MASTER THESIS

Isabell Egloff

IPv6 Scanners: Strategy, Behavior, Intent

Faculty of Engineering and Computer Science
Department of Computer Science

Isabell Egloff

IPv6 Scanners: Strategy, Behavior, Intent

Masterarbeit eingereicht im Rahmen der Masterprüfung
im Studiengang Master Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuender Prüfer: Prof. Dr. Thomas C. Schmidt
Zweitgutachter: Prof. Dr. Franz Korf

Eingereicht am: 20. Dezember 2024

Isabell Egloff

Thema der Arbeit
IPv6 Scanners: Strategy, Behavior, Intent

Stichworte
IPv6, Scanner, Taxonomie, BGP-Announcements, Netzwerk Teleskop

Kurzzusammenfassung
Das Scannen im Internet ist eine weit verbreitete Methode, die sowohl in der Forschung und Wirtschaft als auch von böswilligen Akteuren genutzt wird. Während das Scanverhalten im gesamten IPv4-Adressraum untersucht wird, ist es im IPv6-Adressraum nicht möglich diesen komplett zu beobachten. Aufgrund der enormen Größe des Adressraums ist ein vollständiges Scannen nicht möglich. In dieser Arbeit wird das aktuelle IPv6 Scan-Verhalten analysiert. Hierzu wird der Netzwerkverkehr von vier Netzwerkteleskopen untersucht, die sich durch unterschiedliche Eigenschaften wie Präfixgröße, Reaktivität, DNS-Einträge, Aktivität, Einträge in Hitlisten und BGP-Announcements auszeichnen. Es wird analysiert, wie Scanner auf diese unterschiedlichen Eigenschaften reagieren. Die Ergebnisse zeigen, dass BGP-Announcements von Präfixen sofort deutlich mehr Aufmerksamkeit erregen als nicht annoncierte Subnetze. Darüber hinaus beobachten wir mehrere Zwei-Phasen-Scans. Untersuchungen der Payloads und Ports zeigen, dass die Scans auch Schwachstellenprüfungen enthalten. Die Erkenntnisse dieser Arbeit verdeutlichen, dass die Reaktionen von IPv6-Scannern durch gezielte Maßnahmen beeinflusst werden können, die bei einem Netzwerk Teleskop vorgenommen werden.

Isabell Egloff

Title of the paper
IPv6 Scanners: Strategy, Behavior, Intent

Keywords
IPv6, Scanner, Taxonomy, BGP Announcements, Network Telescope

Abstract
Internet scanning is a widely used method, applied in research and business as well as by malicious actors. While scanning behavior can be examined across the entire IPv4 address space, it is not possible to observe the entirety of the IPv6 address space. Due to the vast size of the address space, a complete scan is not feasible. This study analyzes the current IPv6 scanning behavior. For this purpose, the network traffic of four network telescopes is analyzed, which are characterized by different properties such as prefix size, reactivity, DNS entries, activity, entries in hitlists, and BGP announcements. The study investigates how scanners react to these different properties. The results show that BGP announcements of prefixes immediately attract significantly more attention than non-announced subnets. Furthermore, we observe several two-phase scans. Analyses of payloads and ports reveal that vulnerability checks are part of the scans. The findings of this work show that IPv6 scanners can be influenced by targeted measures applied to a network telescope.

Contents

List of Figures

List of Tables

1 Introduction and Motivation

For research, business interests, or malicious actions, Internet measurements are often conducted. Unlike the IPv4 address space, the IPv6 address space is so large, with 2^{128} addresses, that scanning all of it is impossible. Because of this, IPv6 scanners use specific methods to find active hosts. This raises the question of how these scanners work and whether their behavior can be influenced to attract scanners to scan specific prefixes or addresses.

There are already many scientific studies that address various aspects of scanning activity in IPv4 [13, 36]. Scanning is often performed using well-known tools such as Nmap[1], Masscan [18], or ZMap [14]. The IPv6 address space, on the other hand, is less explored. This is because it is more challenging to draw conclusions when only specific parts of the address space are visible. Therefore, it is reasonable to examine already studied scanning strategies [25, 51, 50, 10, 30]. It is also useful to investigate which effective scanning strategies have been applied in IPv6. There are already studies that focus on such target generation algorithms [15, 39, 12, 27]. In this study, the network traffic of four telescopes with different properties is analyzed. The network traffic of all telescopes is compared in order to understand the impact of their individual properties and evaluate the results. In order to identify and classify scanning behavior, a taxonomy is developed that specifically categorizes this behavior. To examine different scanning behaviors, the traffic from four telescopes with distinct characteristics is analyzed and compared.

One telescope property under investigation is the impact of BGP announcements on scanning activities. To understand these effects, it is important to consider the background of BGP (Border Gateway Protocol) and its role in internet routing. In the late 1990s, efforts began to collect and distribute BGP routing information from multiple backbone networks in near real-time. These projects, RouteViews and RIPE RIS, used BGP Route Monitors (Collectors). Initially only large providers connected to this service, but over time the service was extended to Internet Exchange Points (IXPs), providing important data that is still widely used today by researchers and network operators to monitor and debug network configurations [40].
This study investigates whether announcing a prefix in BGP influences scan traffic and specifically examines how scanners respond to such announcements. Additionally, it ex-

[1] https://nmap.org/

plores whether the size of the announced prefix makes a difference. For instance, does a /32 prefix attract more, less, or roughly the same amount of traffic and scanners compared to a /48 prefix? These investigations can provide insights into how scanners behave in the IPv6 environment.

A widely used scanning technique in IPv4 is the stateless scanning method. Well-known tools for stateless scanning today include Unicorn [31], Masscan, and ZMap. The use of stateless scans in IPv4 has increased significantly with the advent of these tools. Hiesgen *et al.* [20] also observed that many stateless scans are followed by stateful scans if the target responds, often to further investigate potential vulnerabilities. They examined this two-phase behavior using the reactive telescope Spoki developed at HAW Hamburg. Spoki has since been used in several studies [21, 22, 19] for analyzing various attacks, leading to the collection and examination of payloads from both benign and malicious communication partners.
In this work, Spoki was deployed in the IPv6 address space to analyze the behavior of two-phase scanners. These investigations provide deeper insights into IPv6 scanner activity and reveal the types of packets sent during the second phase, which are often not received without a response in the first phase. Accordingly, this study further explores the analysis of two-phase scanners in IPv6 to uncover new insights.

The following research questions are explored in this work.

Research Questions
The analyses are guided by the following questions:

1. How do BGP announcements, prefix size, DNS entries, active subnets, reactivity, and hitlist entries influence scanning behavior?
2. Does the size of the announced prefix affect network traffic?
3. How can scanning behavior be categorized?
4. Can biases arise in the analysis of scanning behavior that can be attributed to the characteristics of the telescopes?
5. Can the representation of the results cover all perspectives of the captured scanning traffic?
6. How do two-phase scanners behave in IPv6?

Outline

Sections 2, 3, and 4 provide the background necessary to clarify key terms and explain the methodology. Section 2 focuses on the problem statement, highlighting the challenge of accurately interpreting findings when IPv6 analyses are limited by the restricted visibility of traffic.
Section 3 presents related work and background information. Following this, Section 4 outlines the methodology and setup established for conducting the study.
In Section 5, an overview of the traffic observed by the four studied telescopes is provided. This section raises several questions based on the analysis, which are explored in greater depth in subsequent sections. Therefore, it offers only a brief initial overview and lays the groundwork for a more detailed analysis later.

After this overview, Sections 7, 8, and 9 delve into more specific investigations. To support this deeper analysis, Section 6 introduces the taxonomy used to classify scanning behavior and clarifies the individual terms of the categorizations, thereby simplifying the categorization of individual scanning behaviors.
Section 7 revisits all telescopes to conduct more focused analyses of open questions. It compares observations across the telescopes and examines the unique characteristics of each telescope in detail.
Section 8 shifts the focus to Telescope T1, which is used for a BGP experiment. This large-scale experiment investigates how IPv6 scanners respond to BGP announcements and their subsequent behavior, analyzing multiple aspects of the resulting traffic.
Section 9 focus on Telescopes T2 and T4, where the reactive network telescope Spoki is deployed. This section particularly examines two-phase scanners in IPv6, conducting extensive analyses to comprehensively map their scanning behavior.

In Section 10, the discussion revisits the telescope properties described in the methodology, connecting them to observations of scanning behavior. The section also raises questions derived from the findings and outlines the challenges faced during the investigations.
Finally, Section 11 summarizes the study and outlines opportunities for future research.

2 Problem Space

Just as it is a significant challenge for scanners to find active addresses, it is also challenging to detect scanning activity. There are many ways to configure a telescope to attract scanners. Each method can attract different scanners because they focus on various aspects to identify targets, which are often unknown. By utilizing activity, reactivity, or attractors in telescopes, as well as prefix announcements, the telescope can become more visible, prompting scanners to send specific packets.

In the following sections, the challenges that need to be considered in this work will be discussed. These include the need to carefully choose the configuration methods and to understand how different scanners respond to various attractors. Additionally, it is crucial to recognize that the effectiveness of these methods can vary significantly based on the underlying behavior and objectives of the scanners involved.

2.1 Impact of Telescope Properties on Scan Behavior

Without measures to encourage scanners to scan the observed network area, very few packets will be received. Conversely, methods that generate high data traffic can bias the results, as they significantly influence the observed traffic and may lead to an inaccurate picture. In addition, the behavior of scanners can be strongly influenced by different factors, which makes precise analysis difficult. Similarly, some scanners have a significant impact on the observed network traffic through packet volume, while others influence it only minimally. This variability makes it difficult to examine overall scanning behavior without inadvertently focusing on the ones that generate the most visible traffic. The goal is to create a balanced view that considers all scanner behaviors without allowing any single type to dominate the analysis.

2.2 BGP Experiment

Special attention is given to the reactions to BGP announcements. One challenge is the difficulty of isolating the immediate effects of BGP announcements, as other factors can also influence network traffic. It is complex to analyze whether certain scanners specifically react to changes or if they are influenced by other factors. Additionally,

characteristics of prefixes, such as prefix size, may affect how scanners respond differently. Which complicates the comparability of the results.

2.3 Impact of the Reactive Network Telescope Spoki

In addition, it is tested whether two-phase scans in IPv6 can be identified using a reactive network telescope. Identifying two-phase scanners can be challenging as IPv6 scanners may perform scans that are distributed across multiple network areas, which can significantly delay the arrival times of packets within the monitored telescope. Unlike IPv4, where the entire address space can be completely scanned in a short time, the scanning process within the IPv6 address space is more complex, making it difficult to determine multiple scan phases.

3 Background and Related Work

The analysis of scanning behavior and the development of effective scanning strategies have become an important research area in recent years. These studies form the foundation for the investigations presented in this work.

3.1 Background

Payload Analysis. Payload analyses have already been carried out extensively for IPv4 traffic. In 2005, Bailey *et al.* [1] collected information on IoT malware, including payload URLs and login credentials. Pang *et al.* [35] analyzed background radiation, breaking it into protocol, application, and often specific exploits. They primarily identified activities related to worms and autorooters. Application-level responders were created based on the traffic volume to gather more information.

Scanning strategies in IPv6. Sometimes, IPv6 addresses were manually configured, as described in RFC 7707 [17], such as when assigning addresses to routers. Instead of randomly choosing IPv6 addresses, specific address structures that were easier to remember were commonly selected. One common approach was to append only a *::1* to a fixed prefix, forming what are known as (*i*) *low-byte* addresses [17]. In these addresses, all bytes of the Interface Identifier (IID), except for the lowest-value bytes, were

set to zero, as in *2001:db8::1*. An address is also considered as a *low-byte* address if the IID is zero except for the last two 16-bit words, as in *2001::db8::1:10*. Another option was the use of (*ii*) *embedded-port* addresses [17], where the port of a running service was embedded into the IID, such as the HTTPS port in *2001:db::443*. (*iii*) *IPv4-embedded* addresses [17], on the other hand, embedded the IPv4 address of the network interface into the IID, as seen in *2001:db8:122:344::192.0.2.33* [2]. (*iv*) So-called *wordy addresses* [53] contained recognizable words, like *2001:db8::cafe* or *2001:db8::affe*. (*v*) *IEEE-based* or *SLAAC* addresses [17] embedding the word *0xfffe* between Organizationally Unique Identifier (OUI). The IID is generated based on the Media Access Control (MAC) address. Another category known as a pattern for target address generation is (*vi*) *ISATAP* (Intra-Site Automatic Tunnel Addressing Protocol). This IPv6 tunneling technique allows IPv6 to connect over an IPv4 network. These addresses are identified by the first two words in the IID being *0000:5efe*, followed by an embedded IPv4 address in the last two words, for example, *2001:db8:1100:1:0000:5efe:8d54:4503* [47] [46]. If destination addresses cannot be assigned to any of the categories, they are classified as (*vii*) randomized [33]. Pattern-based scanning has already been used in the work of Ullrich *et al.* [48]. They developed an algorithm that, starting from a set of seed addresses and a threshold value N, generated targets from IPv6 address spaces of constant size, dependent on N. Over time, many target generation algorithms (TGA) have been developed [52, 8, 28, 29, 25, 51, 50, 24, 10, 11, 9, 41, 30]. These included static TGAs, which generated potential scan candidates based on a fixed training set. In contrast, dynamic TGAs adjusted their training set by immediately evaluating the activity of generated addresses through active scanning [42].

Scanning tools and scan systems in IPv6. Various tools are employed for scanning, each leaving fingerprints in the payloads of the packets. Some scanning tools can, therefore, be detected based on these payloads. Beverly introduced Yarrp [4]. It is a tool designed for rapid Internet topology discovery, which avoids testing each path to individual targets sequentially. Instead, Yarrp uses a random permutation of destination addresses and TTL values to prevent router overload. It operates without storing state information, reconstructing all necessary data from asynchronously arriving ICMP responses, and allowing for high-speed scanning. However, due to the larger address space, Yarrp faces additional challenges in IPv6. To address this, Beverly *et al.* [5] modified Yarrp into Yarrp6, advancing the state of the art in Internet-wide IPv6 active topology mapping. Through their work, they discovered over 1.3M IPv6 interface addresses. Another tool for large-scale Traceroute scans is Flashroute, developed by Huang *et al.* [26]

in 2020. Yang *et al.* [49] develop Trace6. Trace6 is a threat traceback model specifically designed for IPv6 networks, utilizing the large address space and extended address fields of IPv6. It combines user authentication and address verification to reliably link addresses to users, enhancing security with a simpler approach.

3.2 Related Work

In 2005, the analysis of IPv6 scanners began with the work of Ford *et al.* [15]. They examined the background radiation of a /48 prefix. The new Internet protocol (IPv6) was not widely-used then, so they received no more than 12 packets.

In 2023, Ronan *et al.* [39] conducted a new measurement experiment using the same /48 prefix studied by Ford *et al.* [15]. Over six months, they received 5k packets. Of these, 74% were ICMPv6 packets, 21% were TCP packets, and 4% were UDP packets.

In 2013, Czyz *et al.* [12] analyzed the IPv6 background radiation from five announced /12 address blocks assigned to the five Regional Internet Registries (RIRs). They collected exclusively darknet traffic, meaning the packets directed to the subnets (of the /12) that were never assigned or routed. This represented 5% (209M) of all received packets. They observed that for specific RIR prefixes (ARIN and APNIC), as few as one to 2k source IP addresses comprised 90% of the traffic. In contrast, for the other RIRs (AFRINIC and LACNIC), it took over 10k and 100k source IP addresses, respectively, to reach the same 90% coverage of packets.

In 2018, Fukuda *et al.* [16] proposed using DNS backscatter to detect IPv6 scanning behavior. They identified scanners based on DNS reverse lookups performed by routers. The methodology was validated by comparing the scanners with MAWI data and a /37 IPv6 darknet. Over six months, they identified 16 active IPv6 scanners per week. From their analysis of scanning behavior, they derived three scanning methods: (*i*) scanning of *low-byte* addresses, (*ii*) targeting IPs that are discoverable through reverse DNS, and (*iii*) utilizing the TGA 6Gen by Murdock *et al.* [32].

In 2020, Strowes *et al.* [43] analyzed the network traffic of a newly announced /12 prefix. They separately announced four /32 prefixes and four /48 prefixes from a /29 covering prefix. 95% of nearly 5.5M TCP traceroute packets originated from a single Autonomous System (AS). Among the remaining packets, 54% of the remaining packets targeted the reply addresses, one of which was advertised as a test address on a mailing list. The

other 46% of packets were spread more widely across the prefix, with a particular focus on *low-byte* addresses. The authors generated most of the recorded traffic to conduct a more in-depth analysis of route filtering.

Liu *et al.* [27] also analyzed the background radiation of a previously unused /20 prefix in 2021. They observed 2.9M packets over six months. The majority of these packets were ICMPv6 (67%), followed by TCP packets (33%), and finally, they saw the least amount of UDP packets ($< 1\%$). Just ten source addresses accounted for 95% of all received packets.

In 2022, Hiesgen *et al.* [20] developed the reactive network telescope Spoki. It is specifically designed to prompt scanners to start a stateful second phase after a stateless first phase. In this second phase, packets are only sent to targets that were identified as responsive in the first phase. This reduces the complexity of the scans. Spoki reacts specifically to TCP SYN packets and thus triggers the second phase of the scanner. The payload and TCP destination ports in the received packets are then analyzed. Over a period of three months, the two-phase scanners were investigated and a significant number of malicious actors were detected, responsible for a large proportion of the observed events.

Richter *et al.* [37] analyzed IPv6 scanning behavior in 2022 using firewall logs from a major CDN. They excluded ICMP packets, as well as TCP and UDP packets with destination port 80 or 443. Their analysis found that the two most active scanners originated from data center ASes in China. This was followed by a cybersecurity company from the USA then a variety of US and global hosting and cloud providers. The top 5 source ASes accounted for 93% of the scan packets.

Tanveer *et al.* [44] investigated how IPv6 host activities (web crawls, NTP pool servers, public NTP servers, Tor, DNS queries, DNS zones) influence the behavior of scanners in the IPv6. They analyzed the network traffic of a previously unused /56 subnet. Each of the six experiments was conducted on four randomly selected /64 subnets. The results showed that publicly visible active services (NTP, Tor, DNS zones) attracted more scanning activity than communications initiated by telescopes (web crawls and DNS queries). They found that the source addresses either scanned random addresses or focused on *low-byte* addresses. Only one of these two strategies was employed by 65% of all scanners. The *low-byte* scanners generated just 4% of all received packets, while scanners targeting random IIDs accounted for only 5%. The remaining 91% of packets were sent by scanners that utilized a mix of both strategies.

Zhao *et al.* [53] deploy four DNS-based address-exposing methods to validate their effectiveness. They used a previously unused /56 network and published addresses through four methods: (*i*) IPv6 addresses associated with domains that have IPv4 PTR records, (*ii*) PTR records for random addresses, (*iii*) PTR records for *wordy* and *port-embedded* addresses, and (*iv*) IPv6 addresses with popular domain names. Each method was deployed in both a /64 darknet and a /64 honeynet. More than 99.99% of the scans were attracted by the IPv4 reverse method, whereas no associated DNS queries were received for the other methods. In the darknet, scans primarily focused on known addresses (97%). They furthermore, targeted random and unknown addresses. In the honeynet, scanners were less focused, distributing their efforts more evenly across both known and unknown addresses near the known ones. *Low-byte* scans comprised only 0.03% of all scans.

4 Methodology and Setup

The measurement period began on August 24, 2023, and continued until July 2, 2024, during which all packets received by the four telescopes were collected as PCAPs. Each telescope has unique characteristics that can influence the captured traffic differently. For an accurate evaluation, this section defines scan sessions and describes the aggregation of scan sources used throughout this work. Subsequently, the properties of the telescopes are presented, and finally, the setup is explained.

4.1 Scan Sessions and Scan Sources

In this study, traffic is presented almost exclusively through scan sessions or scan sources. It is important to note that some sources send significantly more packets than others. This can result in certain sources being overrepresented in the analysis, as those sending more packets may dominate the results, shifting the focus toward them rather than reflecting the overall traffic distribution. Therefore, analyzing scan behavior based on sources and sessions, rather than total packets, is more meaningful.

Scan sources. A scan source is either an individual address (*e.g.,* /128) or an aggregation of addresses from a network (*e.g.,* /64, /48). Scan sources can be aggregated in

different ways. A source may use multiple /128 source IP addresses to scan. In an analysis that uses /128 aggregation for sources, each of these IP addresses would be identified as a separate source, even though they belong to the same entity. However, if an analysis is done where all sources are grouped into a /64 aggregation, multiple sources might be seen as one, even though they are actually separate entity. Related work also looks at these aggregations. Strowes *et al.* [43] examined individual source IP addresses received. They find that 160k (/128) sources come from 150k /64 sources and 1k ASNs. The high number of /64s suggests that these sources are not highly concentrated. Richter *et al.* [37] generated statistics and compared the results for individual source IP addresses against /64 and /48 aggregations. This study analyzes the influence of different aggregation levels, as discussed in Section 7. Nevertheless, identifying globally connected scanning units is a complex task. This is not part of the work. But the aggregation levels are used in this work to get a better overview of the traffic.

Scan sessions. A scan session is often defined as a sequence of consecutive packets from a single source, where the interarrival time (the time between two consecutive packets) is shorter than the session timeout period. Scan sessions can also be categorized based on different criteria. Zhao *et al.* [53] and Richter *et al.* [37] set the session timeout to one hour. Richter *et al.* also established the criterion that packets must be sent to at least 100 distinct destination addresses within a session. In this study, a session timeout of one hour is applied without additional restrictions, as no packets or sources are intended to be discarded.

4.2 Telescope Properties

Table 4.1 gives an overview of the four network telescopes and their properties. The properties include the prefix size, reactivity, attractors, activity, and entries in the hitlist from the TU of Munich.

Attractors are methods that prompt scanners to scan a prefix. Zhao *et al.* [53] use four DNS-based methods for this purpose, with over 99.99% of the scans being attracted by their *IPv4 reverse method*. Tanveer *et al.* [44] investigate whether host activity types can serve as attractors, and the scanner attention they evoke. The host activity types include web crawls, NTP pool servers, public NTP servers, Tor, DNS queries, and DNS zones. In this study, the attractors include BGP experiments, active subnets within the telescopes, reactivity of the telescopes to TCP requests, and DNS entries.

