Have you SYN me? Characterizing Ten Years of Internet Scanning

Griffioen, Harm; Koursiounis, Georgios; Smaragdakis, Georgios; Doerr, Christian

DOI
10.1145/3646547.3688409

Publication date
2024

Document Version
Final published version

Published in
IMC 2024 - Proceedings of the 2024 ACM Internet Measurement Conference

Have you SYN me? Characterizing Ten Years of Internet Scanning

Harm Griffioen
Delft University of Technology
Delft, The Netherlands

Georgios Koursiounis
Delft University of Technology
Delft, The Netherlands

Georgios Smaragdakis
Delft University of Technology
Delft, The Netherlands

Christian Doerr
Hasso Plattner Institute
Potsdam, Germany

ABSTRACT

Port scanning is the de-facto method to enumerate active hosts and potentially exploitable services on the Internet. Over the last years, several studies have quantified the ecosystem of port scanning. Each work has found drastic changes in the threat landscape compared to the previous one, and since the advent of high-performance scanning tools and botnets a lot has changed in this highly volatile ecosystem.

Based on a unique dataset of Internet-wide scanning traffic collected in a large network telescope, we provide an assessment of Internet-wide TCP scanning with measurement periods in the last 10 years (2015 to 2024). We collect over 750 million scanning campaigns sending more than 45 billion packets and report on the evolution and developments of actors, their tooling, and targets. We find that Internet scanning has increased 30-fold over the last ten years, but the number and speed of scans have not developed at the same pace. We report that the ecosystem is extremely volatile, where targeted ports and geographical scanner locations drastically change at the level of weeks or months. We thus find that for an accurate understanding of the ecosystem we need longitudinal assessments. We show that port scanning becomes heavily commoditized, and many scanners target multiple ports. By 2024, well-known scanning institutions are targeting the entire IPv4 space and the entire port range.

CCS CONCEPTS

• **Security and privacy → Network security**.

KEYWORDS

Network Telescope, Internet Scanning.

ACM Reference Format:
Harm Griffioen, Georgios Koursiounis, Georgios Smaragdakis, and Christian Doerr. 2024. Have you SYN me? Characterizing Ten Years of Internet Scanning. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24), November 4–6, 2024, Madrid, Spain.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3646547.3688409

IMC '24, November 4–6, 2024, Madrid, Spain

ACM ISBN 979-8-4007-0592-2/24/11.
https://doi.org/10.1145/3646547.3688409

1 INTRODUCTION

When a new host connects to a public IP address, it takes only seconds for the first traffic to arrive. This unsolicited data mainly consists of port scanning, probing the machine for any services that might be open to the Internet, and is usually a precursor to a later attempt to exploit vulnerable hosts. With the easy availability of tools and the universal belief that port scanning is the necessary default for computer discovery and exploitation, it is not surprising that by now 98% of unsolicited TCP traffic consists of SYN scans.

This situation can be attributed to better tooling and an increased number of vantage points. When the high-performance scanning tools ZMap [21] and Masscan [26] were released in 2013 and 2014, respectively, algorithmic advances introduced by them enabled users to scan the entire Internet in minutes from a single IP address [36], a process that would have taken days or weeks using established software before. Only soon after, the Internet threat landscape fundamentally changed with the advent of the first IoT botnet Mirai [5], which from the get-go overshadowed previously seen distributed denial-of-service attack (DDoS) volumes by a factor of four. The hundreds of thousands of compromised IoT devices did not only drastically alter DDoS, but Mirai and its siblings made also a landslide shift in port scanning, as each device performs continuous worldwide scanning to spread the infection further [28]. Indeed, when we look at quantifications of Internet-wide scanning over the past decades, we see that the ecosystem has drastically changed. The assessments of Pang et al. in 2004 [44], Wustrow et al. in 2010 [53], Durumeric et al. in 2014 [18], as well as Richter and Berger in 2019 [45] show drastic increases in traffic volume, actors involved and capabilities of these actors. While these papers show a very dynamic ecosystem, they cannot explain the dynamics and actual developments of the threat landscape over the years.

In this paper, we demonstrate this volatility using a dataset that is collected using a large network telescope of three partially populated /16 address blocks over 10 years (2015-2024). Our results uncover the steady increase in unsolicited traffic on the Internet, and show that the port scanning ecosystem is so volatile that quantifications at single moments in time may result in significant under- or overestimations of specific tooling, targeted ports, and actor groups. Using more than *45 billion scan packets* from over *45 million sources* between 2015 and 2024, we quantify the evolution of port scanning and make the following key contributions:

- We provide an overview of 10 years of Internet scanning traffic from 2015 to 2024.
- We show that over the last ten years, Internet scanning has increased 30-fold. Where in 2015 we observed 11 million scanning

packets in our dataset every day, this number has increased to 345 million in 2024.
- We show that the ecosystem of Internet-wide scanning is extremely volatile, with targeted ports and major campaigns drastically changing even at the level of weeks or months. This means that research results on the Internet threat landscape are very temporary, and, depending on the exact moment the quantification was performed, could be largely over- or underestimating certain aspects. Thus, for an accurate understanding of the ecosystem, long-term assessments are essential.
- We demonstrate that although the disclosure of vulnerabilities triggers a sudden influx of port scanning to discover affected devices as reported by [18], *in the long term* these trends do not continue and activity quickly dies down in a matter of weeks.
- We identify an increase in the number of collaborative scans, and scans targeting a large number of ports. By 2024, multiple organizations cover the entire port range, a feat that was not seen in 2015.

2 RELATED WORK

As a de-facto technique for service discovery, Internet-wide port scanning campaigns have existed for many years, and are widely used as a tool both by the academic community as well as malicious actors, albeit for different purposes. In academic research, port scanning is mainly used to collect data on systems vulnerable to certain exploits [20, 30, 41, 42], analyze system configurations [10, 13, 19, 24, 31] and do surveys of which ports are generally exposed to the Internet [16]. According to [32], over 300 papers utilized ZMap. In a parallel work to ours [17], the original creators and developers of ZMap quantify ZMap's adoption since its release in 2013, characterize its usage by researchers and cybersecurity companies, and share lessons and experiences from releasing and maintaining the ZMap code.

Aside from academic questions of studying the Internet, there are also nefarious use cases, the most common one being the discovery of hosts and available services for later exploitation. Several papers have investigated the ecosystem of port scanning in general and characterized the utilized tools and platforms. Lee et al. [35] provide an empirical analysis of scanning behavior and find that 91% of port scanners target IP addresses sequentially. Pang et al. [44] additionally find that port scanning is highly targeted to certain ports. Durumeric et al. [18] show that the high-level metrics like the origin of scans remained constant, but also identify that there are large changes since previous studies such as drastic changes in targeted ports and a major surge in scanning traffic due to the advent of new tools that make Internet scanning more accessible.

Scanning the Internet from one vantage point however used to be impracticably slow. This has changed due to two major developments, which allow that the entire IPv4 space can be enumerated and tested in a matter of hours. First, new tools such as ZMap and Masscan were released that due to several algorithmic improvements – such as direct packet injection into the OS networking stack or by omitting the need to keep local state – can efficiently scan above Gigabit/second speed [1]. Second, scans became massively distributed, either farmed out to compromised PCs as part of a botnet [46], integrated into malware for Internet routers and other IoT devices [5, 28] or through instances run in public clouds. Bou-Harb et al. [11] provide an overview of how these scanning campaigns can work in practice, and the techniques that can be used to perform such distributed scans. Dainotti et al. [15] show that port scanning campaigns are indeed conducted with many hosts to minimize the chance of being detected and maximize the effectiveness of the scan. A large body of research focuses on detection of scanning probes from mainly single-source scanners [4, 6, 22, 23, 50]. Identifying large distributed campaigns is not yet applied at scale [14, 28, 47].

Ghiette et al. [25] identify a large bias in how well-known tools are used along with a large geographical bias in tool usage. Large biases also exist in scans targeting certain ports, with 77% of scans to Microsoft Remote Desktop Protocol (RDP) originating from China in 2014 [18]. Richter and Berger [45] show that only a small fraction of scans actively target the entire IPv4 space, but that these scans account for more than 27% of all scanning traffic due to their size. Durumeric et al. [18] have also identified this imbalance, with 0.28% of scans generating nearly 80% of the traffic. Wan et al. [52] studied how scan origin affects Internet-wide scan results by completing three popular types of scans from geographically and topologically diverse networks.

Researchers have noted temporal differences in the ecosystem when comparing their work against the existing body of knowledge: Scan targets are volatile and change over the years [19, 44, 53], set of countries where scanning traffic originates from changes drastically [18, 45], and new exploits lead to a large increase in the number of probes on a port [18]. As these studies are carried out on different infrastructures in different geographical regions and with varying number of probes, it does not conclusively show the changes in the scanning ecosystem. Our study therefore focuses on the evolution of scanning traffic using a constant vantage point.

Previous work has also shown that individual events have a large impact on the scanning ecosystem, where a major part of all scanning traffic originates from for example a single botnet. Durumeric et al. [18] show that botnets such as the Conficker worm [46] can be primarily responsible for the distribution of scanners in a certain time period, with 41.7% of the scans recorded in [18] directly attributable to the Conficker worm. Similarly, Griffioen and Doerr [28] show that 87% of all telnet traffic could be attributed to variants of the Mirai botnet. These large influences from single causes raise the question of whether port scanning data contains large trends and evolutions, or whether it is a random chain of events.

In 2007, Allman et al. [2] showed trends in scan activity from 1994 and 2006, also identifying temporal scan activities such as scans for the Sasser backdoor, receiving heavy traffic for only a short period. The work by Mazel et al. [38] is the closest work to this study, as the authors analyze 15 minutes of scanning traffic each day from 2001 to 2016 where the authors find an increase of Internet-wide scans, a highly volatile port distribution, but do not report on scanning speeds, coverage, or on the distribution of tools used. As the authors can only probe the network 15 minutes a day, there is a large blind-spot of scans that fall outside of this time period. In our study, we include all scans in a day to better understand the entire ecosystem.

While other works have identified changes in the scanning landscape e.g., [3, 7, 45], over time, a mapping of the ecosystem from

a single vantage point that considers the full lifetime of scans and the evolution of tool usage is currently missing from related works. Additionally, the impact of research scans on the entire scanning landscape is currently unknown, especially the last year that high speed scanning [1] for a large number of ports and protocols is commoditized [32, 33]. We fill this gap by considering a dataset covering ten years of scanning traffic up to 2024, allowing us to understand evolution in scanning speeds and scanning coverage, characterizing differences between tools and the origin of scans, and differentiating between "benign" scans showing the impact of research scanning on the ecosystem.

