This paper appeared in *Proceedings of the 23rd USENIX Security Symposium*, August 2014.

An Internet-Wide View of Internet-Wide Scanning

Zakir Durumeric
University of Michigan
zakir@umich.edu

Michael Bailey
University of Michigan
mibailey@umich.edu

J. Alex Halderman
University of Michigan
jhalderm@umich.edu

Abstract

While it is widely known that port scanning is widespread, neither the scanning landscape nor the defensive reactions of network operators have been measured at Internet scale. In this work, we analyze data from a large network telescope to study scanning activity from the past year, uncovering large horizontal scan operations and identifying broad patterns in scanning behavior. We present an analysis of who is scanning, what services are being targeted, and the impact of new scanners on the overall landscape. We also analyze the scanning behavior triggered by recent vulnerabilities in Linksys routers, OpenSSL, and NTP. We empirically analyze the defensive behaviors that organizations employ against scanning, shedding light on who detects scanning behavior, which networks blacklist scanning, and how scan recipients respond to scans conducted by researchers. We conclude with recommendations for institutions performing scans and with implications of recent changes in scanning behavior for researchers and network operators.

1 Introduction

Internet-wide scanning is a powerful technique used by researchers to study and measure the Internet and by attackers to discover vulnerable hosts *en masse*. It is well known that port scanning is pervasive — including both large horizontal scans of a single port and distributed scanning from infected botnet hosts [5, 14, 15, 28, 39, 45]. However, the past year saw the introduction of two high-speed scanning tools, ZMap [19] and Masscan [23], which have shifted the scanning landscape by reducing the time to scan the IPv4 address space from months to minutes.

In this study, we examine the practice of Internet-wide scanning and explore the impact of these radically faster tools using measurement data from a large network telescope [13, 37, 46]. We analyze scan traffic from the past year, develop heuristics for recognizing large horizontal scanning, and successfully fingerprint ZMap and Masscan. We present a broad view of the current scanning landscape, including analyzing who is performing large scans, what protocols they target, and what software and providers they use. In some cases we can determine the identity of the scanners and the intent of their scans.

We find that scanning practice has changed dramatically since previous studies from 5–10 years ago [5,39,45]. Many large, likely malicious scans now originate from bullet-proof hosting providers instead of from botnets. Internet-scale horizontal scans have become common. Almost 80% of non-Conficker probe traffic originates from scans targeting $\geq$1% of the IPv4 address space and 68% from scans targeting $\geq$10%.

To understand how and why people are conducting scans, we attempt to identify individual large-scale scanning operations. We find that researchers are utilizing new scanning tools such as ZMap to cull DDoS attacks and measure distributed systems, but we also uncover evidence that attackers are using these tools to quickly find vulnerable hosts. In three case studies, we investigate scanning behavior following the disclosure of the OpenSSL Heartbleed vulnerability [36], vulnerabilities in Linksys routers, and vulnerabilities in NTP servers. In each instance, the vast majority of probe traffic originated from large, single-origin scanners. For the Linksys and OpenSSL vulnerabilities, we observed attackers applying ZMap from international bullet-proof hosting providers to complete full scans of the IPv4 address space within 24 hours of public vulnerability disclosure.

We also investigate the defensive mechanisms employed by network operators to detect and respond to scanning. Even in the most favorable case for detection — when repeated, aggressive scan traffic originates from a single IP address and would be trivial to fingerprint — we find that only a minuscule fraction of organizations respond by blocking the probes. When probes are blocked, it is often after operators inadvertently find evidence of scanning during other maintenance, rather than through

automated detection. This may indicate that the vast majority of network operators do not regard scanning as a significant threat. It also validates many recently published research studies based on Internet-wide scanning, as dropped traffic and exclusion requests appear to have minimal impact on study results.

Our findings illustrate that Internet-wide scanning is a rapidly proliferating methodology among both researchers and malicious actors. Maintaining its enormous utility for defensive security research while simultaneously protecting networks from attack is a difficult challenge. Network operators need to be aware that large vulnerability scans are taking place within hours of disclosure, but they should remember that blindly blocking all networks responsible for scanning may adversely impact defensive research. Future work is needed to develop mechanisms for differentiating between benign and malicious scans. In the mean time, we recommend close cooperation between researchers and network operators.

2 Previous Work

Most similar to our work is a study in 2004 by Pang et al. [39], who performed one of the first comprehensive analyses of Internet background radiation. Their study covers many aspects of background traffic, including the most frequently scanned protocols. However, the scanning landscape has changed drastically in the last decade—the Conficker worm [40], a major source of probe traffic, appeared in 2008, and ZMap [19] and Masscan [23] were released in 2013.

In 2007, Allman et al. [5] briefly described historical trends in scan activity between 1994 and 2006. Wustrow et al. [45] again studied Internet background radiation in 2010. They noted an increase in scan traffic destined for SSH (TCP/22) and telnet (TCP/23) in 2007, as well as increased scanning activity targeting port 445 (SMB over IP) in 2009 due to Conficker. We note a different set of targeted services and other changes in scanning dynamics since that time. Czyz et al. [14] explored background radiation in the IPv6 address space. Their work briefly touches on the presence of ICMPv6 probe traffic, but otherwise does not investigate scanning activity; we focus on the IPv4 address space.

There exists a large body of work that focuses on detecting distributed botnet scanning [22, 24, 29, 31, 43]. However, barring few exceptions, this phenomenon has remained largely hypothetical. In one exception, Javid and Paxson [28] unearthed slow but persistent SSH bruteforce attacks in 2013. Similarly, Dainotti et al. analyzed distributed botnet scanning in 2011.

Real-world responses to horizontal scanning have not been previously studied. We briefly discussed reactions to our own scanning in prior work [19], but we perform a more in-depth analysis now. Leonard et al. [32] similarly describe the complaints they received when attempting to build an Internet scanner; however, our analysis is based on a much larger data set. In addition, we perform experiments to detect instances where networks block scan probes without notice.

The dynamics of performing studies on IPv4 darknet traffic have been formally documented by both Moore et al. [37] and Cooke et al. [13]. We utilize both studies when performing calculations in this work.