Table 4.1: Properties of the four network telescopes.

	Size	Attractors	Reactivity	Aktivity	Appearance on hitlist
T1	/32	BGP Announcements	No	No	Non-aliased: since August 29, 2023
T2	/48	• Reactivity • DNS entry • Active /56 subnet	TCP Handshakes starting from December 19, 2023	Active /56 subnet	Non-aliased: Aug. 24, 2023 Aliased: Dec. 21, 2023
T3	/48	No	No	No	Only /29 covering prefix: Non-aliased: Aug. 24, 2023
T4	/48	Reactivity	TCP Handshakes since the start of measurements	No	Only /29 covering prefix: Non-aliased: Aug. 24, 2023

T1: BGP controlled /32 - /48. T1 was announced as an untainted /32 IPv6 prefix. The /32 prefix first appears on the non-aliased hitlist on August 29, 2023 (five days after the announcement). To examine the reactions of IPv6 scanners to BGP announcements, T1 was recursively split into more-specific prefixes during the measurement period. For each split, the prefix is withdrawn for roughly 24 hours. Starting with the /32 prefix announcement, it is progressively decomposed to a /48 prefix. Which is the most-specific prefix size in IPv6 allowed in BGP[2].

In the first twelve weeks, T1 remains completely passive. During this time, intervals for splitting T1 are established. The intervals need to be balanced because if they are too long, the overall measurement period could become excessive. If they are too short, traffic might be too low to effectively analyze scanning behavior. A summary is created to track how quickly new, previously unseen source prefixes emerge over time. Figure 4.1 shows this cumulative increase in new source prefixes.

The largest increase in new source prefixes is seen within the first two weeks. The first two weeks are marked in red in Figure 4.1. Thus, the time interval for the announcements is set to two weeks. Two weeks represent each step in a measurement period totaling 32 weeks, as the /32 prefix is gradually split until it reaches the most-specific /48 prefix. After each announcement period, all announced prefixes are withdrawn from T1, and the next day the currently most-specific prefix is split and these new prefixes are announced along with all the others. We confirmed their visibility with a looking glass [45] and with RIPEstat [34]. The most-specific prefix is always split, so that each of the two new prefixes has a unique *low-byte* address that differs from the previous one. This allows checking whether the scanners respond to the new announcement and scan the new *low-byte* addresses, instead of scanning the already known *low-byte* address. Tanveer *et*

[2]https://blog.apnic.net/2020/06/01/why-is-a-48-the-recommended-minimum-prefix-size-for-routing/

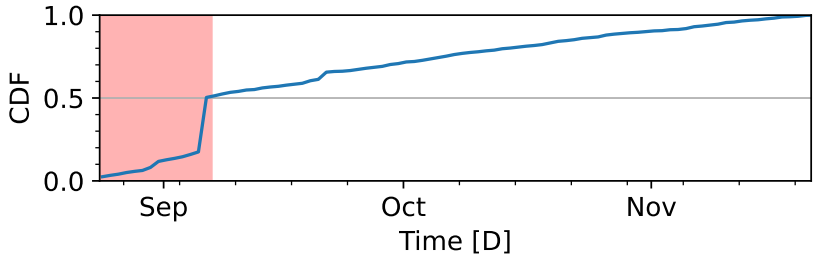


Figure 4.1: Overview of new source prefixes in the first 12 weeks, with the highest increase observed during the first two weeks (indicated by the red marking).

al. [44] also describe *low-byte* scanning as a dominant strategy of the scanners. Therefore, using this strategy when splitting the prefixes seems reasonable.

T2: Partially productive /48. T2 has been continuously announced for 13 years. At the beginning of the experiment, T2 was already included in the non-aliased hitlist. This /48 prefix includes an active /56 subnet within the telescope that is not announced separately. It hosts services such as web servers and IoT devices, some of which have persistent DNS entries. The packets addressed to the /56 subnet are not part of the analysis and are filtered out. After running mostly passive, T2 begins to react to TCP SYN packets starting from December 19, 2023. Additionally, it has been listed in the aliased prefixes since December 21, 2023, two days after it began responding to TCP SYN packets. Furthermore, one address within the /48 prefix, outside the active /56 subnet, has a DNS entry. This name co-exists in IPv4 and is part of the CISCO Umbrella popularity list.

T3: Silent /48. T3 is a /48 network and part of a BGP announced /29 covering prefix. T3 is not separately announced in BGP. At the beginning of the experiment, the /29 covering prefix is included in the non-aliased prefix list. T3 is passive, meaning it does not host any services or have active clients. Furthermore, it is not publicly listed in hitlists.

T4: Reactive /48. This /48 network is, like T3, also a part of the same /29 covering prefix. T4 is also not separately announced in BGP. T4 responds to TCP SYN packets from the beginning of the measurement.

4.3 Setup

This section explains the setup needed for the analyses and specific experiments, including the BGP experiment setup and the extension of the network telescope Spoki for functional application in IPv6.

4.3.1 Setting Up the BGP Experiment

FRR. For the announcements, FRR[3] runs on a Linux server that connects our AS to an IXP and peers with upstream providers. A looking glass [45] and RIPEstat [34] are used for verification.

Route6 object. Route6 objects serve as records that provide details about peering relationships within the RIR database. These records are commonly used in public peering and, on occasion, by upstream providers to verify the legitimacy of the routes received from their peers. To assess the impact, a Route6 object for the non-split /33 prefix is generated four months after the original announcement. The findings can be found in Section 10.

4.3.2 Setting Up the Reactive Network Telescope Spoki

Implementation. We extend the reactive network telescope Spoki [20] to support IPv6. The source code is adjusted so that Spoki accepts IPv6 packets, recognizes the difference between IPv6 and IPv4 packets, processes them correctly, and subsequently sends the matching responses. Based on the C++ Actor Framework (CAF) [7], Spoki uses the IP type implementation in CAF. In addition to the large addresses, the probe module of Spoki now supports Ethernet raw sockets. IPv4 and IPv6 address types can therefore be used. The library libtrace[4] is used for processing network traffic captures. To test the functionality of Spoki in IPv6 after implementation, we verified that TCP SYN packets are captured properly and that responses packets are constructed and sent correctly using Wireshark[5]. We captured network traffic and deployed the packets to test the response of Spoki. Following this, the functionality for both IPv4 and IPv6 packets was confirmed to be correct. As an additional test, we established a TCP connection with

[3] https://frrouting.org
[4] https://github.com/LibtraceTeam/libtrace
[5] https://www.wireshark.org/docs/wsug_html_chunked/

netcat[6] via IPv4 and IPv6. Netcat is especially useful in this case because it allows a TCP connection to be initiated, which Spoki can then accept. This verification confirms that Spoki processes both IPv4 and IPv6 correctly. Despite the modifications, Spoki retains its existing functionalities.

Robustness test. To test for the robustness of Spoki, we sent syntactically incorrect packets and evaluated the responses. This included IP packets with a wrong version number, packets with a wrong payload length, TCP packets with an incorrect window size, packets with an incorrect checksum, packets with incorrect entries in extension headers, and packets with a destination port of zero (reserved port). Since Spoki only checks whether the packets are SYN or ACK and performs no other checks, it does not crash and responds. The only exception occurs with test packets that simultaneously set the SYN, ACK, and FIN flags. These were ignored by Spoki, as these flags are not part of the TCP handshake.
In the future, it might be useful to check for syntactical errors, log them if they occur, but not respond to them.

Performance test. For the performance test of Spoki traffic (TCP SYN packets) was generated using ZMap and sent to usable and unique IPv6 addresses found in RFC 4193 [23]. Two virtual Ethernet interfaces are created. Both requiring free addresses as interfaces for communication between the packet source and Spoki. The packet rate was gradually increased, and the number of packets processed was monitored to check for any delays. If delays are detected, threads are employed to parallelize the work. Subsequently, it is checked whether processing can handle double the number of packets. If delays arise again, additional threads are utilized.

Spoki consists of three components: Ingestion, Core, and Logging. All components are tested individually. The test starts with the Ingestion component, followed by the Core component, and finally, the performance of the Logging component is measured. In the Ingestion component, packets are read from the interface and converted into a data-representative form. In the Core component, decisions are made on how Spoki handles the incoming packets. The packets are either discarded or processed further. Relevant packets include TCP SYN and TCP ACK packets. In the Logging component, these packets and attributes that provide information about the behavior of the counterpart are stored for further use. The packets are forwarded to a pool of shards. The shards

[6] https://www.commandlinux.com/man-page/man1/netcat.1.html

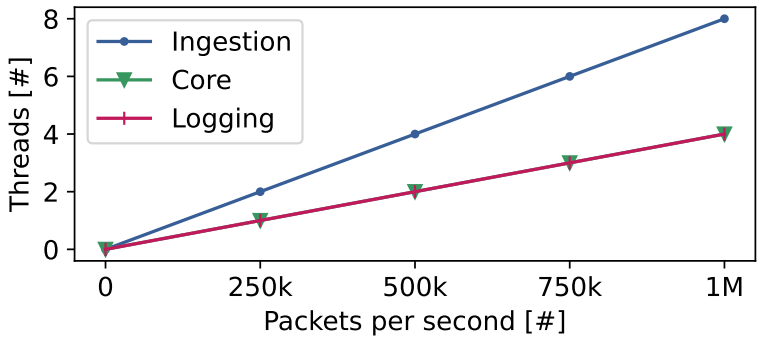


Figure 4.2: Spoki performance test in IPv6.

decide which targets should be probed and forward the data, so that the responses are prepared and sent next. We observe how often the components need to be replicated to distribute the workload. We compare the results with previous results obtained in IPv4. Figure 4.2 shows the results of the performance test in IPv6.

Ingestion. At a rate of 260k packets per second, the rate of processed packets starts to fluctuate. At 270k packets, Spoki can only process about 250k packets. Therefore, the performance of Spoki remains constant at a rate of 250k packets per second, maintaining the same performance results as observed in the previous Spoki tests conducted in IPv4. With two threads, Spoki is able to process twice as many packets, meaning it can handle 500k packets per second. This performance can scale to higher packet rates, so with four threads, Spoki can process one million packets per second. The results from the Ingestion component match Spoki's previous performance.

Core. In this measurement, Spoki is also able to process 250k packets per second. It is also possible to scale the processing up to one million packets per second by distributing the load across four threads.

Logging. For logging, twice the number of threads is required to scale the packet processing to one million packets per second, because the shards are additionally stressed as they also need to send packets.

With this expansion, Spoki is not significantly stressed and maintains the same performance results as those observed in the previous IPv4 tests.

5 Network Traffic Overview

This Section provides an overview of the traffic of all four network telescopes. Since T1 was split after the first 12 weeks and T2 began responding to TCP SYN packets shortly afterward, the measurement period is divided into two time periods for analysis. This helps to better understand the impact of telescope characteristics on IPv6 scanner reactions. During the first period (August 24, 2023 - November 21, 2023), the announcements remained stable, before T1 was split in the second period (November 22, 2023 - July 2, 2024). The characteristics of the other telescopes should also be considered, as they could also impact the observed scanning behavior. An overview of the telescope properties can be found in Section 4.

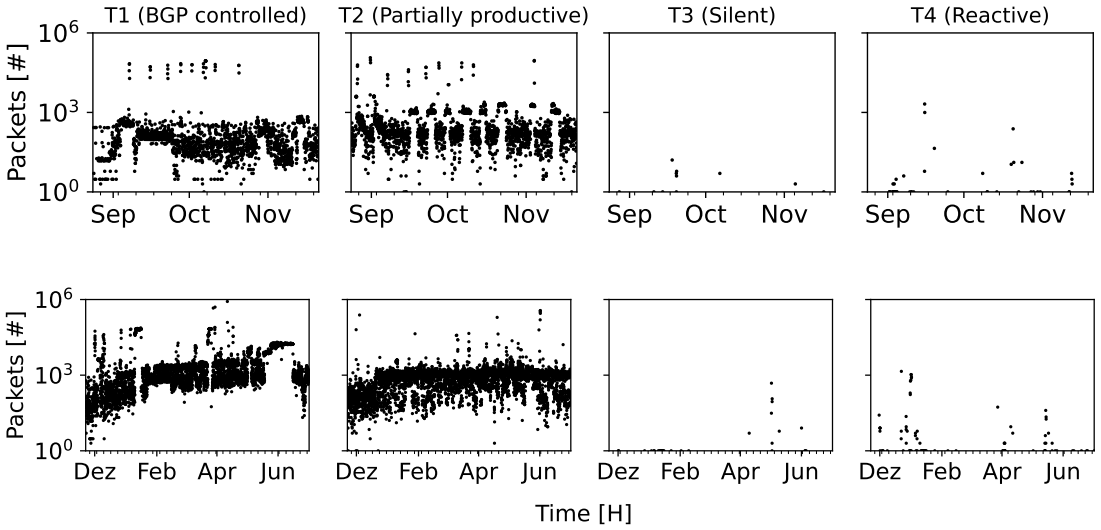


Figure 5.1: Traffic overview of all telescopes before (top row) and during (bottom row) the split period.

Figure 5.1 shows more traffic observed for the two telescopes with BGP announcements compared to the /48 networks without announcements (T3 and T4). T1 (BGP controlled) and T2 (Partially productive) stand out due to the significantly higher volume of packets compared to T3 (Silent) and T4 (Reactive). This could indicate that BGP announcements can significantly affect the number of packets received.
When comparing the time frames before and during the split period, a clear continuous upward trend in packet reception is particularly noticeable in T1. This could indicate that BGP announcements lead to increased network activity. In T2, a rise in the number of received packets is also observed after activating the reactive telescope Spoki (De-

Table 5.1: Comparison of telescopes: Sources, ASNs, and targets (before and during the split period).

	Before the split period				During the split period			
	T1	T2	T3	T4	T1	T2	T3	T4
Src /128 [#]	1387	6648	7	253	11 056	21 518	13	27
Src /64 [#]	1200	2149	6	251	9913	15 031	13	24
Src /48 [#]	1167	1630	6	250	9448	6823	11	19
ASN [#]	417	481	6	9	1766	790	9	13
Target [#]	796 444	714 220	20	1817	35 616 812	2 833 623	288	5146

cember 19, 2023). Further, detailed analyses and observations will be explained in the following sections.

Table 5.1 shows that T3 and T4 receive significantly less traffic from fewer sources compared to T1 and T2. In contrast to the other telescopes, T2 observes many /128 addresses from only a few /64 networks. T4 receives one packet from each of 240 source addresses, which all share the same /28 network. Whether these source IP addresses belong to a single entity cannot be determined. In the further analyses, these sources will be examined in more detail.

A first analysis shows that T1 observes the highest number of its targeted addresses during the split period. By splitting the /32 and announcing more-specific prefixes, scanners may probe more prefixes, covering a larger area of the address space. As shown in Figure 5.1, BGP announcements likely attract IPv6 scanners, resulting in increased traffic across the entire /32. Scanners could also target more addresses when scanning multiple specific prefixes compared to a single /32, as they are then limited to scanning smaller areas of the /32 prefix.

5.1 Transport Protocol and Ports

This subsection presents the distribution of network protocols and destination ports used by scanners, comparing all four network telescopes. It provides an overview of the initial observations and the information that can be derived from these packets to analyze the scanning behavior.

In the following table analyses, e.g., Table 5.2, the sum of the protocol percentages per telescope can exceed 100%, because the percentages represent the share of source

Table 5.2: Comparison of telescopes: Sources per transport protocol (before and during the split period).

	Before the split period								During the split period							
	T1		T2		T3		T4		T1		T2		T3		T4	
Protocol	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
ICMPv6	1114	80.3%	4114	61.9%	7	100%	246	97.2%	8253	74.6%	8604	40.0%	7	53.8%	16	59.3%
TCP	37	2.7%	5354	80.5%	0	0%	6	2.4%	641	5.8%	15 970	74.2%	2	15.4%	2	7.4%
UDP	265	19.1%	1768	26.6%	0	0%	1	0.4%	4180	37.8%	2850	13.2%	4	30.8%	11	40.7%

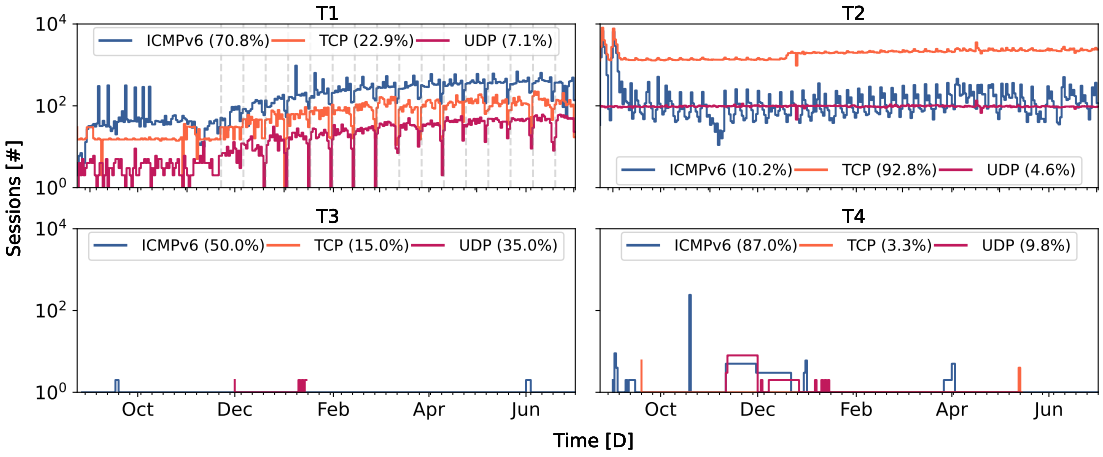


Figure 5.2: Protocol use across sessions per day.

IP addresses per protocol. Sources sometimes use multiple protocols during the measurement period. Therefore, the total percentage is often higher than 100%. The same applies to representations where the percentage of sessions is shown.

Most sources in T1, T3, and T4 use ICMPv6, while TCP is dominant in T2, as shown in Table 5.2. Additionally, there is a noticeable increase in sources sending UDP packets during the split period in T1. In addition, T3 and T4 detect significantly fewer packets than T1 and T2 in both periods. The packets from the 240 sources in T4, which share the same /28 network, are all ICMPv6 packets.
For T1, the number of sessions per protocol increases during the prefix split period. See Figure 5.2. At T2, the number of TCP sessions increases slightly increases when answering TCP requests (December 19, 2023). Overall, the number of sessions per day for TCP and UDP protocols at T2 remains constant throughout the measurement period, while the number of ICMPv6 sessions fluctuates significantly within short intervals. As expected, the plot for T3 shows very few packets. In T4, sources within the same /28

network send multiple packets in a single day, strongly suggesting that they originate from the same entity.

Figure 5.3 shows the top 5 results with the most sessions per port before the split, and Figure 5.4 shows the top 5 results with the most sessions per port during the split period. At T1, TCP destination port 80 (HTTP) and UDP destination ports in the Traceroute range (33434–33523) are most frequently seen, while at T2, a greater variety of destination ports is observed. In addition to the ports 80 and 443 for TCP, the most common UDP destination ports in T2 are 161 (SNMP), 500 (Internet Security Association and Key Management Protocol), 53 (DNS), 123 (NTP), and 3478 (Session Traversal Utilities for NAT). Up to this point, T3 only receives ICMPv6 packets and in T4 only TCP packets with port 443 and UDP packets with port 80 or ports within the Traceroute range are received.

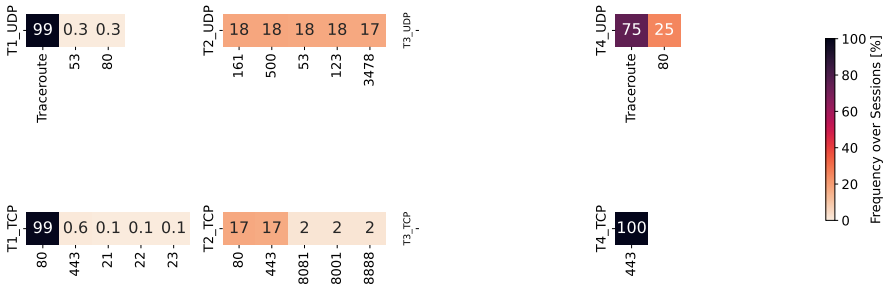


Figure 5.3: Top 5 ports: Sessions per port before the split.

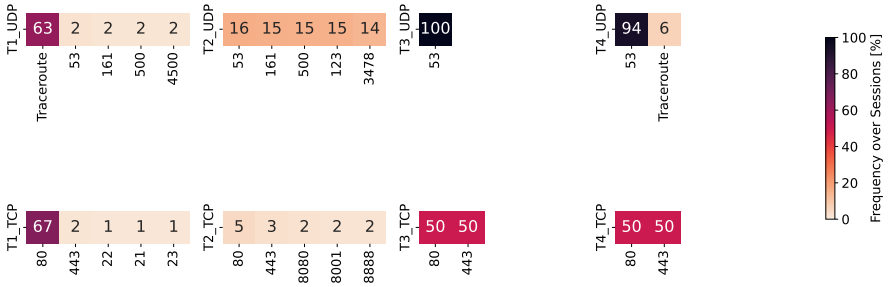


Figure 5.4: Top 5 ports: Sessions per port during the split.

In addition to Traceroute, port 53 (DNS) is frequently found among the UDP ports in the telescopes during the split period. This is shown in Figure 5.4. Another TCP destination port that is observed more regularly next to port 80 is 443 (HTTPS). Furthermore, It is evident that in T2, TCP destination ports 80 and 443 are no longer as dominant as

before, but rather there is a wide distribution of TCP destination port usage. The reason for this lies in the properties of the T2, which will be examined in more detail.

5.2 The Origin of Packets

To gain a better understanding of the origin of the received packets, we will analyze them. First, we collect the ASNs of the received packets. The ASNs are identified using pyasn[7]. The Table 5.3 shows the registered network types from PeeringDB[8]. The analysis reveals that most of the source addresses in T1 come from two entities: (*i*) RIPE-Atlas (55%) and (*ii*) AlphaStrike (33%). A total of 94% of the ASNs in T1 (and 74% of the ASNs in T2) can be attributed to the RIPE Atlas probes based on the payloads within the packets. AlphaStrike is a cybersecurity company that uses many different sources for scanning, which artificially increases the number of unique sources we observe. Nevertheless, only 0.05% of the ASNs can be attributed to AlphaStrike. RIPE Atlas [38] is a distributed measurement infrastructure used by various stakeholders. Therefore, RIPE Atlas constitutes a large portion of the traffic. For clarity, we filter out the ASNs from RIPE Atlas and show the results separately in Table 5.4. Since we observe no RIPE Atlas probes in T3 and T4, we do not include them in this table.

