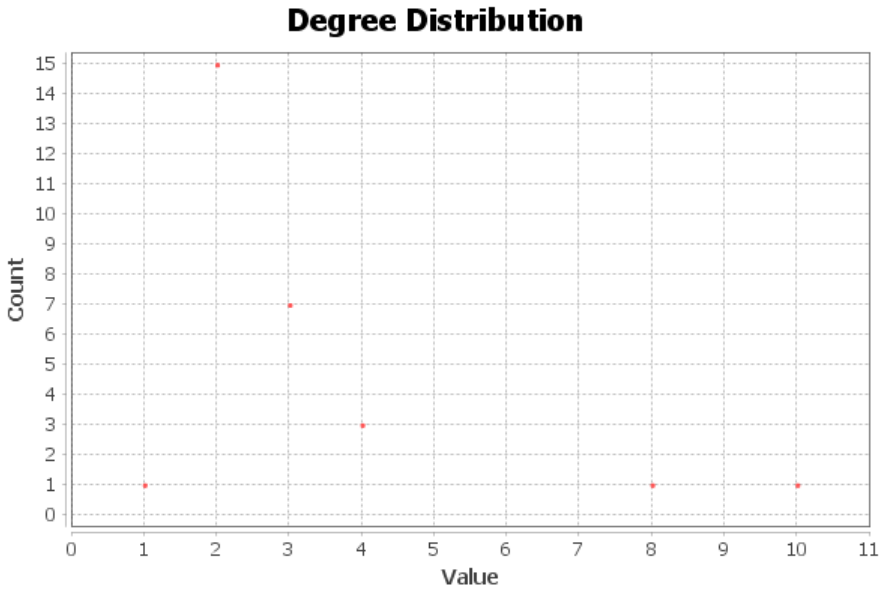

Figure 4.10. AS 4725 IPv6 Graph Degree Distribution.

The IPv4 degree distribution shown in Figure 4.9 follows the same expected pattern with more low degree nodes and decreasing amounts of higher degree nodes. The IPv6 degree distribution is similar, although the small amount of nodes makes the data look less consistent.

4.1.3 T-Mobile Czech Republic (AS 5588)

AS 5588 is the next infrastructure to examine and is registered to T-Mobile in the Czech Republic [45]. 578 out of 1993 IPv4 nodes successfully resolved DNS names. However, only three out of 156 IPv6 nodes successfully resolved. Of those three, none were an exact match for any IPv4 DNS name. Despite this, there were some similarities between the DNS names. For example, backbone-6.ce-colo.viridium.t-mobile.cz appeared in the IPv6 nodes, and backbone.police.viridiuim.t-mobile.cz appeared in the IPv4 nodes. Although these are not the same node (Note the "viridium" and "viriduim" discrepancies showing the unreliability of PTR records), the domain names are close enough to suggest some kind of shared infrastructure.

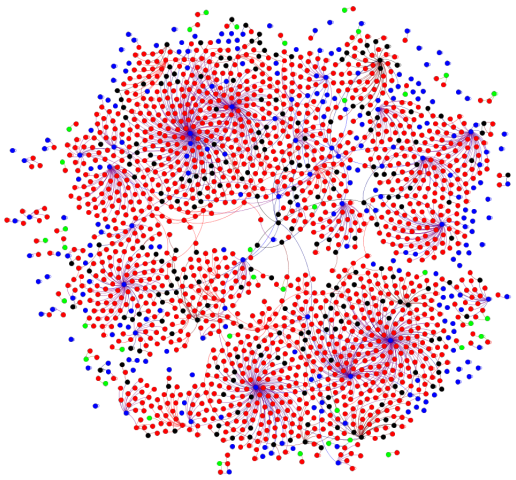

Figure 4.11. AS 5588 IPv4 Graph.

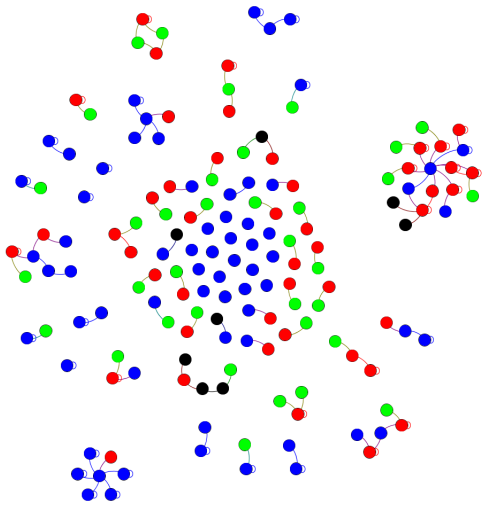

Figure 4.12. AS 5588 IPv6 Graph.

The IPv4 topology shown in Figure 4.11 looks typical of the IPv4 topologies seen so far. Conversely, the IPv6 topology in Figure 4.12 is disconnected and made up of multiple network components. The unconnected blue ingress/egress nodes are where the traceroutes hit either a gap limit or a loop. Counting the connected components that are not these lone nodes, there are 45 components total with 20 of them consisting of only two nodes. These results are summarized in Figure 4.13. Lastly, the Czech Republic only has an 11.5 percent IPv6 adoption rate [44], which is apparent here with how much larger the IPv4 graph is compared to the IPv6 graph.

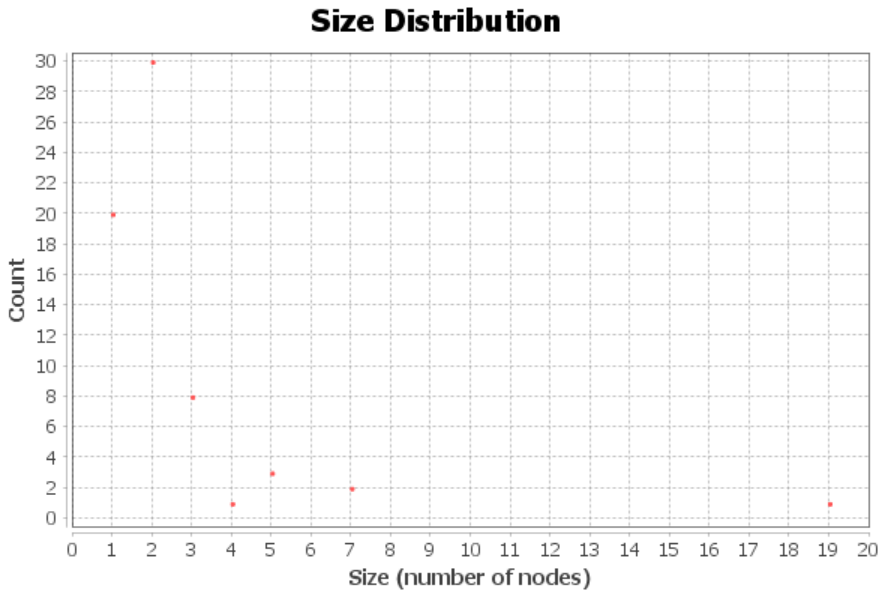

Figure 4.13. AS 5588 IPv6 Graph Size Distribution.

Table 4.4. AS 5588 Statistics.

Metric	**IPv4**	**IPv6**
Total Nodes	1993	156
Total Edges	2192	132
Ingress (Green) Nodes	37	35
Egress (Red) Nodes	1478	46
Ingress/Egress (Blue) Nodes	250	67
Intermediate (Black) Nodes	228	8
Average Degree	2.20	1.69
Average Path Length	7.29	2.29
Network Diameter	18	5

As shown in AS 4725 and in these statistics, the average degree of nodes is somewhat smaller and the average path length and network diameter are drastically smaller when there is a large difference in graph size. An interesting statistic that stems from the disjoint IPv6

graph is that the percent of ingress nodes is much higher than previously seen. Instead of making up less than ten percent of the network, the ingress nodes constitute 22.4 percent of this IPv6 network.

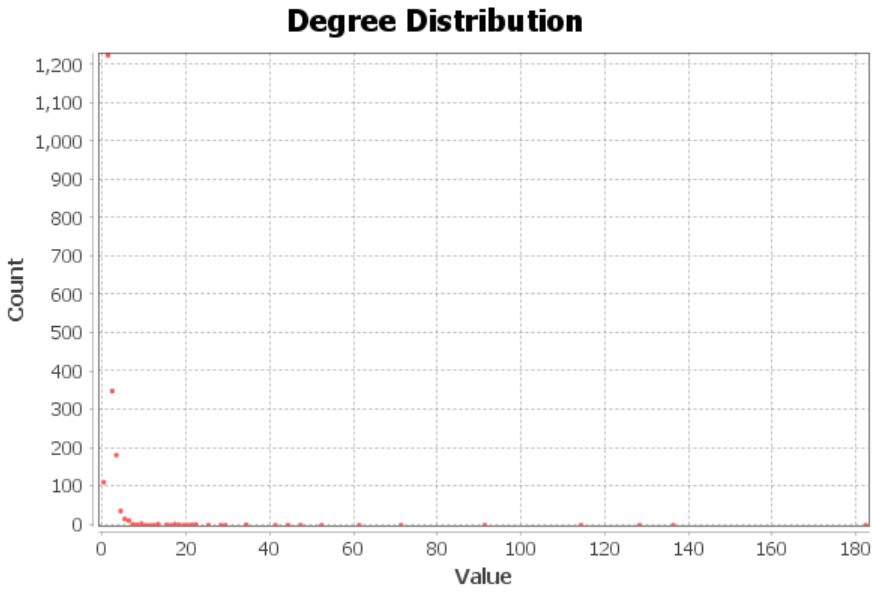

Figure 4.14. AS 5588 IPv4 Graph Degree Distribution.