3 METHODOLOGY

In this section, we present our methodology for collecting data about port scanning activity. Then, we show how we apply our methodology to a large telescope we operate to infer scanning and study the evolution of scanning over time. We also present our methodology to fingerprint different scanning tools using the data collected at our telescope and identify individual scanning campaigns.

3.1 Detecting Port Scanning Activity

To compromise a host and exploit a service, an adversary must first know of its existence. Network services are typically exposed via the transport layer protocols TCP and UDP, but usage of TCP far dominates in practice [7]. To establish a connection, a client would send a TCP control message with a *SYN* to a specific port. If the server has opened this port and accepts new connections, it will respond with a *SYN/ACK* message, or otherwise decline the connection request with a *RST*, instructing the client to reset its internal state. The client confirms the server's *SYN/ACK* with a final *ACK*nowledgement, and the connection is established.

To differentiate whether a port is open and accepting requests or closed, it is, however, not necessary to fully complete the handshake. The client already knows the port status after the server's *SYN/ACK* or *RST* and can, in practice, save overhead by not completing the handshake but stopping after the initial *SYN*, hence the name SYN scan. Various ways exist to understand whether a port is opened or closed, based on protocol and implementation differences in the TCP protocol [34]. For example, servers respond differently when receiving the request to close a non-existent connection (a FIN scan) without an established connection, an acknowledgment (ACK scan) to a packet that has never been sent, an invalid packet with all control bits set (an XMAS scan, as all "candles" are lit), or without any control bits (a NULL scan). In practice, although much of the popular "hacker" folklore frequently refers to these as especially subtle and stealthy forms of port scanning; more than 98% of TCP scans are SYN scans [8], and thus the focus of this paper.

3.2 Network Telescope Data Collection

Core to the work is network telescope data spanning ten years from three partially populated /16 networks, with the IP addresses routed through our network telescope roughly adding up to one full /16 network. As unused IP addresses are void of any user traffic, incoming data is either (a) Internet backscatter of ongoing attacks where the adversary has spoofed one of the telescope's IP addresses to mask the origin of the attack, or (b) scanning traffic from remote parties to identify active hosts and services. As actors are interested in receiving a response to their scanning probes, the addresses in scan probes are by definition not spoofed and thus point to the actual IP address of the perpetrator. We follow the standard practice of selecting only TCP frames with the SYN flag set [53] to separate backscatter from scans.

Over the ten years, we monitored on average 71,536 unrouted IP addresses. With the continuous increase in scanning, each year the collected amount grows, with over one Terabyte of raw network traffic being received per month in 2024. The telescope used for this study has had some outages over the years, ranging from operational windows where traffic was not routed to the telescope range, to data loss due to server failures. For this study, we select for every year a *continuous* range of at most 2 months of data in the first half of the year. This means that the datasets for each year span between 29 and 61 days of uninterrupted data. While this does not constitute to a full longitudinal analysis, it does ensure that trends that are measured in this study are not merely observed due to temporal overlaps in scanning traffic.

Due to operational policies, traffic targeting Samba (445/TCP) and Telnet (23/TCP) are completely blocked at the network ingress of the telescope since the advent of Mirai in 2016 [5]. This means that our dataset does not contain traffic to these two ports from 2017 onwards, and we therefore exclude these from the study. While we are still able to show trends in Internet-wide scanning activity, note that this is a lower bound that is influenced by these organizational policies and the location of the measurement infrastructure. For Mirai-based scanners targeting port 23, it is important to note that many of these would also scan port 2323 [28], meaning that we would see these scanners' activity regardless of the organizational policies.

3.3 Fingerprinting Scanning Tools

Although the concept of port scanning is simple, it is somewhat more complex to engineer a program that can efficiently scan the Internet at a fast pace. Depending on the algorithmic and implementation choices the programmer has made during the design, programs may exhibit slight behavioral differences, also because parts of the Internet and Transport layer protocols offer room for interpretation. For example how the packet is crafted, how certain header fields are populated, how settings about the connection are chosen etc. will mean that each tool has its own behavioral fingerprint on a network. Furthermore, part of the reason why high-speed port scanning tools can send probes at a high speed is that they do not save the state but embed a fingerprint into the outgoing packets to recognize the reply. These deliberate fingerprints are embedded by tools into packets, which we can thus use to fingerprint the tools themselves.

Both aspects provide a reliable signal to link port scanning traffic to a particular tool. In this paper, we will make this link based on the following features from previous literature:

Masscan initializes the IP Identification field of outgoing packets as a function of destination information and TCP header fields, thus for Masscan packets, the following equation holds $IPid = destIP \oplus destPort \oplus SeqNum$ [18].

Unicorn encodes source and destination host information in the TCP sequence number. We can test that two frames were sent by a host using Unicorn, if the following relation holds in two packets: $SeqNum_1 \oplus SeqNum_2 = destIP_1 \oplus destIP_2 \oplus srcPort_1 \oplus srcPort_2 \oplus ((destPort_1 \oplus destPort_2) << 16)$ [25].

NMap recognizes return packets based on embedded information, however it uses a *session secret* to obfuscate this information [25]. Cryptographically speaking, the information is encrypted using a stream cipher. As the stream is however reused (which means that the secret falls out), it is possible to deobfuscate and identify an NMap instance given two frames from the same host, if these two packets match the relationship $(SeqNum_1 \oplus SeqNum_2)\&0xFFFF = ((SeqNum_1 \oplus SeqNum_2) >> 16)\&0xFFFF$. This follows from the relation stated in the original paper [25]: $SeqNum_1 \oplus SeqNum_2 = (nfo_1||nfo_1) \oplus (nfo_2||nfo_2)$.

Mirai uses several specific features for its outgoing packets. Most prominently, it uses the destination IP address as the 32-bit sequence number of the TCP payload [27].

ZMap is a network scanner originally created for research scans, which marks its outgoing frames by setting the IP identification number to 54321 [21]. It is also characterizable by identifying the scan sequence of the probes [39]. Considering the size of our vantage point, calculating these sequences for every source is infeasible. We therefore rely on the fingerprint added by the ZMap authors.

3.4 Identifying Scanning Campaigns

As we would like to quantify the ecosystem of Internet-wide scans, it is necessary to group the individual scan packets received at different destination IP addresses together into the notion of a continuous scan that during its progression has hit the various targets. We will refer to such an activity in the following as a scan *campaign*. To map individual packets to campaigns, we utilize the fact that in order to receive the desired response, scanners will send probes from their actual and not a spoofed IP address.

To classify scan campaigns we extend the methodology and definition introduced by Durumeric et al. [18]: We define a scan as a sequence of probes, originating from one source address, that hit at least 100 distinct destination IP addresses in our network telescope at a minimum Internet-wide hitrate of 100 packets per second (pps). Based on the work by Moore et al. [40], we model our telescope using a geometric distribution to find that a scanner probing random IPv4 addresses at the rate of 100 pps will appear in our dataset within 1 hour with a probability of 99.9%. Thus, we expire scans that do not send any packets after 1 hour. This means that our analysis is sensitive to any campaign that has targeted at least 0.15% of the Internet at a rate of 100 pps, which is an adequate lower bound towards Internet-wide scanning. Previous work has used other bounds, such as capturing scans with 10 pps and expiring after 480 seconds [18]. Given the smaller size of our vantage point, it is crucial to define stricter scanning criteria to avoid noise and ensure that the detected scans reflect genuine Internet-wide behavior.

4 SCANNING ECOSYSTEM OVER THE YEARS

In our study, we observe over 45 billion SYN packets that could be bundled into more than 750 million campaigns originating from over 45 million distinct hosts. In the following, we will report on the developments of scan volume and actors over the years, targeted services, the origins of scans, as well as the tooling used by different types of actors. As related work has reported large temporal differences in scanning traffic, we need to understand how volatile the scanning landscape is to make accurate measurements. This section discusses the scanning ecosystem and the impact of large singular events.

4.1 Scanning traffic increased 30 fold over 10 years

Already when we look at high-level metrics of the scanning threat landscape, we see that over the years many major developments and shifts have taken place. Table 1 describes the ecosystem by several basic metrics, and we can see that over ten years the amount of Internet-wide scans hitting our measurement infrastructure has increased dramatically. The number of scans grew by a factor of 39, but scans got, in general, less intensive, increasing only 30 fold. We can furthermore observe that scanning activity is under significant flux. Indeed, back in 2015 the stock tool NMap was by far the most dominant player but much of Internet scanning was the result of custom-designed tooling. With the advent of IoT botnets, the ecosystem shifted in that Mirai became the dominant platform for scans not only on ports 23 and 2323 (the standard ports of Mirai [5]) but also on ports not directly related to Mirai such as 80, 443, and 8291. While botnets are often held responsible for coordinated Internet scanning and the increase in overall scanning traffic [49], we find that the total number of hosts scanning the Internet has been decreasing since 2017 while overall scanning traffic has increased. We instead find that as IoT botnets slowly started to decline [5, 28], recently developed high-performance scanners that started to rise where Masscan sends 81% of all scanning *traffic* targeting the network telescope. While the number of packets per day remains constant over the last years, the number of scans increases heavily in 2022 and 2024, where the intensity of scans drops proportionally to the growth. Given the increases in Internet speeds and the availability of high-performance tools such as ZMap [18], we would expect scans to become more intensive over the years. While we do see this increase in scanning traffic per single scan between 2015 and 2021, we also see a sudden change in 2022 where the number of scans increases drastically while the amount of packets observed in the network telescope does not increase at the same rate. We see that in the recent years, scans have become increasingly small, even though they are conducted using a high-performance scanning tool. This pattern indicates an increase in collaborating scanners, where multiple devices are used to scan the Internet [27]. In 2024, we find that the number of scans executed with ZMap has increased drastically. To understand whether this is an artifact of a large scan campaign or a shift in the scanning landscape, we do not only look at the average scanning traffic over the measurement period but also identify the amount of scans conducted using ZMap per day. Interestingly, we find that the *minimum* number of scans with ZMap per day in 2024 is 17,122, while this was 3,448 in 2023. Even more, the *maximum* number of scans observed on a single day in 2023 is 9,051, not even close to the number observed in 2024. We have verified these numbers using four single days in the first half

Table 1: Scan volume and five most targeted ports by packets and sources, and top tools and origins between 2015 and 2024. Port 23 and 445 were blocked at the network ingress in 2016 and we do therefore not include these ports in the general statistics.