In Table 5.3, the most frequently observed network type among the source ASNs is *Content*. *Network Service Providers* and *Cable/DSL/ISP* networks are also common.

Table 5.3: The origin of the packets: Network types of ASNs identified via PeeringDB without RIPE Atlas probes. The table is sorted by the total absolute number of network types across all telescopes.

	Before the split period								During the split period							
	T1		T2		T3		T4		T1		T2		T3		T4	
	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
Content	30	19%	35	16%	1	17%	4	44%	34	19%	44	16%	2	22%	3	23%
Cable/DSL/ISP	19	12%	32	14%	0	0%	1	11%	27	15%	48	17%	0	0%	0	0%
NSP	23	15%	33	15%	1	17%	2	22%	22	12%	39	14%	1	11%	1	8%
Educational/Research	17	11%	20	9%	1	17%	0	0%	19	11%	20	7%	0	0%	1	8%
Enterprise	3	2%	5	2%	0	0%	0	0%	4	2%	7	2%	0	0%	0	0%
Non-Profit	2	1%	2	1%	0	0%	0	0%	5	3%	4	1%	0	0%	0	0%
Network Services	1	1%	3	1%	0	0%	0	0%	1	1%	4	1%	0	0%	0	0%
Unknown	62	39%	92	41%	3	50%	2	22%	66	37%	116	41%	6	67%	8	62%

[7] https://catalog.caida.org/software/pyasn
[8] https://www.peeringdb.com

Table 5.4: The origin of the RIPE Atlas probes: Network types of ASNs identified via PeeringDB. The table is sorted by the total absolute number of network types across all telescopes.

	Before the split period				During the split period			
	T1		T2		T1		T2	
	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
Cable/DSL/ISP	87	30%	92	31%	396	24%	145	25%
NSP	45	16%	55	18%	279	17%	114	20%
Content	38	13%	31	10%	173	10%	56	10%
Educational/Research	17	6%	21	7%	135	8%	32	6%
Non-Profit	7	2%	11	4%	74	4%	27	5%
Enterprise	5	2%	5	2%	50	3%	20	4%
Network Services	4	1%	7	2%	33	2%	14	2%
Route Server	0	0%	1	<1%	3	<1%	0	0%
Route Collector	0	0%	0	0%	2	<1%	0	0%
Goverment	0	0%	1	<1%	2	<1%	1	<1%
Unknown	87	30%	74	25%	539	32%	164	29%

Researchers also appear significantly, while network types such as *Non-Profit*, *Enterprise*, and *Network Services* are less commonly encountered.

The RIPE Atlas probes are assigned to many different network types, as shown in Table 5.4. The most common type is *Cable/DSL/ISP* networks. However, the types *Route Server*, *Route Collector*, and *Government* can only be assigned to the ASNs from RIPE Atlas.

In addition, the geographical locations of the source IP addresses are determined using MaxMind[9]. We collect all distinct source IP addresses and aggregate the corresponding geolocations. The results for the period before the split are shown in Table 5.5, while the results for the period during the split are displayed in Table 5.6. Since 55% of the source IP addresses in T1 belong to ASNs that can be assigned to RIPE Atlas probes, these are specifically displayed in Table 5.7.

Source IP addresses from China and Germany are most frequently observed before the split period (Table 5.5). Addresses from the United States are also commonly found in T1 and T2. The other geolocations listed in the tables represent the top 5 telescope geolocations. Compared to T1, T3, and T4, the geolocations of the source IP addresses of T2 are more widely distributed.

[9]https://www.maxmind.com/en/solutions/ip-geolocation-databases-api-services

Table 5.5: MaxMind geolocations of source IP addresses: Top 5 per network telescope without RIPE Atlas probes before the split period. The table is sorted by the total absolute number of geolocations across all telescopes.

	T1		T2		T3		T4	
	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
China	21	2%	928	15%	3	43%	2	1%
Germany	290	33%	65	1%	0	0%	241	95%
United States	136	16%	154	6%	1	14%	0	0%
Singapore	22	3%	101	2%	0	0%	1	0%
United Kingdom	37	4%	68	1%	0	0%	0	0%
Japan	31	4%	53	1%	0	0%	0	0%
Canada	33	4%	30	0%	1	14%	2	1%
Hong Kong	5	1%	34	1%	1	14%	0	0%
Austria	10	1%	12	0%	1	14%	1	0%
Kazakhstan	1	0%	1	0%	0	0%	5	2%

Table 5.6: MaxMind geolocations of source IP addresses: Top 5 per network telescope without RIPE Atlas probes during the split period. The table is sorted by the total absolute number of geolocations across all telescopes.

	T1		T2		T3		T4	
	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
Japan	28	1%	9674	48%	0	0%	0	0%
China	50	1%	3807	19%	4	31%	8	30%
Germany	3678	77%	176	1%	0	0%	0	0%
United States	550	11%	1815	9%	4	31%	8	30%
United Kingdom	42	1%	806	4%	0	0%	1	4%
Canada	74	2%	32	1%	1	8%	1	4%
Singapore	35	1%	34	1%	0	0%	1	4%
South Korea	13	0%	25	0%	2	8%	5	19%
Brazil	16	0%	25	0%	0	0%	1	4%
Austria	8	0%	9	0%	1	8%	1	4%

In T1, Germany and the United States remain the most frequent geolocations during the split period, as shown in Table 5.6. In T2, the frequency of source IP addresses from Japan suddenly increases during this period. The reason for this is further investigated in Section 7.

A majority of RIPE Atlas probes send from Germany and the United States. See Table 5.7. However, other geolocations like France, the Netherlands, and the United Kingdom also appear in the top 5 due to Table 5.6. The geolocations of the RIPE Atlas probes have a noticeable effect on the overall results. Nevertheless, it is important to note that these probes make up 65% of the source addresses in T1, but only 7% in T2. Therefore, their impact is much smaller in T2 than in T1.

Table 5.7: MaxMind geolocations of source IP addresses: Top 5 per network telescope using RIPE Atlas probes only. The table is sorted by the total absolute number of geolocations across all telescopes.

	Before the split				During the split			
	T1		T2		T1		T2	
	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
Germany	1496	23%	342	20%	1432	23%	246	20%
United States	829	13%	227	13%	797	13%	178	14%
France	662	10%	227	13%	648	10%	167	13%
Netherlands	321	5%	95	5%	319	5%	66	5%
United Kingdom	277	4%	66	4%	265	4%	45	4%

Analyzing the factors influencing packet origins. Determining the intentions of scanners is generally a challenging task in scan analysis. Overall, the percentages of T1 and T2 in the network types do not differ significantly. However, for geolocations, there is a greater difference in the telescopes. It is possible that telescope properties influence where we receive packets from. In part, the use of Spoki, as well as the address with the DNS entry, could result in us observing different scanners. T1, with more frequent announcements, might attract different scanners, whereas T2, due to the long period it has already been announced, might not capture these scanners during the measurement period. The impact of the DNS entry will be examined in more detail, among other aspects, in Section 7.

6 A Taxonomy for Classifying Scan Behavior

To analyze and compare scanning behavior, several key factors are considered: *address selection*, *temporal behavior*, and *network selection*. *Address selection* focuses on how target addresses are generated per session. *Temporal behavior* captures the timing of a source's sessions over the measurement period. *Network selection* shows the range of prefix coverage within a source's sessions.

6.1 Address Selection

There are already research studies that classify the *address selection* strategies of scanners, particularly between *random* and *non-random* generation. Tanveer *et al.* [44]

identify *random* and *low-byte* Scanning as dominant strategies. *Low-byte* scanning suggesting a *structured* generation of target addresses. Richter *et al.* [37] used the Hamming weight to assess the randomness of 1-bits in the IID. In this work, we use the test suite from the National Institute of Standards and Technology (NIST) [3]. These tests do not provide absolute certainty but offer strong indicators of randomness.

A test calculates a p-value. A significance level of α is used. This is the recommended value from the documentation [3]. If the p-value is ≥ 0.01, the sequence would be considered to be random with a confidence level of 99%. Conversely, a p-value < 0.01 would lead to the conclusion that the sequence is non-random, also with a confidence level of 99%.

Out of the 15 tests in the NIST test suite, the frequency Test (Monobit test) is applied here. The frequency test checks the balance between ones and zeros in a sequence. If this test fails, it strongly indicates non-random generation.
We define *structured*, *randomized* and *unknown* destination addresses like this:

Structured. A *structured* destination address is an address that follows a recognizable pattern or rule. For example, it could have specific parts that repeat or are arranged in a way that makes it predictable.

Random. A *random* destination address is one that does not follow any predictable pattern. The address is assigned without any specific order or rule. This classification is determined by the frequency test. While randomness cannot be fully proven, it can be excluded if a pattern or predictable behavior is observed. This should be taken into account when analyzing the data.

Unknown. Only sessions with at least 100 packets are analyzed using statistical tests to classify them as *structured* or *random*. For this, the test suite by the National Institute of Standards and Technology (NIST) is used to assess the randomness of target generation. Next, address structures of the destination addresses in sessions with fewer than 100 packets are examined. If only structures like *low-byte*, *embedded-port*, *embedded-IPv4*, etc., then these sessions are also classified as *structured*. The rest is categorized as *unknown*.

6.2 Temporal Behavior

The *temporal behavior* describes the timing of received sessions from a source observed over the measurement period. This behavior can be influenced by the scanners' own schedules or by external events.

Oneoff. If a scan source was only observed for a single session in total within the measurement period, this is referred to as an *oneoff* scanner in this work. In other words, the scanner appears once and then disappears again.

Periodic. A scan source is classified as *periodic* if the time interval between all scan sessions for that source is roughly the same. A scanner in this category must be assigned at least two sessions. For two sessions, the time between the last session and the end of the measurement period must be shorter than the period between the scan sessions. This is because a third session could still occur at the same interval, but would not be visible after the measurement period ends. Period detection is determined using autocorrelation [6].

Intermittent. Unlike *periodic* scanners, *intermittent* scanners do not have a consistent time interval between sessions. An *intermittent* scanner must have at least two sessions, but the interval to a potential third session is non-periodic and may extend beyond the measurement period.

6.3 Network Selection

In the large IPv6 address space, scanners need effective strategies to cover network areas as effectively as possible. To do this, they can obtain information about advertised BGP prefixes. In order to examine the prefix coverage procedure in more detail, four classifications are defined. In this way, the reaction to BGP announcements in particular can also be examined more closely. The classification is based on the density-based clustering algorithm *DBSCAN*. All analyses of *network selection* in this work are only conducted during the split period and are applied only to T1. Before that, it is always a *single-prefix* scan, which can be identified. Additionally, the distribution of packets per prefix is not examined, but rather the number of sessions. This is because, within the taxonomy, the sessions of the scanners are consistently evaluated in all categorizations.

Network-size dependent. If a scanner works *network-size dependent*, it varies the number of sessions based on the network size. This means that they target *more-specific prefixes* with significantly fewer scan sessions. This could happen if a scanner starts a coarse-grained scan and only catches *more-specific prefixes* with fewer scan sessions.

Network-size independent. If a scanner operates *independently* of the network size, the number of scanning sessions per network remains (almost) the same. This can be tested using the T1 experiment by examining how many sessions occur per announcement. If nearly the same number of sessions are received across all announced prefixes, despite their different prefix sizes, the scanning source is classified in this category.

Single-prefix scanning. This type of scanner operates by scanning a single prefix during each announcement period. The specific prefix can vary with each period and can be any of the available prefixes.

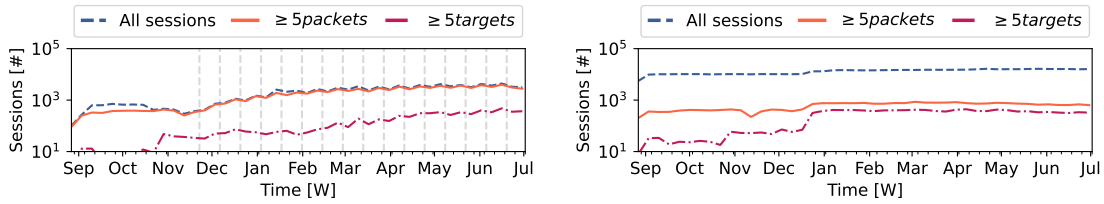
Inconsistent. A scanner that shows *inconsistent* behavior during announcement periods changes the way it operates over time or does not have a clear pattern in how sessions are spread across prefixes.

7 Impact of Telescope Properties on Scan Behavior: A Cross-Telescope Analysis

In this section, the scanning behavior of IPv6 scanners is analyzed based on the observed packets. Different scanning patterns appear across all telescopes, and these distinct characteristics are explored in more detail. Additionally, notable findings from Section 5 are examined further to gain a deeper understanding of the scanning behavior.

7.1 A Closer Look at the Network Telescope Sessions

The analysis of packets per session and targets per session for T1 and T2 indicates unequal scanning behaviors. Figure 7.1 shows that sources in T1 often generate sessions with at least five packets per session, but these packets typically target fewer than five targets. The line representing the number of targets is much lower than the one for the



(a) T1.

(b) T2 without ACK pakets to Spoki.

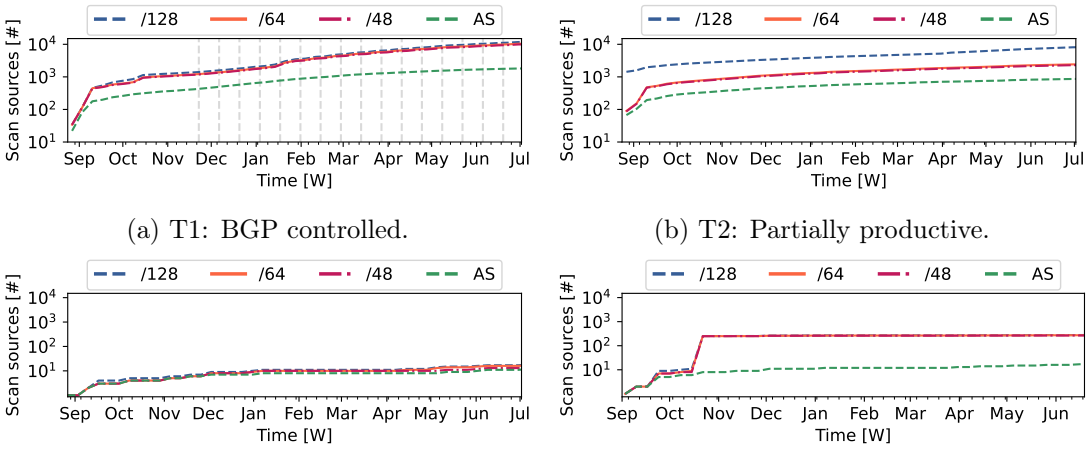
Figure 7.1: Number of sessions per week for all sessions, sessions with ≥ 5 packets, and sessions with ≥ 5 targets.

packet count, but it gradually approaches the packet count line towards the end of the measurement period. In contrast, T2 shows a different pattern. Often, fewer than five packets are sent per session. The number of sessions with at least five targets increases significantly when Spoki is activated. A small portion still sends packets to fewer than five targets per session. When Spoki is activated and accepts a TCP connection, this can quickly lead to an increase in the number of packets per session. Therefore, for this analysis, ACK packets sent back to Spoki are filtered out. T3 and T4 have no separate figures due to the overall low number of packets observed in those cases. In T3, 40 sessions were observed, of which 13 sessions had packet counts ranging from 2 to 256. The remaining 27 sessions, however, consisted of only a single packet each. In T4, over 350 sessions out of 368 contain fewer than 10 packets per session. The largest session includes 3,070 packets.

7.2 Exploring the Effects of Aggregation Levels on Scan Source Identification

Different scanning strategies can already be observed in Figure 7.2. In T1, there are only slight differences between the aggregation levels, while in T2, these differences are much more pronounced. A similar pattern is seen in T3 and T4. In T3, all aggregation levels are almost at the same level, whereas T4 shows a significant difference between the aggregation levels compared to the ASes. From T2, it can already be inferred that many /128 scan sources can be aggregated into a few /64 scan sources. In T4, 240 sources can be aggregated into a /28 scan source.

When analyzing the number of sources compared at different aggregation levels, a significant difference appears in T2 between the number of /128 and /64 sources. Many /128



(a) T1: BGP controlled.

(b) T2: Partially productive.

(c) T3: Silent.

(d) T4: Reactive.

Figure 7.2: Comparison of number of scan sources and ASNs.

Table 7.1: Impact of aggregation levels within T2: Sources per transport protocol, split by time period (before and during the split period) and filtered by whether the packets are associated with AS6939 (only AS6939) or not associated with AS6939 (w/o AS6939).

	Before the split period				During the split period			
	w/o AS6939		only AS6939		w/o AS6939		only AS6939	
Protocol	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
ICMPv6	4100	88%	14	1%	8576	49%	28	1%
TCP	3383	73%	1971	99%	11 865	68%	4105	99%
UDP	301	6%	1467	73%	733	4%	2117	1%

addresses come from just a few /64 sources. AS6939 (*Hurricane Electric LLC*) stands out because 2.7k /128 addresses originate from a single /64 source. Similar patterns are also seen with three other /64 sources from this AS.

This section examines the impact of these /128 sources on the previously presented results and considers the importance of aggregation level choice in presenting the findings. For the analysis, all packets from AS6939 are filtered to get the results for the column *only AS6939*. In contrast, all other packets are excluded to get the results for *w/o AS6939*. The first comparison shows how much the number of /128 sources from AS6939 affects the total number of observed /128 sources. The results are shown in Table 7.1

Table 7.2: Impact of aggregation levels within T2: Top 10 UDP ports by sessions, split by time period and filtered by whether the packets are associated with AS6939 (only AS6939) or not associated with AS6939 (w/o AS6939).

	Before the split period				During the split period			
Rank	Port	w/o AS6939	Port	only AS6939	Port	w/o AS6939	Port	only AS6939
#1	2152	58	3478	1620	53	498	53	4017
#2	2123	54	53	1617	161	374	161	4010
#3	500	41	161	1616	500	368	123	4000
#4	5353	40	123	1616	4500	357	3478	3998
#5	161	40	500	1614	137	349	500	3987
#6	33434	35	33439	1	123	348	33449	3
#7	33439	31	33435	1	1900	342	33444	2
#8	33436	30	33441	1	1434	308	33448	2
#9	33435	30	33434	1	520	307	33445	1
#10	33441	30	33447	1	69	307	33437	1

Table 7.3: Impact of aggregation levels within T2: Top 10 TCP ports by sessions, split by time period and filtered by whether the packets are associated with AS6939 (only AS6939) or not associated with AS6939 (w/o AS6939).

	Before the split period				During the split period			
Rank	Port	w/o AS6939	Port	only AS6939	Port	w/o AS6939	Port	only AS6939
#1	80	31 479	8081	3238	80	24 595	8001	14 356
#2	443	30 059	8001	3236	443	12 124	8090	14 344
#3	22	35	8888	3234	21	1201	8888	14 337
#4	23	26	8080	3232	8080	1083	8080	14 320
#5	21	25	8090	3228	22	452	8081	14 285
#6	465	25	443	1672	23	387	1080	10 598
#7	5001	24	80	1671	1723	382	445	10 590
#8	6001	24	2096	1635	3389	378	25	7379
#9	2082	24	1080	1633	445	373	427	7370
#10	8181	24	179	1622	993	367	1723	7368

The total number of UDP packets per source is strongly influenced by the number of sources from AS6939. If an aggregation level of /64 is chosen, the number of sources from which UDP packets are received is significantly lower compared to other protocols. For T2, TCP is a frequently observed protocol overall, while ICMPv6 is scanned less often by this AS. Another effect appears in the port distribution. Table 7.2 lists the ten UDP ports with the largest number of sessions per port. The number of sessions that can be assigned by the AS6939 is compared with the number of sessions of the other ASNs. The top 5 ports in Figures 5.3 and 5.4 in Section 5 are primarily influenced by source addresses of AS6939, as shown in Table 7.2. In contrast, UDP ports from other sources are more evenly spread, with no single port being notably frequent.

It is different with the TCP ports, as shown in Table 7.3. Ports 80 and 443 are used most often by sessions not from AS6939, while other ports are less common. The TCP ports of sessions from AS6939 have a smaller impact on the top 2 TCP ports. However, as the number decreases significantly from rank 3 onward, their influence becomes stronger.

The results of the geolocations are also examined, but many of these sources from AS6939 do not have geolocations found through MaxMind. As a result, no noticeable influence on the geolocation results can be detected. However, for the protocols and ports, there is a strong influence.
When analyzing scanning behavior, it is important to consider that the results can vary depending on the level of aggregation used. It makes sense to compare the results by applying different levels of aggregation in the analysis.

7.3 Impact of DNS Entry

To provide an overview of how many packets are sent to the IP address with a DNS entry, a comparison of the traffic in T2 is presented. In Table 7.4, the sources are aggregated at the /128, /64, and /48 levels. Additionally, the table shows the number of distinct ASNs and the number of sessions for each protocol. For better comparability with previous results, the traffic is divided into two time periods. The table is also divided into two sections. In the *non-DNS* section, all packets are considered except those sent to the destination address with a DNS entry. In the *DNS* section, only packets received by this destination address are considered. Throughout the entire measurement period, there are only five source IP addresses that scan this destination address at least once, along with others. The remaining source IP addresses scan either exclusively this address or only others.

Before the split period, more /128 and /64 sources scan the destination address with the DNS entry than the rest of the destination addresses. Only the number of ASNs and /64 sources remains lower than for the other sources. In addition, there are many TCP and ICMPv6 sessions scanning the DNS address, and zero sessions using the UDP protocol. In the second period, the number of sources sending packets to the DNS address increases significantly for all three source aggregations compared to the number of other sources. Overall, however, the number of sessions decreases, even though the split period is longer than the period before.

Table 7.4: An overview of the impact of the DNS entry on scan traffic within T2, split by time periods and filtered based on whether the packets target the address with the DNS entry (DNS) or not (non-DNS).