3 State of Scanning

In order to understand current scanning behavior, we analyzed traffic received by a large darknet over a 16-month period. We find that large-scale horizontal scanning—the process of scanning a large number of hosts on a single port—is pervasive and that, excluding Conficker, almost 80% of scan traffic originates from large scans targeting >1% of the IPv4 address space. We find evidence that many scans are being conducted by academic researchers. However, a large portion of all scanning targets services associated with vulnerabilities (e.g. Microsoft RDP, SQL Server), and the majority of scanning is completed from bullet-proof hosting providers or from China. In this section, we describe the dynamics of these scans, including identifying the services targeted, the sources of the scans, and the largest scanning operations.

3.1 Dataset and Methodology

Our dataset consists of all traffic received by a darknet operated at Merit Network for the period from January 1, 2013 to May 1, 2014. The darknet is composed of 5.5 million addresses, 0.145% of the public IPv4 address space. During this period, the darknet received an average of 1.4 billion packets, or 55 GB of traffic, per day. For non-temporal analyses, we focus on January 2014.

In order to distinguish scanning from other background traffic, we define a scan to be an instance where a source address contacted at least 100 unique addresses in our darknet (.0018% of the public IPv4 address space) on the same port and protocol at a minimum estimated Internet-wide scan rate of 10 packets per second (pps). In the case of TCP, we consider only SYN packets.

While we cannot know for sure whether a particular scan covers the entire IPv4 address space, the darknet does not respond to any incoming packets, and the majority of its parent /8 does not host any services. As such, we expect that hosts that send repeated probes to the darknet are scanning naïvely and are likely targeting a large portion of the address space.

Detecting scans Assuming a random uniform distribution of targets, the probability that a single probe packet will be detected can be modeled by a geometric distribution and the number of packets observed by our darknet modeled by a binomial distribution [37]. A scanner probing random IPv4 addresses at the slowest rate we try to detect (10 pps) will appear in our darknet with 99% confidence within 311 seconds and with 99.9% confidence within 467 seconds. We estimate the number of packets sent to the entire IPv4 address space by approximating the binomial distribution with a normal distribution.

We process the darknet traffic using libpcap [27] and apply a single-pass algorithm to identify scans. We expire scans that do not send any packets in more than 480 seconds and record scans that reach at least 100 darknet addresses before expiring. We combine scans originating from sequential addresses in a routed block, as ZMap allows users to scan from a block of addresses. We perform geolocation using the MaxMind GeoIP dataset [35].

Fingerprinting scanners We investigate open-source scanners and fingerprint the probes generated by ZMap [19] and Masscan [23]. In ZMap, the IP identification field is statically set to 54321. In Masscan, probes can be fingerprinted using the following relationship:

$$\text{ip\_id} = \text{dst\_addr} \oplus \text{dst\_port} \oplus \text{tcp\_seqnum}$$

Because the IP ID field is only 16 bits and has a non-negligible chance of randomly being either of these values, we only consider scans in which all packets match one of the fingerprints. We find no easily identifiable characteristics for Nmap [33] probes.

3.2 Scan Dynamics

We detected 10.8 million scans from 1.76 million hosts during January 2014. Of these, 4.5 million (41.7%) are TCP SYN scans targeting less than 1% of the IPv4 address space on port 445 and are likely attributable to the Conficker worm [40]. Excluding Conficker traffic, the scans are composed of 56.4% TCP SYN packets, 35.0% UDP packets, and 8.6% ICMP echo request packets. Only 17,918 scans (0.28%) targeted more than 1% of the address space, 2,699 (0.04%) targeted more than 10%, and 614 (0.01%) targeted more than 50% (see Figure 5). However, after excluding Conficker traffic, we note that 78% of probe traffic is generated by scans targeting $\geq$1% of the IPv4 address space, 62% by scans targeting $\geq$10%, and 30% by scans targeting $\geq$50% (see Figure 4). In other words, while there is a relatively small number of large scans (0.28%), nearly 80% of scan traffic is generated by these scans.

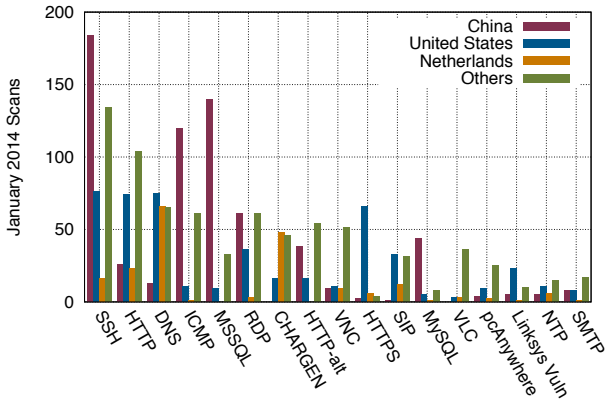

Figure 1: **Large scans ($\geq$10%) by origin country** — Many countries have distinct scanning profiles. For example, the vast majority of MSSQL scanning takes place in China.

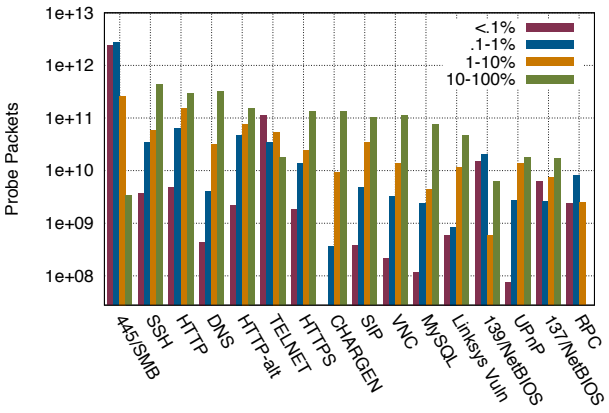

Figure 2: **Targeted ports by scan size** — Small scans target different protocols than large scans. For example, the bulk of port 445 scanning occurs in small scans, whereas port 22 is targeted by larger scans.

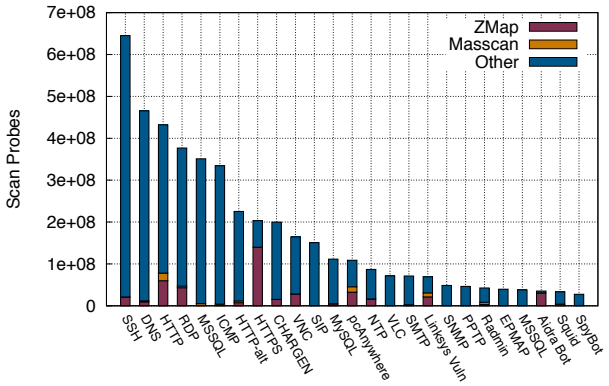

Figure 3: **Large scans ($\geq$10%) by software** — We fingerprint ZMap and Masscan probes and present the breakdown of large scans that use these scanners.