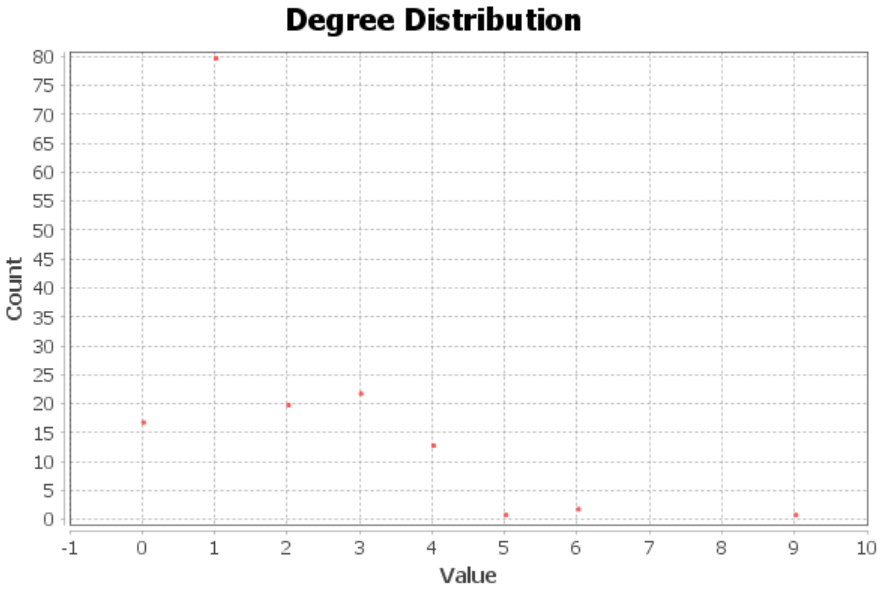

Figure 4.15. AS 5588 IPv6 Graph Degree Distribution.

Figure 4.14 illustrates the typical degree distribution of a network graph. On the other hand, Figure 4.15 displays an unexpected degree distribution. Instead of most nodes being of low degree followed by a sharp decline, the counts trade positions as the values increase. For example, one would expect there to be fewer nodes of degree three than two, but the opposite is true. Following that, one would expect there to be fewer nodes of degree four than three, which is true in this case. Yet the order switches again with more nodes of degree six than degree five. This strange behavior may result from the disjointedness in the IPv6 network, and is an interesting trend to observe.

4.1.4 Verizon Wireless (AS 6167 / AS 22394)

Two Verizon Wireless mobile ASes were identified, AS 6167 [46] and AS 22394 [47]. Both of these ASes are located in the United States. Unfortunately, the same issues exist with using DNS names to determine congruence; both ASes have 100 percent of their IPv4 nodes with resolved DNS names but zero percent of their IPv6 nodes with resolved DNS names.

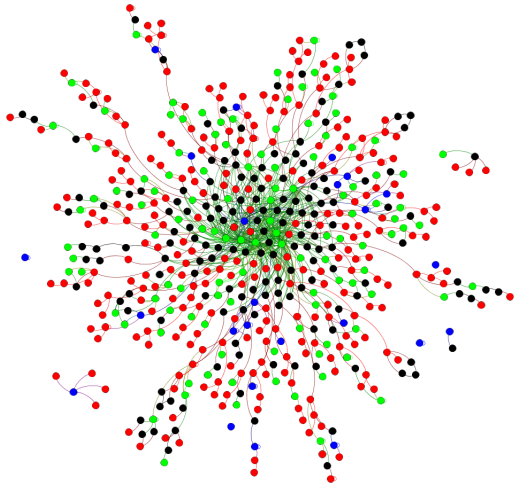

Figure 4.16. AS 6167 IPv4 Graph.

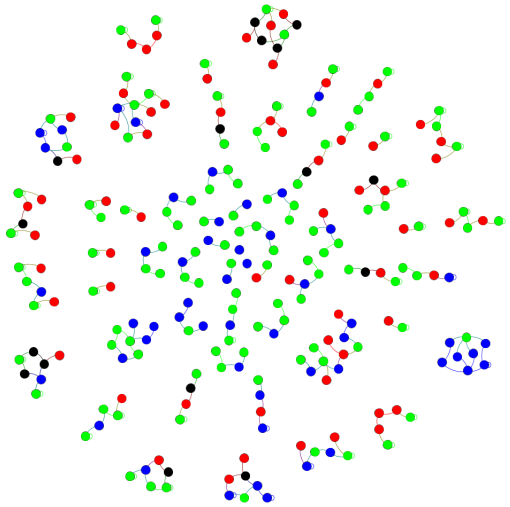

Figure 4.17. AS 6167 IPv6 Graph.

Figures 4.16 shows what one would expect from an IPv4 graph, but with more ingress nodes. In Figure 4.17, another disjoint network graph is shown for an IPv6 network. This suggests that this behavior previously seen in AS 5588 is not a fluke and there must be some underlying reason for it, whether it be in the traceroute behavior or in the actual topology of the network itself.

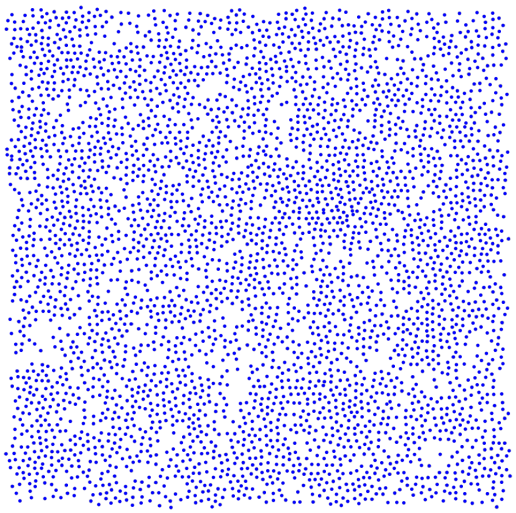

Figure 4.18. AS 22394 IPv4 Graph.

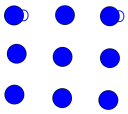

Figure 4.19. AS 22394 IPv6 Graph.

Figures 4.18 and 4.19 show standalone nodes with no edges. Upon closer examination of the traceroutes, the status of these traces was marked as either gap limit or complete. This means that these nodes were either unreachable or were the destination of the traceroute. It seems anomalous for all of these nodes within the same AS to come up as not having any connections, but it could likely be due to filtering of ICMP traffic within Verizon's edge. Future work could see if this behavior is present on traceroutes from different time periods.

Table 4.5. AS 6167 Statistics.

Metric	IPv4	IPv6
Total Nodes	565	245
Total Edges	950	250
Ingress (Green) Nodes	115	114
Egress (Red) Nodes	265	62
Ingress/Egress (Blue) Nodes	24	53
Intermediate (Black) Nodes	161	16
Average Degree	3.36	2.04
Average Path Length	5.45	2.15
Network Diameter	14	8

The statistics for AS 22394 are omitted as the only notable data is the 3987 and nine ingress/egress nodes in IPv4 and IPv6, respectively. Despite AS 6167 having more IPv4 ingress nodes than the previous mobile ASes examined, the IPv6 infrastructure still contains more ingress nodes by percentage of total nodes overall. This appears to be a reoccurring trend within IPv6 topologies.

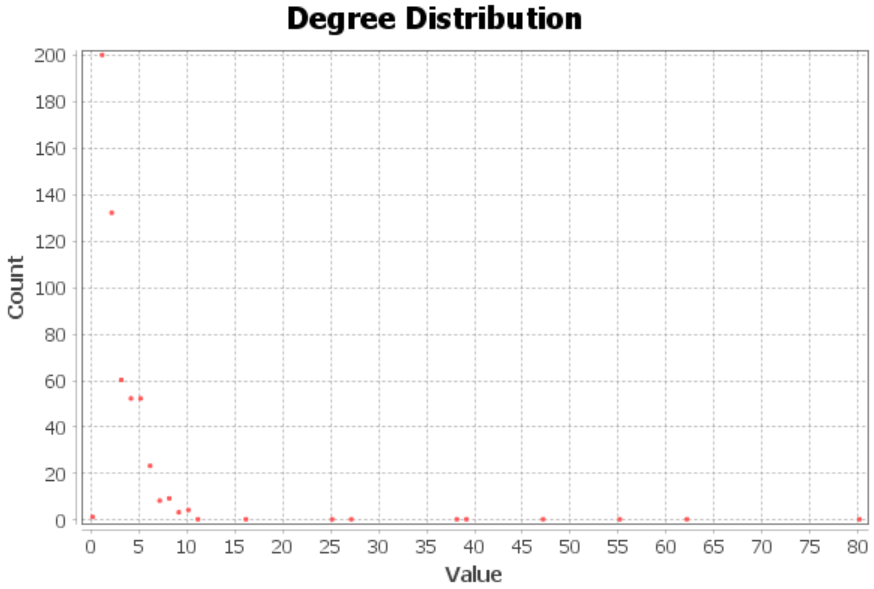

Figure 4.20. AS 6167 IPv4 Graph Degree Distribution.

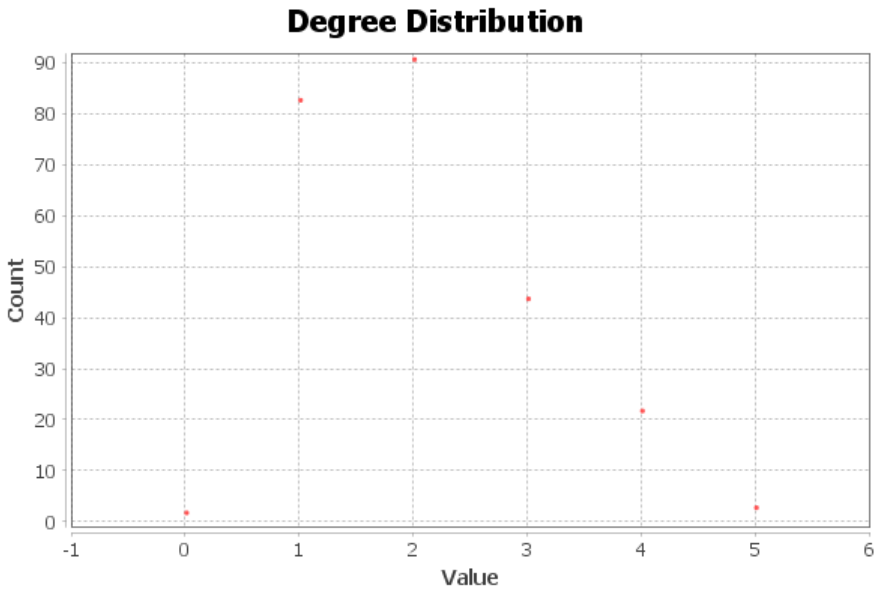

Figure 4.21. AS 6167 IPv6 Graph Degree Distribution.

