Unexpected Applications on IPv6 Ports

Maximilian Ratmeyer

TABLE OF CONTENTS

Table of Contents	**1**
Abstract	**2**
CHAPTER 1. Introduction	**3**
CHAPTER 2. Literature Review	**5**
CHAPTER 3. Project Overview	**9**
CHAPTER 4. Work Plan	**10**
CHAPTER 5. Experimental Setup	**11**
CHAPTER 6. Results and Discussion	**13**
CHAPTER 7. Conclusion	**16**
References	**18**

ABSTRACT

As global IPv6 usage continues to grow, understanding the makeup of services and misconfigurations on the IPv6 address space is crucial to securing the next generation of networks. Prior work on IPv4 has shown that there are many "unexpected" applications running on non-standard ports, but whether this is true in IPv6 remains unexplored. In this thesis, we extend Stanford LZR fingerprinting tool to run IPv6, and integrate it with Georgia Tech's scanv6 scanner and various address generation techniques to perform internet-wide scans of 55 ports. Between March and April 2025, we discovered over 102 million responsive service endpoints and classified them as "Expected," "Unexpected," or "Unknown." We find that IPv6 contains higher rates of expected-port services compared to IPv4, yet many ports contain mis assignments of services. Our results highlight the effectiveness of current IPv6 address generating techniques and their applications in internet scanning. This work lays the groundwork for IPv6 vulnerability scanning and possible best practices for large-scale internet scanning in the IPv6 era.

CHAPTER 1. INTRODUCTION

Network scanning is an important technique used to understand and map modern Internet infrastructure. Specifically, port scanning is a tool that sends probes to computers and identifies what ports are open and responding to traffic. While traditional port scanning can help identify services, the scope of what scans can detect is limited. Misconfigurations and anomalies, which can indicate security or operational issues, are often missed by these scans. These often come in the form of services running on unexpected ports, or using non-standard protocol handshakes. [1]

As IPv6 networks become more widely adopted, a new area of network research has opened up that has yet to be thoroughly explored. Unlike IPv4 which is well studied [2] [3] [4], IPv6 has technical and architectural complexities that have made it historically more difficult to scan. [5] Namely, the vast space of IPv6 addresses makes it impossible to brute force and scan every address like in IPv4. Although extensive research has been done on how to scan IPv6 address space, there exists limited research into existing security vulnerabilities on IPv6. [6] [7]

To narrow the gap in security scanning done in IPv4 and IPv6, we plan on extending previous IPv4 security research into IPv6. Building upon the LZR tool developed by Izhikevich et al. which can efficiently identify unexpected protocol handshakes at scale [8], we plan to extend this capability into the IPv6 address

space. This involves refactoring LZR to be compatible with the IPv6 address space, and making use of other IPv6 services like IPv6 Hitlist [9] and target generation techniques [10] for host discovery.

As IPv6 usage continues to grow [11], critical services such as healthcare systems, financial services, and public infrastructure will be hosted on IPv6. By understanding the current security posture of IPv6, our research will enable researchers to understand what vulnerabilities exist and how current known IPv4 vulnerabilities and behaviors manifest themselves in IPv6. Extending the LZR tool and collecting more data on ports and services accessible in IPv6 allow us to maintain our understanding of IPv4 in the IPv6 ecosystem. This work will be critical to stay ahead of threats as more critical infrastructure is moved to IPv6 in the future.

2. LITERATURE REVIEW

IP addresses are an integral part of Internet infrastructure, and allow every computer on the internet to identify and talk to each other. For much of the Internet's history, networks used IPv4 addresses to identify other machines on the Internet. However, as the amount of devices connected to the Internet grew, IANA started running out of IPv4 addresses and needed a solution to allow more space. [12] Technologies like NAT helped mitigate these capacity issues in IPv4 [13], but a more permanent solution was needed. IPv6 was proposed in 1995 and is structured in a way that allows a significant increase in capacity of addresses available. In fact, these IPv6 address combinations have enough capacity to serve 340 undecillion devices, compared to the approximately 4 billion with IPv4. As of November 2024, 44.75% of users connect to Google via IPv6 [11], with the rate much higher in developing countries.