3.3 Targeted Services

Close to half of all scan traffic (48.9%) targets NetBIOS (TCP/445)—5.4 trillion SYN probes in January 2014 alone. Of these packets, 95.1% originate from small scans—scans targeting <10% of the IPv4 address space—and are likely attributable to Conficker [40, 45]. We note that small scans show different characteristics than large scans. For example, while SSH is the most targeted service in large scans, it is the seventh most targeted in small scans, accounting for only 1.3% of scan traffic.

For the most part, the protocols being targeted are not surprising, although they have shifted from previous studies in 2004 and 2010—we show the differences in Table 3. In both large and small scans, there appear to be a mix of protocols frequently associated with vulnerability scanning (e.g. Microsoft RDP, telnet, Microsoft SQL server, and VNC) as well protocols frequently studied by academic researchers (e.g. HTTP, HTTPS, SSH). We show the differences in Figure 2 and the breakdown of frequently targeted services in Tables 1 and 2.

Despite the fact that most scans originate from large international hosting providers, countries display differences in targeted protocols—particularly China, which performs regular scans against SSH, SQL Server, and Microsoft RDP. For example, while Microsoft Remote Desktop Protocol (RDP) is the fourth most scanned protocol, 77% of scans and 76% of probe packets originate from China. The second most active country (United States) is responsible for only 5.4% of probe traffic. A similar pattern emerges for ICMP echo request scans, MySQL and SSH. We show the differences by country for the top ports in Figure 1.

3.4 Scan Sources

While large scans originate from 68 countries, 76% of scan traffic originates from only five countries: China, the United States, Germany, the Netherlands, and Russia. We list the top countries that performed horizontal scans in Table 4 and the CDF in Figure 7.

While the United States and China have large allocations of address space, Germany and the Netherlands do not. In order to understand why a disproportionate amount of scan traffic is originating from smaller countries, we consider the ASes from which scans are being completed. We find that scans targeting ≥10% of the IPv4 address space occur from only 350 ASes (Figure 8). We manually classify the top 100 ASes, finding that 49 are dedicated hosting services or collocation centers, 31 are Internet service providers, 4 are academic institutions, 3 are corporations, and 13 are unidentifiable networks in China.

In the case of the Netherlands, 93% of probe traffic originates from five hosting providers: Ecatel Network,

2004 [39]	2010 [45]	2014
HTTP (80)	SMB-IP (445)	SMB-IP (445)
NetBIOS (135)	NetBIOS (139)	ICMP Ping
NetBIOS (139)	eMule (4662)	SSH (20)
DameWare (6129)	HTTP (80)	HTTP (80)
MyDoom (3127)	NetBIOS (135)	RDP (3389)

Table 3: **Temporal differences in targeted protocols** — Previous studies on background radiation show a distinct set of most frequently targeted services.

Country	Scans	Country	Scans
China	805 (31%)	Poland	61 (2.3%)
United States	582 (22%)	Korea	61 (2.3%)
Germany	247 (9.5%)	Ukraine	43 (1.7%)
Netherlands	229 (8.8%)	Brazil	34 (1.3%)
Russia	127 (4.8%)	Other	337 (13%)
France	81 (3.1%)		

Table 4: **Large scans (≥10%) by country** — A small number of countries are responsible for the majority of large scans.

Ecatel Network (NL)	Thor Data Center (IS)
Plus Server (DE)	Psychz Networks (US)
Slask Data Center (PL)	ServerStack, Inc. (US)
SingleHop (US)	Amazon.com, Inc. (US)
CariNet, Inc. (US)	LeaseWeb (NL)
SERVER4YOU (DE)	Digital Ocean, Inc. (US)
OVH Systems (UK)	GorillaServers, Inc. (US)

Table 5: **Top providers originating scan traffic** — The majority of scan probes came from large dedicated hosting and colocation providers.

Contact Point	Organizations	
Email listed on website	108	(59.7%)
WHOIS abuse contact	31	(17.1%)
Security office	22	(12.2%)
Specific individuals (e.g. CSO, CIO)	9	(5.0%)
Departmental helpdesk	5	(2.8%)
Other email contacts (e.g. postmaster)	6	(3.3%)
IT help desk phone	2	(1.1%)

Table 6: **Exclusion point of contact** — We track how organizations contacted our research team to request exclusion from future scans.

SMB over IP (TCP/445)	71.8%	SIP (UDP/5060)	0.5%	NetBIOS Helper (TCP/49153)	0.2%
ICMP Echo Request	4.8%	NetBIOS Session (TCP/139)	0.5%	Linksys Vuln. (TCP/32764)	0.2%
Microsoft RDP (TCP/3389)	3.1%	DNS (UDP/53)	0.5%	ASF-RMCP (UDP/623)	0.1%
HTTP (TCP/80)	3.0%	VLC (UDP/1234)	0.4%	SNMP (UDP/161)	0.1%
Telnet (TCP/23)	2.8%	SMTP (TCP/25)	0.2%	CHARGEN (UDP/19)	0.1%
Alt-HTTP (TCP/8080)	1.7%	VNC (TCP/5900)	0.2%	MongoDB (TCP/27017)	0.1%
SSH (TCP/22)	1.3%	Microsoft SSDP (UDP/1900)	0.2%	pcAnywhere (UDP/5632)	0.1%
HTTPS (TCP/443)	0.5%	NetBIOS Name Svc (TCP/137)	0.2%	Other	7.4%

Table 1: **Commonly targeted services for small scans (targeting <10% of the IPv4 address space)**