	Before the split period		During the split period	
	non-DNS	DNS	non-DNS	DNS
/128 Source addr.	3245	3403	6871	14 651
/64 Source addr.	1066	1091	1943	13 106
/48 Source addr.	1006	637	1835	5019
ASN	429	85	724	122
Sessions per ICMPv6	4959	31 142	10 977	20 744
Sessions per TCP	118 032	30 352	441 250	24 880
Sessions per UDP	8672	0	21 738	63

Table 7.5: Impact of DNS entry on TCP port usage by sessions within T2, split by time periods and filtered based on whether the packets target the address with the DNS entry (DNS) or not (non-DNS).

	Before the split period				During the split period			
	non-DNS		DNS		non-DNS		DNS	
Rank	Port	Sessions [#]	Port	Sessions [#]	Port	Sessions [#]	Port	Sessions [#]
#1	8081	3262	443	30 071	8080	14 677	80	20 048
#2	8001	3258	80	30 070	8001	14 412	443	11 688
#3	8888	3257	22	9	8090	14 395	8080	726
#4	8080	3255	8080	1	8888	14 391	22	56
#5	8090	3252	5001	1	8081	14 339	20257	23

Nevertheless, 63 UDP sessions are observed this time. These observations illustrate the strong influence of the address with the DNS entry on the observed traffic.

An analysis of the TCP destination ports in Table 7.5 reveals that traffic to the DNS address significantly impacts the frequency of TCP destination ports 80 and 443. In contrast, when examining the traffic without the DNS address, the TCP port 8081 is present in most sessions with a frequency of 2.5%. It seems like a more even distribution of TCP ports across all sessions, with no few ports standing out with a high amount.

The analysis of the UDP destination ports in Table 7.6 shows that they are only minimally affected due to the low number of UDP packets sent to the DNS address.

Table 7.7 shows that the top 5 source geolocations are also influenced by traffic to the DNS address. In both time periods, the United States, China, Singapore, and the United Kingdom appear in the top 5. The number of sources for traffic related to the DNS address is significantly higher overall. Many sources from Japan during the split period are

Table 7.6: The impact of DNS entry on the number of sessions per UDP port within T2, split by time periods and filtered based on whether the packets target the address with the DNS entry (DNS) or not (non-DNS).

	Before the split period				During the split period			
	non-DNS		DNS		non-DNS		DNS	
Rank	Port	Sessions [#]	Port	Sessions [#]	Port	Sessions [#]	Port	Sessions [#]
#1	161	1656	-	-	53	4512	4500	10
#2	500	1655	-	-	161	4377	500	9
#3	53	1644	-	-	500	4346	443	8
#4	123	1640	-	-	123	4346	161	7
#5	3478	1626	-	-	3478	4013	53	6

Table 7.7: The impact of DNS within T2: Geolocations (MaxMind) from source IP addresses, split by time periods and filtered based on whether the packets target the address with the DNS entry (DNS) or not (non-DNS).

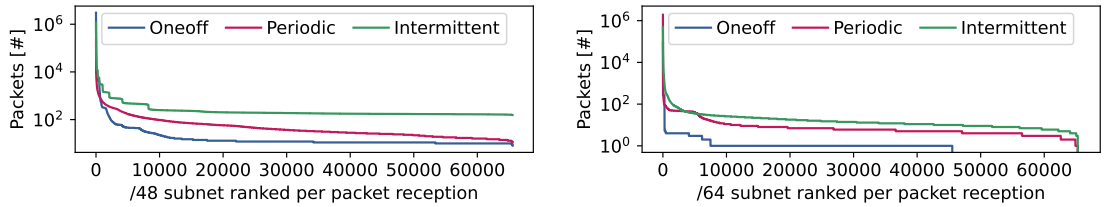
	Before the split period				During the split period			
	non-DNS		DNS		non-DNS		DNS	
Rank	Geo.	Src. [#]	Geo.	Src. [#]	Geo.	Src. [#]	Geo.	Src. [#]
#1	US	154	US	2276	US	1711	Japan	9644
#2	Germany	63	China	892	Germany	145	China	3737
#3	China	36	Singapore	65	China	71	UK	764
#4	Singapore	36	UK	36	UK	42	Singapore	130
#5	UK	32	Hong Kong	29	Singapore	34	US	106

due to traffic to the DNS address. This has a significant impact, as Japan ranks first in the overall analysis because of it.

It can be summarized that a DNS entry can have a significant impact on the received traffic. The related work also shows that DNS entries seem to be efficient attractors. Therefore, when comparing the telescopes, this influence should be considered as well.

7.4 Subnet Coverage

Some scanners scan a wide range of subnets, while others focus on specific subnets. To better understand these behaviors, we analyze the distribution of destination addresses.



(a) /48 subnets for T1.

(b) /64 subnets for T2.

Figure 7.3: Packets categorized by address selection and distributed across the subnets within T1 and T2.

Figure 7.3 illustrates the distribution of destination addresses across the subnets. The analysis focuses on whether the scan sessions are spread evenly across multiple subnets or concentrated on just a few. The destination addresses are sorted from left (many) to right (few) based on the number of received packets per subnet.

T1 (/32) examines the /48 subnets, while T2 (/48) focuses on the distribution across the /64 subnets. Since T2, T3 and T4 have a prefix size of /48 containing 65,536 /64 subnets, analyzing all /64 subnets within the /32 of T1 would result in over 4B subnets. Therefore, for T1, we focus not on all /64 subnets but on the 65,536 /48 subnets, which provides a more manageable scope for the analysis. The packets per subnet are categorized based on the *address selection*. In T1, certain subnets are scanned particularly frequently across all categories. *Intermittent* scanners display a relatively even distribution of packets across the remaining /48 subnets. In contrast, *periodic* scanners show a steady increase along the x-axis, indicating a stronger focus on specific /48 subnets, while others receive significantly fewer packets, resulting in a highly uneven distribution. *Oneoff* scanners have the lowest average number of packets per subnet. The first 10k subnets (left) show a sharp decline in the number of packets per subnet for each category. Most packets from *periodic* and *intermittent* scanners are received in the *0000* subnet, while the majority of packets from *oneoff* scanners are received in the *e000* subnet.
For T2, the line for the *intermittent* scanners demonstrates a continuous increase, whereas the *periodic* scanners show some stepwise increases. Not all /64 subnets receive packets from the *oneoff* scanners, while many subnets have a similarly low number of received packets. Across all categorizations, the majority of packets are always received in the *0000* subnet.

T3 and T4 have only a few /64 subnets that receive packets, which is why no figures are

created. In T3, the subnet *0001* is predominantly scanned, while in T4, the subnet *0000* receives the most packets. In both cases, fewer than 1k /64 subnets receive any packets at all.

7.5 Target Address Generation

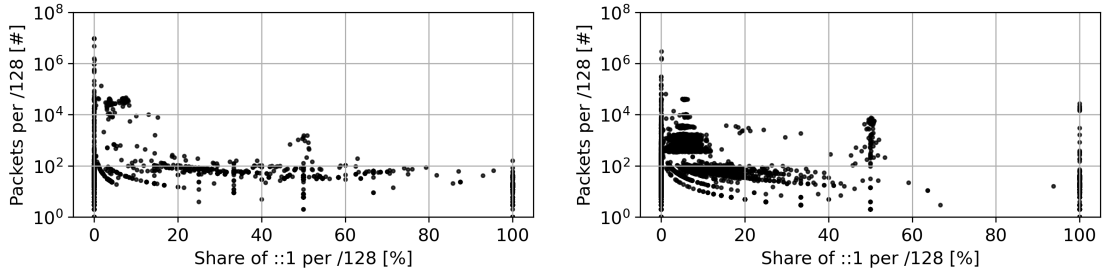
The received destination addresses are categorized based on the address types defined in RFC 7707 [17], as previously described in Section 3. Additionally, destination addresses ending with *::0* are observed and analyzed separately due to their unique structure. The *pattern bytes* category is further clarified and labeled as *3x zero bytes*. For each category, the number of source IP addresses (/128) is calculated. The results of this analysis are presented in Table 7.8.

Most source IP addresses send at least one packet to a *low-byte* address. For T1 and T2, the most common source IPs scanning *low-byte*, *subnet-router-anycast* (*::0*), or unclassifiable addresses (*randomized*). For T4, only a single source IP is observed scanning a *subnet-router-anycast address*. Otherwise, there are no significant findings for T3 and T4, except for a higher proportion of source IPs scanning *low-byte* addresses. Since many sources seem to favor scanning these addresses, a closer look is taken to determine whether this strategy is mostly used by sources sending many or few packets. Figure 7.4 compares the total number of packets per source IP to the number of scanned *low-byte* addresses (with only *::1* appended after the fixed prefix).

Table 7.8: Comparison of telescopes: Interface identifire address types by sources, sorted by quantity.

	Before the split period								During the split period							
	T1		T2		T3		T4		T1		T2		T3		T4	
Address Type	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]	[#]	[%]
low-byte	692	50%	6076	91%	4	57%	89	35%	8433	76%	20 253	94%	3	23%	14	52%
::0	461	33%	471	7%	0	0%	1	0%	834	8%	1059	5%	0	0%	0	0%
randomized	238	17%	160	2%	3	43%	97	38%	1674	15%	367	2%	10	77%	14	52%
3x zero bytes[1]	14	0%	14	0%	0	0%	35	14%	473	4%	21	0%	0	0%	3	11%
embedded-ipv4	35	3%	12	0%	0	0%	34	13%	455	4%	19	0%	0	0%	5	19%
embedded-port	2	0%	12	0%	1	14%	1	0%	46	0%	39	0%	2	15%	1	4%
ieee-derived	2	0%	8	0%	0	0%	0	0%	11	0%	8	0%	0	0%	0	0%
isatap	0	0%	0	0%	0	0%	0	0%	2	0%	0	0%	0	0%	0	0%

[1]No other address type could be found, and it contains at least three zero bytes in its IPv6 IID.



(a) T1. (b) T2.

Figure 7.4: The ratio of packets sent to the low-byte address (ending with ::1) per source IP address (y-axis), compared to the percentage of packets sent to ::1 relative to all packets sent by the individual source IP address (x-axis), in T1 and T2.

The focus is on this *low-byte* address because it is scanned most frequently. As a result, the analysis does not consider all *low-byte* addresses and can instead focus specifically on one address, simplifying the investigation.

Many source IP addresses sending a large number of packets either do not scan this specific target address or only partially do so. For some sources, 50% of the scanned destination addresses are the *low-byte* address. Some sources exclusively send packets to the *low-byte* address. In T1, these sources send just over 100 packets at most, whereas in T2, some sources send over 10k packets exclusively to this *low-byte* address. In both figures, curves can be seen among the scatter points with percentage shares between 0 and 20%, although the analysis shows that these packets originate from multiple ASes, indicating that they likely come from different scanners. Despite this, the curves still show a structure as if they came from a single source. It is important to note that only this single *low-byte* address was analyzed, and the BGP Experiment of T1 as well as the activation of Spoki could have influenced the results.

7.6 Heavy Hitters

We define scanners as heavy hitters if they account for at least 10% of the total scan traffic. In T1, four heavy hitters are observed, three in T2, and two each in T3 and T4. Considering all scanners and packets across all telescopes, there are a total of three heavy hitters responsible for at least 10% of all packets received during the entire measurement period. These three heavy hitters were all observed in T1 and are analyzed in more detail

in this subsection. The analysis includes all sessions received over the entire measurement period. The scanners generate one or more sessions, which take place at different times. To better understand how scanners distribute their scan sessions within T1, we examine how many of the /48 subnets in T1 receive packets within sessions. Additionally, we provide an overview of how the number of sessions is distributed across the announced more-specific prefixes (/33 - /48), revealing insights into the *network selection*. Each heavy hitter sends over 9M packets.

AS53667 Ponynet. By the end of the measurement period, the scanner with the most packets originates from AS53667, sending between 143 and 159 packets to each /48 subnet of T1, with an average of 147 packets per subnet. AS53667, or *Ponynet*, is named by the U.S. cybersecurity company CUJO AI[10] as a malware distribution center. *CUJO AI* provides cybersecurity and device management solutions for network operators. *Ponynet* is part of *Frantech Solutions*, a company known for providing *bulletproof* hosting services. *Bulletproof* hosting providers allow customers to upload and distribute various illegal content, including phishing campaigns and malware. *BuyVM*, owned by *Frantech Solutions*, operates data centers in the USA and is linked to many attacks. This scanner scans with a total of 4 sessions at the beginning of the year 2024. During the analysis of the scanner, DNS requests are observed within the payloads. Apparently, these DNS requests are sent to around 150 randomly selected destination addresses. No structure can be detected in the IIDs of the destination addresses. Only the /48 subnet area seems to be specifically generated by an algorithm. It could be a scan intended to explore the network and monitor for any responses to the DNS requests.

AS12816 Leibniz supercomputing centre. The scanner with the second most packets observed is assigned to AS12816 and *the Leibniz supercomputing centre* as the AS organization. This scanner sends packets to over 9M destination addresses, including 200k to already scanned targets. However, the scanner only scans 4096 /48 subnets within T1. As a result, 61k /48 subnets do not receive any packets from this scanner. The /48 subnets that do receive packets get different amounts. A maximum of 3M packets are received within a /48 subnet, while the subnet with the fewest packets receives only 34. These scans do not seem to focus on the same breadth as *Ponynet*. Instead, they seem to target specific areas within the telescope. There is exactly one session in which all packets are received by the announced /36 prefix, which has already been announced multiple times by that point. This could be a research scan, aimed at focusing on a specific prefix among the announced ones.

[10] https://cujo.com/blog/threat-alert-krane-malware/

AS38272 Cernet. The third-largest scanner is AS38272, which is assigned to the AS organization China Education and Research Network (*Cernet*). *Cernet* is China's first and largest national academic Internet backbone. The network infrastructure mainly serves universities, institutes, colleges, and schools across China. The end users are professors, researchers, and students[11]. All /48 subnets of T1 receive at least 7 and at most 44K packets from this scanner, with an average of 141 packets per subnet. This shows a different strategy compared to the other two scanners. All /48 subnets within T1 are scanned, but the focus is on the prefixes that are announced at the time. The scans are observed starting from May 1, 2024, so the most-specific prefixes are two /44 prefixes that are announced at that time.

Table 7.9: Overview of package distribution by sessions.

Prefix	AS53667	AS12816	AS38272
2001:db8::/33	4	0	7
2001:db8:8000::/34	4	0	7
2001:db8:c000::/35	4	0	7
2001:db8:e000::/36	4	1	7
2001:db8:f000::/37	4	0	7
2001:db8:f800::/38	4	0	7
2001:db8:fc00::/39	4	0	7
2001:db8:fe00::/40	4	0	7
2001:db8:ff00::/41	4	0	7
2001:db8:ff80::/42	4	0	7
2001:db8:ffc0::/43	4	0	7
2001:db8:ffe0::/44	4	0	7
2001:db8:fff0::/45	4	0	7
2001:db8:fff8::/46	3	0	7
2001:db8:fffc::/47	3	0	7
2001:db8:fffe::/48	3	0	7
2001:db8:ffff::/48	3	0	7

Categorization of heavy hitters. AS53667 (*Ponynet*) scans with four sessions at the beginning of 2024. At that time, the most-specific prefixes currently announced are two /36. In Table 7.9, it can be seen that the sessions are evenly distributed across the announced prefixes. However, the /46 to /48 prefixes receive a slightly lower number of sessions. This scanner is categorized as *intermittent* and *network-size independent*.

AS12816 (Leibniz supercomputing centre) behaves quite differently. This scanner is only seen on one day (May 29, 2024) with exactly one session, and it scans only the /36 prefix that has already been announced for almost five months. This scanner is categorized as *oneoff* and *single-prefix*.

[11] https://www.edu.cn/english/cernet/introduction/200603/t20060323_158626.shtml

The scanner from AS38272 (*Cernet*) behaves in a similar way to AS53667 (*Ponynet*). It starts scanning on May 22, 2024, and is seen with a total of seven sessions until June 15, 2024. At that time, first the /46 and then the /47 are the most-specific prefixes that are announced. All prefixes receive a similar number of sessions. This scanner is categorized as *intermittent* and *network-size independent*.
Different and similar approaches can be observed among the scanners. This shows that scanners may react differently to the announcement of prefixes.

7.7 Taxonomic Analysis

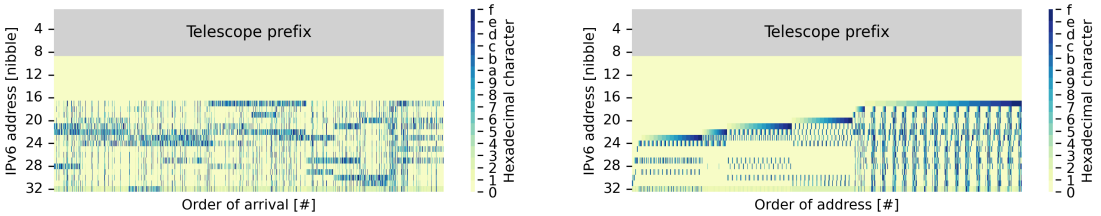
The following presents the results of the taxonomy. First, the approach to the classification will be explained. Finally, an overview of the results will follow. The terms of the taxonomy are already explained in Section 6.

7.7.1 Address Selection

Figure 7.5 and Figure 7.6 illustrate these strategies (*random* and *structured*) in the form of heatmaps, where the destination addresses of a session are sorted in hexadecimal notation by arrival time or by address value. Each destination address is represented vertically, with the y-axis displaying all 32 hexadecimal characters of a destination address using a color scale. The x-axis in Figure 7.5a and 7.6 show the arrival time of the corresponding packets from left to right. In Figure 7.5b, the x-axis sorts the destination addresses by address value. The fixed prefix is grayed out.

In Figure 7.5a, an example of a *structured* target address generation is shown, based on a scan session from AS132203 with 151k packets. Many areas of the destination addresses are filled with zeros. However, there are also block-like color changes that indicate bit changes at certain positions within the addresses.

The same destination addresses from Figure 7.5a are also shown in Figure 7.5b. However, here the destination addresses are sorted not by arrival time, but by the lexicographical value of the address, making the progression within a subnet more visible. Together, both figures show that the scanner scans dense address regions within the prefix, with the destination addresses remaining largely unchanged.



(a) Sorted by the arrival time at which the packets were received.

(b) Sorted by the lexicographical value of the address.

Figure 7.5: Hexadecimal representation of structured target address generation within a session.

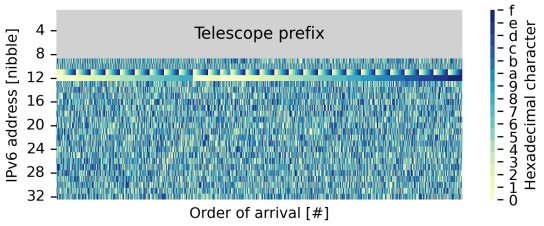


Figure 7.6: Hexadecimal representation of random target address generation within a session.

In contrast, Figure 7.6 shows an almost *random* structure. Here, a specific pattern of the hexadecimal characters is visible at the 11th and 12th nibbles. This indicates a scan through multiple subnets, suggesting distributive subnet coverage. This scan session from AS53667 comprises 113k packets and indicates that the scanner targets multiple subnets, while the IID of the destination addresses appears to be *randomly* generated, as no clear structure can be observed within the IID of the destination addresses.

7.7.2 Temporal Behavior

Packets are grouped into scan sessions based on an interarrival time. If a packet from the same source arrives within an hour of the previous one, it is part of the current session. If the next packet arrives after more than an hour, a new session starts. The shortest sessions observed consist of just one packet, while the longest are compared in Table 7.10.

Longer sessions sometimes last for several days. For example, in T1, a session from AS26832 lasted 25 days.

Table 7.10: The longest scan session observed during the entire measurement period.

T1	T2	T3	T4
25 days 01:57:00	7 days 18:23:00	0 days 01:13:26	0 days 03:15:08

Table 7.11: Total number of source IP addresses and sessions per temporal behavior before the split period.

	T1		T2		T3		T4	
Category	Src [%]	Session [%]	Src [%]	Session [%]	Src [%]	Session [%]	Src [%]	Session [%]
Oneoff	51.9%	11.3%	22.7%	0.9%	57.1%	44.4%	97.6%	86.4%
Periodic	14.2%	62.9%	33.2%	35.1%	0%	0%	0.8%	7.7%
Intermittent	33.9%	25.8%	44.1%	64%	42.9%	55.6%	1.6%	5.9%

For the classification of *temporal behavior*, the categories *periodic* and *intermittent* are assigned using autocorrelation [6], which mathematically examines regular patterns in the occurrence of sessions per source. Since this involves real data, non-periodic noise often occurs. Autocorrelation serves as a tool for assessing *temporal behavior*. The result of the autocorrelation provides a float value that acts as a performance metric for the model.

Tables 7.11 and 7.12 show the categories into which the *temporal behavior* of sources and sessions can be classified. Since the number of sessions can vary significantly from the number of sources and has a considerable impact on the observed traffic, both categorizations are presented separately. Additionally, the results for the first 12 weeks (before the split period) and the following weeks (during the split period) are shown separately, each in its own table.

The results show that T1 observes many *oneoff* sources but fewer *periodic* sources before the split period, while it sees more *periodic* sessions and fewer *oneoff* sessions. During this time, T1 is still passive. During the split period, the number of *periodic* sources and sessions increases slightly, possibly due to the influence of every two weeks BGP announcements. Reactions to BGP announcements are analyzed in Section 8. Overall, T1 shows that most scanners do not return and generate only one session. 43% of the *oneoff* sources during the split period are from RIPE Atlas, which scans across a wide range of targets, significantly contributing to the high number of *oneoff* sources.
T2 mostly observes *intermittent* sources and sessions and fewer *oneoff* sources before the split period. During the split period, this changes, and *oneoff* sources appear more often. About 85% of these *oneoff* sources (during the split period) target the address

Table 7.12: Total number of source IP addresses and sessions per temporal behavior during the split period.