To understand what systems, services, and vulnerabilities are running on the Internet, researchers conduct measurement studies using tools like Zmap [14], or leveraging services such as Censys [15]. Zmap implements certain networking optimizations around single packet scanning and minimal state tracking that make it significantly more efficient to scan than previous work like Nmap. With powerful enough hardware, the range of IPv4 addresses is small enough that it's possible to look at every possible IPv4 address in under an hour, making scanning

the IPv4 world for data relatively easy. Censys is a service that maintains an active network of scanners and a database with historical results found from scans. This allows researchers to search for specific systems online, without performing new scans every time. Censys makes scanning much faster and reduces the amount of scanning traffic on the Internet.

While fast internet scanning is possible in IPv4, IPv6 it would take trillions of years to brute force every possible address, due to the increased address size. This makes it difficult to do scans on IPv6, because a scanner does not know which of the trillions of addresses actually hold active servers. To adapt to this, researchers have developed techniques to track and predict IPv6 addresses that can be scanned.

One common technique developed to perform scans are IPv6 hitlists [9]. These are passive datasets that can be gathered from sources like DNS entries, networking logs, or machine learning models, and create pre-built centralized lists of IPv6 addresses. However, the effectiveness of these lists is dependent on the quality of the data sources. Because IPv6 addresses cannot be brute forced, the dataset used will bias the IPs available in the hitlist. For example, a hitlist built on DNS entries may inherently contain primarily web or file servers. This would ignore a large swath of the Internet that also exists, like personal routers that could be a beneficial target to scan for vulnerabilities.

To explore more addresses, we need to go beyond finding addresses from data sources towards predicting never before seen addresses that are responsive. To accommodate for this, researchers started to build machine learning models that can predict new addresses in IPv6 space that may be valid. Initial attempts utilized patterns in the active addresses [16] and starting from known seed addresses [6]. Later proposed methods leveraged clustering regions of valid addresses and generating target addresses within those regions. [17] In addition, many more tree-based algorithms are designed to predict new candidate IPv6 addresses. [18] [19] [20] [21] 6Sense, a Georgia Tech project uses a combination of reinforcement learning and machine learning to generate IPv6 addresses, finding 3.6x more hosts than previous methods. [5]

Much research has been done exploring the landscape of security vulnerabilities and server misconfigurations on the IPv4 Internet. However, as IPv6 usage grows, understanding how these vulnerabilities and patterns manifest in IPv6 is important. IPv6 usage looks fundamentally different, with developing countries like India and Malaysia having around 70% of Google usage coming from IPv6 systems [11]. Initial investigations show that IPv6 systems have more open security policy compared to IPv4, with ports SSH, Telnet, and SNMP being almost twice as open. [7] Understanding the IPv6 world is critical to knowing what the

layout of the Internet looks like, what systems and vulnerabilities are out there, and what threat actors and targets exist.

Hosted services on the internet are assigned ports, which are virtual entries for network traffic on systems. Most applications have designated ports that host specific applications. Services that are hosted on non-designated ports are generally less secure and can indicate misconfigured systems. To make scanning for these applications easier, Stanford researchers developed a tool called LZR, which is able to scan the Internet for handshakes running on ports. [8] In their study across the IPv4 address space, researchers found that only 3% of HTTP services ran on the assigned port, and a lot of these services were running vulnerable applications. LZR is an open source tool that as currently released does not work in IPv6.

Network measurement is a critical area of cybersecurity. Understanding the layout of the Internet is critical to understanding where vulnerable targets are, how Internet censorship is carried out across the world, and what future attack vectors may exist.

3. PROJECT OVERVIEW

We are modifying LZR to support IPv6 networking, and recreating LZR's port scanning experiments in IPv6. The goal of this scan is to understand the distribution of services on ports within the IPv6 ecosystem, and how many services are running on mis-assigned ports.