Both of the degree distributions in Figures 4.20 and 4.21 follow the expected network distribution. This is actually noteworthy because the IPv6 graph of AS 6167 is even more disjoint than AS 5588 (53 components in AS 6167 versus 45 components in AS 5588, both excluding lone nodes), yet does not have the same abnormal behavior where there are more high-degree nodes in the degree distribution. This suggests that there might not be a strong correlation between network connected components and degree distribution. For the sake of completeness, Figure 4.22 shows the IPv6 size distribution of AS 6167.

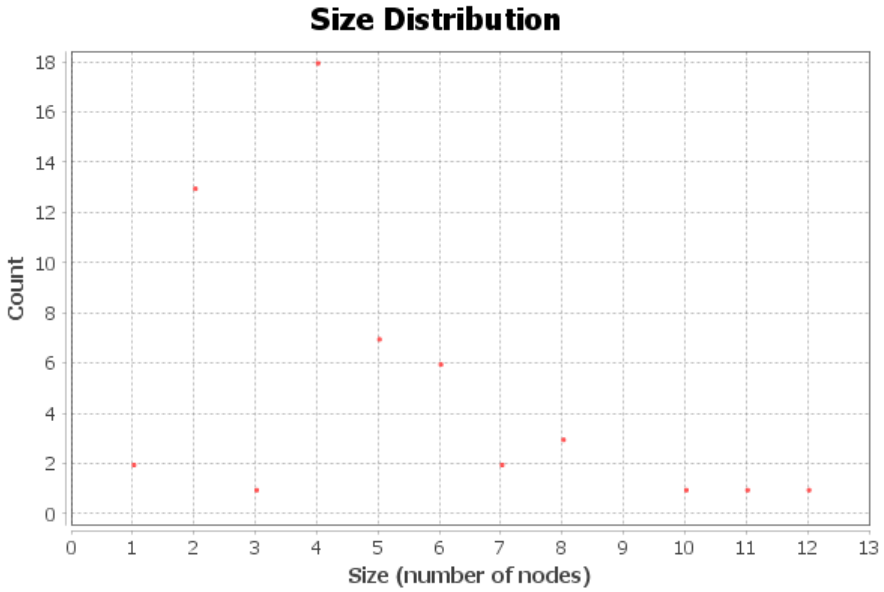

Figure 4.22. AS 6167 IPv6 Graph Size Distribution.

4.1.5 China Mobile (AS 9808 / AS 56040)

Like the Verizon Wireless analysis, there are also two ASes to analyze for China Mobile. These are AS 9808 [48] and AS 56040 [49]. According to Akamai, China's IPv6 adoption rate is relatively low at 15.3 percent [44]. This point is being mentioned now as the IPv4 infrastructure is too large to display with a reasonable amount of detail, but the IPv6 topology is of a similar size to ones that have been previously seen.

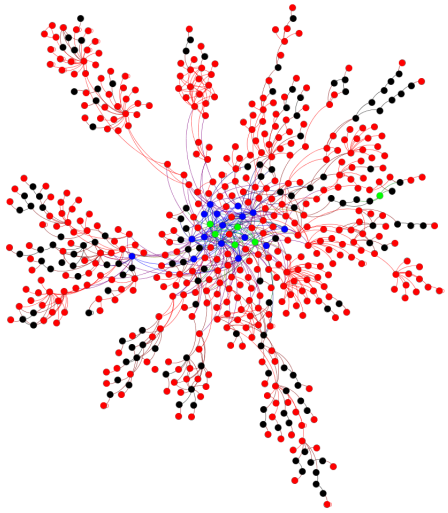

Figure 4.23. AS 9808 IPv6 Graph.

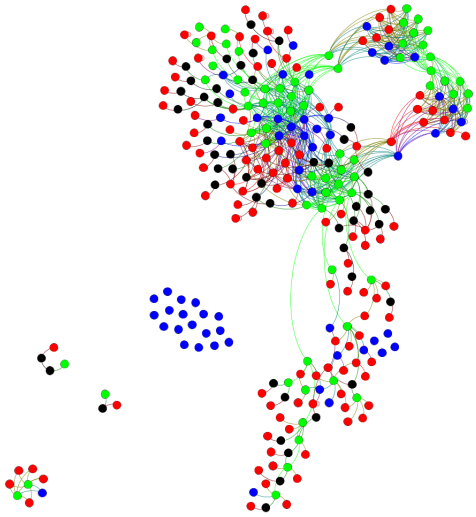

Figure 4.24. AS 56040 IPv6 Graph.

Both Figures 4.23 and 4.24 look like previously seen networks. Although the AS 56040 IPv6 graph has a few different connected components, the network as a whole is mostly connected.

Table 4.6. AS 9808 and AS 56040 Statistics.

Metric	**IPv4 (9808)**	**IPv6 (9808)**	**IPv4 (56040)**	**IPv6 (56040)**
Total Nodes	17867	575	3818	293
Total Edges	28810	927	7137	812
Ingress (Green) Nodes	75	6	156	69
Egress (Red) Nodes	16430	415	2004	114
Ingress/Egress (Blue) Nodes	208	16	1416	60
Intermediate (Black) Nodes	1154	138	242	50
Average Degree	3.23	3.22	3.74	5.54
Average Path Length	5.41	5.81	4.58	4.85
Network Diameter	14	16	13	12

The relationships between the IPv4 and IPv6 topologies of both ASes is something that has not been observed previously. For one, the average path length and network diameter of AS 9808 are greater in IPv6 than in IPv4. Typically, that relationship has been reversed with IPv4 having higher values in those statistics. Interestingly enough, a similar relationship holds true for the other China Mobile AS. In AS 56040, the IPv6 average degree and average path length are both larger than their IPv4 counterparts. This suggests that these relationships are not static and there is not a one-size-fits-all solution in predicting the statistics of a mobile topology.

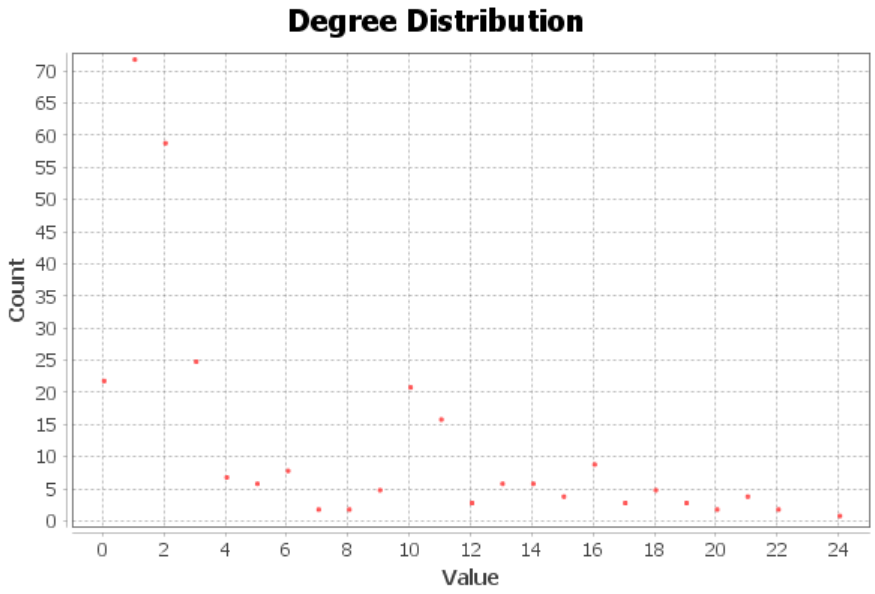

Figure 4.25. AS 56040 IPv6 Graph Degree Distribution.

Of the four degree distributions for these two ASes, only the IPv6 graph degree distribution of AS 56040 stood out; the other three followed the standard degree distribution. Figure 4.25 looks to be the most erratic seen so far. These types of degree distributions have been observed to be more prominent for IPv6. Although that is not enough to make a definitive statement correlating this degree distribution and IPv6, it is the start of a pattern.

4.1.6 Reliance Jio Infocomm Limited (AS 55836)

The final case study examines AS 55836, registered to Reliance Jio Infocomm Limited in India [50]. Out of all the countries Akamai surveyed, India has the highest rate of IPv6 adoption at 59.0 percent [44]. This leads to an interesting reversal in a common trend that has been observed thus far.

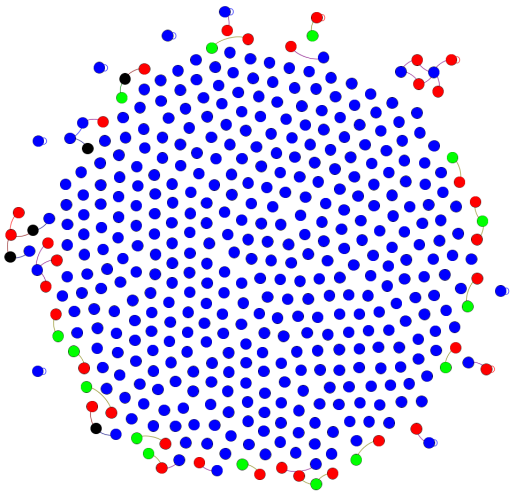

Figure 4.26. AS 55836 IPv4 Graph.

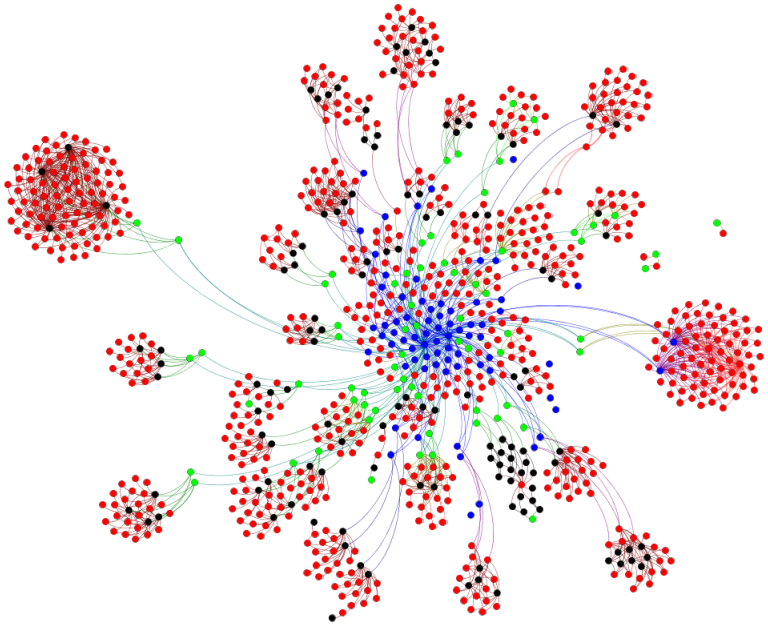

Figure 4.27. AS 55836 IPv6 Graph.