SSH (TCP/22)	12.5%	CHARGEN (UDP/19)	3.9%	Linksys Vuln. (TCP/32764)	1.3%
DNS (UDP/53)	9.0%	VNC (TCP/5900)	3.2%	SNMP (UDP/161)	1.0%
HTTP (TCP/80)	8.4%	SIP (UDP/5060)	2.9%	Micorosft PPTP (TCP/1723)	0.9%
Microsoft RDP (TCP/3389)	7.3%	MySQL (TCP/3306)	2.2%	Radmin (TCP/4899)	0.8%
SQL Server (TCP/1433)	6.9%	pcAnywhere (TCP/5631)	2.1%	DCOM SCM (TCP/UDP/135)	0.8%
ICMP Echo Request	6.5%	NTP (UDP/123)	1.7%	MS SQL Server (UDP/1434)	0.7%
Alt-HTTP (TCP/8080)	4.4%	VLC (UDP/1234)	1.4%	Aidra Botnet (TCP/4028)	0.7%
HTTPS (TCP/443)	4.0%	SMTP (TCP/25)	1.4%	Other	16.2%

Table 2: **Commonly targeted services for large scans (targeting $\geq$10% of the IPv4 address space)**

LeaseWeb, WorldStream, Datacenter, Nedzone, and TransIP. We note that Ecatel was one of the hosting providers that Hurricane Electric stopped peering with in 2008 due to spam traffic and malware hosting [12]. In Germany, PlusServer was responsible for 45% of probe traffic. In the United States, scanning was present from 440 ASes, but a small handful of hosting providers were responsible for 39% of scan traffic[1]. We list the hosting providers and collocation centers responsible for the most scan traffic in Table 5.

3.5 Regularly Scheduled Scans

We investigate the 25 most aggressive scanners and find several examples of both academic research scans and likely malicious groups performing repeated scans. In many of the cases where scans were performed from an academic network, researchers provided information on the purpose of their scanning. However, most scans take place from bullet-proof hosting providers or from China and provide no identifying information.

The academic and non-profit scans primarily focus on protocols used for DDoS amplification and studying cryptographic ecosystems (e.g. HTTPS and SSH). All of the groups we identified explained the purpose of their scanning and allow operators to request exclusion. Similarly, several security companies also completed scans. The Shodan Search Engine [34] was the only security group that we were able to detect that did not provide information over the web on scan addresses.

The University of Michigan performs regular ZMap scans for HTTPS hosts in order to track the certificate authority ecosystem [18, 19, 25, 47]; their data is available online at https://scans.io [17]. Ruhr-Universität Bochum completes weekly scans on ports 53, 80, 123, 137, 161, and 1900 in order to measure amplification attacks [42]. The Shadow Server Open Resolver Scanning Project [4] performs daily scans for DNS servers (UDP/53); their scanning machines are hosted by AOL. One of their hosts generated the most probes of any source in our sample — an estimated 97 billion packets in January 2014 alone. Similarly, the Open Resolver Project [3] completes weekly scans for DNS (UDP/53) and NTP (UDP/123) servers. All these institutions provide information on scan intent and how to request exclusion on a simple website at the scan source IPs.

Shodan completed 2,294 scans targeting 53 ports, sending an estimated 209 billion probes from six servers[2] in January 2014. The scans most frequently targeted ports 443, 80, 53, 32764, 1900, 23, 623, 27017, 161, and 137. Errata Security executed 89 scans of common ports using their Masscan tool. Rapid7 performed 13 scans of common ports using ZMap; their datasets are publicly available at https://scans.io [17].

There are two daily ICMP echo request scans from Guangzhou, China that jointly target an average estimated 77% of the IPv4 address space[3]. The hosts only appear to be used for these ICMP scans. A second host in

[1] CariNet (13.0%), SingleHop (11.4%), Hosting Solutions International (4.37%), Versaweb, LLC (3.46%), Psychz Networks (2.2%), Amazon.com (2.1%), and Leaseweb USA (2.0%)

[2] 198.20.69.98, 198.20.69.74, 198.20.70.114, 66.240.192.138, 71.6.5.200, and 71.6.167.142

[3] 113.108.2.117, 159.253.146.141, 220.177.198.034, and 59.46.161.130

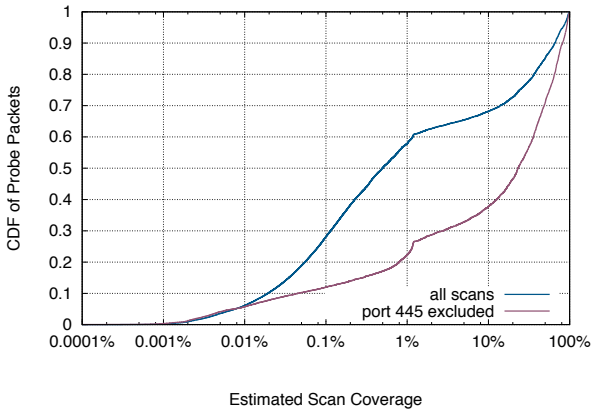

Figure 4: **CDF of scan traffic** — 40% of probes originated from scans targeting ≥1% of the IPv4 space and 30% from scans targeting ≥10%.

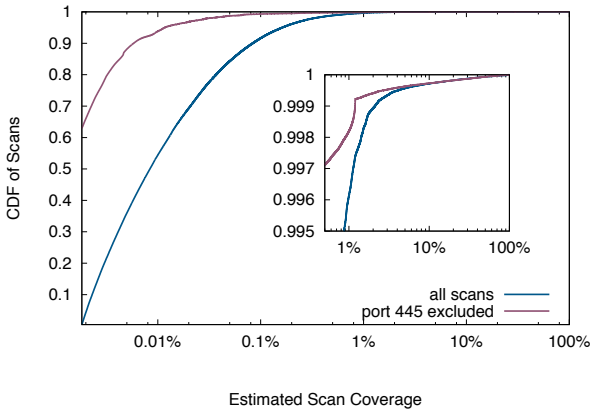

Figure 5: **CDF of scan coverage** — 45.5% of scans achieved 0.01% coverage, 8.37% achieved 0.1% coverage, and 0.38% achieved ≥1% coverage.

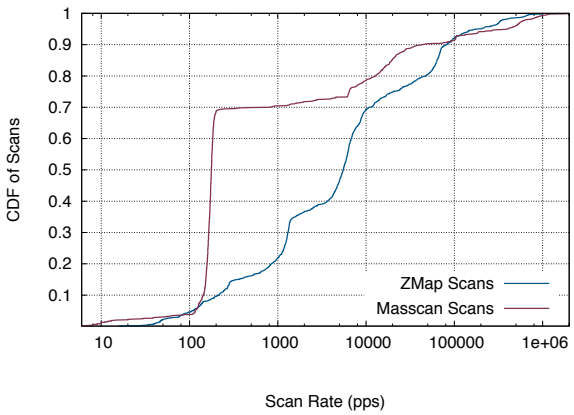

Figure 6: **CDF of scan rate** — The fastest scans operated at 2.2 Mpps (about 1.5 Gbps). However, less than 10% of scans exceeded 100 Mbps.