		2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Packets/day		11 million	19 million	45 million	133 million	117 million	283 million	281 million	285 million	402 million	345 million
Top ports by packets	1	22 (15.0%)	22 (8.2%)	5358 (14.4%)	22 (3.1%)	22 (2.9%)	80 (1.0%)	6379 (1.4%)	22 (2.7%)	22 (1.8%)	3389 (2.2%)
	2	8080 (8.7%)	80 (6.0%)	7574 (12.1%)	8545 (1.4%)	80 (2.0%)	3389 (0.9%)	22 (1.3%)	80 (1.4%)	8080 (1.5%)	22 (1.8%)
	3	3389 (7.1%)	3389 (4.5%)	22 (11.2%)	3389 (1.1%)	8080 (1.8%)	81 (0.9%)	80 (1.1%)	443 (1.3%)	80 (1.5%)	80 (1.5%)
	4	80 (7.0%)	1433 (3.5%)	2323 (9.2%)	80 (0.9%)	81 (1.7%)	22 (0.8%)	3389 (0.8%)	2375 (1.3%)	3389 (1.3%)	443 (1.2%)
	5	443 (6.0%)	8080 (2.3%)	6789 (6.2%)	8080 (0.9%)	3389 (1.6%)	8080 (0.8%)	8080 (0.8%)	2376 (1.2%)	443 (1.1%)	8080 (1.2%)
Top ports by sources	1	10073 (33.0%)	21 (10.2%)	7545 (38.8%)	8291 (38.8%)	80 (30.4%)	80 (35.9%)	80 (46.0%)	80 (48.5%)	80 (30.6%)	80 (37.4%)
	2	3389 (11.3%)	3389 (9.6%)	2323 (25.3%)	2323 (10.4%)	8080 (30.3%)	8080 (30.4%)	8080 (42.0%)	8080 (41.9%)	8080 (27.1%)	8080 (29.0%)
	3	80 (5.82%)	20012 (5.2%)	5358 (11.5%)	21 (9.8%)	2323 (18.8%)	81 (13.2%)	5555 (13.5%)	5555 (13.0%)	52869 (17.7%)	443 (16.2%)
	4	8080 (2.7%)	80 (3.3%)	22 (8.0%)	21 (7.00%)	5555 (11.7%)	5555 (11.0%)	81 (9.8%)	81 (10.2%)	60023 (17.4%)	2323 (12.1%)
	5	22555 (2.0%)	8080 (1.4%)	23231 (7.4%)	22 (7.3%)	5900 (8.2%)	2323 (9.1%)	8443 (8.3%)	8443 (7.7%)	2323 (11.5%)	5900 (10.5%)
Top ports by scans	1	3389 (23.4%)	3389 (19.9%)	7547 (29.5%)	8291 (19.2%)	80 (20.2%)	80 (16.0%)	80 (13.6%)	80 (4.4%)	2323 (0.13%)	22 (0.10%)
	2	10073 (23.4%)	21 (6.8%)	2323 (25.1%)	21 (6.7%)	8080 (19.2%)	8080 (13.8%)	8080 (12.4%)	8080 (3.9%)	80 (0.12%)	80 (0.81%)
	3	80 (4.1%)	20012 (5.4%)	5358 (9.1%)	2323 (6.3%)	2323 (6.9%)	81 (4.6%)	5555 (3.0%)	5555 (1.0%)	443 (0.11%)	3389 (0.73%)
	4	8080 (2.7%)	80 (3.8%)	22 (5.7%)	22 (4.3%)	5555 (5.5%)	5555 (4.1%)	81 (1.8%)	81 (0.7%)	22 (0.10%)	443 (0.72%)
	5	443 (1.9%)	22 (1.9%)	6789 (5.4%)	3389 (4.1%)	5900 (3.9%)	2323 (2.8%)	8443 (1.6%)	8443 (0.5%)	8080 (0.10%)	8080 (0.72%)
Scans/month		33 K	38 K	252 K	137 K	238 K	222 K	290 K	777 K	727 K	1.3 M
Tools by scans	Masscan	0.5%	1.5%	0.7%	20.9%	21.9%	20.5%	25.1%	9.9%	0.2%	0.2%
	NMap	31.7%	12.8%	2.6%	3.2%	3.6%	5.0%	6.8%	2.3%	0.004%	0.006%
	Mirai-like	-	-	46.5%	19.2%	16.2%	14.9%	2.4%	1.0%	39%	5.3%
	ZMap	2.1%	9.1%	1.1%	4.7%	2.7%	13.1%	9.2%	3.7%	22%	59%

of 2023 and 2024 that are not part of the continuous measurement period for this study, and find the same relation in number of ZMap scans. While there is a significant increase in number of scans, we do not see the same for the number of packets sent in total in these scans. Instead, this number decreases while the number of hosts participating in these scans increases from 25,809 in 2023 to 41,038 in 2024. These findings can be caused by an increasing number of scanners being distributed over multiple hosts, which is one of the features of ZMap called "sharding" [1], where multiple hosts are used to conduct a single scan.

4.2 Scanning no longer focuses on typical targets nor originates from the well-known countries

The same shifts have also taken place in what is being targeted and by whom. While commonly used, well-known ports such as SSH (22/TCP) and HTTP (80/TCP and 8080/TCP) constantly score high, but the landscape diversifies: while in 2015 these three ports accounted for more than one-third of all scanning packets, eight years later this share has dropped to below 3%. As we show later in detail, scanners diversified and started targeting lesser-known ports as these are also used as an alias to host another service. For example, Izhikevich et al. [32] find that only 3.0% of all HTTP services are located on their standard port. Scanners do however not take this into account as a high percentage of sources is scanning just for port 80 for the last six years. While ports 80 and 8080 are constantly on the top of ports in terms of scanning sources, the amount of traffic received for these ports is much lower. While investigating these, we hypothesize that many of these scans originate from "benign" scanners such as webcrawlers owned by search engines, as they might have large networks to circumvent IP blocking. While this is partially true, the largest contributing factor to the number of sources scanning a port are actually botnets that adopt part of the Mirai source code. This has two reasons: (1) botnet operators extend their arsenal with new exploits, growing their network and scanning an increasing number of ports [48, 51], and (2) botnet infections are often in residential network spaces where DHCP churn is more likely to occur, inflating the number of sources measured in studies [9].

Diversification is also visible in where the scanning originates from: while in the beginning more than 30% was originating from China alone, we see constantly increasing activity from everywhere. When we normalize traffic by metrics such as the number of IP addresses or citizens in a particular country, countries which were historically linked to aggressive scanning no longer stand out. Now, the Netherlands is the odd one out – its high activity in cybercrime is attributed to high-speed Internet connectivity, cheap hosting [37], and bulletproof hosting [17], and we show later that scanning has shifted to better-performing platforms over the years.

4.3 Scanning does not have a memory, the Internet forgets fast

One of the reasons for these major changes in target ports is that following a vulnerability disclosure, adversaries massively trawl the Internet for hosts running that vulnerable software. Durumeric et al. [18] report in 2014 that following such a vulnerable release, interest in a particular port is sparked and from that time on becomes a continuous target for scanning. When we look a couple of years later, this behavior has stopped: Figure 1 shows the increase in scanning activity from the yearly average measured in days after the vulnerability disclosure and increasing interest for ten major events. Soon after disclosure, we can see that while activity skyrockets, the issue is by and large as quickly forgotten and the scanning distribution returns to "normal" which we verify using the Kolmogorov Smirnov (KS) test for the events in the figure. Scanning has thus become much more opportunistic and driven by fast trends.

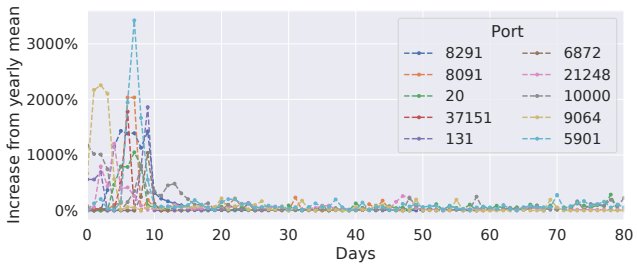

Figure 1: Large scanning events after vulnerability disclosures stop receiving traffic quickly.

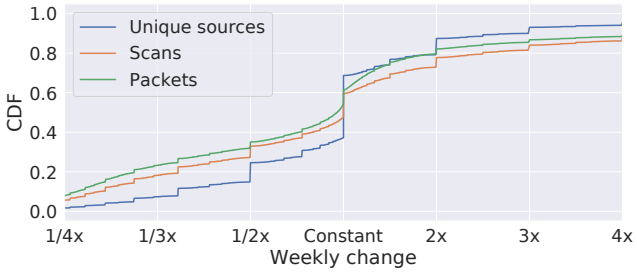

Figure 2: Weekly change of scans calculated per /16 net block, showing volatility of the ecosystem.

4.4 The ecosystem is volatile, 50% changes by a factor of 2 or more every week

We can also see this volatility in the way scans are launched. Figure 2 shows the weekly difference in the number of IPs that participated in scanning, the volume of scan campaigns launched as well as the number of packets sent in a cumulative density function over the entire continuum of /16 netblocks. We see that only 20-30% of the netblocks in the world are stable in terms of scanning and do more or less the same week after week, which cannot only be explained by IP churn [29]. The activity in the bulk of the netblocks on the Internet is thus highly volatile, for example in more than 50% of the /16s scanning either increased by a factor of 2 or more (or decreased by more than half) on average from one week to the other and for more than one third the increase/decrease was even larger than threefold.

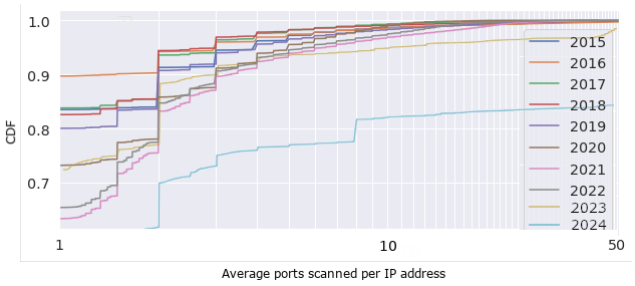

Figure 3: Ports scanned averaged over all scans per IP address, showing increase in block scans.

This means that when looking at the issue of Internet-wide port scanning it entirely depends on *when* we look: a study observing the Internet at or shortly after a main vulnerability release will get an equally biased view on scanning as if the measurement took place during the brief moment an actor launched a major temporary scan campaign using disposable infrastructure. To accurately understand the ecosystem of Internet-wide scanning, long-term measurements are thus needed. Furthermore, blocklists consisting of known (malicious) Internet scanners will be quickly outdated and can as such only be used as a real-time feed and not as a collection of IP addresses to block.