	T1		T2		T3		T4	
Category	Src [%]	Session [%]	Src [%]	Session [%]	Src [%]	Session [%]	Src [%]	Session [%]
Oneoff	68.1%	8.8%	65.8%	2.8%	53.9%	22.6%	59.3%	19.5%
Periodic	15.8%	73.5%	17.5%	32.1%	30.8%	64.5%	14.8%	30.5%
Intermittent	16.2%	17.7%	16.8%	65.1%	15.4%	12.9%	25.9%	50%

with the DNS entry. In this analysis, these sources often do not initiate a second session, which has a significant impact on the observed results. Since the prefix is on the aliased hitlist, this might also explain why it is not scanned again frequently.
T3 remains passive in both periods. Before the split period, we observe seven sources, while during the split period, there are 13. Due to the small number of sources, changes in the *periodic* and *intermittent* categories have a larger effect. However, it remains roughly balanced whether scanners appear only once or multiple times.
T4 is reactive in both periods and sees significantly less traffic compared to T1 and T2. During the split period, *intermittent* and *periodic* sessions increase. One reason is that before the split period, 240 of the 253 sources are identified as *oneoff* sources, each sending only one packet. This is observed only before the split period, explaining the high number of *oneoff* sources and sessions at that time.
The categorizations for T1 and T2, which see much more traffic, show that telescope characteristics influence which scanners target the prefix. Additionally, *periodic* actions, such as the BGP experiment, can trigger specific scan behaviors, including *periodic* scanning patterns.

7.8 Taxonomic Results: Scan Behavior in the First 12 Weeks

The results of the classification according to the taxonomy for the first 12 weeks are presented in Figure 7.7. We have decided to focus on the first 12 weeks for the taxonomic analysis across all telescopes. This is because, during this period, Spoki is only active in T4, and T1 is only announced as a /32 and is passive. The reaction to Spoki and the BGP announcements should be analyzed separately, as including them in an already complex taxonomic analysis would introduce additional complexity.

Figure 7.7 is designed to summarize all classifications according to the taxonomy across all telescopes. Each column represents evaluations for a single telescope. The rows group

sessions based on the *temporal behavior* of the sources. The classification of *address selection* per session is shown using bars in each field. For the labels of the telescopes, the following characteristics are used: BGP-controlled (T1), Partially productive (T2), Silent (T3), and Reactive (T4).
Structured approaches are more frequently chosen. *Oneoff* scans are less common in T1 and T2. Most of the sources observed during the split period reappear (*intermittent*: 41%, *periodic*: 29%). None of the sessions of T3 and T4 are classified as *random*.
T3 aligns with its name, *Silent*, as only a few sessions are observed, while T4 has a somewhat higher number of sessions. T2 records slightly more sessions than the passive T1. In the first 12 weeks, T1 remains completely passive, while T2 has already been announcing for a longer time and may have attracted more scanners into the prefix due to its active /56. Furthermore, the ratio of *oneoff* sessions is similar for T1 and T2.

Figure 7.7: Classification of scanners per telescope based on temporal behavior, with a further subdivision of sessions per classification according to address selection, before the split period.

Overall, parallels in scan behavior can be observed. Nevertheless, the characteristics of the telescopes also show effects that suggest differences in scan behavior.

7.9 Summary of Findings

This subsection summarizes the key findings observed during the analysis of the scanning behavior across all telescopes. It is observed that sessions across all telescopes often focus on fewer than five targets. Additionally, the majority of source IP addresses exclusively scan the address with the DNS entry and do not target any other addresses, thus focusing on just one target. Only five scanners (0.02%) in the entire measurement period scan both this address and others. The scanners focusing on the address with the DNS entry primarily target ports 80 and 443. Furthermore, while some scanners perform broad scans, covering many subnets, others concentrate on specific destination addresses. *Low-byte* addresses are scanned particularly frequently. It can also be observed that scanners show differences in their scanning behavior within a telescope. Additionally, certain telescope characteristics (DNS Entry, Spoki, BGP Experiment) have an impact on the scan traffic. Two of these telescope characteristics will be examined in more detail in the following two sections.

8 BGP Experiment: Analyzing Scanner Behavior in Response to Prefix Announcements

It is observed that T3 and T4, which are not announced in BGP, receive significantly fewer packets compared to T1 and T2, which are announced in BGP. This difference motivates a closer investigation of BGP announcements and their impact on scan traffic.

Starting from November 22, 2023, T1 will be subdivided every two weeks. Then the new most-specific prefixes and the other will be announced. This subdivision begins with the initially announced /32 prefix and ends on July 2, 2024, with the last withdrawal of prefixes from /33 to /48 prefixes. This section focuses on the reactions of scanners to the BGP announcements and the scanning behavior in T1.

Figure 8.1 provides an overview of the cumulative total packets per day. It can be seen that during the split period, the number of packets increases more significantly

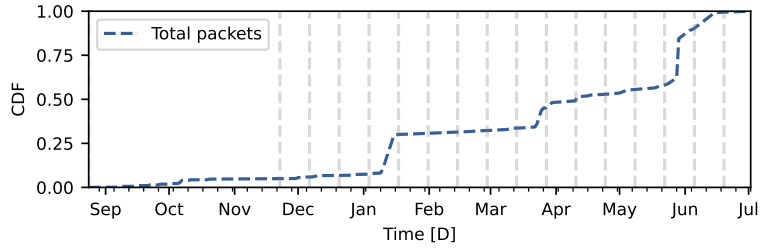


Figure 8.1: Cumulative number of total packets per day during the split period.

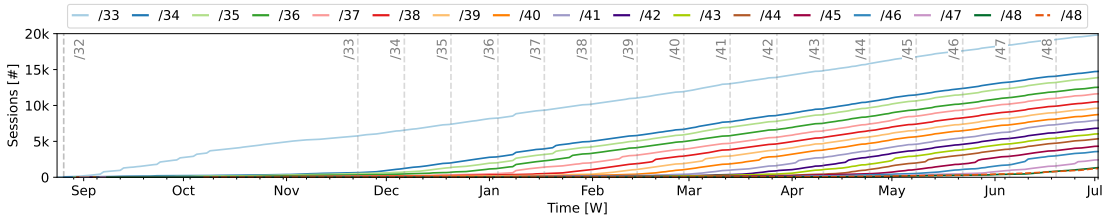


Figure 8.2: Cumulative number of scan sessions per most-specific prefix.

than before. This may indicate that scanners are reacting to these announcements and scanning the corresponding prefixes.

Figure 8.2 shows the number of scan sessions per prefix. It becomes evident that the number of sessions per prefix steadily increases after each announcement. This indicates that BGP announcements lead to a rise in traffic, as seen both in the packet reception (Figure 8.1) and the number of scan sessions (Figure 8.2). In contrast, subnets that are not announced receive minimal traffic. For instance, one of the /48 subnets accounts for only 0.00042% of the total packets on November 21, 2023, the day before the first split period begins. However, two weeks after the /48 prefix is announced on July 2, 2024, this packet share increases significantly to 0.19% (*i.e.,* a 452-fold increase).

The observations already indicate that there are IPv6 scanners that respond to BGP announcements.

8.1 Network Selection

During the split period, the classification of *network selection* is used to examine how the split experiment impacts the scanning behavior of IPv6 scanners.

The classification for *network selection* categorizes sessions of the sources into *network-size dependent*, *network-size independent*, *single-prefix*, or *inconsistent*. The definitions are explained in more detail in Section 6.

As a result, nearly 50% of the sessions fall into the *inconsistent* category, 30% into the *network-size independent* category, 18% into the *single-prefix* category, and 2% into the *network-size dependent* category. The *network-size dependent* strategy is shown to be rare, while *network-size independent* and *single-prefix* strategies are observed more frequently. *Inconsistent* behavior also suggests that no uniform strategy is observed. This could also mean that multiple strategies were applied during the measurement period.

8.2 Overview of T1 Taxonomic Results

Figure 8.3 shows the result of the complete classification according to the taxonomy based on the scanning behavior within T1 during the split period.
The classification of the sessions from T1 indicates that most sessions are categorized as *periodic*. *Oneoff* sessions are the least frequent. *Random* sessions are frequently found in *periodic* sessions within the categories *network-size independent* and *network-size dependent*, while some *random* scans are also observed in these *intermittent* and *oneoff* categories. *Single-prefix* scans are generally classified as *structured* sessions. *Inconsistent* sessions, especially in connection with *oneoff* sessions, were not identified.

Intermittent and *periodic* sessions are observed more frequently in total. *Oneoff* sessions, on the other hand, are seen less often overall, but *oneoff* sessions primarily target *single-prefixes*. In addition, fewer sessions are observed from *prefix-size dependent* scanners.

This reveals varying scanning behaviors. However, fewer sources determine their number of sessions based on prefix size, and less randomization is observed in the IID of the destination addresses.

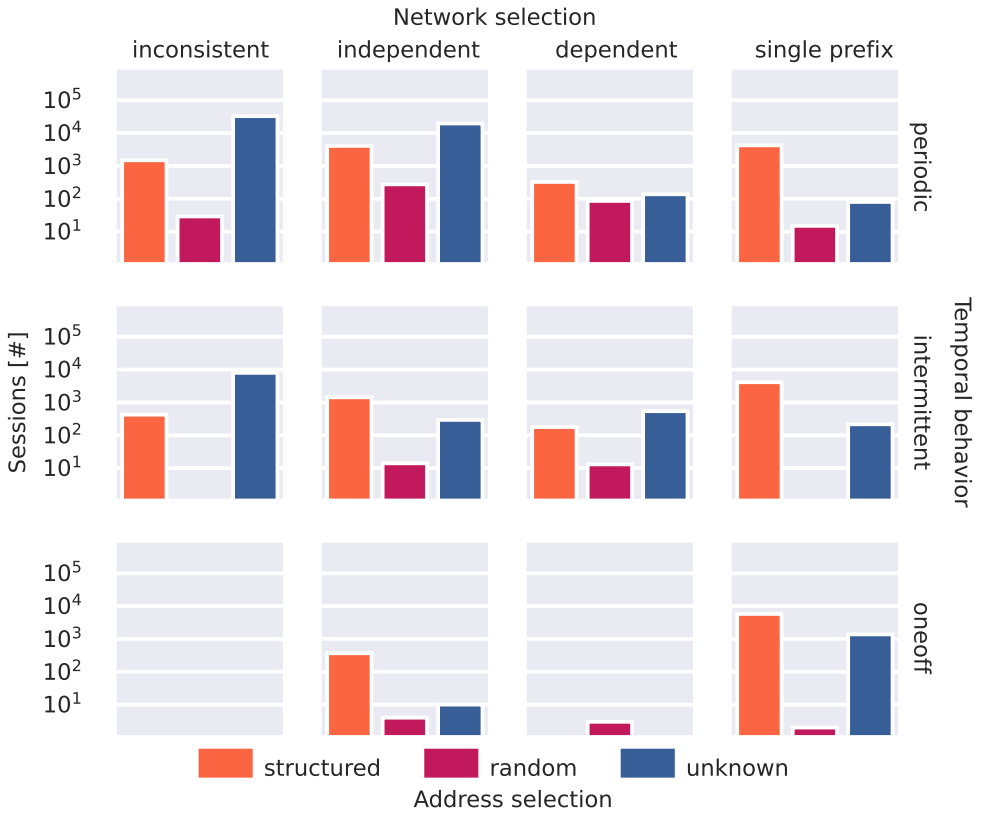
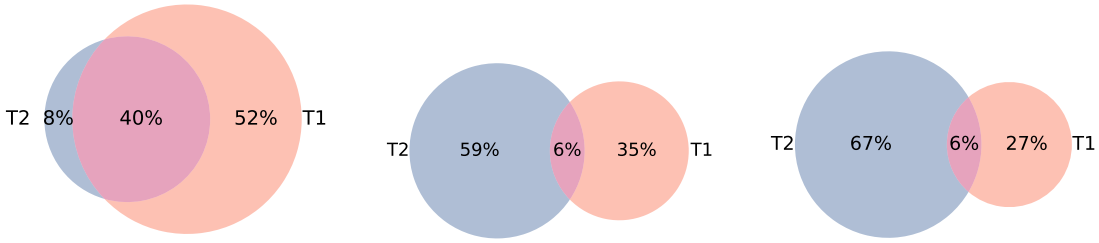


Figure 8.3: The classification of scanners from T1 during the split period based on temporal behavior, with a further subdivision according to network selection. The sessions within each subclassification are then further divided based on address selection.

8.3 Known and New Sources

The analysis of the BGP experiment also includes a comparison of the ASNs and sources from T1 with those of the other telescopes. This makes it possible to compare the origin of packets from T1 with those from the others and observe potential differences triggered by the BGP announcements. Since T1 is announced in BGP from the start of the measurement period, the results of the entire measurement period are included in the analysis. The results are shown in Figure 8.4.

Figure 8.4a shows a much larger exclusive proportion for T1 compared to Figures 8.4b and 8.4c. T2 observes many /128 and /64 sources that T1 does not see. However, both



(a) ASN (b) Source addresses /64 (c) Source addresses /128

Figure 8.4: Overlaps between the ASN and sources of T1 and T2 throughout the entire measurement period.

observe their sources frequently from the same AS, with T1 seeing a few more sources from different ASes that T2 does not observe. Of the 52% of ASNs that are only seen in T1, 99% are ASNs that can be attributed to RIPE Atlas. Nevertheless, there is a high overlap (40%) even though T2 has other characteristics that could prompt other scanners to scan the telescope, and it was already announced much over 13 years ago.

Through the BGP experiment with T1, there are new announcements every two weeks.

When comparing T1 and T3, there is an overlap of 11 ASNs (0.61%). Therefore, T1 observes 99% that are not seen in T3. For the source addresses, there is an overlap of 15 ASNs (0.13%). Thus, T1 only observed 99.8% of the source IP addresses. The share is equally large for the /64 sources.

In T4, slightly more sources are observed. There are 19 overlapping ASNs (1%) between T1 and T4. Therefore, T1 observes 98.95% of the ASNs. For the source IP addresses, 36 (0.3%) overlap, and for the /64 sources, 31 (0.3%) overlap. In T1, 97% of these sources are observed that are not seen in T4.

Each telescope observes both unique and overlapping ASNs and sources. When examining the overlaps between T1 and T2, new and unknown ASNs seem to appear more frequently in T1, which are not observed by T2. These are primarily ASNs that can be attributed to RIPE Atlas. However, the /64 and /128 sources differ significantly between T1 and T2.

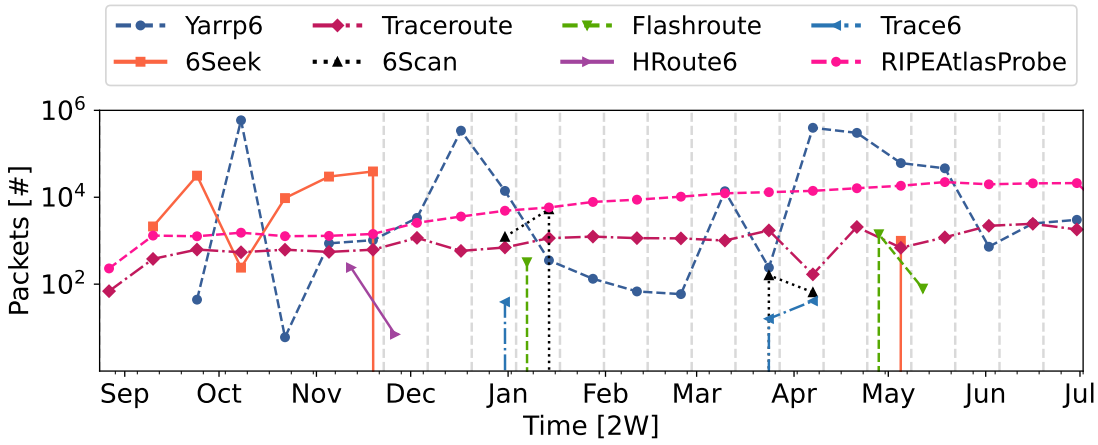


Figure 8.5: Total packets per scan tool (aggregated over 2 weeks).

8.4 Payload Analysis

In the previous subsection, it is shown that T1 observes traffic from more ASNs compared to the other telescopes. Since T1 draws attention through announcements in BGP, it makes sense to analyze the scan traffic of the T1 scanners in more detail. It is important to note that the announced prefixes by T1 also appear on the non-aliased hitlist, and it cannot be ruled out that the scanners find the prefixes through this list. Results on the origin of the packets are already presented in Section 5. In this subsection, the payload of the scanners is examined to gain additional insights.
Some *fingerprints* can be identified in the payloads of the received packets. For example, the ASCII text *yrp6* (Yarrp6 [5]) can be decoded from the payload *79727036*, and in some cases, a URL can also be found, such as from *6Seeks*[12]. This information help to better understand the origin and behavior of T1 scanners. Additionally, it is highlighted how scanners behave before the split period and during the split period. This can also be observed through the scantools and scan system *fingerprints*.

Figure 8.5, shows the number of packets over two weeks originating from these identified scanning tools and systems, such as RIPE Atlas. However, no definitive scientific evidence exists for the tool *HRoute6*, which can be conclusively attributed to this tool.

[12]https://6seeks.github.io

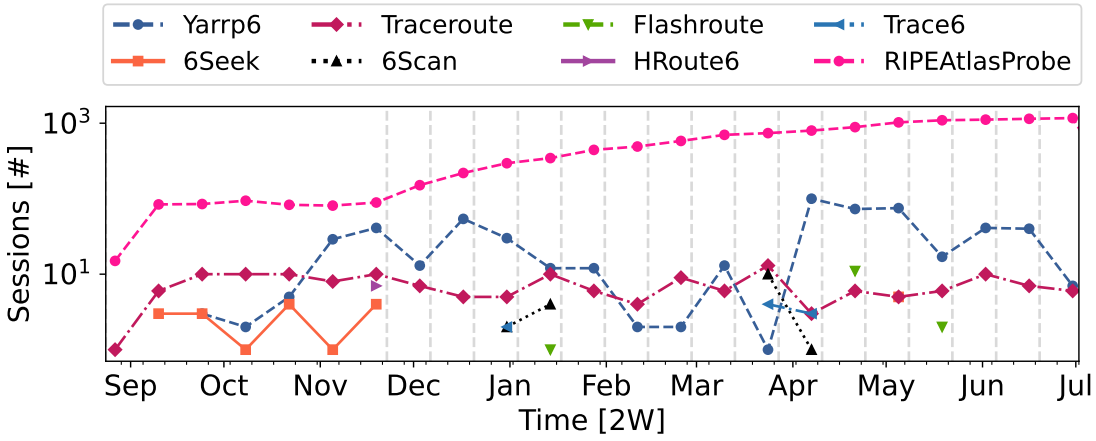


Figure 8.6: Total sessions per scan tool (aggregated over 2 weeks).

It can be seen that packets from the RIPE Atlas probes [38] as well as Traceroute[13] are very consistent and frequently observed. Starting from the split period, the number of packets from the RIPE Atlas probes continues to increase every two weeks, while the number of Traceroute packets remains rather constant, with infrequent downward fluctuations. *Yarrp6* payloads are also frequently observed, although with highly variable packet volumes. *6Seeks* packets are mainly observed before the split period. Tools like *6Scan*[25], *Flashroute*[26], *HRoute6*, and *Trace6* [49] appear less frequently in the payloads.

In Figure 8.6, the sessions of the scanning tools over a two-week period are shown. Similarities to Figure 8.5 can be observed. The sessions from the RIPE Atlas probes and Traceroute are consistently seen throughout the measurement period, while *Yarrp6* appears frequently, but with varying session amounts. The other scan tools are observed only sometimes. However, this Figure 8.6 clearly illustrates how many sessions originate from the RIPE Atlas probes compared to the other scanning tools. Furthermore, it appears that the RIPE Atlas probes react to the announcements, as the number of sessions per announcement period continuously increases.

8.5 Reaction to BGP Signals

BGP announcements seem to attract scan traffic. When choosing the prefix size, the question arises whether there is a difference between announcing a /32 or a /48 prefix in

[13] https://elinux.org/Traceroute_-_Tracing_Route

Table 8.1: Distinct ASNs, source IP addresses and targets during the first two weeks of annoucement for each prefix.

Prefix	ASNs [#]	Src. IPs [#]	Targets [#]
2001:db8::/33	125	184	192 285
2001:db8::/34	120	178	7274
2001:db8::/35	128	175	20 658
2001:db8::/36	164	418	603 329
2001:db8::/37	124	162	2616
2001:db8::/38	127	167	2580
2001:db8::/39	123	168	1000
2001:db8::/40	125	176	19 470
2001:db8::/41	113	168	794 324
2001:db8::/42	117	171	2211
2001:db8::/43	126	222	2542
2001:db8::/44	122	270	49 615
2001:db8::/45	128	234	21 283
2001:db8::/46	123	210	78 344
2001:db8::/47	119	203	38 933
2001:db8:fffe::/48	115	158	246
2001:db8:ffff::/48	114	210	268

BGP. To investigate this, we compare the number of scan sources, ASNs, and destination addresses for each announced prefix. Only the announced prefixes on June 19, 2024 (/33 - /48), are considered in this analysis.

Relevance of prefix size. Table 8.1 shows the number of distinct ASNs, source IP addresses, and destination addresses within the first two weeks after the initial announcement of each prefix. Between the ranges of ASNs within the column (min: 113 and max: 164), there is a difference of 51. However, this larger difference is due to the /36 prefix, which observe more ASNs than the other prefixes. Without the /36 prefix, the difference would only be 15. Therefore, there is no significant difference in the number of ASNs. For the source IP addresses, there is a similar distribution across the prefixes, with a slight majority for the /36 prefix. However, there are noticeable differences in the scanned destination addresses. For the /41 and /36 prefixes, significantly more distinct destination addresses are scanned. In contrast, very few distinct targets are scanned by sources in the /48 prefix. In the remaining prefixes, sources are observed scanning between 1k and 200k targets.

Based on the observations, it seems that the /36 prefix sees more sources in the first two weeks after it was firstly announced, which also scan many destination addresses. Therefore, it can be suggested that there might be scanners that prefer scanning /36 prefixes. However, this conclusion is based solely on the observed traffic.

Table 8.2: Overlaps of the scan sources within the network prefixes during the split period.