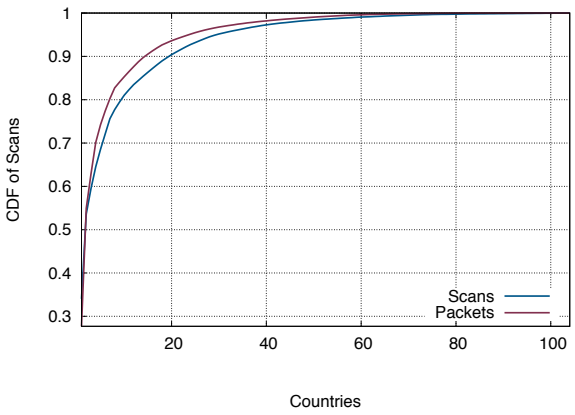

Figure 7: **CDF of scanning countries** — 76% of scans targeting ≥10% of the IPv4 address space originated from only five countries.

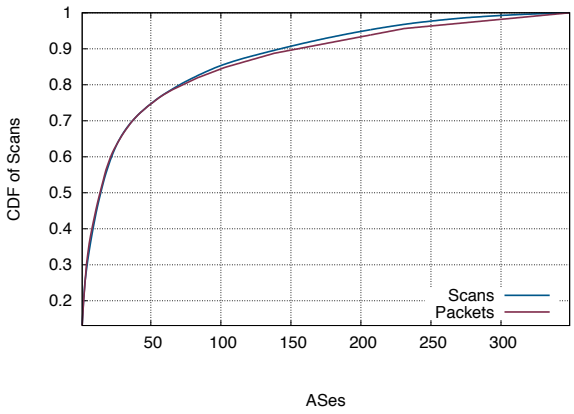

Figure 8: **CDF of scanning ASes** — Scans targeting ≥10% of the address space originated from only 350 ASes, many of them large hosting providers.

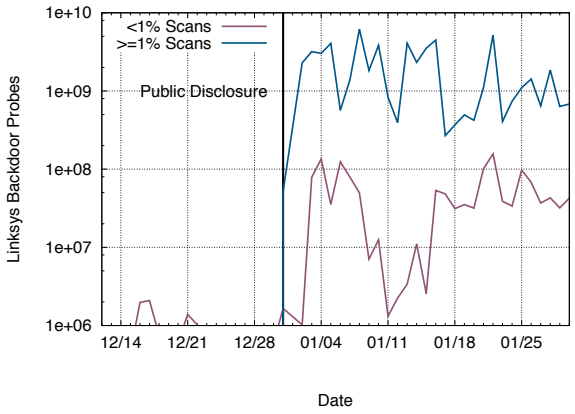

Figure 9: **Linksys backdoor** — Widespread scanning of port 32764 started within hours of the disclosure of a Linksys backdoor on that port.

Guangzhou (113.108.21.16) performs regular daily SYN scans of TCP/0, and a host in Changzhi (218.26.89.179) performs similar scans targeting SSH (TCP/22). We note that while TCP/0 is reserved, it is frequently used for fingerprinting network stacks and because it is not possible to block the port on some firewalls.

The remaining hosts in the top 25 most active scanners repeatedly scanned well-known ports and were hosted from large hosting providers in Germany, Iceland, Romania, Poland, Russia, and China. None of the hosts provided any identifying information in WHOIS records, reverse DNS records, or websites.

3.6 ZMap and Masscan Usage

The majority of scans targeting $\geq$10% of the IPv4 address space used neither ZMap nor Masscan. However, as scan coverage increases, the probability that a scanner uses ZMap steeply increases. ZMap was utilized for 133 (21.7%) of the 614 scans of more than 50% of the IPv4 address space in January 2014; Masscan was used for 21 (3.4%). Of the 242 ZMap scans targeting $\geq$10% of the address space, 70 (30%) targeted HTTP (TCP/80) and HTTPS (TCP/443) and were conducted by academic institutions and other clearly identifiable researchers. We show a breakdown of what scans used various scanners in Figure 3.

3.7 Estimated Scan Rate

In order to estimate the resources that scanners have available, we consider the estimated scan rate observed from ZMap and Masscan scans. We choose to utilize these as our metric for scan rate because the randomization algorithms are approximately uniformly random. We find that hosts are scanning between 13 pps and 1.02 million pps using ZMap and between 5 pps and 2.2 million pps—slightly more than 1.5 Gbps—using Masscan. While both tools support scanning at over 1 Gbps, all but a handful of scans were operated at much lower speeds. As shown in Figure 6, more than 90% of scans operate at under 100 Mbps, and over 70% are operated at under 10 Mbps.

4 Case Studies

Recent advances in high-speed scanning have altered the security landscape, making it possible for attackers to complete large-scale scans for vulnerable hosts within hours of a vulnerability's disclosure. In this section, we analyze scanning related to three recent vulnerabilities that affected Linksys routers, OpenSSL, and NTP servers. We find that likely attackers are taking advantage of new tools: they have started to use ZMap and Masscan from bullet-proof hosting providers instead of using distributed botnet scans. In the cases of the Linksys backdoor and the Heartbleed vulnerability, attackers began scans within 48 hours of public disclosure. We note that while conducting single-origin scans from bullet-proof hosting providers may lower the burden for attackers, it may also allow defenders to more easily detect and block scanning activity and identify the malicious actors.

4.1 Linksys Backdoor

In late December 2013, Eloi Vanderbeken disclosed a backdoor in common Cisco, Linksys, and Netgear home and small business routers [44]. The backdoor allowed full, unauthenticated, remote access to routers over an undocumented ephemeral port, TCP/32764. While there was previously only negligible traffic to the port, traffic spiked on January 2, 2014 when news sources began to cover the story [1,11,21]. There remained an average, sustained 1.98 billion estimated probe packets and 99.55 GB of traffic per day through the end of January (Figure 9).