Unlike all the IPv4 and IPv6 topologies seen in the other case studies, this is the first one where the IPv6 network topology eclipses the IPv4 network topology in terms of size. Unfortunately most of the IPv4 nodes were unresponsive in the traceroute data which makes the statistical comparison not as fully realized.

Table 4.7. AS 55836 Statistics.

Metric	**IPv4**	**IPv6**
Total Nodes	438	962
Total Edges	54	1998
Ingress (Green) Nodes	15	69
Egress (Red) Nodes	34	693
Ingress/Egress (Blue) Nodes	384	73
Intermediate (Black) Nodes	5	127
Average Degree	0.25	4.15
Average Path Length	1.53	4.99
Network Diameter	4	12

Table 4.7 shows IPv6 average degree, average path length, and network diameter to be comparable to mobile IPv4 topologies where IPv4 is the dominant protocol. This suggests that a trend exists where IPv6 mobile topologies have a greater average degree, path length, and network diameter in countries where IPv6 has a higher adoption rate. As other case studies demonstrate, this trend may be reversed in countries where IPv6 adoption is minimal.

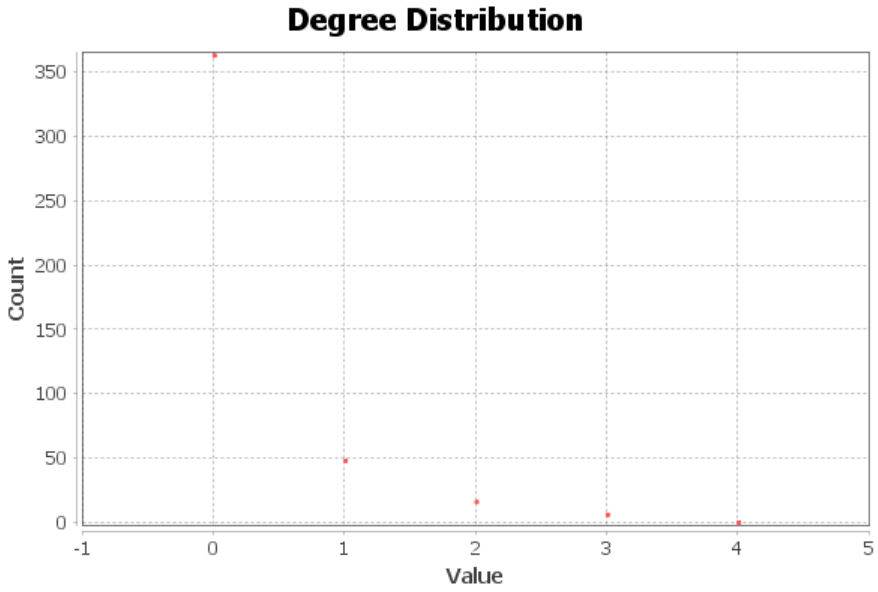

Figure 4.28. AS 55836 IPv4 Graph Degree Distribution.

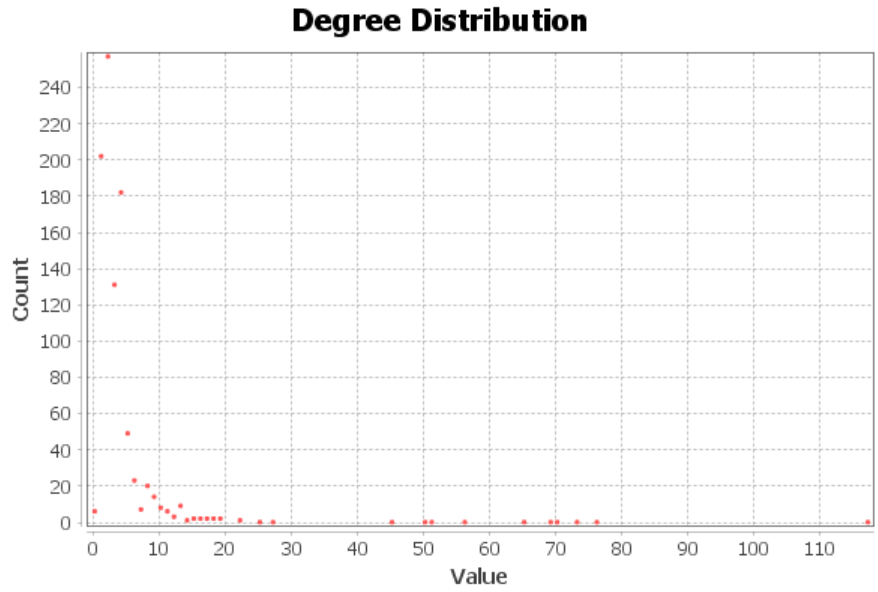

Figure 4.29. AS 55836 IPv6 Graph Degree Distribution.

The final interesting observation to make relates to the degree distribution. The IPv4 degree distribution matches the expected distribution. Although the IPv6 degree distribution is

close to what is expected, there is some abnormal behavior. This, along with previous case studies, suggests that IPv6 degree distributions tend to be more varied and do not necessarily follow expected trends.

4.2 Other Mobile ASes

Of course, there are more mobile ASes than the ones studied in this chapter. This may lead one to ask why other well-known providers were not examined. For example, studying AT&T, Sprint, or T-Mobile cellular ASes located in the United States would have provided a great point of comparison with Verizon. Each of these providers reported 57-93 percent of their customers using IPv6 in 2018 within the United States [6], and their mobile ASes were successfully identified (AT&T: AS 20057, Sprint: AS 10507, T-Mobile: AS 21928). However, analysis of the traceroute data showed no or very few nodes to be within these ASes. This is likely an issue with the way the data is collected. The probes are sent from vantage points located outside these mobile networks, and thus none of the traceroutes get routed through them. If traceroute data existed that originated from within these mobile networks, their mobile topologies could be mapped in the same manner as the previous ones. Additionally, the lack of probes sent from within mobile networks may also cause the graphs of the case-study ASes to not necessarily reflect their true topology.

4.3 User Agent Correlation

The last interesting piece of work performed is user agent (UA) string correlation which looks at HTTP web browser UAs. Although the MaxMind data proved helpful in determining if an IP address was mobile, an alternative method to identify and verify these mobile IP addresses would increase the integrity of the data. After all, MaxMind claims to have 95 percent accuracy [23] but how that data is actually derived is unknown. However, the proposed method is not without its own problems; mobile UAs could be connecting from a wired network, and non-mobile UAs could be connecting from a cellular hotspot.

The idea carried out was to use CMAND's access log [26] with data from September 5, 2015 to July 8, 2020 and implement the following algorithm:

1. Find the set of unique client IPv6 addresses contained in the access log. This step ensures that multiple HTTP accesses from a single client do not distort the

inference.

2. Maintain a data structure that counts the number of mobile and non-mobile hits per distinct IPv6 /48 prefix.
3. For each unique IPv6 client, use its UA string to determine if it used a mobile or non-mobile device.
4. For each unique IPv6 client, increment the mobile or non-mobile count in the data structure for its corresponding /48 and UA.
5. Compute the percentage of mobile hits as a fraction of total hits for each distinct /48.

In this experiment, a device is considered mobile if the client's UA string contains "Android" or "Phone", "Phone" of which covers both iPhones and Windows Phones. Note that there are bias in the set of networks accessing the CMAND web server, and many networks are under-represented, or not present at all. As such, we ignore inferences for /48 prefixes that contain fewer than 11 unique IPv6 clients. There were 3551 total network prefixes in the data. Of the 112 prefixes that had >10 hits, three prefixes were considered cellular and 109 were considered non-cellular. Table 4.8 displays these three cellular prefixes followed by three sample non-cellular prefixes and corresponding data.

Using our data and algorithm, we compared the data to see if prefixes with a high percentage of mobile hits are within a known MaxMind cellular prefix, and prefixes with a low percentage of mobile hits are not contained within a MaxMind prefix. The results proved promising when using /48s. The mobile hits for what MaxMind considers cellular networks ranged from 63-73%. The non-cellular networks had a mobile hit rate between 0-37%, which suggests that UA string correlation has merit. We again reiterate that sample size could be an issue here, and future work should apply the algorithm to larger and more complete access logs.

Table 4.8. UA Prefix Information.

Prefix	ASN	Name	Mobile Hits	Non-Mobile Hits	Percent Mobile Hits
2600:387:6::/48	20057	AT&T Mobility LLC	30	15	66%
2600:387:3::/48	20057	AT&T Mobility LLC	11	4	73%
2600:387:2::/48	20057	AT&T Mobility LLC	7	4	63%
2a03:2880:30ff::/48	32934	Facebook, Inc.	4	13	23%
2600:6c44:5f7f::/48	15169	Google LLC	1	38	2%
2a01:4f8:211::/48	24940	Hetzner Online GmbH	0	12	0%

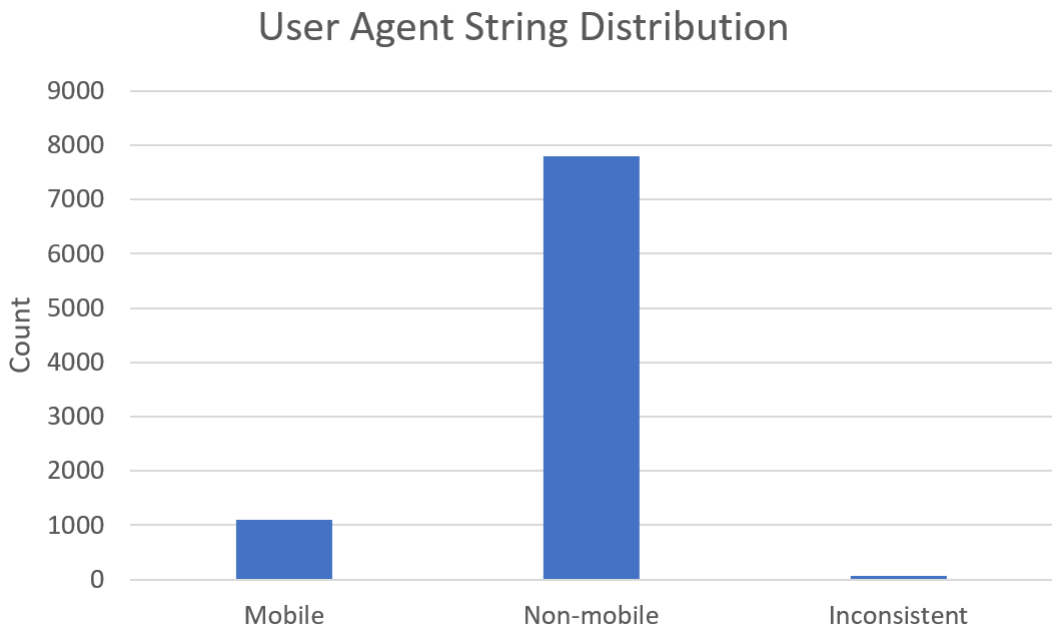

Figure 4.30. UA Distribution within CMAND Dataset.