	First two weeks		Complete announcement period	
Overlap	Count [#]	Share of total sources [%]	Count [#]	Share of total sources [%]
1	1651	14,93%	7332	66.32%
2	170	1.54%	1578	14.27%
3	35	0.32%	686	6.2%
4	16	0.14%	335	3.03%
5	8	0.07%	86	0.78%
6	8	0.07%	40	0.36%
7	7	0.06%	59	0.53%
8	2	0.02%	50	0.45%
9	1	0.01%	28	0.25%
10	0	0.00%	22	0.20%
11	2	0.02%	55	0.50%
12	1	0.01%	90	0.81%
13	2	0.02%	65	0.59%
14	2	0.02%	53	0.48%
15	2	0.02%	50	0.45%
16	3	0.03%	49	0.44%
17	58	0.50%	71	0.64%

Overlaps of the scan sources and ASNs. Table 8.2 shows, in the first left column, the number of prefixes with overlapping sources. The columns labeled *Count* show the number of source IP addresses found in the number of prefixes listed in the first column. For example, 1,651 source IP addresses are observed exclusively in one prefix, while 58 sources appear in all prefixes. The other column shows the percentage frequency of these source IP addresses from the *Count* column in relation to all source IP addresses during the split period. In addition to the number of source IP addresses that appear only in the first two weeks after the initial announcement of the individual prefix, the number of source IP addresses is also allocated accordingly, which are observed throughout the entire period of the prefix's announcement.

Overall, most source IP addresses are observed in only one subnet. However, a source is often found in two subnets. The number of sources observed in exactly 10 different subnets is the smallest. The number of sources decreases from an overlap of 1 to 10 overlaps. Then, a few more sources are found that appear in all subnets. When looking at only the first two weeks after the announcement of a subnet, fewer sources are found in this period than in the entire duration of the respective subnet announcements. This suggests that scanners do not always react immediately to the announcements. Some scanners could focus on one prefix first and then scan others later. However, there are

also scanners that respond to the announcements within the first two weeks and scan all prefixes.

Non-distributed scanners. The previous results already show that many scanners focus on a single prefix. In Table 8.3, all more-specific prefixes (/32 - /48) are ordered by the number of ASNs. Only the ASNs that are observed in exactly one prefix (Overlap == 1, see Table 8.2) during the respective period are considered. This is because the focus from this observation is on non-distributed scanners that do not scan across multiple prefixes, but rather concentrate on a single prefix. Overall, it can be observed that the number of ASNs in the first two weeks differs slightly compared to all prefixes. However, when considering the entire announcement period, more noticeable differences can be observed. It must be noted that the more-specific prefixes were announced later, thus having a shorter observation period. Therefore, it is not surprising that fewer ASNs are observed for these prefixes. The results show that it does not make a substantial difference whether a /32 or a /48 prefix is announced in BGP. Only the majority of sources for the /36 prefix are noticeable again.

Table 8.3: Number of ASNs appearing in exactly one announced prefix during the split period.

	First two weeks		Complete announcement period	
Rank	Prefix	Number of ASNs [#]	Prefix	Number of ASNs [#]
#1	2001:db8::/36	43	2001:db8::/34	74
#2	2001:db8::/43	27	2001:db8::/33	60
#3	2001:db8::/33	25	2001:db8::/36	55
#4	2001:db8::/34	24	2001:db8::/38	49
#5	2001:db8::/35	23	2001:db8::/35	43
#6	2001:db8::/44	23	2001:db8::/40	36
#7	2001:db8::/40	22	2001:db8::/37	36
#8	2001:db8::/46	22	2001:db8::/39	32
#9	2001:db8::/47	22	2001:db8::/41	32
#10	2001:db8::/38	21	2001:db8::/42	29
#11	2001:db8::/42	21	2001:db8::/43	18
#12	2001:db8::/37	20	2001:db8::/45	13
#13	2001:db8:ffff::/48	20	2001:db8::/46	12
#14	2001:db8::/41	20	2001:db8::/44	10
#15	2001:db8::/39	17	2001:db8::/47	7
#16	2001:db8:fffe::/48	17	2001:db8:fffe::/48	6
#17	2001:db8::/45	16	2001:db8:ffff::/48	5

Distributed scanning systems. During the first two weeks after the initial announcement of each prefix, a total of 58 source IP addresses and 45 ASNs are observed across all prefixes (Overlap == 17, see Table 8.2). Among these ASNs, 14 belong to the network type *Educational/Research*, seven to *Cable/DSL/ISP*, six are *Network Service Provider*,

three are *Hoster*, and two belong to *Non-Profit*, as identified via PeeringDB. These 58 source IP addresses are very likely reacting to BGP signals.
Additionally, there are distributed scanning systems like RIPE Atlas that use multiple source IP addresses for their scans. The RIPE Atlas fingerprint is not found in the payload of any source IP address that scans all prefixes within the first two weeks after their announcement. However, it does appear in the payload of sources that scan only a single prefix. This suggests that RIPE Atlas performs distributed scanning using multiple source IP addresses. This highlights a challenge: Without knowing whether multiple source IP addresses come from the same source, it is difficult to reliably identify all distributed scanners.

8.6 Summary of Findings

We identify distributed scanners that scan all prefixes during the first two weeks after their announcement. This suggests that these scanners actively respond to BGP announcements. Additionally, we observe distributed scanning systems (such as RIPE Atlas) where there are multiple source IP addresses that scan individual prefixes and also respond to BGP announcements.
We also detect non-distributed scanners that scan exactly one prefix within the first two weeks of a new announcement. However, it is not certain that all of these are truly non-distributed scanners, since we only observe traffic from T1 and cannot definitively associate all source IP addresses with the same entity or determine if they truly belong together.
Across all scanners, we observe different methods, including *single-prefix* scanning, *prefix-size independent* scanning, *prefix-size dependent* scanning, and *inconsistent* scanning behavior. Based on the observed traffic, it can be observed that the prefix size does not have a significant impact. However, slightly more ASNs, sources, and targets are observed in the first two weeks after the announcement of the new /36 prefix.

9 Impact of the Reactive Network Telescope Spoki

Since stateless scanning and two-phase scans are established techniques in IPv4, it is worth exploring whether these methods are also applied in IPv6 scans. For this reason, such an analysis is examined in more detail in this section. In addition to understanding

how these methods are applied, it is also important to investigate the motivations for these scans using the payloads in the packets. Previous studies in IPv4 revealed several instances of malicious downloads embedded in the payloads of packets, and security checks were observed.

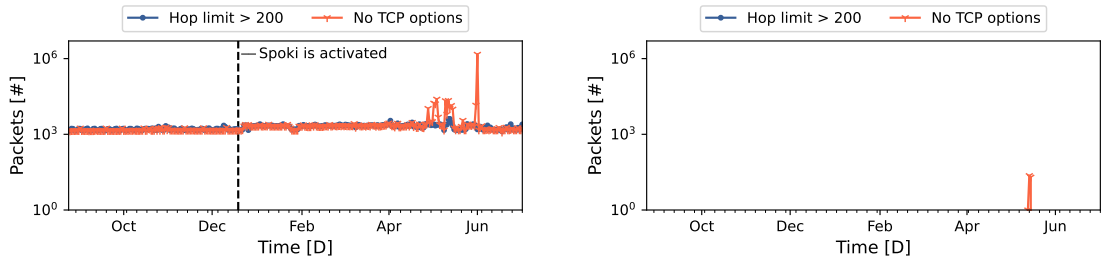
9.1 Two-Phase Scanning in IPv4 and IPv6

The scalable reactive network telescope Spoki was extended to enable the analysis of two-phase scanners in IPv6. In previous work [20], analysis of network traces and open-source code of popular scanning tools shows that stateless scans in the first phase often use irregular TCP SYN packets as probes. For stateless scanning, ZMap [14] is preferably used. Due to the stateless character of its architecture, ZMap achieves high scanning speeds. Furthermore, the utilization of the cyclic multiplicative group ensures that no additional memory is required, as all targets are systematically reached. In the stateful second phase, scanners connect only with those targets that responded in the first phase. This efficient approach facilitates the scanning process. In the first phase, irregular TCP SYN packets are considered, as outlined in [20], which possess at least one of the following properties: (*i*) TTL > 200, (*ii*) no TCP options, (*iii*) an IP ID of 54321. This IP ID is known to be used by ZMap. Two of the three properties are used in this work as a criterion for irregular SYN packets. However, the IP ID is not part of IPv6 headers. In addition, no such identifier is assigned by default in the extended ZMap version for IPv6 scans[14].

9.2 Irregular SYN Behavior in T2 and T4

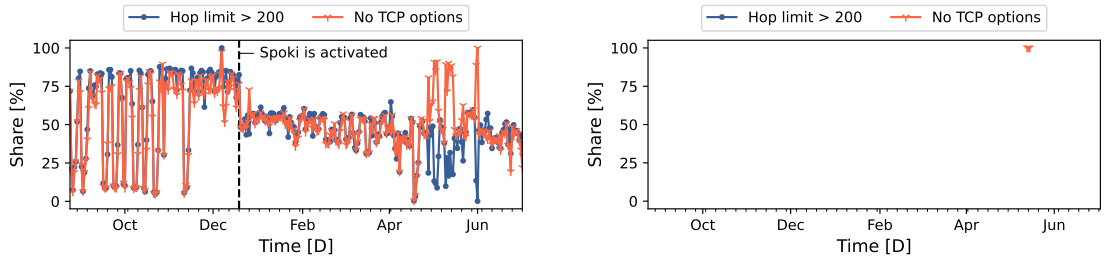
The two-phase scans can have various objectives, for example, research projects or malicious activities. Different results have already been achieved in the investigation of the IPv4 address space [21]. In this work, Spoki is deployed in T2 and T4 to analyze two-phase scanners and their packets in the IPv6 address space. T4 and the /29 covering prefix does not appear in the aliased prefix list of the TU of Munich during the measurement period. T2 appears there on December 21, 2023, two days after the deployment of Spoki.

[14] https://github.com/tumi8/zmap/blob/6c4585ba10926870c17c7bd2ad9f3872cc34c892/src/probe_modules/packet.c



(a) T2. (b) T4.

Figure 9.1: Overview of TCP SYN packets with a hop limit greater than 200 or without TCP options.



(a) T2. (b) T4.

Figure 9.2: Overview of TCP SYN packets with a hop limit greater than 200 or without TCP options, in relation to the total number of TCP SYN packets.

In Figure 9.1, an overview of the number of TCP SYN packets with a hop limit greater than 200 or without TCP options, received by T2 and T4, is shown. The entire measurement period is displayed, including a time period when Spoki was not active in T2. In T4, relatively few of such packets are received, whereas in T2, a very constant value of these SYN packets is observed. From late April to early June, there are some fluctuations in the values. Therefore, the percentage share of these packets compared to all TCP SYN packets received by T2 will also be examined (see Figure 9.2).

During the measurement period when Spoki is active, T2 receives over 2M irregular SYN packets from 5k source IP addresses directed at 1.5M destination addresses. In contrast, T4 only records 44 irregular SYN packets from two source IP addresses from AS2637, sent to 14 destination addresses. A subsequent second phase is not observed in T4, which is why the analysis of two-phase scanners focuses on T2.

In Figure 9.2a, various fluctuations of the two lines can be observed. Before the ac-

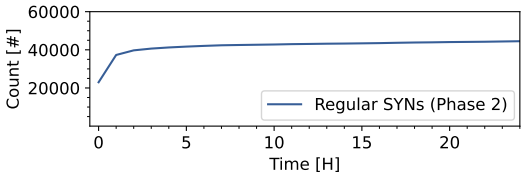


Figure 9.3: Number of regular SYN packets across different timeframes.

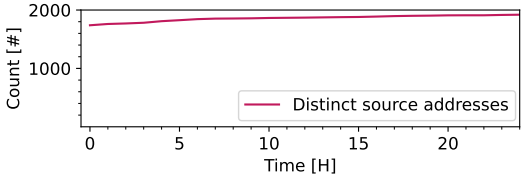


Figure 9.4: Number of distinct source addresses across different timeframes.

tivation of Spoki, packets with high hop limits or without TCP options were sometimes sent, but their share often dropped to 0%. After Spoki was activated, the proportion of irregular SYN packets remained relatively constant until April 2024. Starting from April 2024, several changes have been made to the ZMap tool for IPv6. On April 16, 2024, a standard TTL of 64 was set instead of the previous maximum TTL from ZMap to align with Ubuntu. By April 18, the share of packets with high hop limits over 200 and without TCP options decreases. From then on, further changes were made, including the parsing of TCP options, leading to an increase in the reception of SYN packets without TCP options. We do not know definitively whether these Commits[15] had such effects on the two values, but it is possible that they partially influenced these fluctuations.

Two-phase scanner. When determining the time frame from irregular SYN packets (phase 1) to regular SYN packets (phase 2), an analysis is conducted to assess how much the number of regular SYN packets increases with the addition of another hour compared to the previous time frame. The results of this analysis are presented in Figure 9.3. Furthermore, the increase in distinct source addresses is examined in Figure 9.4. It becomes evident that the number of source addresses does not change significantly. However, the number of regular SYN packets shows a significant increase within the first hour compared to the rest of the extended time frame. Therefore, the time frame after the occurrence of irregular SYN packets is set to 60 minutes. If a regular SYN packet is received within this 60-minute timeframe, a second phase is initiated.

The second phase only ends when no new packets are received within five minutes after the last received packet. If a new packet is received within these five minutes, the five-minute frame is restarted.

[15] https://github.com/zmap/zmap/compare/v4.1.0-RC2...main

ASCII and binary payloads. Using this two-phase scan analysis, the payloads of 66k packets can be decoded from UTF-8 to ASCII. 59k packets remain undecoded and are categorized as binary payloads, while 57k packets contain no payloads. The binary payloads are decoded for this work using Wireshark.

Downloader. In the investigations with Spoki in the IPv4 address space, ASCII-decoded payloads were identified that trigger downloads using the commands wget or curl. Therefore, a targeted search for these commands in the ASCII-decoded payloads was conducted, but no downloaders were found.

9.3 Spoki Analysis per AS

Many packets need to be analyzed to assess the payloads of the two-phase scanners. The packets received within the first five minutes after the regular SYN are grouped by their respective ASNs and analyzed individually. This approach provides a clearer understanding of the methods and behaviors of different scanning sources. In total, seven ASNs are identified, with detailed observations presented in the following subsections.

9.3.1 AS6939 Hurricane Electric LLC

Hurricane Electric[16] operates a global Internet backbone and provides IP transit services. It is connected to over 250 major exchange points worldwide. During the observation period, 3 million packets were received from 4k source IP addresses, targeting 59 destination addresses, with 570k sessions identified. Packets were sent to addresses ending with *::0* and *randomized* addresses. As soon as Spoki is active, 291k ACK packets and 576k SYN packets are received from AS6939.

In Figure 9.5, before the deployment of Spoki (December 19), there is a high proportion of sessions, but a low proportion of sessions with at least five packets and at least five targets. Only 65 sessions are observed, each containing at least five packets. This changes abruptly when Spoki is active. At T2, there is a sharp increase in sessions with at least five packets and at least five targets. This pattern remains consistent throughout the entire measurement period. Figure 9.6 shows the packets per day identified in both

[16] http://he.net

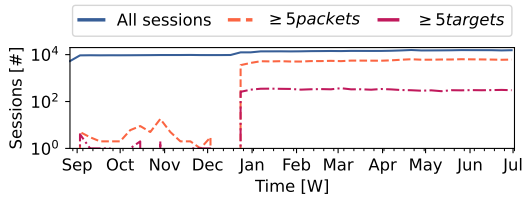


Figure 9.5: Number of sessions per week for all sessions, sessions with ≥ 5 packets, and sessions with ≥ 5 targets (AS6939).

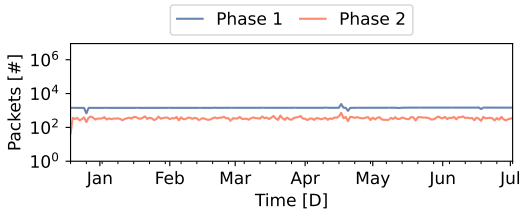


Figure 9.6: Comparison of daily packet counts for phase 1 and phase 2 (AS6939).

Table 9.1: Overview of ASCII payload (AS6939).

Payload prefix	Share[%]	Number of distinct ports [#]
GET / HTTP/1.1	78.1	33
SSH-2.0-	7.4	3
xml version	2.7	1
POST /	2.5	1
MQTT	2.3	1
Other	7	3

phases (two-phase scans), where the number of packets remains very constant. Comparing the two figures, it becomes clear that these two-phase scans, originating from the sources of AS6939, are related to the sudden increase in packets per session and targets per session.

Since scanners in the first phase only target responsive addresses and gather more information through additional packets in the second phase, the payloads contained in the packets received after the second SYN packet are analyzed. ASCII and binary payloads are examined separately. A total of 58k packets are assigned to the second phase, which are received after the regular SYN. Of these packets, 33% of payloads can be decoded in ASCII, 29% remain binary payloads, and 38% have no payload. 18 different destination addresses each receive around 3k packets during the second phase and 75 to 78 destination ports being used per destination address. This indicates a very structured approach. During the second phase, a total of 79 destination ports are observed, along with nearly 7k different payloads within the packets.

ASCII payload prefixes are grouped and displayed along with their frequency and the number of distinct destination ports used for these packets in Table 9.1.

The destination addresses most frequently receive *GET* requests, particularly on port 8080, though many other ports are also used. In addition to *HTTP* requests, *SSH*, *XML-Jabber*, *POST-HTTP-1.1*, and *MQTT* payloads are also observed. *SSH* packets use ports 22, 2222, and 212, while port 5222 is associated with the *XMPP* chat protocol (Jabber). All *POST* payloads contain the string *zgrab* and a User-Agent, matching the default value in the *ZGrab* source code [17]. These requests often use destination port 631, which corresponds to the *Internet Printing Protocol (IPP). MQTT* requests are sent to port 1883.
For most payloads that cannot be decoded as ASCII, they are *TLS Client Hello* payloads.

Table 9.2: Overview of top 5 destination ports with binary payload (AS6939).

Packet type	Ports [#]	Share[%]
TLS Client Hello	443, 10443, 2031	10
SMB Negotiate	445	5.4
RDP[1] NMAP Negotiate Request	3389	3.4
LDAP Bind Request <ROOT>[2]	389	3.4
SNPP[3]	444	3.3
Other	-	74.5

[1]Remote Desktop Protocol (RDP).
[2]To authenticate a user, the client sends a bind request.
[3]Simple Network Paging Protocol.

Table 9.2 shows that a wide variety of payloads were sent, with even the top 5 payloads accounting for only a small portion of the total. In addition to *TLS Client Hello*, there are also *Server Message Block (SMB)* requests, *Remote Desktop Protocol (RDP)* requests, *LDAP Bind* requests, and *Simple Network Paging Protocol(SNPP)* under the top 5 packet types.

9.3.2 AS4134 ChinaNet-Backbone

AS4134 belongs to the organization *ChinaNet*, which operates the largest Internet backbone in China[18] and is a network service provider. During the entire measurement period, packets were sent to both *low-byte* addresses and addresses with *embedded-port*. In total, we observed 115k packets from almost 3k source IP addresses from AS4134, which were sent to 42k destination addresses within over 8k sessions. When Spoki becomes ac-

[17]https://github.com/zmap/zgrab2/blob/76d09b59c5ec1b20fcc0a172d84df99802865250/modules/http/scanner.go#L46
[18]https://www.ctamericas.com/products/internet/chinanet-peering/

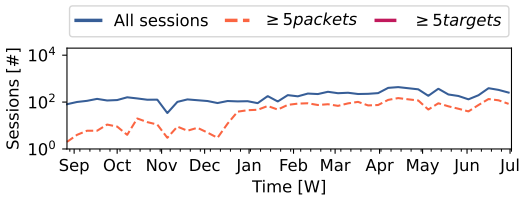


Figure 9.7: Number of sessions per week for all sessions, sessions with ≥ 5 packets, and sessions with ≥ 5 targets (AS4134).

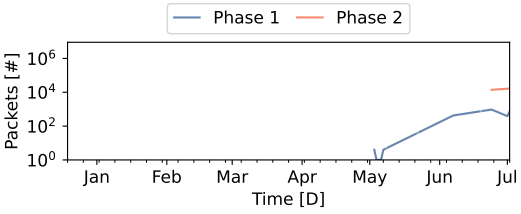


Figure 9.8: Comparison of daily packet counts for phase 1 and phase 2 (AS4134).

tive, AS4134 emits 12k ACK packets and 17k SYN packets primarily targeting *low-byte* addresses.

All packets, after the regular SYN (second phase), were sent to the destination address that has a DNS entry. Figure 9.7 shows that there are usually at least five packets per session. However, since packets are sent to only one destination address, the line showing the number of targets per session is not visible in the Figure. The two-phase scans, as seen in Figure 9.8, occur relatively late. Irregular SYN packets are received starting in May 2024, while the second phase with a regular SYN begins in late June 2024. During the time when Spoki is active, but no first phase has been initiated yet (December 2023 - May 2024), TCP SYN packets are received from AS4134. However, these are not irregular SYNs, as the packets either contain TCP options or have a hop limit of at most 200.

In total, 24k packets are received in the second phase after the regular SYN. Of these, 36% contain binary payloads, 27% are converted to ASCII payloads, and 37% have no payload. In addition, 368 destination ports and 725 payloads are observed.

Table 9.3 shows that the ASCII-decoded payloads from AS4134 frequently contain *GET /HTTP/1.0* requests, with a wide range of destination ports. *MQTT*, *EHLO (SMTP)*, and other payloads are less common. There is a large diversity of payloads and destination ports. When analyzing the ASCII payloads and comparing with other two-phase scanners, it is often observed that a *GET HTTP* request is received first. This is usually followed by other packets with different payloads. These subsequent packets often use the same destination port that was observed in the *GET HTTP* request. This pattern

Table 9.3: Overview of ASCII payload (AS4134).

Payload prefix	Share[%]	Number of distinct ports [#]
GET / HTTP/1.0	42.9	366
MQTT	0.2	1
EHLO (SMTP)	0.2	2
xml version	0.1	2
Other	56.6	366

is seen in many scanners. For this reason, the number of distinct destination ports in Table 9.3 is the same as in the row *Other* (366).

Similarly, the binary payloads in Table 9.4 show high variability. Among the top 5 payloads are packets with *DRDA (Distributed Relational Database Architecture)*, which is used to communicate with *IBM* clients. Target ports like 50000 and 9090 are used to extract information from database servers that support the *DRDA* protocol[19]. Here, a *DRDA EXCSAT* command packet (Exchange Server Attributes) is sent, and the response is analyzed.
Additionally, 0.6% of the packets have destination port 1433, which is used for *Microsoft SQL*. This port could be linked to an exploit for privilege escalation in *SIMATIC* controllers, possibly indicating tests of this security vulnerability[20].