Key Findings:

- Internet-wide scanning activity increased drastically between 2015 and 2020. After 2020, the growth has slowed down and is no longer exponential, but the scans are increasingly spread out over the entire port range.
- The scanning landscape is highly volatile, with significant fluctuations in activity week-to-week. Scanning behavior is increasingly driven by short-term trends, with interest in new vulnerabilities spiking quickly and then dissipating.
- Blocklists and similar defensive measures must be updated in real-time to remain effective, given the rapid changes in the scanning ecosystem.

5 SCANNING DYNAMICS

As shown in the previous section, overall scanning traffic has greatly increased over the years. Where in 2015 we record 11 million TCP scanning packets per day, in 2022 we record 285 million scanning packets each day, which is a 2,591% increase in eight years time. Between 2022 and 2024 the growth has however slowed down significantly. From Table 1 we find that there is a disproportional growth in the number of scans as opposed to the amount of traffic. While the number of scans initially increased at a lower pace than the number of packets between 2015 and 2020, we see the opposite between 2021 and 2024. One explanation for this could be that when between 2015 and 2020 the scanning rate grew, only a small portion of scanners was actively throttling their application leading to an increase in scanning speed due to Internet rates and connectivity improvements. However, between 2021 and 2024, scanning campaigns became more diverse and scanners do not use their maximum throughput. Large scanning operations leveraging thousands of hosts have been identified in previous works [15, 27] and in the recent years the amount of scans conducted by more than one host has increased based on our analysis. This section discusses the changes in scanning dynamics we observed over the years.

5.1 Coverage of the entire port space increases

While in 2015 only 31% of the privileged ports were probed (above a 1% noise floor level), ten years later such selectivity was no longer the case: instead of only targeting well-known default ports of common services, scanners blanketed almost the entire privileged port space. Although the popular ports are still scanned magnitudes more than the rest, all ports are receiving more than 1,000 probes per day by 2022, and this number increases to over 1,500 by 2024. This is noteworthy because previous research performing vertical scans finds only a small number of ports being in use on servers [54]. On the other hand, Izhikevich et al. find that services can be located across many different ports on the Internet [32]. It therefore seems that the perpetrators command over enough resources to not shy at this overhead and low expected return. These observations are supported by the fact that organizations scanning the Internet such as Censys are increasing the number of ports scanned, which by 2024 has reached all 65,536 ports as we will discuss later on in this paper. When looking at the targeted ports within single scans, we find that scans are increasingly targeting multiple ports that might be running the same protocol, with for example in 2015 18% of scans targeting port 80 were also targeting port 8080, this has increased to 87% in 2020 and plateaus since then. We find the same increasing trend for other protocols such as SSH and HTTPs. This means that the commonly told practice of moving services to non-standard ports (typical patterns are 23 $\longrightarrow$ 2323, 443 $\longrightarrow$ 1443, 80 $\longrightarrow$ 8080, 22 $\longrightarrow$ 2222) as a method to lower the attack probability and keep logs cleaner is much more futile than one would think - scanners have in practice no problem covering a large number of ports to discover alternative configurations.

This trend is not only visible for select groups or only privileged ports but occurs across the entire spectrum of scanners and the entire port range. Figure 3 shows in a cumulative density function the number of different ports targeted by all source IPs in our study. While in 2015 83% of all scanners focused on exactly one port only, this percentage has dropped to 74% by 2020 and 65% in 2022. While at the beginning of our study only 2% of all scanners targeted 5 destination ports or more, eight years later this practice is common to 10% of all sources and we find a large statistically significant increase in the percentage of scans targeting 3 or more ports per year with a Pearson correlation of $R = 0.88, p < 0.05$. To understand whether actors scan ports proportionally to the number of services operating on a port we perform a complete vertical scan against a random sample of 100,000 IP addresses and compare the distribution of open ports against scanning intensities and find that there is no relation between the number of services and the number of scans targeting a service ($R = 0.047, p < 0.01$). Scanners thus do

not always target the ports where they can find most services, but may have different goals.

In 2024, we find a large deviation in the amount of ports targeted in a scan, with 15% of all scans targeting more than 10 distinct ports. As we will show later, we see that large scanning institutions have greatly increased their scanning activity and have started to target the entire port range, leading to a large increase in targeted ports.

5.2 The number of vertical scans is increasing

While it would be infeasible to perform an Internet-wide vertical scan with a scanning tool such as NMap, which is estimated to only scan the Internet for one port in 62.5 days [21], tools such as ZMap and Masscan perform much faster scans and thus allow more vertical scans. We find that over time we indeed record more vertical scans, with only one scan campaigns targeting more than 10,000 ports in 2015 as opposed to 2,134 in 2020. The percentage of scans targeting more than 100 ports is also increasing, but still only accounts for less than 0.5% of all scans recorded each year. The largest scans target almost the entire TCP port range, in 2020 a scan probed 54,501 (83%), but these large scans are extremely rare and we have only observed 20 scans (0.0005%) targeting more than 10,000 ports in 2022. The 406 (0.01%) scans targeting more than 1,000 ports in 2022 scan on average with a speed of 0.3 Gbps, significantly higher than the overall average scanning speed of 14 Mbps.

5.3 Scans used to get more intensive and take longer, but are increasingly spread out

Actors can likely afford to also target unconventional ports because they have ample resources to spend on their search. Indeed, scan volume increases by 63% per annum from 2015 to 2020 and the speed of a scan positively correlates with the number of ports being targeted ($R = 0.88, p < 0.05$). The obvious (but incorrect) hypothesis for this drastic increase is that Internet connectivity is getting increasingly performant, but this is not the case: on average, *scanning speed* is largely remaining constant. The increased coverage and frequency of Internet-wide scanning is the result of three components: first, the influx of a large number of devices, second, active scanners send more probes, and third, the scans generally take longer.

From 2020 onwards, the scan volume does not follow the same trend. While constant between 2020 and 2022, scan volume shows again an increase in 2023 but drops again in 2024. We are not able to identify what causes the scan traffic to drop again, as the obvious trend throughout the years is for the traffic to grow, but do note a drastic increase in the number of scans between 2021 and 2022, and 2023 and 2024. These signify major changes in the way scans are conducted, as scanning *sources* become less intrusive, but scan *campaigns* are growing.

5.4 Origin country specific scanning

As reported in [18], certain parts of the world used to be leading the scoreboard in port scanning. While China and the US together accounted for more than half of all scans in 2016, by 2020 the US is home to only 3.2% of scan sources. Internet scanning has diversified and is spread over the entire world now. Surprisingly though, exactly what is targeted is not evenly distributed, but a clear bias exists between the targeted port and where the scan is taking place from. While scanning for HTTPs on 443/TCP is predominantly a US-based endeavor (which we can link to institutional research scanners), targeting MySQL (3306/TCP) or the Remote Desktop Protocol (3389/TCP) is essentially happening from China. While overall we find that scanning traffic geographically diversifies over time, China has originated more than 80% of all scanning traffic on 14,444 unique ports, the traffic for 666 unique ports originates for more than 80% from the US, for Brazil this is the case for 221 ports, for Taiwan 59, and Iran 57 unique ports in 2022.

While we find major biases remaining constant over many years, we again notice large volatility in the ecosystem with scans on port 5555 shifting heavily away from the original distribution in 2017. Similar trends are observed in ports 8080 and 8545. Even on large popular protocols such as HTTP (80/TCP) we find that the distribution can largely change between individual measurements with the US being very active in 2016-2018, but in 2019 almost completely abandoning that protocol. While we are unable to attribute large biases or swings to specific actors, the presence of these biases indicates that a large amount of scanning originates from specific areas and is not performed by botnets located around the world.

Key Findings:

- Scanning activity shows significant geographic and protocol shifts over time, with major changes in scanning sources and targeted protocols, such as the US reducing HTTP scanning after 2018.
- The number of vertical scans (scans targeting many ports) has significantly increased, with 2,134 campaigns targeting more than 10,000 ports in 2020, compared to just one in 2015.
- There is a positive correlation between scan speed and the number of ports targeted, highlighting that faster scans tend to cover more ports.
- There are clear biases in protocol targeting based on geographic origin, such as scans originating from China predominantly targeting MySQL and Remote Desktop Protocol, while scans originating from the US focus on HTTPs.

6 SCANNING TOOLS USAGE

As discussed in Section 3, we fingerprint well-known scanning tools in the network traffic. These tools together generate 95% of Internet-wide *scanning traffic* (number of packets) in the first months of 2022. In 2024 this number has lowered as scanning organizations do not use the version of ZMap that is easily fingerprintable with a static IP identification number anymore, to under 40%. In this section, we will identify the evolution in tool usage and investigate tool adoption.

6.1 Standard tool adoption heavily fluctuates over the years

Table 1 shows the change in tool usage over the years, where we find an increased adoption of common tools. Where in 2015 34%

of scans originate from one of the common tools, we find that in 2020 this has increased to 54%. In the case of packets sent, we find that in 2015 25% originated by the 5 most well-known tools, while in 2020 this share increased to 92% of all scanning probes. In 2022, the number of scans that use common tools has dropped significantly, while the amount of packets sent using these common tools remained high. In 2015 only a small portion of all scans are performed with high-speed network scanning tools, but over time high-speed scanning tools are adopted by a wider audience to scan the Internet. Other than [25], we find no evidence of Unicorn being used for Internet-wide scanning and instead record in total only 2 distinct IP addresses ever using the Unicorn scanning tool.

Figure 4 shows the top 10 ports receiving the most traffic per year and the distribution of tools where the traffic originates from and shows that tool adoption differs for ports. For example, NMap is used less over time, but a small portion of scanners probing SSH (22/TCP), HTTP (80/TCP) and RDP (3389/TCP) consistently consists of NMap-based scanners while HTTPS (443/TCP) receives at most half a percent of NMap-based traffic. While in 2020, 14.9% of scans are based on the Mirai scanning routine, but they only account for 3.3% of all traffic received by us. In 2021 and 2022, traffic originating from Mirai-based scanning routines drops even further below 1%. NMap and ZMap are respectively sending 0.5% and 6.9% and Masscan generates 81% of all incoming probes. Only 7.9% of all probes sent in Internet-wide scanning campaigns in 2020 originate from different tools than the common tools we have fingerprinted. In 2022, the four tracked tools are responsible for over 95% of all scanning traffic. In the last years, we find that the tracked tools decrease in traffic volume again with 39% of all scanning traffic in 2024 being directly relatable to the four tools. This can mean two things: (1) scanners are changing their fingerprint, for example by changing the simple identifiers such as the IP identification number of ZMap or the sequence number of Mirai, or (2) the ecosystem is again becoming more diverse as Internet scanners start creating their own network scanning tools, like we observed being the case in 2015. While tools such as ZMap and Masscan are great for academic research, the ability to perform high-speed Internet-wide scanning from a simple laptop has set the bar so low that everyone can scan for and consecutively exploit services on the Internet. The homogeneity in scans caused by the standardization of these tools, however, allows organizations to distinguish most scanning activity from normal traffic by using the same fingerprinting methods as we do in this paper, and block these scanning probes, reducing the reconnaissance activities against networks and reducing alert-fatigue in security operation centers.