There is also the issue of UA consistency, which is when a single IP address maps to multiple types of UA strings. In this dataset, there were 8901 unique IPv6 addresses. 1097 of these IPv6 addresses had mobile UA strings, and 7804 had non-mobile UA strings. 476 of these IP addresses mapped to more than one UA string. This was mostly a non-issue as the overall device would stay the same (e.g., an IP address might map to UA strings with different

browsers but be on the same type of mobile device). In this experiment, this behavior was not considered to be inconsistent. However, there were 67 IPv6 addresses that were shared between mobile and non-mobile UA strings. As an example, an IP address might first be seen on an iPhone, then seen on a Macintosh. These addresses are considered to be inconsistent UAs. Despite this, these addresses were included in the analysis as they only make up 0.7 percent of the dataset and they affect the mobile and non-mobile counters equally. These findings are summarized in Figure 4.30. As an alternative way of finding mobile IP addresses/verifying MaxMind data, this method seems to be worth using. However, having more IPv6 web server log entries to pull data from would yield more accurate results.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:
Conclusions and Future Work

This chapter concludes the thesis by delving into the questions answered and insights gained via the work in Chapter 4. Further research and work to be done in this area are also suggested.

5.1 Conclusions

While the case studies provided interesting insights, they have some limitations. For one, the sample represents eight mobile ASes of significant size, but there exist more mobile ASes worth examining. Although the analysis revealed some trends within mobile ASes, pulling in more data could potentially highlight other aspects of these topologies. Additionally, responses to traceroute probes from numerous, worldwide vantage points through network infrastructure of many mobile ASes simply did not appear in the data. This likely occurred for two different reasons: the routers within these ASes did not respond to traceroute probes, and traceroute probes were sent from outside these mobile networks rather than from within. These issues greatly restricted the choice of mobile topologies to map.

Despite these limitations, the questions that the case studies set out to answer have indeed been determined to some degree. As an overall trend, IPv4 mobile topologies typically have a greater number of nodes and edges. This proved not to be true in the single case of Reliance Jio, but the IPv6 adoption in India has surpassed IPv4 usage which may explain the inconsistency. Additionally, IPv6 infrastructure also tends to have a higher percentage of ingress nodes compared to IPv4. This correlates to a more robust and resilient IPv6 infrastructure compared to IPv4 as more nodes need to be taken offline to impact service. In two of the eight cases, the IPv6 mobile topology exhibits disjoint network characteristics and contains more connected components than the IPv4 topology.

When examining other metrics such as average degree, average path length, and network diameter, the results show no correlation between these values and whether a given mobile topology is IPv4 or IPv6. In most cases these metrics are greater in IPv4 compared to IPv6, yet China Mobile and Reliance Jio reveals that these metrics were higher in IPv6 than

IPv4. This could have potentially been attributed to a greater IPv6 adoption rate based on Reliance Jio, but China's IPv6 adoption rate is below average. As another observation, the prefix sizes as seen in global routing tables found in mobile infrastructure typically ranged from /16 to /24 in IPv4, and from /32 to /48 in IPv6. Relative to their respective address spaces, the IPv4 prefixes encompass a larger fraction of the whole despite the IPv6 prefixes having a larger absolute value.

Next is the comparison of degree distribution of nodes within mobile ASes. Ignoring nodes with a degree of one due to traceroute artifacts, the IPv6 degree distribution is more erratic compared to IPv4. Normally, there are consistently fewer nodes as degree increases. This is true in mobile IPv4 with only some exceptions. However, more often than not this behavior proved inconsistent in IPv6. For example, nodes of degree x would have a given count, followed by nodes of degree $x + 1$ with a lower count, followed by nodes of degree $x + 2$ with a higher count than the previous. This erratic up-and-down behavior in degree distribution may be an indicator of a mobile IPv6 AS. This type of degree distribution may also increase network robustness and resilience; instead of having a few very high degree nodes that can be brought down, the availability of the network is spread across a greater number of nodes.

The last question this thesis sought to answer was whether topological congruence existed between mobile IPv4 and IPv6 infrastructures. The most reliable method of determining this is using domain names attributed to each node. Congruence was discovered in the Vodafone Group network with the highest degree IPv4 and IPv6 nodes resolving to the same DNS name, and fifty percent of IPv6 nodes sharing a DNS name with IPv4 nodes. Unfortunately, DNS names were never resolved with that much completeness for the remaining seven ASes and congruence was unable to be determined. Even so, the case of Vodafone suggests that the practice of sharing mobile IPv4 and IPv6 infrastructure exists and potentially occurs in other mobile ASes. This implies that attacks on IPv4 infrastructure may affect IPv6 infrastructure as well, and vice versa.

5.2 Future Work

As mentioned previously, this thesis aimed to provide an exploration of the IPv6 mobile topology rather than completely solve it. As such, there is much work left to be done in

expanding the research. The first and greatest priority is to get measurement data using probes sent from within these mobile networks. Currently, the probes sent from outside the mobile network do not get routed through much mobile infrastructure; having a larger, more relevant dataset would do much to enhance this work. On the subject of data collection, having an IPv6 web server that serves many clients and logs their UAs would enhance the validity of UA string correlation in determining mobile IP addresses.

While using UA string correlation to determine mobile IP addresses proved promising, there is more work to be done in this area. In our experiment, the mobile IP addresses belonged to cellular devices, while the IP addresses in the traceroute data belonged to routers. Finding some methodology to differentiate between mobile infrastructure IP addresses and the end-user devices is worth pursuing.

Furthermore, there are certainly underlying assumptions in the methodology that could be tweaked. For example, the loosest definition of a mobile AS was used, where only one mobile prefix was necessary to label the AS as mobile. Although non-mobile nodes within these ASes were excluded and not mapped, a more stringent definition could see different trends. Of course, the aforementioned dataset with measurements from within mobile networks would help provide for a strict definition. We created these topology graphs with a best-effort approach given the limitations faced, and obtaining validation from the cellular providers could highlight strengths and weaknesses in our methodology.

The topology graphs focus on a router-level view of mobile infrastructure, but there are other types of connectivity worth investigating. For example, how does a cellular provider's infrastructure connect to other networks, ASes, or Internet exchange points (IXPs), and how does this differ between IPv4 and IPv6? We also know that middleboxes exist in mobile IPv4 and IPv6 infrastructure [7], but the extent of their prevalence in these topologies remains undetermined. The issue of forwarding paths also remains. We observed congruence in some ASes, which warrants a deeper analysis of the traceroute data to see how IPv4 and IPv6 forwarding paths compare.

There is also the concept of address agility specific to IPv6, which is how often nodes change IP addresses according to the privacy extensions standard [51]. Understanding both how these mobile IP addresses are changing and how often they do so has implications from a security perspective. These routers could be using EUI-64-based IPv6 addresses, which

reveal the make and model of a device and allows malicious actors to carry out targeted attacks [52].

Lastly, traceroute data for this study was sourced only from a single point in time. This point in time could perhaps have been anomalous, and examining other time periods could reveal different trends even among the same mobile ASes studied. This thesis also took a global view of the topology where vantage points from all over the world were used as sources of traceroute data. Further research could carefully choose vantage points to create specific views of the mobile topology, which has use cases such as comparing how mobile topologies appear in different countries. Ultimately, there are numerous future directions that this early research could take.

APPENDIX A:
Mobile Topology Graphs Ranked by Degree

The network topologies shown in Chapter 4 highlighting ingress, egress, ingress/egress, and intermediate nodes paint one idea of the topology, but other ways of representing the topology have equal merit. This appendix is dedicated to showing the same graphs in Chapter 4 using a degree ranking instead.

Figure A.1. AS 4725 IPv4 Graph Ranked by Degree.

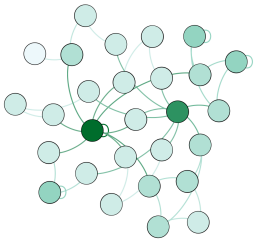

Figure A.2. AS 4725 IPv4 Graph Ranked by Degree.