After the disclosure, 22 hosts completed 43 scans targeting port 32764 on $\geq$1% of the IPv4 address space. Shodan [34] started scanning on December 31, 2013, within 48 hours of the disclosure, and continued to scan throughout January, approximately daily. Within one week, security consulting groups began scanning: Errata Security on January 7, M5 Computer Security on January 13, and Rapid7 on January 22. Two academic institutions, Katholieke Universiteit Leuven and Naukowa i Akademicka Sieć Komputerowa completed scans on January 3 and 6, respectively. Between January 14–16, two Chinese hosts (AS4808/China169 Beijing Province Network) completed scans. The remaining scans were performed from dedicated hosting providers[4]. No identifying information was found on any of the scanning hosts.

All non-Shodan scans utilized ZMap (71%) or Masscan (29%). Surprisingly, 98% of the probes targeting port 32764 were part of large scans targeting $\geq$1% of the IPv4 space, and 79% of probes were part of scans targeting $\geq$10%. In other words, scan traffic was not from a large number of distributed botnets hosts, but rather a small number of high-speed scanners.

While we cannot definitively determine the intent of the hosts in colocation centers, several of the providers have reputations for hosting malware and spammers, and for turning a blind eye to malicious behavior [12]. Assuming that customers of these providers are malicious, this implies that attackers completed comprehensive scans within 48 hours of disclosure using ZMap and Masscan from bullet-proof hosting providers.

[4]Hetzner Online AG (DE), UrDN/Ukranian Data Network (Ukraine), Ecatel Network (NL), Kyiv Optic Networks (Ukraine), root (Luxembourg), Digital Ocean, (US), Cyberdyne (Sweden), and Enzu (US)

4.2 Heartbleed Vulnerability

The Heartbleed Bug is a vulnerability in the OpenSSL cryptographic library [7] that was discovered in March 2014 and publicly disclosed on April 7, 2014 [36]. The vulnerability allows attackers to remotely dump arbitrary private data (e.g. cryptographic keys, usernames, and passwords) from the memory of many popular servers that support TLS, including Apache HTTP Server and nginx [36].

In the week following the disclosure, we detected 53 scans from 27 hosts targeting HTTPS. In comparison, in the week prior to the disclosure, there were 29 scans from 16 hosts. Unlike the Linksys vulnerability, there was not a sustained increase in scanning behavior. However, scan traffic was temporarily more than doubled for several days following the public disclosure.

While we do not know whether the scanners intended to exploit the vulnerability, we can detect which hosts began scanning for the first time following the disclosure. Of the 29 HTTPS scans seen prior to the disclosure, seven were daily scans from the University of Michigan, one was executed as part of Rapid7's SSL Sonar Project, and one belonged to the Shodan Project. A Chinese host (218.77.79.34) also performed daily scans. The remaining scans were operated out of bullet-proof hosting providers in the US, Great Britain, Poland, France, Iceland, and the Netherlands; none of them provided any identifying information.

Only 5 of the 27 hosts found scanning after the disclosure had previously been seen scanning on port 443, and only 3 had performed any scanning in 2014. The only recognizable organizations scanning in the week following the disclosure were the University of Michigan, Technische Universitaet Muenchen, Rapid7, Errata Security, and Nagravision. The remainder of the scans were completed from China and bullet-proof hosting providers. Within 24 hours of the vulnerability release, scanning began from China — 20 of the 53 scans (38%) originated from China. The remaining scans occurred from Rackspace, Cyberdyne, SingleHop, CariNet, Ecatel, myLoc, and Amazon EC2. 74% of the scans used ZMap; 21% used Masscan. Only three scans (6%) used other software.

4.3 NTP DDoS Attacks

Network Time Protocol (UDP/123) is a protocol that allows servers to synchronize time. In December 2013, attackers began to use NTP to perform denial-of-service amplification, in a similar way to how DNS had been abused in the past. Traffic from NTP servers began to rise around December 8, 2013 [2] and in February 2014, attackers attempted to DDoS a Cloudflare customer with over 400 Gbps of NTP traffic — one of the largest ever DDoS attacks [41].

The scanning behavior surrounding NTP is similar to what we observed for the Linksys backdoor and the Heartbleed vulnerability. Specifically, 97.3% of probe traffic destined for NTP was part of large scans (targeting $>1\%$), rather than from distributed botnet scanning. In January 2014, 29 scans from 19 hosts targeted NTP (UDP/123); 8 of the hosts used ZMap; 1 used Masscan. Three groups completed regular scans: Ruhr-Universitaet Bochum completed weekly scans, Shodan performed daily scans, and Errata Security completed one scan.

Three hosts in China completed full scans. The remaining 14 scans occurred from otherwise anonymous hosts in several hosting providers, including Ecatel, OVH Systems, FastReturn, Continuum Data Centers, and ONLINE S.A.S. One of the IPs hosts a website for the "Openbomb Drone Project" and also hosts the website http://ra.pe; the scan from the host only achieved 3% coverage; another one of the IPs hosts a site stating "#yolo"; one server had a reverse PTR record of "lulz".

As with the other vulnerabilities, there is no way to ascertain the intent of the scanners with certainty. However, the names and sites hosted on the IPs do not instill confidence that the hosts are maintained by responsible researchers rather than attackers.

5 Defensive Measures

In the previous two sections, we showed that Internet-wide scanning is widespread and that likely-attackers are scanning for vulnerabilities within 48 hours of disclosure. However, it is equally important to consider the reactions and defenses of those being scanned. Not only does this help us understand the defensive ecosystem, it also provides important data to calibrate the results from scanning research. In this section, we analyze networks' reactions to scanning, including which networks detect scan activity, drop traffic from repeat scanners, and report perceived network misuse.

Despite the fact that a large number of scans are occurring from unique source IPs and could be easily detected and blocked by network intrusion detection systems, we find that only a minuscule number of organizations block scan traffic or request exclusion. Our scan subnet at the University of Michigan is responsible for the third most aggressive scanning campaign on the Internet, yet we find that only 0.05% of the IP space is inaccessible to it. Similarly, only 208 organizations have requested that we exclude their networks from our scans, reducing the IPv4 address space for study by only 0.15%.

We further uncover evidence that networks are not detecting scans proactively, but are instead stumbling upon scans after *years* of consistent scanning — most likely during other troubleshooting or maintenance. While this lack of attention paints a dismal picture of current defensive

measures, the lack of blocking and exclusion also validates many of the recent research studies that utilize active Internet-wide measurements [8–10, 16, 18–20, 25, 26, 30, 38, 42, 47], as blacklisting does not appear to significantly bias scan results.