6.2 The scanning routine of Mirai has been adopted in many different programs

In August 2016 the Mirai botnet was discovered, and shortly thereafter the source code of the botnet was shared on the Internet leading to several new and competing variants [27]. While Mirai originally scanned for Telnet to spread itself, the botnet was later augmented to serve as a platform to target other ports and in 2020 we observe scans with the Mirai fingerprint targeting 65,286 (99.6%) of all TCP ports. When we look at the ports other than Telnet (23/TCP) used for its self-propagation, we see that in 2017 more

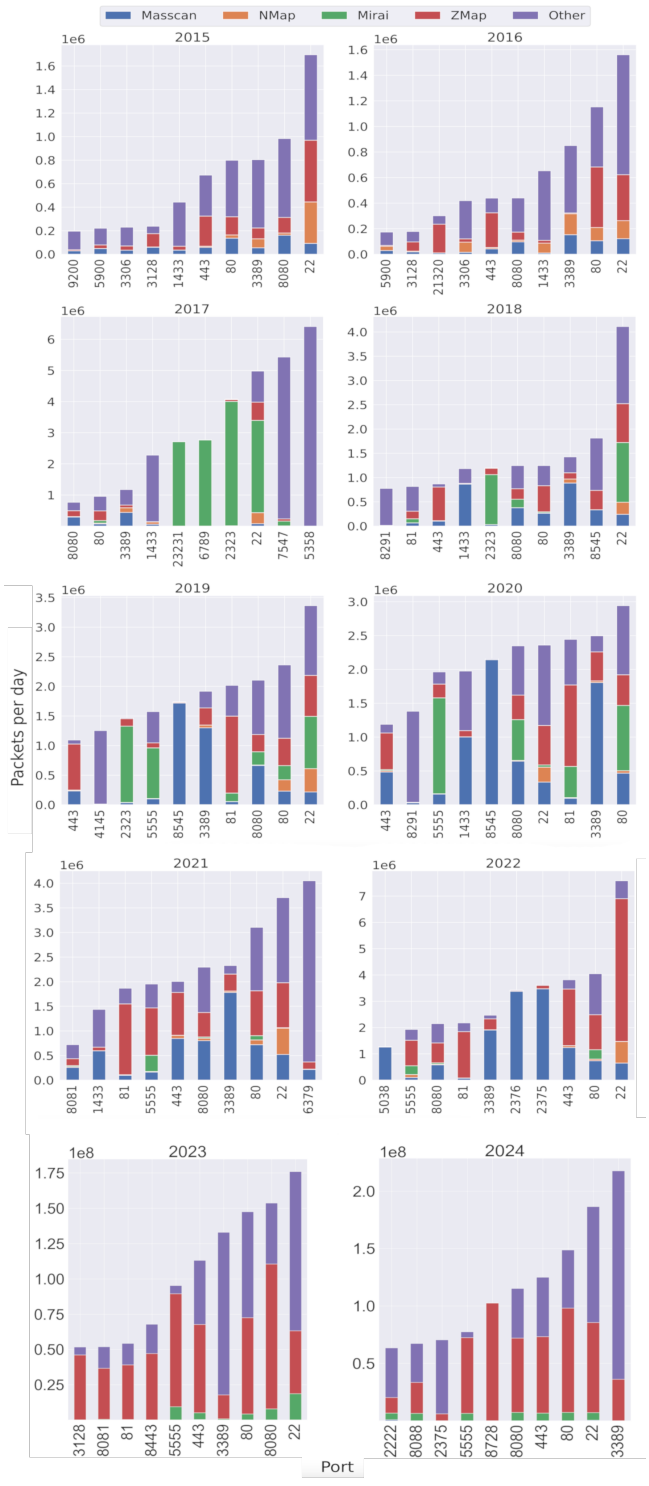

Figure 4: Top 10 ports in terms of traffic and scanning tools used unveil increasing tool adoption.

than half of all scans originated from Mirai. We see from figure 4 that Mirai heavily dominates five of the top ports in 2017, but is much less pronounced in the years after. The constant modification by a variety of actors [5], however, makes the original botnet code a widely adopted tool that is still responsible for more than 25% of the scans in 2020. In 2022 this share drops to less than 10% of all scans and below 1% of all incoming probes. In 2023, we see a spike in the number of scanning sources exhibiting the Mirai fingerprint, but this does not significantly increase the number of packets sent by devices infected with some strain of Mirai. We do not find a clear explanation for this temporal effect, which does not signify a trend as the number of scans has significantly dropped again in 2024.

6.3 Overall scanning speed decreases over the years

We find an overall decrease in scanning speed over the years, which is surprising as the average global Internet speed has increased in the last decade. When we look at individual scanning tools however, we find that scans performed using ZMap are the fastest on average, but only a small portion of these scans exceed speeds of 1 Gbps in practice, while the tool is capable of scanning speeds much beyond this [1]. While NMap is usually looked at as a slow tool as opposed to Masscan and ZMap, we find that NMap was actually used to scan the Internet at faster rates than Masscan. Moreover, NMap is the only tool where we find an increasing trend in scanning speed, even though the overall increase in speed is minimal ($R = 0.12, p < 0.01$), which indicates that the scanners that are successful in scanning the Internet keep using their tool, whereas others divert to other tools, explaining the decrease in overall NMap usage. Curiously, the advantages of high-speed scanning tools such as ZMap or Masscan over NMap are not cashed in, and the speed advantage is only realized by a select few at the very high end (beyond 10^5 packets per second). Looking over the entire spectrum, on average hosts using NMap even consistently realize faster performance than those relying on Masscan. As Mirai-based scanning would mostly originate from embedded devices with limited processing capabilities, it is unsurprising that this scanning is the slowest.

While overall the scanning speed is decreasing, there are developments at the top end. The speed of the top 100 fastest scans significantly increases over the years with a Pearson correlation of $R = 0.356, p < 0.001$. While increasingly more hosts would be capable of scanning at speeds over 1 Gbps and tools such as Masscan and ZMap allow for these scanning speeds, the number of hosts performing scans at these speeds does thus not significantly increase. As the resources available to scanners would generally be higher than the average speed with which they operate, we expect a large number of scans to be actively throttled.

6.4 Scan coverage is stable

We can estimate the coverage of a scan by extrapolating the amount of destination IP addresses scanned in our telescope over the entire IPv4 space. By doing so, we find that large scans are rare, and only NMap scanners increase the coverage of their operations per source IP address over time. Not only are Internet-wide scans from single sources rare, their percentage also decreases over time. While more than 20% of all Masscan scans in 2016 targeted the entire IPv4 space, this number drops in subsequent years. We however only find a statistically significant decreasing trend in average scan coverage of ZMap. ZMap and Masscan are also the only tools showing concrete evidence of logical slicing of the target space. For instance, if one would use 256 sources that jointly scan the entire Internet, one would expect coverage values of $2^8/2^{32}$ for all of these devices, visible in the plot as a vertical increase at a particular value. As more of this coordination would take place, this will result in a mode at that value. Indeed, we observe modes for ZMap, for example a pronounced peak at around 0.65% IPv4 coverage where we find a /24 subnet of (academic) scanners collaborating to scan the entire IPv4 space. For institutional scanning this is expected as we find subnets being used to perform coordinated Internet scans. While we find these indications of scanner distribution, we did not cluster these scanners in this paper. However, we note that we find an increasing number of scans being split over multiple hosts over the years.

6.5 Tool usage is largely geographically biased

Previously, we stated that while scanning is a global phenomenon, what is targeted has clear geographical preferences. The same is also true for the tool in use. While most incoming traffic originated from a small number of scans from China in 2015, we have seen increasing adoption of tools in other countries over the years. Even though we see increased adoption, still large biases exist for most tools, with ZMap being almost exclusively used from China and the US. NMap is, on the other hand, more globally distributed, but we find that in 2019 and 2020, there has been an increased disproportional adoption from countries such as Indonesia and Iran. Large scanning campaigns can skew the country distribution greatly, with Russia performing more than 80%, of all Masscan scans in 2018, heavily impacting not only Masscan's country distribution but the overall country distribution as well. The impact of increasing scans using a single tool from a single country thus has major effects on the ecosystem. In 2023 and 2024, we find a large decrease in the usage of "standard" tooling as a whole, which is not relatable to a single country.

6.6 Scanners do not come back, except for institutional ones

Over time, the services exposed to the Internet on an IP address are subject to change because of servers moving around, different services being used, better firewall policies, or even IP churn. Therefore, it is vital to update the knowledge about open ports and thus re-scan the Internet. When looking at the recurrence of scanners overall, we do not see a visible trend. When looking at scans from known institutions however, where we know that scans are recurring, we do however see a clear distinction. To visualize this, we assign a label to incoming scans to distinguish scanning from institutions, hosting providers, residential connections, or the autonomous systems of large enterprises. To classify which type of origin a particular source IP address belongs to, we rely on a dataset of the commercial intelligence provider Greynoise, which provides a service to label IP addresses to organizations known to perform scanning for research and commercial purposes, such as the University of Michigan, Censys, or Rapid7. We classify these

Table 2: Unique IP addresses, scans and packets recorded for the different scanning types.

Scanner type	Sources	Scans	Packets
Hosting	0.87%	5.61%	18.52%
Enterprise	6.71%	15.75%	3.85%
Institutional	0.16%	7.45%	32.63%
Residential	54.92%	46.12%	23.39%
Unknown	37.33%	25.07%	21.61%

scans as "institutional" if they are from an organization that publicizes that they are scanning. Scans originating from the ASN of a hosting provider or enterprise that are not directly linked to scanning are classified accordingly. For residential hosts, we rely on a matching between netblocks of telecom providers that are classified as residential space using the methodology described in [29]. Scans that could not be classified are labeled as "Unknown". The result of this classification is shown in Table 2. Interestingly, institutional scanners overall contribute almost a third of all packets observed in the data, while these account for only 0.16% of all sources observed scanning the Internet.