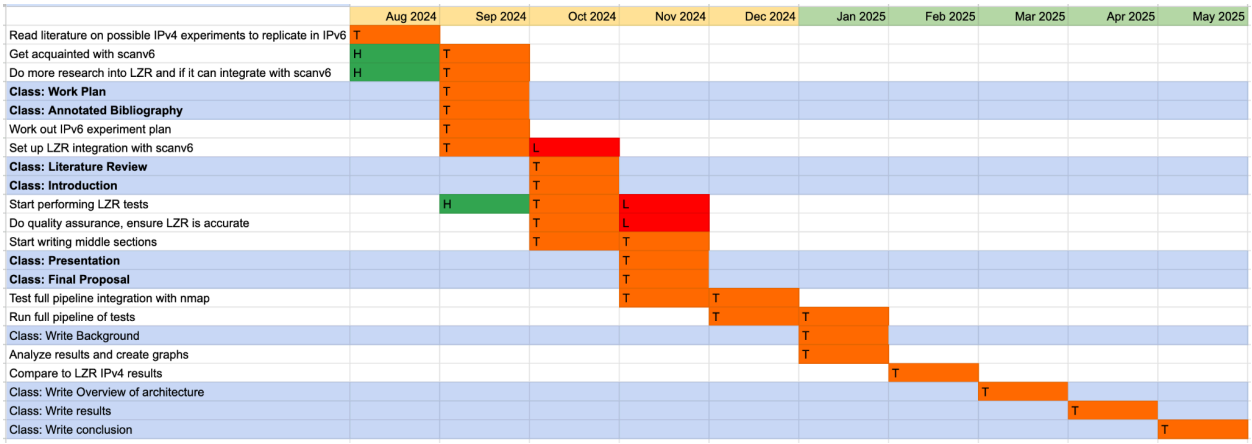
We will use 6Sense to generate a list of valid de-aliased IPv6 addresses to scan. This will serve as our base addresses to perform our scans. Using scanv6, we will check each address against a predetermined list of ports. An open port indicates that the port is accepting packets, but doesn't say anything about which service in particular. To determine this, we process the scanv6 results and feed the open IPs and ports into LZR, which will try to identify specific services being hosted on each system.

Due to IPv6's usage in developing countries and in newer systems, this research can provide unique insight into how these systems are set up, what vulnerabilities are out there, and how these stack up to existing setups in IPv4. Findings from this work can help determine systemic differences in how IPv4 and IPv6 services are used and deployed, especially in these regions of the world where much analysis hasn't been done.

4. WORK PLAN

Topic: Analyzing Misconfigurations and Vulnerabilities in IPv6

Problem: Lots of research has been done into misconfigurations and vulnerabilities in IPv4. IPv6, which is a newer protocol, has not been explored as much, and leaves lots of questions about how systems are laid out and what configurations and vulnerabilities they have. Especially since IPv6 is used more in developing regions, how systems across these networks are configured remains to be unexplored.



	Aug 2024	Sep 2024	Oct 2024	Nov 2024	Dec 2024	Jan 2025	Feb 2025	Mar 2025	Apr 2025	May 2025
Read literature on possible IPv4 experiments to replicate in IPv6	T									
Get acquainted with scanv6	H	T								
Do more research into LZR and if it can integrate with scanv6	H	T								
Class: Work Plan		T								
Class: Annotated Bibliography		T								
Work out IPv6 experiment plan		T								
Set up LZR integration with scanv6		T	L							
Class: Literature Review			T							
Class: Introduction			T							
Start performing LZR tests		H	T	L						
Do quality assurance, ensure LZR is accurate			T	L						
Start writing middle sections			T	T						
Class: Presentation				T						
Class: Final Proposal				T						
Test full pipeline integration with nmap				T	T					
Run full pipeline of tests					T	T				
Class: Write Background						T				
Analyze results and create graphs						T				
Compare to LZR IPv4 results							T			
Class: Write Overview of architecture								T		
Class: Write results									T	
Class: Write conclusion										T

5. Experimental Setup

While most studies assume internet services to be hosted on their IANA-assigned ports, Izhikevich et al. showed that vast discrepancies exist between services and the actual port hosted on. Their study revealed that only 3% and 6% of all detected HTTP and TLS services are on ports 80 and 443 respectively, revealing a significant amount of missed area in traditional internet scans. While these findings changed our perspective on the IPv4 ecosystem, it is unknown whether or not they hold true in IPv6. Thus, in this section, we aim to scan for services on unexpected ports in IPv6, and compare it to the results in IPv4.