5.1 Detecting Blocked Traffic

In order to detect networks that are dropping scan traffic, we completed simultaneous ZMap scans from our scan subnet at the University of Michigan (141.212.121.0/24) and from a subnet that had never previously been used for scanning at the Georgia Institute of Technology. These scans took place on Wednesday, February 5, 2014 between 1:00 PM EST and 23:20 EST.

While our subnet at Michigan is used for multiple ongoing scanning effort, it has primarily been used for scanning the HTTPS ecosystem [18]. Between April 2012 and February 2014, we completed 390 scans on port 443 (HTTPS). The Michigan subnet was responsible for the third most scan traffic in January 2014. The scanning hosts all have corresponding DNS PTR records, WHOIS entries, and a simple website that describes our scanning, the data we collect, recent publications, and how to request exclusion from future research scans [19]. Despite these steps, we expected that some fraction of networks had detected our scanning and opted to silently drop traffic from our subnet.

For the simultaneous scans, we chose to scan port 443 at 100,000 pps in order to compare against our historical data on HTTPS. Both hosts used Ubuntu 12.04 and ZMap 1.2.0, and both had access to a full 1 Gbps of upstream bandwidth. We performed the two scans using ZMap, selecting identical randomization seeds such that the probes from both subnets arrive at approximately the same time.

There exists the likely possibility that some hosts were lost due to random packet drop and not intentional blocking — previous measurements on our network have shown a packet loss rate of approximately 3% [19]. In order to ensure that missing hosts are inaccessible due to blacklisting and not dropped packets, we immediately completed a secondary scan from the Michigan subnet, sending three SYN probes to each missing host, and removing hosts that were missed due to random packet drop. Previous work shows that sending three packets achieves a 99.4% success-rate [19].

We analyzed the set of hosts that appeared in scans from the "clean" subnet at Georgia Tech but not in scans from the "dirty" subnet at Michigan. We aggregate inaccessible hosts by routed block and find that there are two categories of missing hosts: (1) entire routed blocks that drop all traffic and (2) sporadic hosts and small networks belonging to large ISPs that are generally unidentifiable.

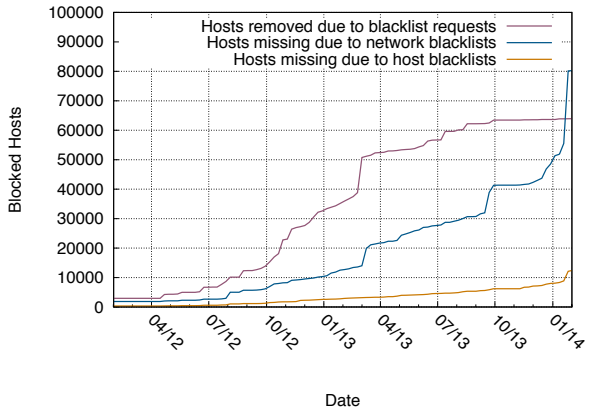

Figure 10: **Impact of blacklisting on HTTPS results** — The impact of external blacklisting and requests to be excluded from scans continue to grow over time rather than plateau.

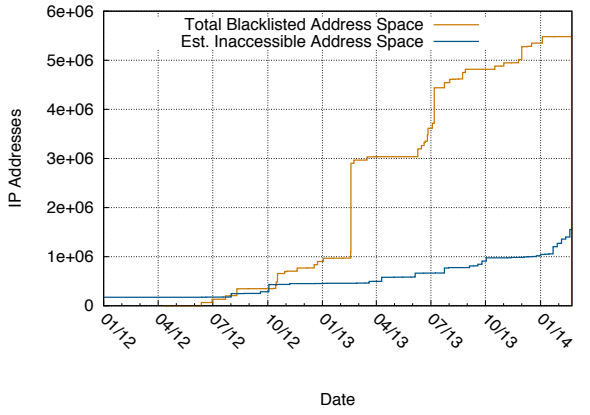

Figure 11: **Estimated impact** — We estimate inaccessible address space based on the total size of inaccessible routed blocks.

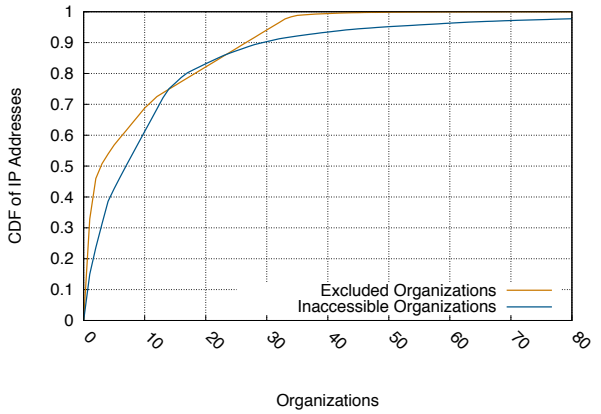

Figure 12: **CDF of blacklisting organizations** — 60% of inaccessible IPs were due to only 10 organizations.

We consider any routed block with more than three hosts in the clean subnet's scan and zero responses from the dirty subnet's scan to have blocked traffic. We find that 99,484 hosts from 612 routed blocks, 198 ASes, and 194 organizations belong to first category; 67,687 hosts belong to the second.

However, these numbers do not represent the total address space that is inaccessible to the dirty subnet, but rather the difference in hosts that respond on port 443. In order to estimate the total inaccessible address space, we consider the size of the routed blocks that appear to drop all traffic and find that these routed blocks comprise a total of 1.55 million addresses. In aggregate with the individual addresses that dropped scan traffic, we find a total of 1.62 million addresses (0.05% of the public IPv4 address space) are no longer accessible. We note that this is a lower bound of inaccessible address space as many of the individual IP addresses that we were unable to classify may represent larger, inaccessible networks. However, ultimately, only a minuscule number of organizations are detecting and blocking scan traffic.

It is important to consider not just the raw number of hosts that are inaccessible, but also the impact on the research that was being conducted by Internet-scale scanning — in our case, what percentage of the HTTPS ecosystem we are unable to measure. We compare the number of unavailable hosts to the most recent results in our HTTPS dataset, which contained TLS handshakes with 27.9 million hosts. The 167,171 inaccessible hosts would have resulted in a 0.4%–0.6% change in the result set, depending on the number of unavailable hosts that successfully completed a TLS handshake.