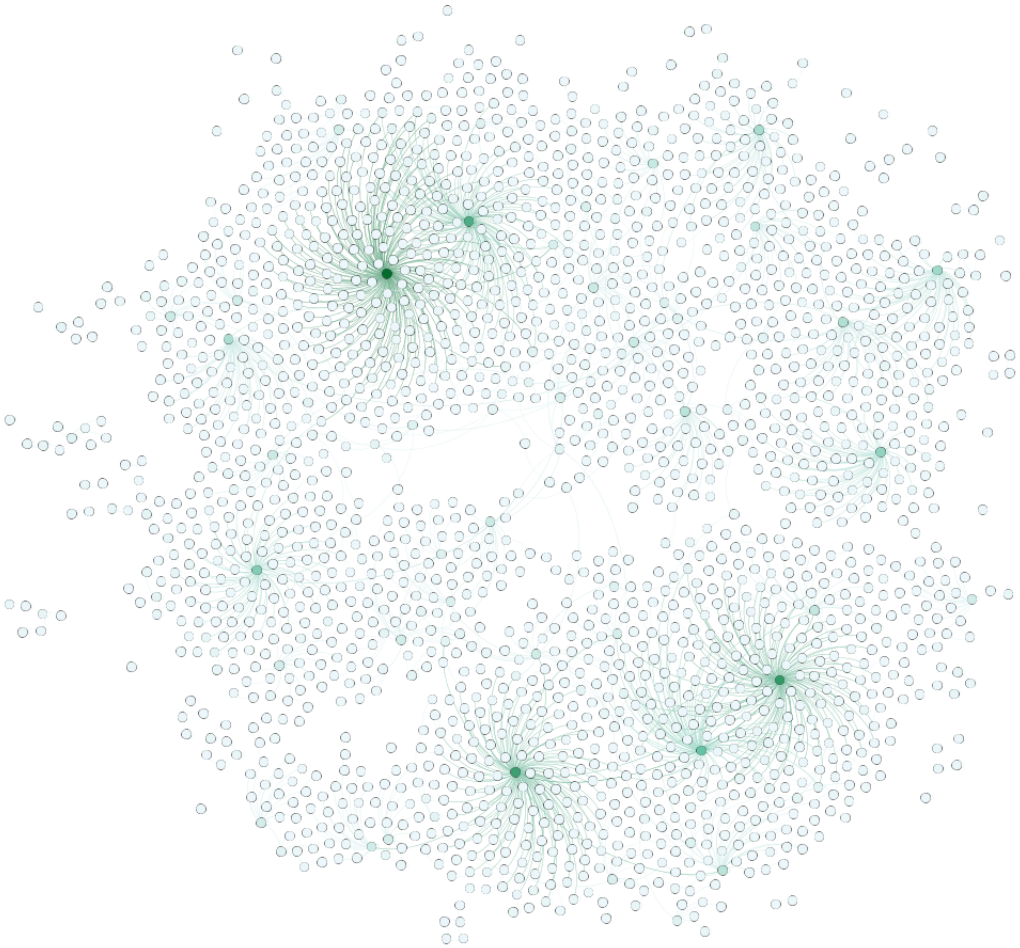

Figure A.3. AS 5588 IPv4 Graph Ranked by Degree.

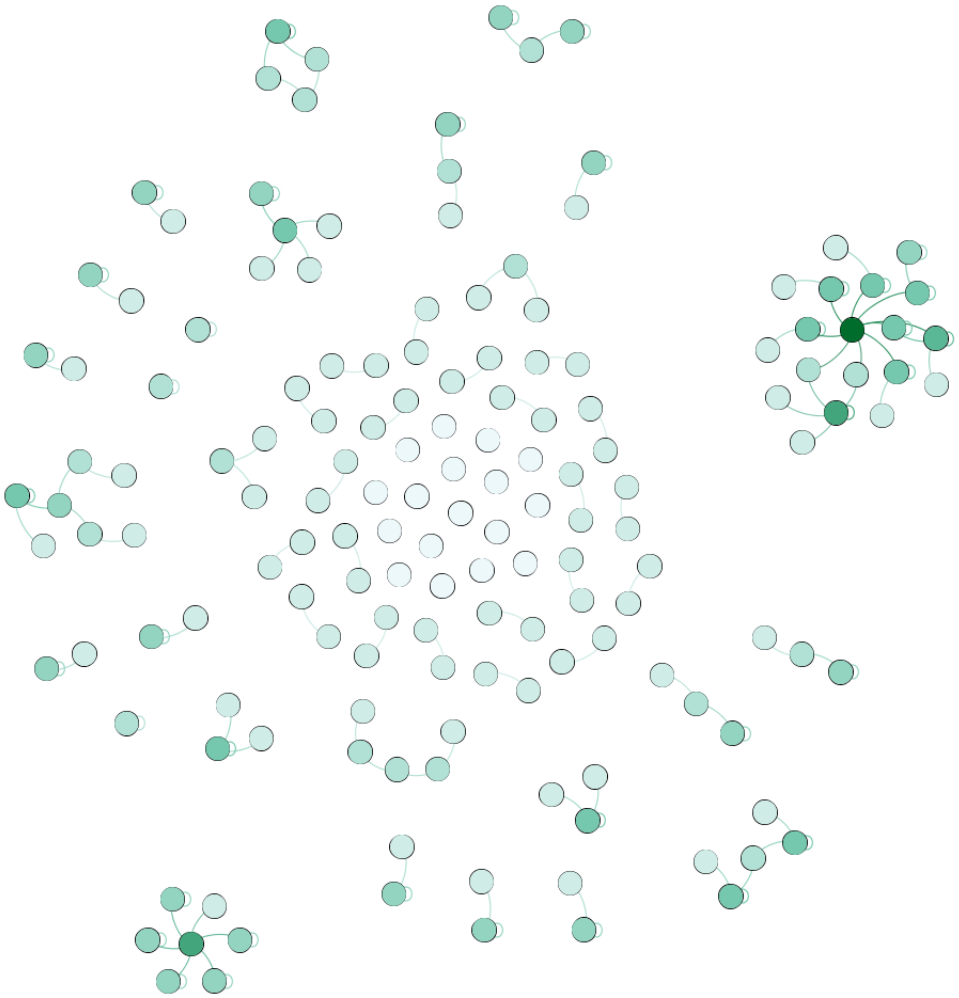

Figure A.4. AS 5588 IPv6 Graph Ranked by Degree.

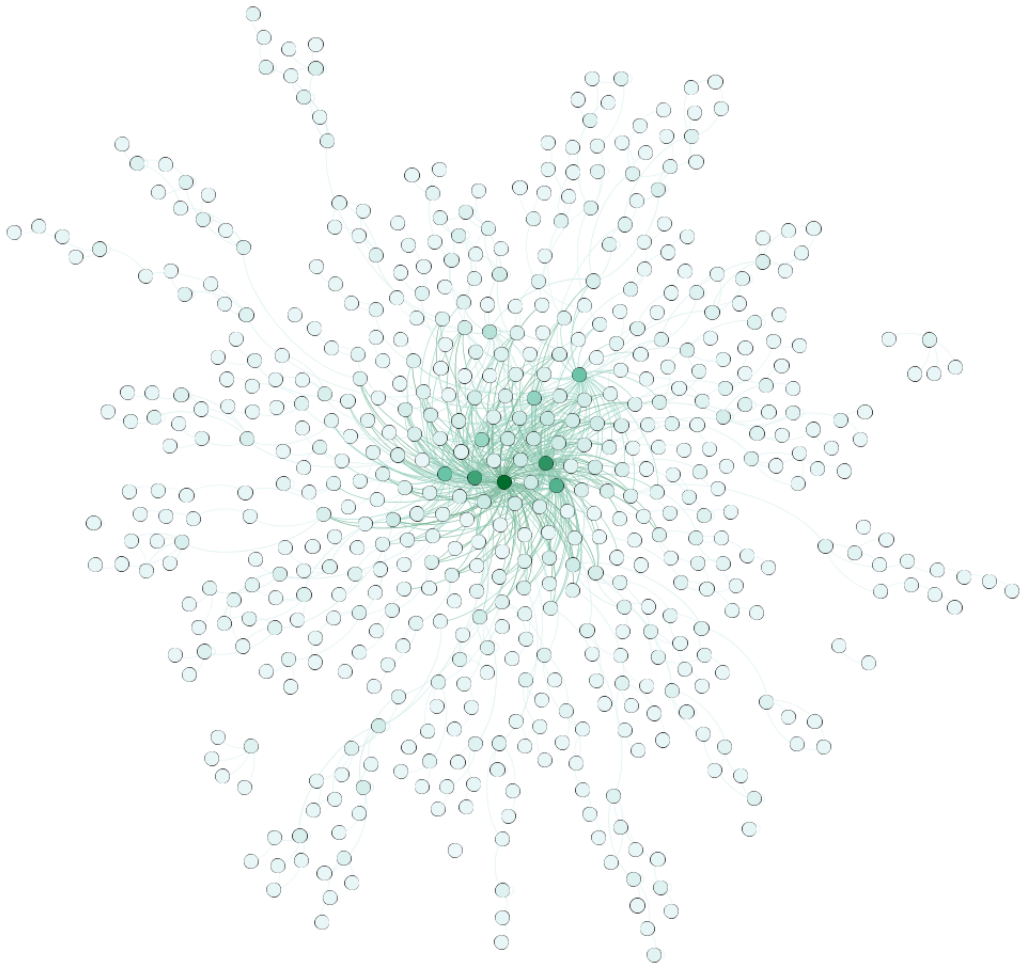

Figure A.5. AS 6167 IPv4 Graph Ranked by Degree.

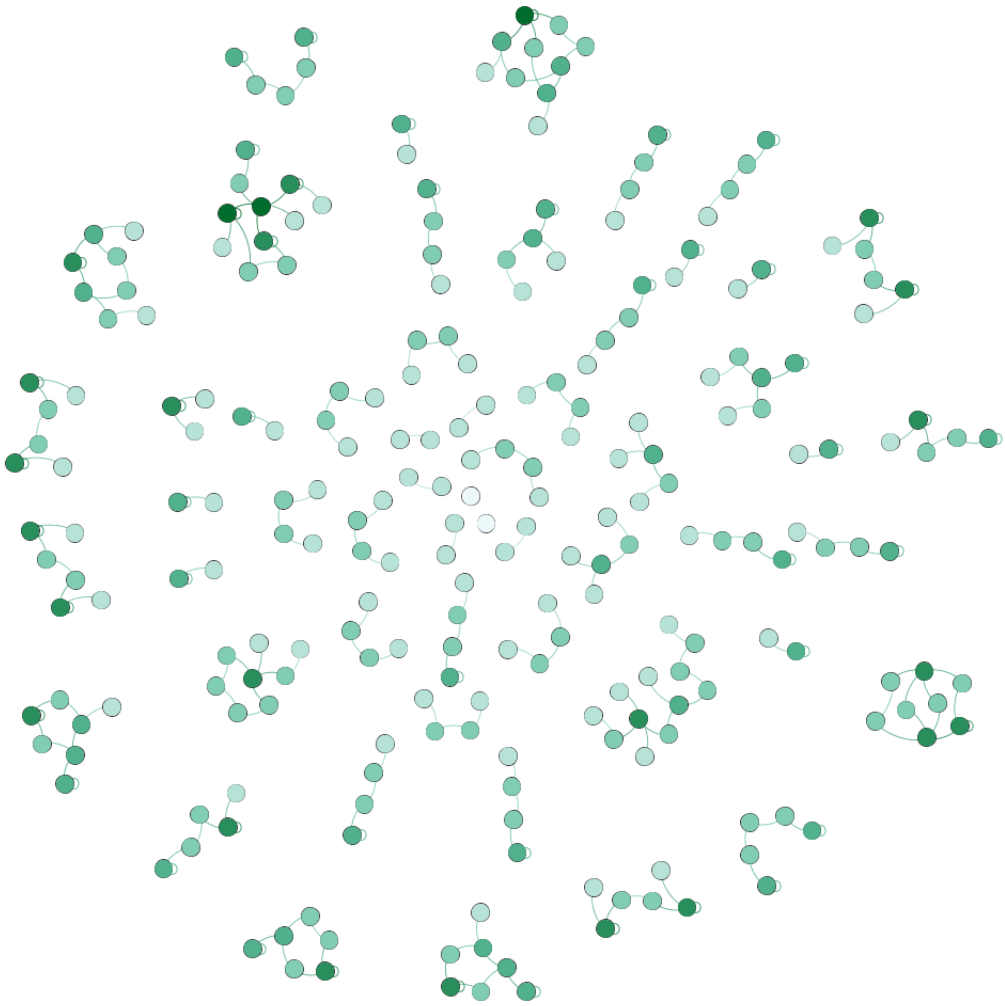

Figure A.6. AS 6167 IPv6 Graph Ranked by Degree.