Figure 6 shows the CDF of the number of times a source IP address has been observed to scan the Internet, as well as a CDF of the average downtime of a recurrent scanner before scanning again. We find that overall only a small portion of scanners return to re-scan the Internet, basically the exception are research-based scanners where a large share performs more than 100 separate campaigns. While it is expected from residential IP addresses to be less likely to return due to for example a large presence of Mirai which has been shown to be very volatile [28], it is surprising to find that scanners located at enterprises are similarly non-persistent. We find that for the scanners that do come back to scan, most scanners repeat within one day of the end of the last scan. For institutional scanners, there is a large mode of scanning IP addresses that consistently scan the Internet every day. For the other scanning types, we do not find similar modes, indicating that the scanners that do come back are not consistently scanning the Internet every day, week, or month.

This has important implications for detecting and mitigating port scanning. As non-institutional IP addresses are basically not significantly reused across scans and either deliberately (hosting/cloud) or intrinsically (residential) burned, collecting and sharing lists of IP addresses observed to have participated in scanning (like for example IP addresses used in spamming, or brute-forcing) with the aim of blocking their traffic at the perimeter would in practice be relatively ineffective, as these source addresses have a very short lifetime. By the time a list is distributed a scanning IP address would have already vanished for good. If malicious actors would then also deliberately spread their activities out and rely on a large number of endpoints each contributing part of the overall scan like we have seen in this paper, effectively suppressing such reconnaissance scans will be very difficult.

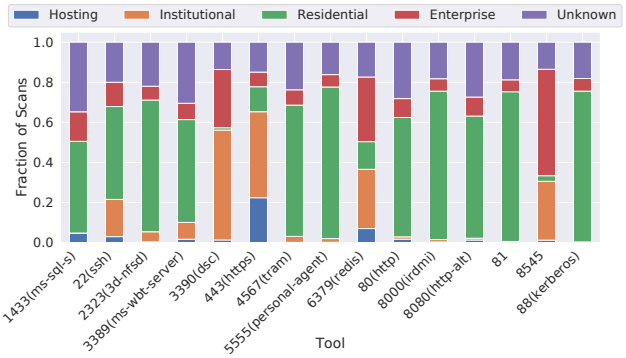

Figure 5: Distribution of scanner types (top 15 ports).

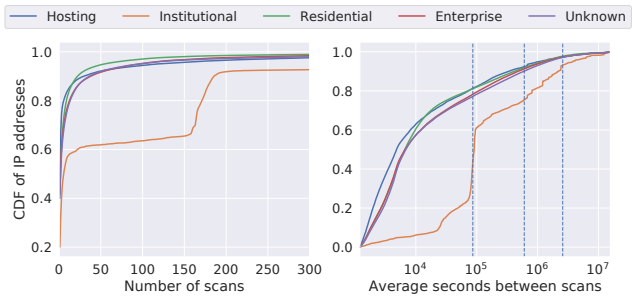

Figure 6: Scanner recurrence down-time between scans. Lines show a day, week, and month.

6.7 Scanner types differ

Figure 5 shows the distribution of scanner types over the 15 most targeted ports, where we find large deviations in which port is being targeted by different groups of hosting types. While for most ports incoming scans originate mainly from residential sources, some ports are more favored by scans from institutions or scanners located at hosting providers. For HTTPs (443/TCP) only 15% of all incoming scans originate from residential IP space, while institutional scans sent 41% of this traffic. For DSC (3390/TCP) this is even starker, as institutional sources generate half of all traffic, and residential IPs only sent 2%. The port running JSON-RPC (8545/TCP) – commonly used together to run services related to the Ethereum cryptocurrency – is disproportionally targeted by IP addresses located in autonomous systems related to enterprises, especially from ASN 18403 (FPT-AS-AP The Corporation for Financing & Promoting Technology). While we would expect most cybercrime to originate from residential and hosting IP addresses as these would be more likely to be taken over by botnets or rented out to malicious actors, we do not observe these groups scanning the JSON-RPC protocol. This imbalance between scanner types is surprising as we would expect large amounts of malicious scanning on this port because of the potential monetary gain due to exposed Ethereum wallets [43].

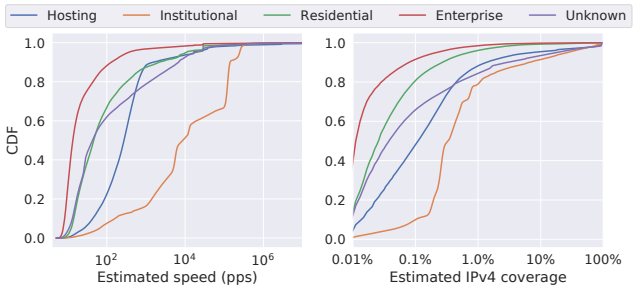

Figure 7: Speed and coverage of scanner types.

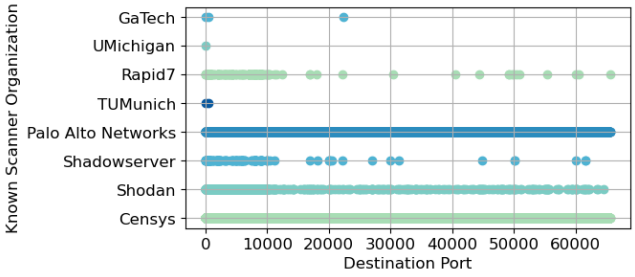

Figure 8: Port coverage of some well-known Internet-wide scanning projects in 2024.

6.8 Institutional scanners have the largest footprint in the scanning landscape

The capabilities of scanning types highly vary, as small residential IP addresses will not possess the same networking capabilities as a large research institution. While many scans would be throttled and not use the line-rate available to the host machine, we find that institutions with large capabilities are scanning magnitudes faster than residential sources. Only 12% of residential IP addresses exceed speeds of 0.06 Mbps (1,000 packets per second), whereas 84% of institutional scanning exceed 1,000 pps. Figure 7 shows the estimated speed and IPv4 coverage of scanning types averaged per source IP address. While scanners located at hosting providers are scanning faster than residential scanners, the top residential scanners are scanning at the same rate as the top scanners located at hosting providers. Scans at hosting providers however last longer, providing better coverage of the IPv4 space. While enterprises generally have more networking capabilities than residential scanners, we find that scanners based in enterprises are heavily throttled to scan at the slowest rate and have the least coverage of all scanner types, except for JSON-RPC (8545/TCP) where the fastest scanners originate from enterprise IP ranges. Institutional sources – research institutes, universities, and commercial entities with legitimate scanning interests such as Censys or Shodan – are key players in the scanning ecosystem and largely eclipse the activities of much of the rest, scanning on average 92 times faster than the average scanner. Figure 8 shows how these different well known Internet scanners cover the entire port range in 2024. Various scanners such as Censys and Palo Alto indeed cover all TCP ports in their scans, while Shadowserver and Rapid7 are not yet scanning all available ports for services. Organizations are also rapidly expanding the number of ports targeted in their scans, with for example the scanner *Onyphe* scaling up their operations between 2023 and 2024 from targeting less than half of all ports to targeting the entire port range. While many institutions are thus scaling up their operations, universities conducting scans are scanning at a much lower pace, targeting only a few ports. For universities, we also do not see a growth in ports targeted over the years. The large footprint of institutional scanning can be fully attributed to a handful of organizations. Full figures showing scanning activity from known scanning organizations in 2023 and 2024 are added in the Appendix A.

At speeds and scan coverages orders of magnitude larger than the other sources of scanning, it is paramount that these source IPs are filtered out before doing analyses. Otherwise, papers quantifying the Internet – by passive collection of scans – are essentially looking into the mirror, and describing the activities of other researchers in the field rather than studying those of malicious actors.

Key Findings:

- The use of common scanning tools has fluctuated significantly over the years. From 2015 to 2020, the adoption of these tools increased, with 54% of scans in 2020 originating from known tools, up from 34% in 2015. By 2022, the number of scans using common tools decreased, but the proportion of packets sent using these tools remained high. By 2024, scanning traffic attributable to four tracked tools dropped to under 40%.
- Initially targeting Telnet, the Mirai botnet expanded to scan nearly all TCP ports by 2020. Despite a significant presence in 2020, its share of total scans decreased to less than 10% by 2022 and dropped further in 2024.
- Large-scale scans targeting the entire IPv4 space from a single scanning source are rare and are becoming less common. This indicates that scanners are increasingly spreading their scan campaigns over multiple hosts.
- Institutional scanners are more likely to perform recurring scans, often scanning the Internet daily. Non-institutional scanners, especially from residential IPs, rarely return, making IP-based blocking ineffective.
- The types of scanners (institutional, residential, etc.) targeting specific ports vary significantly. For example, HTTPS is predominantly targeted by institutional scanners, while residential sources dominate traffic to other ports.

7 FUTURE WORK

Identifying scanners with benign intent: The volume of scanning originating from institutional sources is significant and highly different from the rest of the ecosystem, and the abundance of scans by research institutions or businesses can largely bias the view we

have as a community on the Internet-wide scanning landscape. It is important that as a research community we identify these scanners having *legitimate intent* to generate reliable conclusions, as measurements could be off by over 30%.

Long term measurements are paramount: The ecosystem of Internet-wide scanning is highly volatile and in constant flux. Depending on when and how long assessments of it are taking place can severely over- or underestimate our quantification of it, such as what is targeted, by whom, and using which resources. We find that these developments only intensify over the years. This means that while in the past it was sufficient to focus on shorter measurement periods, we advocate that future research studies investigating the scanning landscape should incorporate longer-term data as a default method.

Combating alert-fatigue in organizations: Like any other technology, naturally also port scanning software is dual-use. While tools such as ZMap or Masscan have revolutionized port scanning and opened the ability for a variety of Internet security research to take place at Internet-scale, we find that these high-performance tools developed by the academic community are not only used by academics, they also make the activities of non-friendly scanners easier. In 2020, 92.1% of all scanning traffic originated from 4 known tools. Much of the scanning activity targeting a network can thus be blocked by detecting these tools, reducing successful reconnaissance.

Comparing vantage points: In this paper, we rely on a single vantage point to characterize Internet-wide scanning. This inherently biases the study towards scans that are geographically targeted, and might over- or underestimate Internet-wide phenomena. Data from multiple vantage points should be considered in a long-term study to verify that these results are generalizable over the entire Internet.

8 ETHICS

In this paper, we measure Internet background radiation from a set of IP addresses that are routed but unused. While this traffic will include data sent by infected devices of unknowing users, the traffic collected in this dataset is not linked to specific individuals. We enrich the data on connecting IP addresses by looking up the Autonomous System and country for every incoming IP address, but do not report specific data of single Autonomous Systems or companies, and instead only report on country-level statistics. The only exception to this are organizations that actively communicate their Internet scanning efforts, as for these organizations it is already clear that they are scanning the Internet.