Table 9.4: Overview of top 5 destination ports with binary payload (AS4134).

Packet type	Ports [#]	Share[%]
DRDA EXCSAT	9090, 50000	1.1
TLS Client Hello	443	0.6
TDS7[1] Pre-login	1433	0.6
Amazon AWS TLS[2]	9092	0.6
LDAP Bind Request <ROOT>[3]	389	0.6
Other	-	96.5

[1]Tabular Data Stream Protocol (TDS) used by Microsoft SQL.
[2]Amazon AWS MSK uses these TCP port 9092 for TLS communication.
[3]To authenticate a user, the client sends a bind request.

9.3.3 AS10439 CariNet, Inc.

AS10439 is the ASN of *CariNet, Inc.*, a company that operates as a hosting provider, also known as Fiber Alley Data Center based in the United States. This ASN transmits a total of 63k packets from four source addresses, all using a *low-byte* address structure

[19]https://github.com/nmap/nmap/blob/master/scripts/drda-info.nse
[20]https://www.cisa.gov/news-events/ics-advisories/icsa-14-205-02a

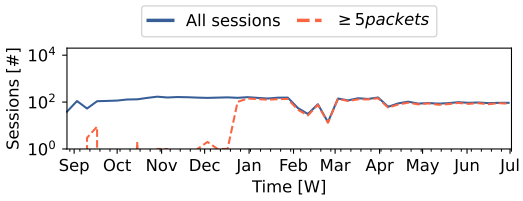


Figure 9.9: Number of sessions per week for all sessions and sessions with ≥ 5 packets (AS10439).

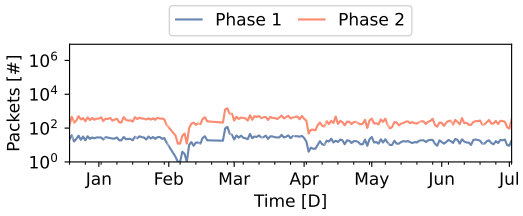


Figure 9.10: Comparison of daily packet counts for phase 1 and phase 2 (AS10439).

with *::1* appended after the prefix. When Spoki becomes active, AS10439 emits 9k ACK packets and 13k SYN packets.

Since there is only one target, Figure 9.9 displays only the count of five or more packets per session. It can be observed that from the point when Spoki becomes active, the number of sessions with at least five packets rises sharply, evident across almost all sessions. Figure 9.10 shows that both the irregular SYN packets from phase 1 and the packets from phase 2 are received less frequently starting in early February 2024, but their counts increase again shortly afterward. Overall, both lines show a similar trend, although the number of phase 2 packets remains consistently higher over the entire period. The reason for this is that the number of packets in the second phase (*GET HTTP* request, *Client Hello* packets, *etc.*) is higher than the packets from the first phase (irregular TCP SYN packets).

The 45k packets received in the second phase consist of 42% with binary payload, 37% with ASCII payload, and 21% without any payload.

A majority of the ASCII payloads, as shown in Table 9.5, contain *GET / HTTP /1.1* requests, targeting over 100 different destination ports. Additionally, payloads of the type *SSH-2.0-* are observed. It is unclear why *SSH* is found in the payloads. Normally, one would expect *SSH* to respond at this point, so the reason for the request remains unclear. For the remaining payloads, nothing stands out except for one payload containing *OPTIONS rtsp://<destination address>:554 RTSP/1.0*. After *rtsp://*, the destination address and the destination port are specified. Such packets are also received with the destination ports 554, 8554 and 10554. The abbreviation *rtsp* stands for Real-Time Streaming Protocol and it seems to be a check of the ports performed during the scans.

Table 9.5: Overview of ASCII payload (AS10439).

Payload prefix	Share[%]	Number of distinct ports [#]
GET / HTTP/1.1	86.6	127
SSH-2.0-	7.1	9
Other	6.3	11

The binary payloads are shown in Table 9.6, with TCP destination port 3389 being the most frequently identified. These are *Remote Desktop Protocol (RDP)* requests. Overall, the low-byte address receives packets with 133 different destination ports and 4k unique binary payloads. This suggests that the AS may be conducting targeted scans to identify various security vulnerabilities.

Table 9.6: Overview of top 5 destination ports with binary payload (AS10439).

Packet type	Ports [#]	Share[%]
RDP[1] Negotiate Request	3389	7.3
TLS Client Hello	33389	7
SMB Negotiate	445	2.3
TDS7[2] Pre-login	1433	1.4
PSQL	5432	1.4
Other	-	80.6

[1]Remote Desktop Protocol (RDP).
[2]Tabular Data Stream Protocol (TDS) used by Microsoft SQL.

9.3.4 AS14061 DigitalOcean

AS14061, also known as *DigitalOcean*, is based in the United States and operates as a hosting provider. The company offers a range of cloud-computing solutions[21]. In total, 80k packets are received from AS14061, originating from 629 source IP addresses and directed to 9k destination addresses across 708 sessions. These packets target both *low-byte* addresses and addresses ending with *::0*. When Spoki becomes active, AS14061 emits 10k ACK packets and 16k SYN packets. The regular SYN packets and the packets of the second phase are sent only to the *low-byte* address ending with *::1*.

Figures 9.11 and 9.12 show significant differences in the number of sessions, packets per session, targets, and packet volume in phase 1. While session counts gradually increase over time, phase 1 packet volumes are initially stable (January 2024, see Figure 9.12) but begin to vary significantly from mid-April onward. In contrast, phase 2 packet volumes

[21]https://www.digitalocean.com/about

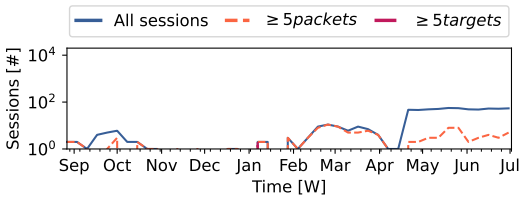


Figure 9.11: Number of sessions per week for all sessions and sessions with ≥ 5 packets (AS14061).

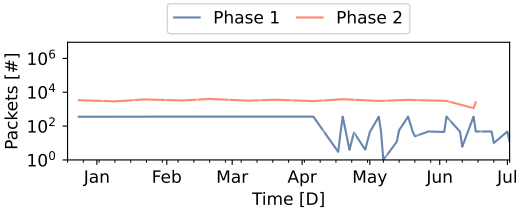


Figure 9.12: Comparison of daily packet counts for phase 1 and phase 2 (AS14061).

remain steady. The early occurrence of two-phase scans (December 2023, see Figure 9.12) suggests that scanners increasingly react to Spoki's responses, leading to a growing number of sessions.

Overall, AS14061 emits 36k packets in the second phase following the regular SYN. Among these, 14% contain binary payloads, 50% of payloads can be decoded in ASCII, and 36% have no payload at all. During this phase, the scans exclusively targeted the *low-byte* address ending with *::1*. A total of 352 destination ports and 915 distinct payloads were observed.

Table 9.7: Overview of ASCII payload (AS14061).

Payload prefix	Share[%]	Number of distinct ports [#]
GET / HTTP/1.1	81.3	349
SSH-2.0-	0.9	4
POST /	0.8	4
xml version	0.7	3
OpenTelnet	0.2	1
GET /spotifyconnect	0.2	1
EHLO (SMTP)	0.2	1
MQTT	0.2	1
Other	15.5	68

Most packets received from AS14061 contain decodable ASCII payloads that include *GET / HTTP/1.1* requests. As shown in Table 9.7, these packets are also sent over a variety of different destination ports. Among the requests are potentially malicious ones, such as *OpenTelnet:OpenOnce*, targeting port 9530. It has been claimed that a backdoor may have been built into surveillance devices[22]. Additionally, a connection to port 9530 is established, and randomly generated session keys are exchanged. Subsequently, an *OpenTelnet:OpenOnce* request is sent to the device to instruct it to open a Telnet service.

[22]https://www.theregister.com/2020/02/04/dvr_nvr_backdoor/

Table 9.8: Overview of top 5 destination ports with binary payload (AS14061).

Packet type	Ports [#]	Share[%]
ISO-TSAP COTP	102	3.3
Nessus Daemon Detection[1]	3001	1.7
TDS7[2] Pre-login	1433	1.6
AJP13[3] Request:GET	8009	1.4
DRDA EXCSAT	50000	1.4
Other	-	90.6

[1]Standard port for the Nessus daemon (network and vulnerability scanner).
[2]Tabular Data Stream Protocol (TDS) used by Microsoft SQL.
[3]Apache JServ Protocol (AJP) and CVE-2020-1938 Ghostcat.

If successful, a Telnet daemon is started on TCP port 9527, and a connection to the remote service can be established using the username *root* and the password *123456*. In the payloads, only *OpenTelnet:OpenOnce* is found, without any further requests or password entries.

The 5k packets with binary payloads listed in Table 9.8, contain a significant variability in both the TCP destination ports and the payloads. A total of 118 different destination ports are seen, and 788 different payloads are transmitted. Among the most common destination ports, Port 3001 stands out, which is used for the Nessus daemon, a tool for scanning networks and vulnerabilities[23].

Additionally, Port 8009 is also noted among the most common destination ports. An analysis with Wireshark reveals an Apache JServ Protocol (AJP) request in the payload. This protocol is associated with a known vulnerability affecting the AJP connector, which operates on Port 8009. It is identified as vulnerability *CVE-2020-1938*[24], also known as *Ghostcat*. This vulnerability in Apache Tomcat allows attackers to access sensitive data or execute remote code. The issue can be resolved through an update or by disabling the AJP connector.

9.3.5 AS16509 Amazon.com, Inc.

Amazon.com, Inc. or *Amazon Web Services (AWS)*, based in the United States, offers AWS, including globally available cloud solutions. *Amazon.com, Inc.* manages a total of 4,055 IP netblocks[25]. During the entire measurement period, T2 receives 104k packets

[23]http://www.di-srv.unisa.it/ ads/corso-security/www/CORSO-0304/nessus/index.htm
[24]https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2020-1938
[25]https://whoisrequest.com/ip/AS16509

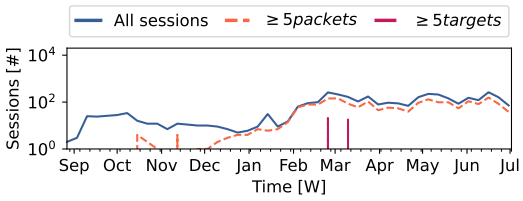


Figure 9.13: Number of sessions per week for all sessions and sessions with ≥ 5 packets (AS16509).

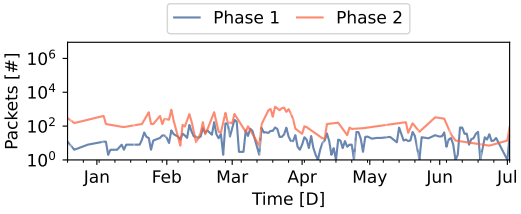


Figure 9.14: Comparison of daily packet counts for phase 1 and phase 2 (AS16509).

from 1k source IP addresses. These packets are sent to various destination addresses, including *low-byte* addresses and destination addresses ending with *::0*. A total of 203 destination addresses receive at least one packet within 3k sessions. When Spoki becomes active, AS16509 emits 14k ACK packets and 19k SYN packets.

Only the address with the DNS entry receives packets from two-phase scans from AS16509. In Figure 9.13, it is observed that starting in December 2023, most sessions receive at least 5 packets per week, often with fewer than five destination addresses involved. However, temporarily, more destination addresses are scanned at the end of February and the beginning of March.

It remains unclear whether the activation of Spoki has influenced the increased number of packets, as there was already an increased rate of packets per session at the beginning of December. Nevertheless, no significant decline is evident after this period.
As shown in Figure 9.14, both irregular and regular SYN packets are observed until the end of the measurement period. Within the packets of the second phase, following the regular SYN, only the destination ports 80, 8080, and 443 are monitored.

In the second phase of the scans after the regular SYN packets, AS16509 registers a total of 17.7k packets. Of these, 32% of the payloads can be decoded in ASCII, while 50% remain binary. 18% of the packets contain no payload. In addition, three destination ports are observed and nearly 2k distinct payloads within the packets.

Table 9.9 shows that all ASCII payloads (5.6k packets) include *GET / HTTP 1.1* requests. Only the destination ports 80 and 8080 are used. However, it is important to note that these do not represent the same payload prefixes. In nearly 94% of cases, a specific path is inserted between *GET /* and *HTTP/1.1*. Examples of these paths in-

Table 9.9: Overview of ASCII payload (AS16509).

Payload prefix	Share[%]	Number of distinct ports [#]
GET / HTTP/1.1	6.4	2
GET /<*path*> HTTP/1.1	93.6	2

clude *admin/index.html, manage/account/login, index.html, cgi-bin/login.html, ...* The scanner might specifically search for these paths to determine whether access to certain areas is possible.

The binary payloads of the 8.9k packets in Table 9.10 consist of *HTTP* packets and *TLS Client Hello* requests.

Table 9.10: Overview of top 5 destination ports with binary payload (AS16509).

Packet type	Ports [#]	Share[%]
HTTP	8080, 80	78.5
TLS Client Hello	443	21.5

9.3.6 AS396982 Google LLC

The AS396982 is associated with the *Google Cloud Platform* and is owned by *Google LLC*, based in the United States. The organization *Google LLC* allocates 46 IP netblocks under AS396982, encompassing a total of 131,072 IP addresses.
We receive 5k packets from AS396982, originating from 64 source IP addresses and directed to 69 destination addresses. These packets are sent within 2k scan sessions. When Spoki becomes active, AS396982 emits 294 TCP ACK packets and 2k TCP SYN packets. It is only at the beginning of 2024 that sessions sometimes receive five or more packets per week. Overall, however, the sessions are very irregularly distributed, showing no consistent trend in the number of sessions per week. Instead, they vary. This is evident in Figure 9.15. Figure 9.16 shows that irregular SYN packets were not received until February. Additionally, a second phase initiated by the regular SYN is only observed between February and the end of March.

A total of 1k packets from the second phase after the regular SYN are received, of which 27% of the payloads can be decoded in ASCII. Therefore, 52% remain as binary payloads and 21% have no payload. In the second phase, only packets sent to the address with the DNS entry are observed from AS396982. The scans reveal that these packets have 126

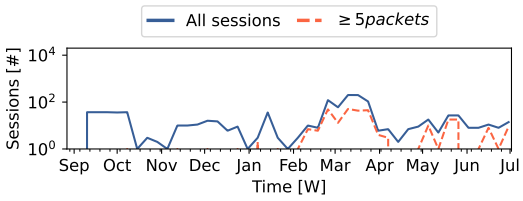


Figure 9.15: Number of sessions per week for all sessions and sessions with ≥ 5 packets (AS396982).

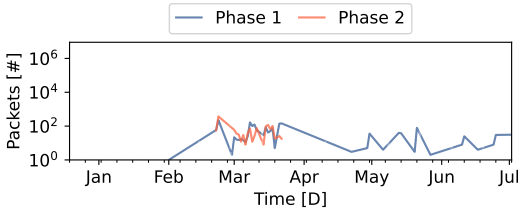


Figure 9.16: Comparison of daily packet counts for phase 1 and phase 2 (AS396982).

destination ports and contain 192 distinct payloads. Table 9.11 shows a comparatively high number of TCP destination ports in the *GET / HTTP/1.1* requests. In contrast, *SSH-2.0-* requests are directed to destination port 22 and *MQTT* requests to destination port 1883, while only 8 distinct destination ports are used in the other packets.

Table 9.11: Overview of ASCII payload (AS396982).

Payload prefix	Share[%]	Number of distinct ports [#]
GET / HTTP/1.1	77.5	46
SSH-2.0-	2.8	1
MQTT	1.4	1
Other	18.3	8

A total of 548 packets with binary payloads are recorded, distributed across 61 different destination ports and containing 134 distinct binary payloads. When examining the top destination ports, seven different ports associated with the *TLS Client Hello* payload (see Table 9.12), where some ports can be grouped together based on the same packet type. Additionally, two destination ports are identified that were used for *HTTP* packets. Overall, the distribution of destination ports among the received packets from AS396982 appears very even, with no destination port being used significantly more often than others. However, the proportion of *Client Hello* packets among the analyzed payloads is the highest.

Table 9.12: Overview of top 5 destination ports with binary payload (AS396982).

Packet type	Ports [#]	Share[%]
TLS Client Hello	8003, 8089, 54528, 42713, 52590, 2080, 8066	20.4
HTTP	80, 8080	5.8
SMTP Data Fragment	587	2.9
PSQL	5432	2.2
XMPP (jabber) chat protocol	5222	1.5
Other	-	67.2

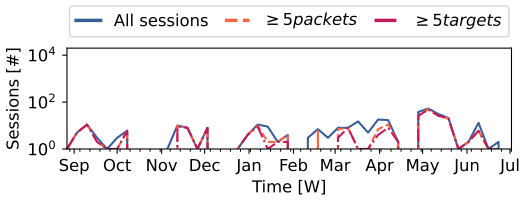


Figure 9.17: Number of sessions per week for all sessions and sessions with ≥ 5 packets and with ≥ 5 targets (AS2637).

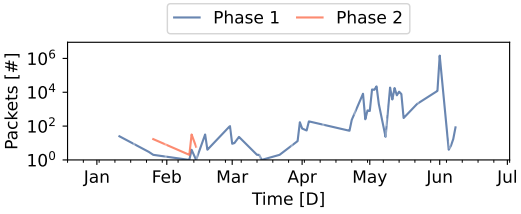


Figure 9.18: Comparison of daily packet counts for phase 1 and phase 2 (AS2637).

9.3.7 AS2637 Georgia Institute of Technology

AS2637 is associated with the *Georgia Institute of Technology* in Atlanta, Georgia. It is an educational institution. AS2637 emits a total of 1.7M packets from 7 source IP addresses within 337 sessions. In this process, 1.4M destination addresses receive at least one packet. Various address types are scanned, including *low-byte*, *::0*, *embedded-port*, *embedded-ipv4*, *pattern-bytes*, randomized and ieee-derived. As soon as Spoki is active, 14 TCP ACK packets and 1.6M TCP SYN packets are received from AS2637.

In contrast to most other ASes, Figure 9.17 shows that both the number of packets per session (≥ 5) and the number of targets (≥ 5) are relatively high and appear in most sessions of the respective week.

Figure 9.18, however, clearly shows that while many irregular SYN packets are received, packets from the second phase only follow for a short period, from late January to mid-February. It is possible that the two-phase scan was not the main focus here. However, as many TCP SYN packets were sent, this could indicate that AS2637 is focussing on TCP scans.

A total of 42 packets were received in the second phase following the regular SYN. Of these, 24 packets (57%) contain payloads that can be decoded in ASCII (Table 9.13). Eight packets (19%) still have a binary payload (Table 9.14), while ten packets (24%) contain no payload. In this phase, only one destination address is targeted, which seems to be chosen at random. No clear reason or structure can be identified for the selection of this destination address. Furthermore, only four destination ports (80, 7547, 443, 22) are identified among the ASCII-decoded payloads in this phase. Port 22 is again

Table 9.13: Overview of ASCII payload (AS2637).

Payload	Share[%]	Number of distinct ports [#]
GET / HTTP/1.1	50	3
SSH-2.0-	50	1

associated with the payload *SSH-2.0-*, while the other payloads primarily occur in the *GET* requests. All the *GET* requests utilizing the *ZGrab* user agent[26].

In the remaining eight packets with binary payloads, only port 443 is observed as the destination port with the payload *TLS Client Hello*.

Table 9.14: Overview of top 5 destination ports with binary payload (AS2637).

Packet type	Ports [#]	Share[%]
TLS Client Hello	443	100

9.4 Summary of Findings

The identified two-phase scanners in IPv6 display both similarities and differences in their scanning behavior. Most scanners focus on a single destination address, often targeting *low-byte* addresses or the address with the DNS entry. However, some scanners distribute their scans across multiple destination addresses. During these scans, a wide range of destination ports is often targeted, and numerous payloads are sent in the second phase. While no malicious downloaders are observed, several vulnerability checks are detected, suggesting potential attempts to identify weaknesses.

10 Discussion

This section further explores scanner reactions to the telescope properties, which have not yet been fully examined or were only briefly discussed previously. Additional thoughts and considerations regarding the results will also be discussed in this section. Furthermore, the key findings will be revisited and explained in more detail.

[26]https://github.com/zmap/zgrab2/blob/76d09b59c5ec1b20fcc0a172d84df99802865250/modules/http/scanner.go#L46

10.1 Influence of the Telescope Properties

This subsection addresses the question of which telescope properties truly influence scanning behavior. Since some aspects have received little attention so far and have not been directly compared, this subsection aims to provide a focused analysis and comparison.

10.1.1 Activity

The fact that the active /56 subnet in T2 has been active for many years likely leads to scanners scanning outside of T2's subnet as well. To estimate this, a comparison could be made before the split period when Spoki was not yet active and T1 was completely passive. T2 appears to see significantly more sources than T1. Even though T1 is significantly larger, without considering the destination address with the DNS entry, 3245 source IP addresses are scanning T2, while only 1387 source IP addresses are scanning T1. However, many /128 source addresses in T2 come from fewer /64 sources. In fact, T1 captures more /64 sources than T2 when the address with the DNS entry is not taken into account. Still, T2 identifies more ASNs than T1. 796k destination addresses of T1 receive at least one packet, while 714k destinations of T2 receive at least one packet. Since T1 was completely passive at this time, it still receives packets from a similar number of /64 sources and ASNs as T2. Additionally, more destination addresses receive packets, which may be due to the fact that T1 has many more destination addresses, as it is a /32. Therefore, it can be concluded that an active subnet likely has an influence on attracting traffic in the remaining prefix, although this impact is difficult to analyze and determine precisely.

10.1.2 Prefix Size

The results of the *network selection* indicate that only 2% of scan sessions are dependent on prefix size. Therefore, it can be concluded that while some IPv6 scanners do operate based on prefix size, this approach is relatively uncommon in the observed traffic.