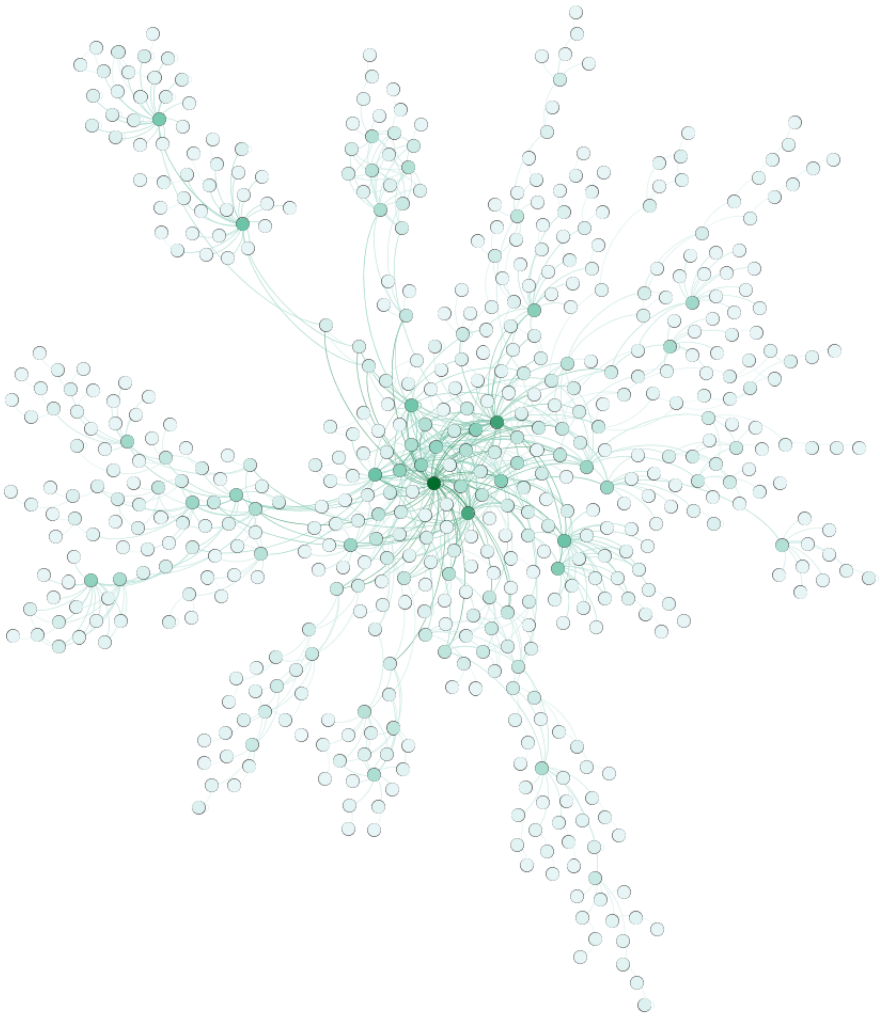

Figure A.7. AS 9808 IPv6 Graph Ranked by Degree.

Figure A.8. AS 56040 IPv6 Graph Ranked by Degree.

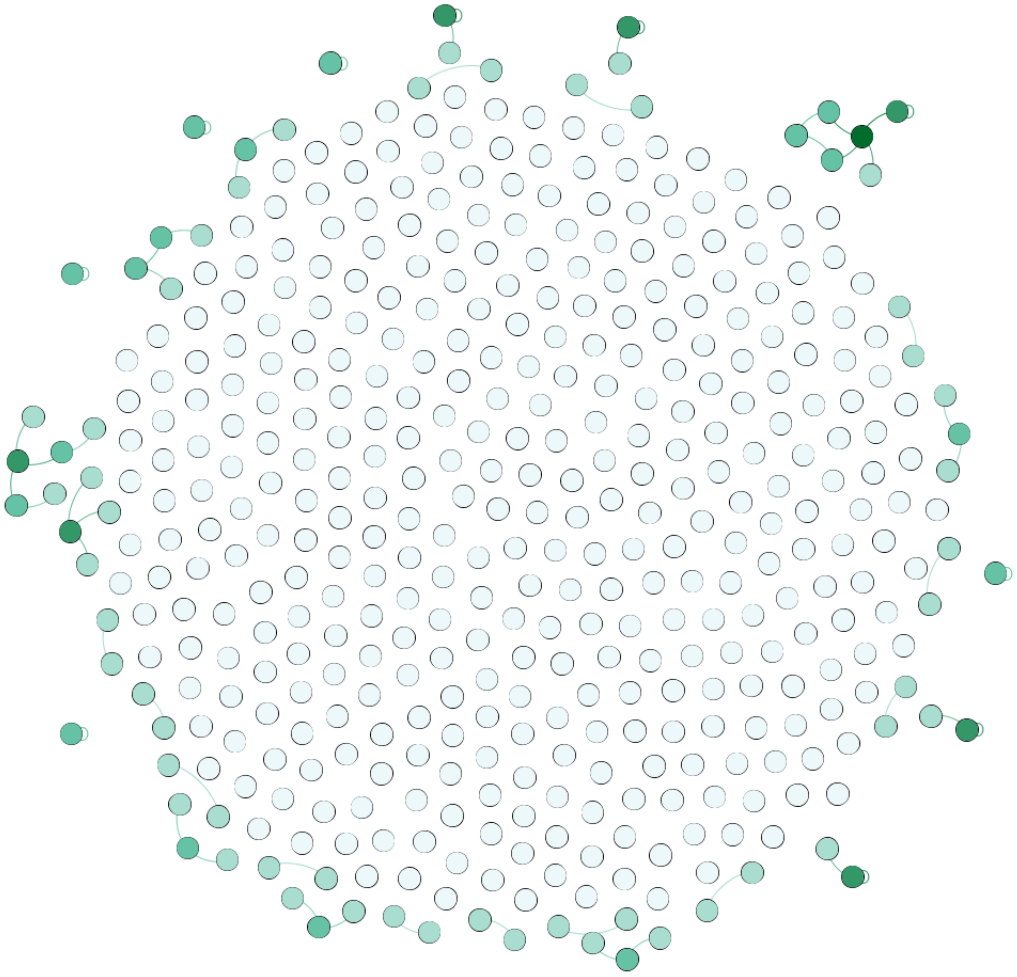

Figure A.9. AS 55836 IPv4 Graph Ranked by Degree.

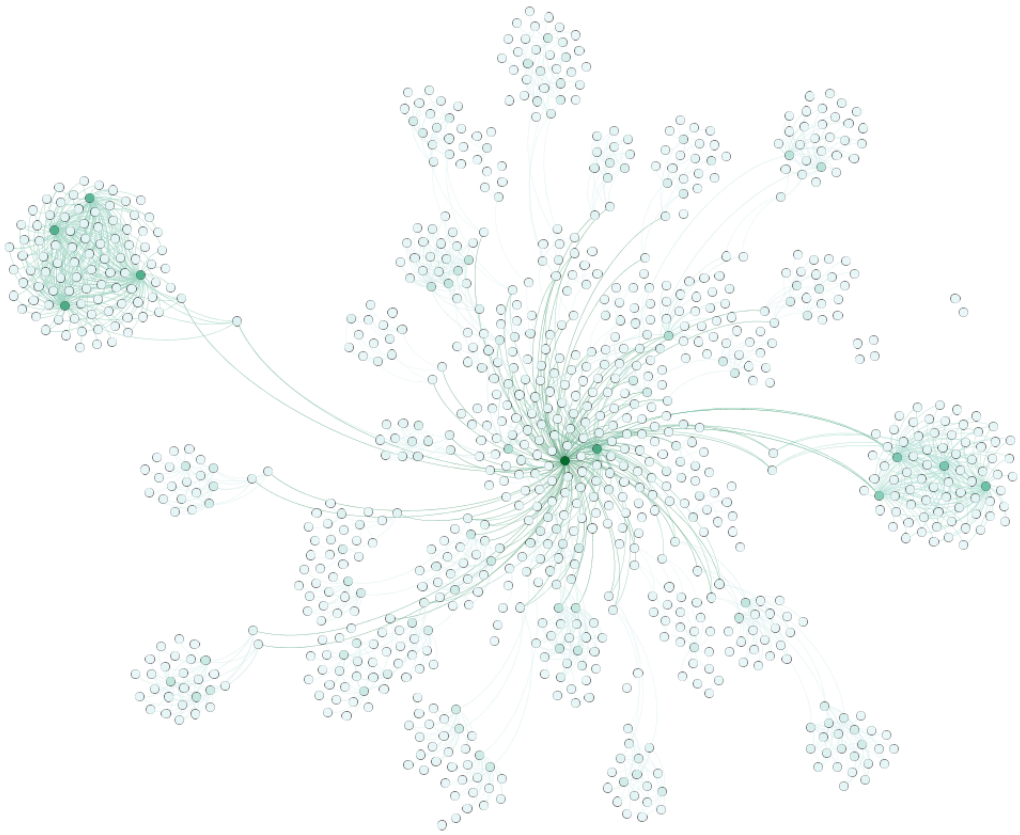

Figure A.10. AS 55836 IPv6 Graph Ranked by Degree.