To facilitate an accurate comparison between scans in IPv4 and IPv6, we created an experiment using a new scanner based on scanv6 [42, 43], and a modified version of LZR [22]. We modified scanv6 to support adaptive rate limiting based on autonomous systems, which enables us to increase the scanning speed of scanv6 without overprobing autonomous systems. To source the autonomous systems list we use to detect individual AS's, we collect a 'date' snapshot of the pfx2as dataset. We also tweaked the LZR client and built a version compatible with IPv6. This included networking fixes and some optimizations in the way LZR handled worker threads.

To start, we compile a list of known active IPv6 addresses based on Seeds, 6Sense, 6Tree, and ICMP error. We utilize the scanv6-based scanner to check the

activeness of the target IP addresses across 55 ports. We chose these 55 ports as we wanted to direct comparisons with previous experiments done by Izhikevich et al. showing expected/unexpected relationships on IPv4. After running the scanv6 scans, we calculate all IP addresses that have more than 85% of ports checked open, and mark these as proxy IPs to not be further included.

Using the output data from the scanv6 scans, filtering for proxy IP addresses we now run a series of LZR scans. We start by sending a LZR handshake for wait, http, and tls as the LZR study showed X% of protocols can be covered for the configuration. For the remaining protocols that require a handshake of the correct protocol, we send handshakes for each remaining protocol.

To identify the protocol being hosted, we aggregate the protocols identified by LZR for every IP-port pair. First, if LZR finds the default protocol, it will mark the IP-port as using that protocol. Next, we use a hierarchy to determine if any specialized services are being hosted, before defaulting to the most common fingerprint.To scan IPv4 we use ZMap [33] with LZR to conduct 100% scans

6. Results and Discussion

We scanned and analyzed large scale internet data across the IPv6 address spaces. Our scans were performed in March and April of 2025, and identified 102,143,608 recognized IPv6 endpoints across 55 ports.

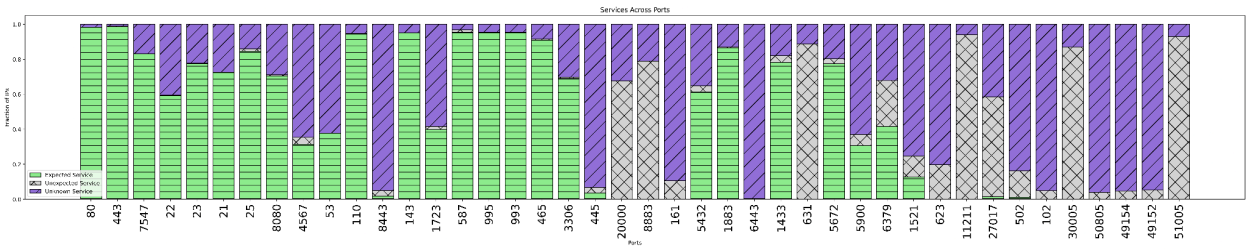


Figure 1: Distribution of Types of Services on IPv6

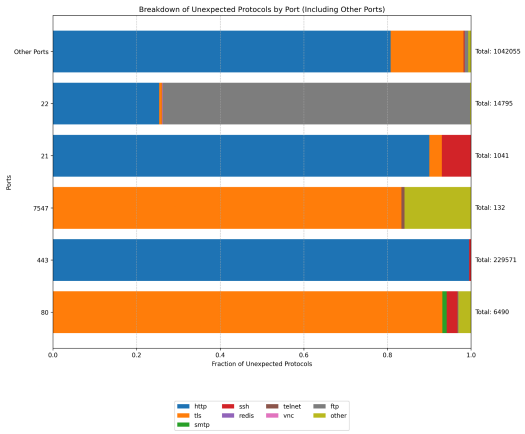


Figure 2: Distribution of Unexpected Protocols on IPv6

6.2.1 IPv6 Service Classification. We first analyzed the distribution of services as "Expected", "Unexpected", or "Unknown" based on the identification methodology described in [22]. Figures 1 and 2 illustrate the aggregate

classifications across the 55 ports. For example, drilling into port 23 on the graph, 77.59% of services were identified as Telnet, and thus considered “Expected” as they are running on the default port. 22.34% of services were “Unknown,” meaning that LZR was not able to classify the service under an existing handshake. For some ports like 51005, despite being made up of 92.87% HTTP connections, it is classified as “Unexpected” because it is an ephemeral port that doesn’t have an assigned protocol.