9 CONCLUSION

In this work, we have surveyed Internet-wide scanning traffic over ten years. We report an overall exponential growth of scanning activity year after year that halts in 2020. We also notice that large scanning events have a major, but temporary impact on the scanning landscape. We analyze the evolution of the scan landscape over multiple years and see that the ecosystem is highly volatile, for example, 50% of the /16 netblocks change their activity in terms of active sources, campaigns launched, and the number of packets sent by at least a factor of 2 on a weekly basis. This explains the significant deviations in research findings across previous studies conducted over the past years. As the ecosystem of Internet-wide scanning is highly volatile and in constant flux. Depending on when and how long assessments of it are taking place can severely over- or underestimate our quantification of it, such as what is targeted, by whom, and using which resources. We find that these developments only intensify over the years. This means that while in the past it was sufficient to focus on shorter measurement periods, we advocate that future research studies investigating the scanning landscape should incorporate longer-term data.

The volume of scanning originating from institutional sources is significant and highly different from the rest of the ecosystem, and the abundance of scans by research institutions or businesses can largely bias the view we have as a community on the Internet-wide scanning landscape. It is important that as a research community we identify these scanners having *legitimate intent* to generate reliable conclusions.

Finally, the large impact of events on the scanning landscape and the impact of scanners with legitimate intent have to be accounted for in studies surveying the scanning ecosystem to avoid systematic biases. While tools such as ZMap or Masscan have revolutionized port scanning and opened the ability for a variety of Internet security research to take place at Internet-scale, we find that these high-performance tools developed by the academic community are not only used by academics, they also make the activities of non-friendly scanners easier. While in 2020, 92.1% of all scanning traffic originated from 4 known tools, making it trivial to detect and block these scanners, in 2024 the number of scans being easily identifyable has dropped significantly. In the last years the total amount of scanning probes has not increased. The number of scans on the other hand is steadily rising, indicating that scanning campaigns are increasingly spread out over many different hosts. Counting scans as "single-source", will therefore largely bias measurements, and future work should take this into account in measurements.

Acknowledgments

The authors thank the anonymous IMC reviewers and the paper shepherd, Paul Barford, for their reviews and insightful comments. They also want to thank Zakir Durumeric for useful feedback regarding ZMap and Censys. This work was supported by the European Commission under the Horizon Europe Programme as part of the project SafeHorizon (Grant Agreement no. 101168562). The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

REFERENCES

[1] David Adrian, Zakir Durumeric, Gulshan Singh, and J Alex Halderman. 2014. Zippier Zmap: Internet-wide Scanning at 10 Gbps. In *8th USENIX Workshop on Offensive Technologies (WOOT 14)*.

[2] Mark Allman, Vern Paxson, and Jeff Terrell. 2007. A brief history of scanning. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. 77–82.

[3] Aniket Anand, Michalis Kallitsis, Jackson Sippe, and Alberto Dainotti. 2023. Aggressive Internet-wide Scanners: Network Impact and Longitudinal characterization. In *Companion of the 19th International Conference on emerging Networking EXperiments and Technologies*. 1–8.

[4] Evgeny V Ananin, Arina V Nikishova, and Irina S Kozhevnikova. 2017. Port scanning detection based on anomalies. In *2017 Dynamics of Systems, Mechanisms and Machines (Dynamics)*. IEEE, 1–5.

[5] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J Alex Halderman, Luca Invernizzi, Michalis Kallitsis, et al. 2017. Understanding the mirai botnet. In *26th USENIX security symposium (USENIX Security 17)*. 1093–1110.

[6] Soniya Balram and M Wiscy. [n. d.]. Detection of TCP SYN scanning using packet counts and neural network. In *2008 IEEE International Conference on Signal Image Technology and Internet Based Systems*. IEEE, 646–649.

[7] Shehar Bano, Philipp Richter, Mobin Javed, Srikanth Sundaresan, Zakir Durumeric, Steven J Murdoch, Richard Mortier, and Vern Paxson. 2018. Scanning the internet for liveness. *ACM SIGCOMM Computer Communication Review* 48, 2 (2018), 2–9.

[8] Norbert Blenn, Vincent Ghiëtte, and Christian Doerr. 2017. Quantifying the spectrum of denial-of-service attacks through internet backscatter. In *Proceedings of the 12th International Conference on Availability, Reliability and Security*. 1–10.

[9] Leon Böck, Dave Levin, Ramakrishna Padmanabhan, Christian Doerr, and Max Mühlhäuser. [n. d.]. How to Count Bots in Longitudinal Datasets of IP Addresses.

[10] Joppe W Bos, J Alex Halderman, Nadia Heninger, Jonathan Moore, Michael Naehrig, and Eric Wustrow. 2014. Elliptic curve cryptography in practice. In *International Conference on Financial Cryptography and Data Security*. Springer, 157–175.

[11] Elias Bou-Harb, Mourad Debbabi, and Chadi Assi. 2013. Cyber scanning: a comprehensive survey. *Ieee communications surveys & tutorials* 16, 3 (2013), 1496–1519.

[12] M. Patrick Collins, Alefiya Hussain, and Stephen Schwab. 2023. Identifying and Differentiating Acknowledged Scanners in Network Traffic. In *2023 IEEE European Symposium on Security and Privacy Workshops*. 567–574. https://doi.org/10.1109/EuroSPW59978.2023.00069

[13] Andrei Costin, Jonas Zaddach, Aurélien Francillon, and Davide Balzarotti. 2014. A Large-Scale Analysis of the Security of Embedded Firmwares. In *23rd USENIX Security Symposium (USENIX Security 14)*. 95–110.

[14] Mehiar Dabbagh, Ali J Ghandour, Kassem Fawaz, Wassim El Hajj, and Hazem Hajj. 2011. Slow port scanning detection. In *2011 7th International Conference on Information Assurance and Security (IAS)*. IEEE, 228–233.

[15] Alberto Dainotti, Alistair King, Kimberly Claffy, Ferdinando Papale, and Antonio Pescapé. 2014. Analysis of a "/0" stealth scan from a botnet. *IEEE/ACM Transactions on Networking* 23, 2 (2014), 341–354.

[16] Zakir Durumeric, David Adrian, Ariana Mirian, Michael Bailey, and J Alex Halderman. 2015. A search engine backed by Internet-wide scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 542–553.

[17] Zakir Durumeric, David Adrian, Phillip Stephens, Eric Wustrow, and J. Alex Halderman. 2024. Ten Years of ZMap. In *Proceedings of the ACM Internet Measurement Conference*.

[18] Zakir Durumeric, Michael Bailey, and J Alex Halderman. 2014. An Internet-wide view of Internet-wide Scanning. In *23rd USENIX Security Symposium (USENIX Security 14)*. 65–78.

[19] Zakir Durumeric, James Kasten, Michael Bailey, and J Alex Halderman. 2013. Analysis of the HTTPS certificate ecosystem. In *Proceedings of the 2013 conference on Internet measurement conference*. 291–304.

[20] Zakir Durumeric, Frank Li, James Kasten, Johanna Amann, Jethro Beekman, Mathias Payer, Nicolas Weaver, David Adrian, Vern Paxson, Michael Bailey, et al. 2014. The Matter of Heartbleed. In *Proceedings of the Internet Measurement Conference*. 475–488.

[21] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide scanning and its security applications. In *22nd USENIX Security Symposium (USENIX Security 13)*. 605–620.

[22] Wassim El-Hajj, Fadi Aloul, Zouheir Trabelsi, and Nazar Zaki. [n. d.]. On detecting port scanning using fuzzy based intrusion detection system. In *2008 International Wireless Communications and Mobile Computing Conference*. IEEE, 105–110.

[23] Wassim El-Hajj, Hazem Hajj, Zouheir Trabelsi, and Fadi Aloul. 2011. Updating snort with a customized controller to thwart port scanning. *Security and Communication Networks* 4, 8 (2011), 807–814.

[24] Adrienne Porter Felt, Richard Barnes, April King, Chris Palmer, Chris Bentzel, and Parisa Tabriz. 2017. Measuring HTTPS adoption on the web. In *26th USENIX Security Symposium (USENIX Security 17)*. 1323–1338.

[25] Vincent Ghiëtte, Norbert Blenn, and Christian Doerr. 2016. Remote identification of port scan toolchains. In *8th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 1–5.

[26] Robert David Graham. 2014. MASSCAN: Mass IP port scanner. *URL: https://github.com/robertdavidgraham/masscan* (2014).

[27] Harm Griffioen and Christian Doerr. 2020. Discovering Collaboration: Unveiling Slow, Distributed Scanners based on Common Header Field Patterns. In *NOMS IEEE/IFIP Network Operations and Management Symposium*. IEEE, 1–9.

[28] Harm Griffioen and Christian Doerr. 2020. Examining Mirai's Battle over the Internet of Things. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*. 743–756.

[29] Harm Griffioen and Christian Doerr. 2020. Quantifying Autonomous System IP Churn using Attack Traffic of Botnets. In *International Conference on Availability, Reliability and Security (ARES)*.

[30] Marcella Hastings, Joshua Fried, and Nadia Heninger. [n. d.]. Weak Keys Remain Widespread in Network Devices. In *Proceedings of the Internet Measurement Conference*. 49–63.

[31] Ralph Holz, Johanna Amann, Olivier Mehani, Matthias Wachs, and Mohamed Ali Kaafar. 2015. TLS in the wild: An Internet-wide analysis of TLS-based protocols for electronic communication. *arXiv preprint arXiv:1511.00341* (2015).

[32] Liz Izhikevich, Renata Teixeira, and Zakir Durumeric. 2021. LZR: Identifying Unexpected Internet Services. In *30th USENIX Security Symposium (USENIX Security 21)*.

[33] Liz Izhikevich, Renata Teixeira, and Zakir Durumeric. 2022. Predicting IPv4 Services Across All ports. In *Proceedings of the ACM SIGCOMM 2022 Conference*. 503–515.

[34] Postel John. 1981. Transmission Control Protocol. *RFC 793* (1981).

[35] Cynthia Bailey Lee, Chris Roedel, and Elena Silenok. 2003. Detection and characterization of port scan attacks. *Univeristy of California, Department of Computer Science and Engineering* (2003).

[36] Derek Leonard and Dmitri Loguinov. 2010. Demystifying service discovery: implementing an internet-wide scanner. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 109–122.

[37] Rutger Leukfeldt, Sander Veenstra, and Wouter Stol. 2013. High volume cyber crime and the organization of the police: The results of two empirical studies in the Netherlands. *International Journal of Cyber Criminology* 7, 1 (2013), 1.