THIS PAGE INTENTIONALLY LEFT BLANK

List of References

[1] S. Deering and R. Hinden, "Internet Protocol, version 6 (IPv6) specification," RFC, July 2017. [Online]. Available: https://tools.ietf.org/html/rfc8200

[2] S. Deering and R. Hinden, "Internet Protocol, version 6 (IPv6) specification," RFC, December 1998. [Online]. Available: https://tools.ietf.org/html/rfc2460

[3] ICANN, "Available pool of unallocated IPv4 internet addresses now completely emptied," February 3, 2011. [Online]. Available: https://www.icann.org/en/system/files/press-materials/release-03feb11-en.pdf

[4] Google, "IPv6 adoption," July 2020. [Online]. Available: https://www.google.com/intl/en/ipv6/statistics.html

[5] C. Byrne, "464XLAT: breaking free of IPv4," June 2014. [Online]. Available: https://archive.nanog.org/sites/default/files/wednesday_general_byrne_breakingfree_11.pdf

[6] ISOC, "State of IPv6 deployment 2018," June 6, 2018. [Online]. Available: https://www.internetsociety.org/wp-content/uploads/2018/06/2018-ISOC-Report-IPv6-Deployment.pdf

[7] F. Li, A. M. Kakhki, D. Choffnes, P. Gill, and A. Mislove, "Classifiers unclassified: An efficient approach to revealing IP traffic classification rules," in *Proceedings of the 2016 Internet Measurement Conference* (IMC '16). New York, NY, USA: Association for Computing Machinery, 2016, p. 239–245. [Online]. Available: https://doi.org/10.1145/2987443.2987464

[8] 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

[9] J. P. Rula, F. E. Bustamante, and M. Steiner, "Cell spotting: Studying the role of cellular networks in the Internet," in *Proceedings of the 2017 Internet Measurement Conference* (IMC '17). New York, NY, USA: Association for Computing Machinery, 2017, p. 191–204. [Online]. Available: https://doi.org/10.1145/3131365.3131402

[10] D. B. Johnson, J. Arkko, and C. E. Perkins, "Mobility support in IPv6," RFC, July 2011. [Online]. Available: https://tools.ietf.org/html/rfc6275

[11] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, and R. Teixeira, "Avoiding traceroute anomalies with Paris traceroute," in *Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement* (IMC '06). New York, NY, USA: Association for Computing Machinery, 2006, p. 153–158. [Online]. Available: https://doi.org/10.1145/1177080.1177100

[12] J. Hawkinson and T. Bates, "Guidelines for creation, selection, and registration of an Autonomous System (AS)," RFC, March 1996. [Online]. Available: https://tools.ietf.org/html/rfc1930

[13] K. Keys, Y. Hyun, M. Luckie, and K. Claffy, "Internet-scale IPv4 alias resolution with MIDAR," *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 383–399, Apr 2013.

[14] M. Luckie, R. Beverly, W. Brinkmeyer, and K. Claffy, "Speedtrap: Internet-scale IPv6 alias resolution," in *ACM Internet Measurement Conference (IMC)*, Oct 2013, pp. 119–126.

[15] K. Keys, "IP alias resolution techniques: Technical report," Cooperative Association for Internet Data Analysis (CAIDA), Tech. Rep., Dec 2008. [Online]. Available: https://www.caida.org/publications/papers/2008/alias_resolution_techreport/alias_resolution_techreport.pdf

[16] G. Miao, J. Zander, K. W. Sung, and S. Ben Slimane, *Fundamentals of Mobile Data Networks*. Cambridge, England: Cambridge University Press, 2016.

[17] Mozilla, "User agent - MDN web docs glossary: Definitions of web-related terms | MDN," Accessed July 20, 2020. [Online]. Available: https://developer.mozilla.org/en-US/docs/Glossary/user_agent

[18] CAIDA, The CAIDA UCSD IPv4 Routed /24 Topology Dataset - January 27, 2019, 2020. [Online]. Available: http://www.caida.org/data/active/ipv4_routed_24_topology_dataset.xml

[19] CAIDA, The CAIDA UCSD IPv6 Topology Dataset - January 27, 2019, 2020. [Online]. Available: http://www.caida.org/data/active/ipv6_allpref_topology_dataset.xml

[20] CAIDA, The CAIDA UCSD IPv4 Routed /24 DNS Names Dataset - January 27, 2019, 2020. [Online]. Available: http://www.caida.org/data/active/ipv4_dnsnames_dataset.xml

[21] CAIDA, The CAIDA UCSD IPv6 DNS Names Dataset - January 27, 2019, 2020. [Online]. Available: http://www.caida.org/data/active/ipv6_dnsnames_dataset.xml

[22] CAIDA, The CAIDA UCSD Routeviews Prefix to AS Mappings Dataset (pfx2as) for IPv4 and IPv6 - January 27, 2019, 2020. [Online]. Available: https://www.caida.org/data/routing/routeviews-prefix2as.xml

[23] MaxMind, GeoIP2 Connection Type Database - May 5, 2020, 2020. [Online]. Available: https://www.maxmind.com/en/geoip2-connection-type-database

[24] RouteViews, MRT format RIBs and UPDATEs - January 27, 2019, 2020. [Online]. Available: http://archive.routeviews.org/bgpdata/

[25] RouteViews, v6 MRT format RIBs and UPDATEs - January 27, 2019, 2020. [Online]. Available: http://archive.routeviews.org/route-views6/bgpdata/

[26] CMAND, CMAND.org Access Log - September 5, 2015 to July 8, 2020, 2020. [Online]. Available: https://www.cmand.org/index.php

[27] B. Huffaker, M. Fomenkov, and K. Claffy, "Internet topology data comparison," Cooperative Association for Internet Data Analysis (CAIDA), Tech. Rep., May 2012. [Online]. Available: https://www.caida.org/publications/papers/2012/topocompare-tr/topocompare-tr.pdf

[28] J. Czyz, M. Allman, J. Zhang, S. Iekel-Johnson, E. Osterweil, and M. Bailey, "Measuring IPv6 adoption," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 4, p. 87–98, Aug 2014. [Online]. Available: https://doi.org/10.1145/2740070.2626295

[29] V. Giotsas, M. Luckie, B. Huffaker, and K. Claffy, "IPv6 AS relationships, cliques, and congruence," in *Passive and Active Network Measurement Workshop (PAM)*, Mar 2015, pp. 111–122.

[30] R. Almeida, O. Fonseca, E. Fazzion, D. Guedes, W. Meira, and Í. Cunha, "A characterization of load balancing on the IPv6 Internet," in *International Conference on Passive and Active Network Measurement*. Springer, 2017, pp. 242–254.

[31] E. Carisimo, C. Selmo, J. Alvarez-Hamelin, and A. Dhamdhere, "Studying the evolution of content providers in IPv4 and IPv6 Internet cores," *Elsevier Computer Communications Journal*, vol. 145, pp. 54–65, Sep 2019.

[32] S. Jia, M. Luckie, B. Huffaker, A. Elmokashfi, E. Aben, K. Claffy, and A. Dhamdhere, "Tracking the deployment of IPv6: Topology, routing and performance," *Computer Networks*, vol. 165, no. 106947, Dec 2019.

[33] CAIDA, "AS rank: AS1239 (Sprint)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/1239

[34] Sprint, "Sprint.net IPv4/IPv6 looking glass," Accessed July 20, 2020. [Online]. Available: https://www.sprint.net/lg/lg_start.php

[35] M. Luckie, "Scamper: A scalable and extensible packet prober for active measurement of the Internet," in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (IMC '10). New York, NY, USA: ACM, 2010, pp. 239–245. [Online]. Available: http://doi.acm.org/10.1145/1879141.1879171

[36] M. J. Schultz, "py-radix - PyPi," October 9, 2017. [Online]. Available: https://pypi.org/project/py-radix/

[37] T. Kiso, "GitHub - t2mune/mrtparse: MRT format data parser," May 4, 2020. [Online]. Available: https://github.com/t2mune/mrtparse

[38] Gephi, "GDF Format," Accessed July 20, 2020. [Online]. Available: https://gephi.org/users/supported-graph-formats/gdf-format/

[39] CAIDA, "Vela MIDAR API," Accessed July 20, 2020. [Online]. Available: https://www.caida.org/projects/ark/vela/midar-api/

[40] A. Marder, M. Luckie, A. Dhamdhere, B. Huffaker, J. Smith, and K. Claffy, "Pushing the boundaries with bdrmapIT: Mapping router ownership at Internet scale," in *ACM Internet Measurement Conference (IMC)*, Nov 2018, pp. 56–69.

[41] CAIDA, "AS rank: AS1273 (Vodafone Group PLC)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/1273

[42] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software," *PloS one*, vol. 9, no. 6, p. e98679, 2014. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679

[43] CAIDA, "AS rank: AS4725 (SOFTBANK MOBILE Corp.)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/4725

[44] Akamai, "State of the Internet IPv6 adoption visualization | Akamai," Accessed July 20, 2020. [Online]. Available: https://www.akamai.com/us/en/resources/our-thinking/state-of-the-internet-report/state-of-the-internet-ipv6-adoption-visualization.jsp

[45] CAIDA, "AS rank: AS5588 (T-Mobile Czech Republic a.s.)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/5588

[46] CAIDA, "AS rank: AS6167 (Cellco Partnership DBA Verizon Wireless)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/6167

[47] CAIDA, "AS rank: AS22394 (Cellco Partnership DBA Verizon Wireless)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/22394

[48] CAIDA, "AS rank: AS9808 (China Mobile)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/9808

[49] CAIDA, "AS rank: AS56040 (China Mobile Communications Corporation)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/56040

[50] CAIDA, "AS rank: AS5836 (Reliance Jio Infocomm Limited)," Accessed July 20, 2020. [Online]. Available: https://asrank.caida.org/asns/55836

[51] D. T. Narten, R. P. Draves, and S. Krishnan, "Privacy extensions for stateless address autoconfiguration in IPv6," RFC, Sep. 2007. [Online]. Available: https://tools.ietf.org/html/rfc4941

[52] E. C. Rye, J. Martin, and R. Beverly, "Eui-64 considered harmful," *arXiv preprint arXiv:1902.08968*, 2019. [Online]. Available: https://arxiv.org/abs/1902.08968

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California