Across the port numbers, IPv6 had significantly more “Expected” services than seen in previous IPv4 research [8]. This trend was particularly pronounced for mail protocols. For example 95.38% and 95.44% of port 993 and 995 correspondingly were identified by their expected protocols on IPv6.

6.2.2 Analysis of Services on Unexpected Ports. To further investigate the distribution of services, we next examined the specific protocols making up the “Unexpected” category. We choose to examine ports 21, 22, 80, 443, and 7547. Figure 2 shows the distribution on IPv6

On port 80 where HTTP is expected, TLS makes up the vast majority of unexpected ports, sitting at 93.03% of unexpected ports overall. The second largest unexpected protocol for port 80 on IPv6 is SSH, sitting at a distant 2.59%.

For port 22, we see one of the biggest differences between IPv4 and IPv6. Although SSH is the expected protocol, FTP makes up 73.50% of unexpected protocols on IPv6. This can be partially attributed to the specific collection of IPv6 addresses included in the seed database.

For example, digging into port 27017 we see that one AS, 51167, makes up 66.89% of unknown ports. Similarly, port 631 is 58.54% of AS 51167.

7. Conclusion

In this thesis, I set out to answer two questions:

1. How does the prevalence of unexpected and unknown services compare in IPv6 versus IPv4?
2. Are there any unique considerations to internet scanning on IPv6 versus IPv4?

To address these, we built a scanning setup based on scanv6 and modified LZR to support IPv6 addresses. We performed scans across 55 ports between March and April 2025.

Our study found 102,143,608 recognized IPv6 endpoints. Overall, IPv6 had a higher conformity of services to IANA-assigned port numbers than IPv4. For example, 95.38% of IMAP services on port 993 and 95.44% of POP3 services on port 995 appeared on their standard ports. On the other hand, for port 53, only 37.58% were identified as DNS, with 62.38% being identified as unknown.

While our results may be impacted by factors like seed-list composition and proxy filtering, this work establishes a reputable methodology for large-scale IPv6 internet scanning. Future efforts may focus on further diversifying address sources or integrating active vulnerability probes.

By showing how services are mapped onto ports in IPv6, this thesis lays the groundwork for more robust IPv6-compatible security tools and best practices. As

the internet continues to shift towards IPv6, continued measurement at this scale is important to understand shifting dynamics in the internet landscape.

REFERENCES

[1] Liz Izhikevich, Renata Teixeira, and Zakir Durumeric. 2022. Predicting IPv4 Services Across All Ports. In ACM SIGCOMM 2022 Conference (SIGCOMM '22), August 22–26, 2022, Amsterdam, Netherlands. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3544216.3544249

[2] Keegan Ryan, Kaiwen He, George Arnold Sullivan, and Nadia Heninger. 2023. Passive SSH Key Compromise via Lattices. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS '23), November 26–30, 2023, Copenhagen, Denmark. ACM,* New York, NY, USA, 15 pages. https://doi.org/10.1145/3576915.3616629

[3] Sullivan, George Arnold, et al. "Open to a fault: On the passive compromise of TLS keys via transient errors." *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, 2022.

[4] Antonakakis, Manos, et al. "Understanding the Mirai Botnet." 26th *USENIX Security Symposium (USENIX Security 17)*. USENIX Association, 2017.

[5] Williams, Grant, Mert Erdemir, Amanda Hsu, Shraddha Bhat, Abhishek Bhaskar, Frank Li, and Paul Pearce. "6Sense: Internet-Wide IPv6 Scanning and its Security Applications." *33rd USENIX Security Symposium (USENIX Security 24).* USENIX Association, 2024.