[38] Johan Mazel, Romain Fontugne, and Kensuke Fukuda. 2017. Profiling internet scanners: Spatiotemporal structures and measurement ethics. In *2017 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 1–9.

[39] Johan Mazel and Rémi Strullu. 2019. Identifying and characterizing ZMap scans: a cryptanalytic approach. *arXiv preprint arXiv:1908.04193* (2019).

[40] David Moore, Colleen Shannon, Geoffrey M Voelker, Stefan Savage, et al. 2004. *Network telescopes*. Technical Report. Technical Report CS2004-0795, CSE Department, UCSD.

[41] Marcin Nawrocki, Thomas C Schmidt, and Matthias Wählisch. 2020. Uncovering Vulnerable Industrial Control Systems from the Internet Core. In *NOMS IEEE/IFIP Network Operations and Management Symposium*. IEEE, 1–9.

[42] Jamie O'Hare, Rich Macfarlane, and Owen Lo. 2019. Identifying vulnerabilities using Internet-wide scanning data. In *IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3)*. IEEE, 1–10.

[43] P Paganini. 2018. Hackers steal 20 million from Ethereum clients exposing interface on port 8545. https://securityaffairs.co/wordpress/73436/digital-id/ethereum-scanning-port-8545.html.

[44] Ruoming Pang, Vinod Yegneswaran, Paul Barford, Vern Paxson, and Larry Peterson. 2004. Characteristics of internet background radiation. In *Proceedings of the 4th ACM SIGCOMM conference on Internet measurement*. 27–40.

[45] 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*. 144–157.

[46] Seungwon Shin, Guofei Gu, Narasimha Reddy, and Christopher P Lee. 2011. A large-scale empirical study of conficker. *IEEE Transactions on Information Forensics and Security* 7, 2 (2011), 676–690.

[47] Himanshu Singh. 2009. Distributed Port Scanning Detection. (2009).

[48] Simon Nam Thanh Vu, Mads Stege, Peter Issam El-Habr, Jesper Bang, and Nicola Dragoni. 2021. A survey on botnets: Incentives, evolution, detection and current trends. *Future Internet* 13, 8 (2021), 198.

[49] Sadegh Torabi, Elias Bou-Harb, Chadi Assi, ElMouatez Billah Karbab, Amine Boukhtouta, and Mourad Debbabi. 2020. Inferring and investigating IoT-generated scanning campaigns targeting a large network telescope. *IEEE Transactions on Dependable and Secure Computing* (2020).

[50] Hung Nguyen Viet, Quan Nguyen Van, Linh Le Thi Trang, and Shone Nathan. 2018. Using deep learning model for network scanning detection. In *Proceedings of the 4th International Conference on Frontiers of Educational Technologies*. 117–121.

[51] Benjamin Vignau, Raphaël Khoury, Sylvain Hallé, and Abdelwahab Hamou-Lhadj. 2021. The evolution of IoT Malwares, from 2008 to 2019: Survey, taxonomy, process simulator and perspectives. *Journal of Systems Architecture* 116 (2021), 102143.

[52] Gerry Wan, Liz Izhikevich, David Adrian, Katsunari Yoshioka, Ralph Holz, Christian Rossow, and Zakir Durumeric. 2020. On the Origin of Scanning: The Impact of Location on Internet-wide Scans. In *Proceedings of the ACM Internet Measurement Conference*. 662–679.

[53] Eric Wustrow, Manish Karir, Michael Bailey, Farnam Jahanian, and Geoff Huston. 2010. Internet background radiation revisited. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. 62–74.

[54] Vinod Yegneswaran, Paul Barford, and Johannes Ullrich. 2003. Internet intrusions: Global characteristics and prevalence. *ACM SIGMETRICS Performance Evaluation Review* 31, 1 (2003), 138–147.

APPENDIX

A KNOWN SCANNERS IN 2023 AND 2024

In order to identify known scanners reaching our telescope, we collect and aggregate data from diverse sources, including the Scanner Repository by Collins et al. [12], Greynoise, Censys API, IPinfo API and IPinfo AS, as well as reverse DNS using massDNS and OSINT. For the latter, we perform exhaustive online research to identify ASes, IP addresses and network prefixes based on the Greynoise list of benign actors.

To integrate data, we employ a three-phase data warehousing and analytics process, called ETL (Extract, Transform, Load). The first step in ETL refers to the extraction of data from each data source. The staging area includes the transformation step and comprises two phases: IP-based matching (Phase-1) and IP-keyword-based matching (Phase-2). Keyword-based matching is necessitated since, in some sources, there is no direct link between the IP address and the owning entities/actors. Therefore, data needs to be scraped to extract meaningful information. Our keyword list is composed of known scanner keywords extracted from actors during IP-based matching in Phase-1, enriched with manual additions.

During IP-based matching (Phase-1), we match the source IP addresses appearing in the Darknet with those of the data sources. Phase-2 requires customized data processing for each data source. Having compiled the keyword list, we search for keywords in four datasets: Censys API, IPinfo, and reverse DNS. We extract the following fields from Censys data: WHOIS network handle, network name, organization name, WHOIS admin and abuse emails, response header location, forward and reverse DNS names and service banners. Fields are ordered from the most important to the least important one. Next, we extract domain names from IPinfo and reverse DNS and match with the keyword list. Having transformed the datasets and matched the IP addresses, we load the result files into the warehouse. Next, we launch an analytics phase where transformed data are selected for analysis.

The aggregated subset of matched IPs in 2023 pinpoints to 36 organizations, which correspond to 0.36% of the total source IP addresses and account for 51.31% of the total telescope traffic. In 2024, we identify 40 organizations, which correspond to 0.62% of the total source IP addresses and contribute 50.86% of the total telescope traffic.

Known scanners employ Internet scanning to serve certain purposes or deliver specific products. First, Stretchoid focuses on identifying online services of organizations. A search engine for Internet-connected devices is provided by Shodan and Censys. Large-scale Internet measurements are carried out by Internet Census Group to assess security performance and trends across industries. LeakIX scans and indexes web services monitoring for leaks. Intrinsec offers vulnerability management and cyber threat intelligence among its cybersecurity services. A framework for retrieving and examining DNS data is provided by bufferover.run. Palo Alto Networks provides an attack surface management solution via Cortex Xpanse. Adscore's goal is to classify website traffic that is originally generated or purchased by their client companies. CyberResilience.io provides insights into security flaws. Driftnet.io offers footprint discovery so their client companies can assess the level of their services' exposure to the Internet. Rapid7 is a cybersecurity company running Project Sonar to facilitate security research. SecurityTrails LLC offers a broad spectrum of services such DNS history, brand protection, threat hunting etc. Alpha Strike Labs performs global scans and collaborates with governmental agencies and national Computer Emergency Response Teams (CERTs). Bit Discovery is part of the Tenable attack surface management service for cyber risk management. Criminal IP offers an OSINT-based search engine for cyber threat intelligence, and an attack surface management tool. Leitwert.net, Hadrian.io and DataGrid Surface offer Threat Intelligence Data as a Service.

Non-profit security organizations like the Shadowserver Foundation are also listed as known scanners. Known scanners also include academic institutions such as UCSD, University of Michigan, TU Munich, etc. which focus on Internet measurement research aiming to improve security.

In Figures 9 and 10 we plot the port scan activity by known ("institutional") scanner in June 2023 and February 2024, respectively, as observed in our telescope. We notice that across these consecutive years, the activity per known scanner has not changed significantly. However, we notice striking differences across known scanners. Although some known scanners only scan a small number of ports, e.g., TU Munich, RWTH Aachen, and Stanford University only focus only on a few ports, there are other enterprises, e.g., Censys, Palo Alto Networks that scan all the ports, and some other that scan a very large number of all 64k ports, e.g., Criminal IP, Shodan.

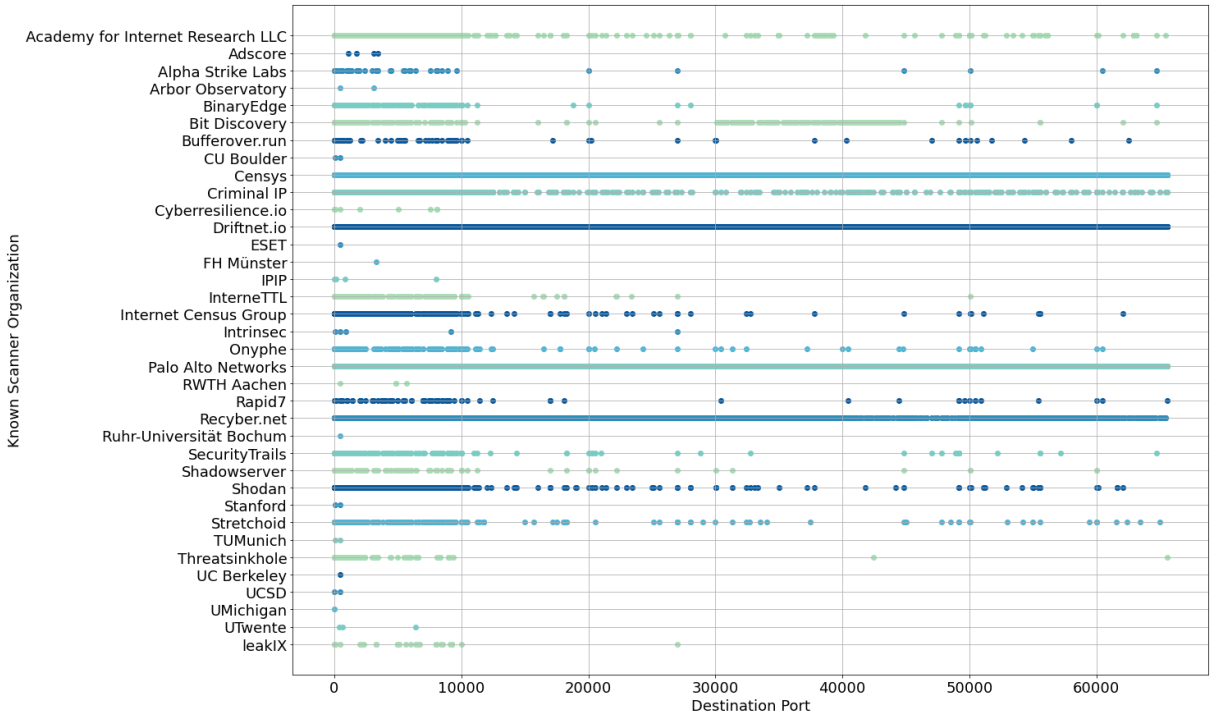

Figure 9: Ports scanned by known scanners in 2023.

Figure 10: Ports scanned by known scanners in 2024.