10.1.3 DNS Entry

During the measurement period, 18k (68%) of source IP addresses exclusively scan the destination address with the DNS entry, while 8k (32%) of sources do not send any

packets to this address at all. Only five IP addresses (0.02%) send packets both to this address and to other destination addresses. Addresses that only target the DNS address make up a large part of the traffic. Packets from 37 source IP addresses target this DNS address across at least five different TCP destination ports. 31 source IP addresses even use at least 30 different TCP destination ports. Of these, 13 source addresses use at least 100 TCP destination ports, seven IPs use at least 350 destination ports, and one IP address from AS37963 sends packets with a total of 1k different TCP destination ports to this address. These results show that an address with a DNS entry can attract targeted network traffic. Scanners are attracted that specifically target such destination addresses, and among them are scanners that use many different TCP destination ports and payloads. The analyses also show that scanners look for different targets. Specifically, they search for either addresses with DNS entries or other attractors.

10.1.4 Reactivity

T2 has been responding to TCP requests since December 19, 2023, while T4 has responded to TCP requests since the beginning of the measurement period. Both telescopes use the reactive network telescope Spoki. No two-phase scans are observed at T4. However, T2 shows a clear reaction from scanners to this feature. After the activation of Spoki, certain scanners increase the number of packets they send, including some that initiate both a first (stateless) phase and a second (stateful) phase to intensify their scans. This group includes scanners that use a wide range of target ports and payloads. Although the exact reasons for this behavior are unclear, some requests appear to be related to security vulnerability tests. Comparing the time before the activation of Spoki at T2 with the same period afterwards reveals an increase in packets, source IP addresses, and sessions. The detailed results are shown in Table 10.1.

Table 10.1: Impact of Spoki on network traffic.

Time range	Total packets [#]	Source IP Addresses[#]	Sessions [#]
116 days before Spoki is active in T2	2 894 046	7391	202 661
116 days during which Spoki is active in T2	2 931 354	10 824	260 693

10.1.5 BGP Announcements

Since the beginning of the measurements, it has become apparent that T1 and T2, which are announced in BGP, receive significantly more packets compared to T3 and T4, which are not separately announced. The split and announcement experiment with T1 shows that the number of packets and sessions increases with each additional announcement. The *network selection* also reveals different scanning strategies. Some scanners target multiple prefixes with the same number of packets, while others focus on just one prefix per announcement phase (two weeks). There are also scanners where the size of the prefix determines the amount of received packets. From these results, it can be concluded that there are IPv6 scanners that specifically target prefixes announced in BGP.

10.1.6 Route6 Objekt

No noticeable changes are seen after creating the Route6 object. There is no significant increase in sources for the /33 prefix.

10.1.7 Appearance on Hitlist

After a prefix is announced, it appears on the public TUM hitlist around two weeks later. However, no noticeable changes in scanner behavior can be observed after its appearance on the TUM hitlist. Since more traffic is observed even before the publication on the hitlist, the BGP announcements seem to have a stronger influence on traffic than the hitlist entry itself.

10.1.8 Open Questions

There are clear indications of reactions to BGP announcements, the reactive network telescope, and the address with the DNS entry. Regarding BGP announcements, the /36 prefix appears to attract slightly more activity, in terms of observed ASNs and scanned targets, compared to other announced prefixes. This raises the question of whether scanners are specifically targeting these prefixes or if these observations are just a coincidence in this study.

Another investigation could involve deploying Spoki without being included on the aliased

prefix list. Instead of replying with all addresses in the prefix, responses could be limited to specific *low-byte* addresses to determine whether this attracts more two-phase scanners or reveals additional insights about them.

A further experiment could involve reversing the BGP announcement approach. This would mean initially announcing /33 to /48 prefixes and, every two weeks, combining the most-specific prefixes. Such an approach would help assess whether scanners persist in targeting previously unannounced prefixes or adjust their behavior based on new announcements, effectively *forgetting* the unannounced ranges.

10.2 Limited Perspective: Constraints of IPv6 Telescope Observations

The telescopes only capture a heuristically determined small part of the IPv6 address space and do not have a complete overview of the address space. Therefore, only the scanners that are actually observed can be analyzed. Statements about the telescope properties can also only be made for the specific context in which they were used. Additionally, analysis is conducted within the measurement period. Different results could be observed at other times. The properties of the telescopes can have various influences that lead to the attraction of certain types of traffic. For example, reacting to TCP traffic can lead to attracting more TCP traffic than if this was not the case. If effective methods to attract scanners are not employed, it is very likely that only minimal traffic will be observed.

In this study, all data are considered in the analyses. This has advantages and disadvantages. Although showing all packets gives a complete picture, heavy hitters can greatly affect the results. For this reason, the analyses often present the number of sessions or the number of source addresses. However, many sessions can originate from a single source, or many source IP addresses could come from a /64 network, which can also have a considerable impact on the data. Therefore, the source addresses are aggregated for some analyses and compared in various ways. It should be noted that a /64 only represents a standardized semantics. There are different semantics in many policies. Given the many representations, it is not easy to consider all influencing factors in a plot or table, and this should always be taken into account.

When creating the taxonomy, it is essential to ensure that it provides a clearer picture

of scanning behavior. However, it is challenging to define fixed categories for varying behaviors. For example, in *temporal behavior*, a scanner could send packets regularly within an hour over multiple days, resulting in only one session overall. Another scanner could send packets in the same timeframe but generate multiple sessions. According to established criteria, the scanner with one session would be classified as *oneoff*, while the other would not. If a different timeframe were set, the categorization might change. So, how should this timeframe be correctly determined? Another example is scanners that probe many prefixes simultaneously, which can lead to extended timeframes for packet reception per prefix. This factor should also be considered when determining the timeframe. The *random* category in *address selection* presents a different challenge. Even with standardized NIST tests, true randomness cannot be proven with 100% certainty. Randomness can be ruled out but not confirmed with certainty. Therefore, while these tests are helpful for gaining a better overview of scan traffic, they cannot guarantee a completely reliable result.

Observations in telescopes are not unbiased, as scanners often do not scan randomly. Instead, their behavior is influenced by external factors that direct them to target specific network areas. Such factors can include BGP announcements, hitlists, or DNS entries. Scanners that respond to these triggers cause biases in the recorded network traffic, as they are directed to scan particular areas.

11 Conclusion and Outlook

This study examines how the properties of four network telescopes influence the observed network traffic. The telescopes analyzed vary in their visibility, activity, reactivity, and attractors. We observe that large prefix sizes are not necessarily required. Instead, it is more important to announce prefixes in BGP and prepare them to the specific type of scanner being analyzed. Furthermore, a specialized analysis identifies two-phase scanners in IPv6 that frequently target specific hosts (*i*) low-byte address, (*ii*) address with DNS entry and send packets with different payloads to them. This often reveals security checks.

These findings suggest new measurement tasks for the future. (*i*) It should provide a clearer view of biases in network telescopes caused by attractors. (*ii*) Other attractors should be identified, and their impact on network traffic patterns should be further studied. (*iii*) Expanding the BGP experiment could provide more insights by first keeping

all announced prefixes (/32 to /48) and then combining the most-specific prefixes into larger ones every two weeks. For instance, two /48 prefixes could be combined into a /47. Using this method, it could be determined whether scanners still scan the previously announced prefix or, following the new announcement, shift their focus to the newly announced prefixes, effectively *forgetting* the previous one. (*iv*) Spoki could be used in the IPv6 area for behavioral analysis in scanning campaigns.

References

[1] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, and David Watson. 2005. The Internet Motion Sensor: A Distributed Blackhole Monitoring System. In *Proc. Of NDSS*. Internet Society, San Diego, CA, USA, 1–13.

[2] C. Bao, C. Huitema, M. Bagnulo, M. Boucadair, and X. Li. 2010. *IPv6 Addressing of IPv4/IPv6 Translators*. RFC 6052. IETF. https://doi.org/10.17487/RFC6052

[3] Lawrence Bassham, Andrew Rukhin, Juan Soto, James Nechvatal, Miles Smid, Elaine Barker, Stefan Leigh, Mark Levenson, Mark Vangel, David Banks, Nathanael Alan Heckert, James Dray, and San Vo. 2010. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. Special Publication NIST SP 800-22. National Institute of Standards & Technology, Gaithersburg, MD, US.

[4] Robert Beverly. 2016. Yarrp'ing the Internet: Randomized High-Speed Active Topology Discovery. In *Proc. of ACM IMC*. ACM, New York, NY, USA, 413–420. https://doi.org/10.1145/2987443.2987479

[5] Robert Beverly, Ramakrishnan Durairajan, David Plonka, and Justin P. Rohrer. 2018. In the IP of the Beholder: Strategies for Active IPv6 Topology Discovery. In *Proc. of ACM IMC (IMC '18)*. ACM, New York, NY, USA, 308–321. https://doi.org/10.1145/3278532.3278559

[6] Tim Breitenbach, Bartosz Wilkusz, Lauritz Rasbach, and Patrick Jahnke. 2023. On a Method for Detecting Periods and Repeating Patterns in Time Series Data with Autocorrelation and Function Approximation. *Pattern Recognition* 138 (2023), 1–22.

[7] Dominik Charousset, Raphael Hiesgen, and Thomas C. Schmidt. 2016. Revisiting Actor Programming in C++. *Computer Languages, Systems & Structures* 45 (April 2016), 105–131. http://doi.org/10.1016/j.cl.2016.01.002

[8] Xue Chen, Weiwei Shi, Jieyan Liu, Mengshu Hou, and Yujun Li. 2023. 6Community: An Active IPv6 Address Detection Method Based On Community Discovery Algorithm. In *ADMIT 2023*. ACM, New York, NY, USA, 114–119. https://doi.org/10.1145/3625403.3625426

[9] Tianyu Cui, Gaopeng Gou, and Gang Xiong. 2020. 6GCVAE: Gated Convolutional Variational Autoencoder For IPv6 Target Generation. In *2020 24th PAKDD*. Springer, Springer International Publishing, Cham, 609–622.

[10] Tianyu Cui, Gaopeng Gou, Gang Xiong, Chang Liu, Peipei Fu, and Zhen Li. 2021. 6GAN: IPv6 Multi-Pattern Target Generation Via Generative Adversarial Nets With Reinforcement Learning. In *IEEE INFOCOM 2021*. IEEE, IEEE, Piscataway, NJ, USA, 1–10.

[11] Tianyu Cui, Gang Xiong, Gaopeng Gou, Junzheng Shi, and Wei Xia. 2020. 6VECLM: Language Modeling In Vector Space For IPv6 Target Generation. In *ECML PKDD 2020*. Springer, Springer International Publishing, Cham, 192–207.

[12] Jakub Czyz, Kyle Lady, Sam G. Miller, Michael Bailey, Michael Kallitsis, and Manish Karir. 2013. Understanding IPv6 internet background radiation. In *Proc. of the ACM IMC* (Barcelona, Spain). ACM, New York, NY, USA, 105–118. https://doi.org/10.1145/2504730.2504732

[13] Zakir Durumeric, Michael Bailey, and J. Alex Halderman. 2014. An Internet-Wide View of Internet-Wide Scanning. In *Proc. of the 23rd USENIX Conference on Security Symposium* (San Diego, CA). USENIX Assoc., Berkeley, CA, USA, 65–78.

[14] Zakir Durumeric, Eric Wustrow, and J. Alex Halderman. 2013. ZMap: Fast Internet-Wide Scanning and its Security Applications. In *Proc. of the 22nd USENIX Security Symposium*. USENIX Assoc., Berkeley, CA, USA, 605–620.

[15] Mat Ford, J. Stevens, and John Ronan. 2006. Initial Results from an IPv6 Darknet. In *Proc. of the ICISP*. IEEE, Piscataway, NJ, USA, 13–13. https://doi.org/10.1109/ICISP.2006.14

[16] Kensuke Fukuda and John Heidemann. 2018. Who Knocks at the IPv6 Door? Detecting IPv6 Scanning. In *Proc. of IMC* (Boston, MA, USA) *(IMC '18)*. ACM, New York, NY, USA, 231–237. https://doi.org/10.1145/3278532.3278553

[17] F. Gont and T. Chown. 2016. *Network Reconnaissance in IPv6 Networks*. RFC 7707. IETF. https://doi.org/10.17487/RFC7707

[18] Robert Graham. 2013. MASSCAN. https://github.com/robertdavidgraham/masscan.

[19] Raphael Hiesgen, Marcin Nawrocki, Marinho Barcellos, Daniel Kopp, Oliver Hohlfeld, Echo Chan, Roland Dobbins, Christian Doerr, Christian Rossow, Daniel R. Thomas, Mattijs Jonker, Ricky Mok, Xiapu Luo, John Kristoff, Thomas C. Schmidt, Matthias Wählisch, and KC Claffy. 2024. The Age of DDoScovery: An Empirical Comparison of Industry and Academic DDoS Assessments. In *Proc. of ACM Internet Measurement Conference (IMC)*. ACM, New York, 259–279. https://doi.org/10.1145/3646547.3688451

[20] Raphael Hiesgen, Marcin Nawrocki, Alistair King, Alberto Dainotti, Thomas C. Schmidt, and Matthias Wählisch. 2022. Spoki: Unveiling a New Wave of Scanners through a Reactive Network Telescope. In *Proc. of 31st USENIX Security Symposium*. USENIX Association, Berkeley, CA, USA, 431–448. https://www.usenix.org/system/files/sec22-hiesgen.pdf

[21] Raphael Hiesgen, Marcin Nawrocki, Thomas C. Schmidt, and Matthias Wählisch. 2022. The Race to the Vulnerable: Measuring the Log4j Shell Incident. In *Proc. of Network Traffic Measurement and Analysis Conference (TMA)* (Enschede, Netherlands). IFIP, Laxenburg, MD, Austria, 1–9. https://tma.ifip.org/2022/wp-content/uploads/sites/11/2022/06/tma2022-paper40.pdf

[22] Raphael Hiesgen, Marcin Nawrocki, Thomas C. Schmidt, and Matthias Wählisch. 2024. The Log4j Incident: A Comprehensive Measurement Study of a Critical Vulnerability. *IEEE Transactions on Network and Service Management (TNSM)* (2024). https://doi.org/10.1109/TNSM.2024.3440188

[23] R. Hinden and B. Haberman. 2005. *Unique Local IPv6 Unicast Addresses*. RFC 4193. IETF. https://doi.org/10.17487/RFC4193

[24] Bingnan Hou, Zhiping Cai, Kui Wu, Jinshu Su, and Yinqiao Xiong. 2021. 6Hit: A Reinforcement Learning-Based Approach To Target Generation For Internet-Wide IPv6 Scanning. In *IEEE INFOCOM 2021*. IEEE, IEEE, Piscataway, NJ, USA, 1–10.

[25] Bingnan Hou, Zhiping Cai, Kui Wu, Tao Yang, and Tongqing Zhou. 2023. 6Scan: A High-Efficiency Dynamic Internet-Wide IPv6 Scanner With Regional Encoding. *IEEE/ACM Transactions on Networking* 31, 4 (2023), 1870–1885.

[26] Yuchen Huang, Michael Rabinovich, and Rami Al-Dalky. 2020. FlashRoute: Efficient Traceroute on a Massive Scale. In *Proc. of ACM IMC (IMC '20)*. ACM, New York, NY, USA, 443–455. https://doi.org/10.1145/3419394.3423619

[27] ChenHuan Liu, ShanShan Hao, QianKun Liu, CongXiao Bao, and Xing Li. 2021. IPv6-Network Telescope Network Traffic Overview. In *2021 IEEE 11th ICEIEC*. IEEE, Piscataway, NJ, USA, 1–4. https://doi.org/10.1109/ICEIEC51955.2021.9463724

[28] Ning Liu, Chunbo Jia, Bingnan Hou, Changsheng Hou, Yingwen Chen, and Zhiping Cai. 2023. 6Search: A Reinforcement Learning-Based Traceroute Approach For Efficient IPv6 Topology Discovery. *Computer Networks* 235 (2023), 1–10.

[29] Qiankun Liu and Xing Li. 2023. 6Former: Transformer-Based IPv6 Address Generation. In *ISCC 2023*. IEEE, IEEE, Piscataway, NJ, USA, 1142–1148.

[30] Zhizhu Liu, Yinqiao Xiong, Xin Liu, Wei Xie, and Peidong Zhu. 2019. 6Tree: Efficient Dynamic Discovery Of Active Addresses In The IPv6 Address Space. *Computer Networks* 155 (2019), 31–46.

[31] Jack C. Louis. 2004. Unicornscan. https://sourceforge.net/projects/osace/files/unicornscan/.

[32] Austin Murdock, Frank Li, Paul Bramsen, Zakir Durumeric, and Vern Paxson. 2017. Target Generation for Internet-wide IPv6 Scanning. In *Proc. of IMC*. ACM, New York, NY, USA, 242–253. https://doi.org/10.1145/3131365.3131405

[33] T. Narten, R. Draves, and S. Krishnan. 2007. *Privacy Extensions for Stateless Address Autoconfiguration in IPv6*. RFC 4941. IETF. https://doi.org/10.17487/RFC4941

[34] RIPE NCC. 2024. RIPEstat. https://stat.ripe.net/ui2013/

[35] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. 2004. Characteristics of internet background radiation. In *Proc. of the 4th ACM SIGCOMM Conference on Internet Measurement (IMC '04)*. ACM, New York, NY, USA, 27–40. https://doi.org/10.1145/1028788.1028794

[36] Philipp Richter and Arthur Berger. 2019. Scanning the Scanners: Sensing the Internet from a Massively Distributed Network Telescope. In *Proceedings of the Internet Measurement Conference* (Amsterdam, Netherlands) *(IMC '19)*. Association for Computing Machinery, New York, NY, USA, 144–157.

[37] Philipp Richter, Oliver Gasser, and Arthur Berger. 2022. Illuminating Large-Scale IPv6 Scanning in the Internet. In *Proc. of the ACM IMC*. ACM, New York, NY, USA, 410–418. https://doi.org/10.1145/3517745.3561452

[38] RIPE NCC. 2010. What is RIPE Atlas? https://atlas.ripe.net/about/

[39] John Ronan and David Malone. 2023. Revisiting and Revamping an IPv6 Network Telescope. In *2023 34th ISSC*. IEEE, Piscataway, NJ, USA, 1–6. https://doi.org/10.1109/ISSC59246.2023.10162033

[40] Matthew Roughan, Walter Willinger, Olaf Maennel, Debbie Perouli, and Randy Bush. 2011. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *IEEE Journal on Selected Areas in Communications* 29, 9 (2011), 1810–1821.

[41] Guanglei Song, Jiahai Yang, Zhiliang Wang, Lin He, Jinlei Lin, Long Pan, Chenxin Duan, and Xiaowen Quan. 2022. DET: Enabling Efficient Probing Of IPv6 Active Addresses. *IEEE/ACM Transactions on Networking* 30, 4 (2022), 1629–1643.

[42] Lion Steger, Liming Kuang, Johannes Zirngibl, Georg Carle, and Oliver Gasser. 2023. Target Acquired? Evaluating Target Generation Algorithms for IPv6. In *Proc. of TMA*. IEEE, Piscataway, NJ, USA, 1–10.

[43] Stephen D Strowes, René Wilhelm, Florian Obser, Riccardo Stagni, Agustín Formoso, and Emile Aben. 2020. Debogonising 2a10::/12 Analysis of one week's visibility of a new /12. In *Proc. of TMA*. IFIP, Laxenburg, MD, Austria, 1–9.

[44] Hammas Bin Tanveer, Rachee Singh, Paul Pearce, and Rishab Nithyanand. 2023. Glowing in the Dark: Uncovering IPv6 Address Discovery and Scanning Strategies in the Wild. In *Proc. of the USENIX Security Symposium*. USENIX Association, Anaheim, CA, 6221–6237. https://www.usenix.org/conference/usenixsecurity23/presentation/bin-tanveer

[45] Telia. 2024. Telia Looking Glass. https://lg.telia.net/

[46] F. Templin. 2013. *Operational Guidance for IPv6 Deployment in IPv4 Sites Using the Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)*. RFC 6964. IETF. https://doi.org/10.17487/RFC6964

[47] F. Templin, T. Gleeson, and D. Thaler. 2008. *Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)*. RFC 5214. IETF. https://doi.org/10.17487/RFC5214

[48] Johanna Ullrich, Peter Kieseberg, Katharina Krombholz, and Edgar Weippl. 2015. On Reconnaissance with IPv6: A Pattern-Based Scanning Approach. In *10th International Conf on Availability, Reliability and Security*. IEEE, Piscataway, NJ, USA, 186–192.

[49] Chaoqiang Yang, Liancheng Zhang, Yi Gou, Wenhao Xia, Ming Hu, and Jichang Wang. 2023. Trace6: A Practical Threatener Traceback Model in IPv6 Network. In *2023 19th International Conference on Mobility, Sensing and Networking (MSN)*. IEEE, Washington, DC, USA, 780–785. https://doi.org/10.1109/MSN60784.2023.00114

[50] Tao Yang, Zhiping Cai, Bingnan Hou, and Tongqing Zhou. 2022. 6Forest: An Ensemble Learning-Based Approach To Target Generation For Internet-Wide IPv6 Scanning. In *IEEE INFOCOM 2022*. IEEE, IEEE, Piscataway, NJ, USA, 1679–1688.

[51] Tao Yang, Bingnan Hou, Zhiping Cai, Kui Wu, Tongqing Zhou, and Chengyu Wang. 2022. 6Graph: A Graph-Theoretic Approach To Address Pattern Mining For Internet-Wide IPv6 Scanning. *Computer Networks* 203 (2022), 1–12. https://doi.org/10.1016/j.comnet.2021.108666

[52] Li Zhang and Shaowei Fu. 2024. 6MCBLM: Multi-scale CNN and BiLSTM-Attention Hybrid Model for IPv6 Target Generation. In *NNICE 2024*. IEEE, IEEE, Piscataway, NJ, USA, 499–505.

[53] Liang Zhao, Satoru Kobayashi, and Kensuke Fukuda. 2024. Exploring the Discovery Process of Fresh IPv6 Prefixes: An Analysis of Scanning Behavior in Darknet and Honeynet. In *Proc. of PAM (LNCS, Vol. 14537)*. Springer, Berlin Heidelberg, 95–111. https://doi.org/10.1007/978-3-031-56249-5_4

Erklärung zur selbstständigen Bearbeitung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne fremde Hilfe selbständig verfasst und nur die angegebenen Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quellen kenntlich gemacht.

Ort	Datum	Unterschrift