[6] Foremski, Paweł, et al. "Entropy/IP: Uncovering Structure in IPv6 Addresses." *Proceedings of the 2016 Internet Measurement Conference (IMC '16),* ACM, 2016, pp. 167-181.

[7] Czyz, Jakub, et al. "Don't Forget to Lock the Back Door! A Characterization of IPv6 Network Security Policy." *NDSS '16: Network and Distributed System Security Symposium*, Internet Society, 2016, pp. 1-15, http://dx.doi.org/10.14722/ndss.2016.23047.

[8] Izhikevich, Liz, and Renata Teixeira. “LZR: Identifying Unexpected Internet Services.” *30th USENIX Security Symposium*, 2021, https://doi.org/https://www.usenix.org/system/files/sec21-izhikevich.pdf.

[9] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Kor-czyński, Stephen D. Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters

in the Expanse:, Understanding and Unbiasing IPv6 Hitlists. In 2018 Internet Measurement Conference (IMC ’18), October 31-November 2, 2018, Boston, MA, USA. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3278532.3278564

[10] Austin Murdock, Frank Li, Paul Bramsen, Zakir Durumeric, and Vern Paxson. Target Generation for Internet-wide IPv6 Scanning. In Proceedings of IMC ’17, London, UK, November 1-3, 2017, 12 pages. https://doi.org/10.1145/3131365.3131405

[11] “IPv6.” *Google*, Google, www.google.com/intl/en/ipv6/statistics.html. Accessed 25 Nov. 2024.

[12] Davies, Kim. "An Introduction to IANA: Presentation Notes." *ICANN*, 29 Sept. 2008.

[13] Johnson, Daryl G., and Bruce H. Hartpence. "A Re-examination of Network Address Translation Security." *Rochester Institute of Technology*, 2010.

[14] Durumeric, Z., Wustrow, E., & Halderman, J. A. (2013). ZMap: fast Internet-wide scanning and its security applications. USENIX Security Symposium, 605–620. https://zmap.io/paper.pdf

[15] Durumeric, Z., Adrian, D., Mirian, A., Bailey, M., & Halderman, J. A. (2015). A Search Engine Backed by Internet-Wide Scanning. CCS'15. https://doi.org/10.1145/2810103.2813703

[16] Ullrich, Johanna, et al. "On Reconnaissance with IPv6: A Pattern-Based Scanning Approach." *SBA Research*, Vienna, 2014.

[17] Austin Murdock, Frank Li, Paul Bramsen, Zakir Durumeric, and Vern Paxson. Target Generation for Internet-wide IPv6 Scanning. *In Proceedings of IMC '17*, London, UK, November 1-3, 2017, 12 pages. https://doi.org/10.1145/3131365.3131405

[18] Zhizhu Liu, Yinqiao Xiong, Xin Liu, Wei Xie, Peidong Zhu, 6Tree: Efficient dynamic discovery of active addresses in the IPv6 address space, Computer Networks, Volume 155, 2019, Pages 31-46, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2019.03.010.

[19] B. Hou, Z. Cai, K. Wu, J. Su and Y. Xiong, "6Hit: A Reinforcement Learning-based Approach to Target Generation for Internet-wide IPv6 Scanning," *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, Vancouver, BC, Canada, 2021, pp. 1-10, doi: 10.1109/INFOCOM42981.2021.9488794.

[20] B. Hou, Z. Cai, K. Wu, T. Yang and T. Zhou, "6Scan: A High-Efficiency Dynamic Internet-Wide IPv6 Scanner With Regional Encoding," in *IEEE/ACM Transactions on Networking*, vol. 31, no. 4, pp. 1870-1885, Aug. 2023, doi: 10.1109/TNET.2023.3233953.

[21] Yang, Tao, et al. "6Graph: A Graph-Theoretic Approach to Address Pattern Mining for Internet-Wide IPv6 Scanning." *Computer Networks*, vol. 203, 2022, p. 108666. https://doi.org/10.1016/j.comnet.2021.108666