5.2 Organizations Blocking Scan Traffic

We identify and categorize the organizations that own each of the inaccessible routed blocks (Table 7). We note that this categorization is skewed towards organizations that are large enough to control an entire AS. Unfortunately, when attempting to classify individual IPs that blacklisted addresses, we find that most do not expose any identifying information.

As shown in Figures 10 and 11, the removal of a small number organizations resulted in large changes in the aggregate inaccessible address space — only ten organizations[5] are responsible for 60% of dropped traffic (Figure 12).

We note a bias in the countries that have blocked traffic, which we show in Table 10. However, we note that when considering the percentage of blacklisted addresses per

[5]Enzu, Corespace, Internode, Fidelity National Information Services, AR Telecom, Western Australia Department of Finance, State of Tennessee, Hershey Chocolate & Confectionery Corporation, DFN (German National Research and Education Network), and Research Organization of Information and Systems National Institute of Informatics (Japan)

Type	Organizations	Hosts
Internet service provider	73	389,120
Corporation	36	448,000
Hosting provider	34	344,832
Government	22	299,008
Academic institution	12	255,232
Small/medium business	12	63,232
Unknown	6	1,792
Total	195	1,801,216

Table 7: **Organizations that filter scans** — We categorize the organizations that blacklist scan traffic.

Type	Organizations	Hosts
Small/medium business	45	391,358
Individual	39	102
Corporation	30	671,060
Academic institution	19	1,654,401
Government	13	926,210
Internet service provider	6	1,838,827
Unknown	5	32,772
Total	157	5,514,730

Table 8: **Organizations that request exclusion** — We classify the organizations that have requested exclusion from future scans.

Country	Organizations	
United States	129	(63.0%)
United Kingdom	15	(7.4%)
Germany	12	(5.9%)
Australia	9	(4.4%)
Canada	7	(3.4%)
Other	32	(15.0%)

Table 9: **Excluded addresses by country** — We geolocate the organizations that have requested exclusion and find that the majority are in the United States.

Country	Orgs	Hosts	% Addr Space
United States	96	1,029,632	0.07%
Korea	8	43,008	0.03%
Canada	7	25,344	0.04%
Austria	7	225,024	0.40%
Great Britain	5	1,536	0.001%
Romania	5	3,072	0.03%
France	5	133,120	0.17%
Portugal	5	80,640	1.1%
India	4	1,280	0.002%
Russia	4	8,192	0.01%

Table 10: **Inaccessible hosts by country** — We geolocate the routed blocks that are no longer accessible to scanning hosts.

country, a different pattern emerges, because the removal of a single AS can greatly impact the availability within the region. For example, while only one organization in Nigeria blacklisted our subnet, this single rule blocked more than 1% of the country's IP space. A similar situation appears in Portugal, Ireland, Luxembourg, Honduras, Argentina, and Lithuania.

5.3 Organizations Requesting Exclusion

Another indicator of scan detection can be found in the scan exclusion requests that we receive. Over the course of our HTTPS scanning, we have received 208 exclusion requests—resulting in the removal of 5.4 million addresses from our study—0.15% of the public IPv4 address space. Of the excluded hosts, 1.46 million (28%) had previously been seen hosting HTTPS. In comparison, only 1% of IPv4 hosts respond on port 443. We present the types of organizations that have requested exclusion in Table 8 and countries in Table 9. As with the organizations that dropped scan traffic, the majority of requests originated from the United States. We only received four requests from Asia and Africa: one each from Taiwan, India, South Africa, and Japan.

In our prior work [19], we suggest that researchers post a website that explains the purpose of their scanning and that they coordinate with their local network administrators. In order to understand whether this information was useful to network operators and to revise our recommendations, we tracked how network operators contacted us. We find that almost 60% of emails were sent directly to our research team via the site hosted on the scan IPs, 17% were sent to the WHOIS abuse contact, and 12% were sent to our institution's security office (e.g. security@umich.edu). We show a breakdown of contact points in Table 6.

Our informational page has been viewed by 6,600 unique users with an average of 357 visitors per month. More than 90% of visitors used common web browsers (Chrome, Firefox, Internet Explorer, Safari, or Opera). Viewers primarily geolocated to the United States, Germany, United Kingdom, Canada, and Japan. The ratio of page views to complaints (approximately 1:30) suggests that many organizations are cognizant of our scanning activity and do not object to it.

5.4 Blacklisting Scope

While we expected that a small number of organizations would block our scan hosts, it is not immediately clear what network segment organizations would block. We scanned from an additional, unrelated /24 in our institutional AS and found that 38,648 (39%) of the hosts that we could not reach on port 443 are also unavailable from the unrelated /24 in our AS. In other words, 39% of organizations that blocked our dirty subnet blocked the entire /16 in which our scan subnet is located or blocked our entire AS. In terms of estimated total inaccessible address space, 338,944 addresses (18.7% of the addresses inaccessible in our scan subnet) are possibly unavailable from the entire AS.

5.5 Temporal Analysis of Scan Detection

We initially hypothesized that our scanning would cause observant networks to immediately blacklist our network or contact our research team. If this were the case, we would expect that network exclusion requests would plateau after several scans. Instead, we find that organizations are slowly continuing to blacklist our scan subnet or request exclusion more than two years after we began regular scanning. In order to estimate when users detected scanning and blacklisted the scan subnet, we analyzed our historical data on the HTTPS ecosystem and recorded the last time any IP address in each routed block responded.

As shown in Figure 10, there is no plateau in the number of blacklisted hosts or in the number of organizations that have requested removal. Instead, we find that organizations continue to freshly notice the scanning behavior and to blacklist us or request exclusion. Further, more than half of the organizations began starting dropping traffic after more than a year of daily scans. We suspect that the organizations that request exclusion or begin blocking traffic years later are not proactively noticing scan traffic, but rather happening upon log entries during other maintenance and troubleshooting.

5.6 Scan Detection Mechanisms

In order to understand how organizations detect scans, we categorized the emails requesting exclusion or alerting us of potential abuse. In 64 cases (31%), network operators included evidence that was copied directly from log files or otherwise explained how they detected our scanning.

In 50% of cases, network operators noticed scans in their firewall or IDS logs. However, in 22% of reports, operators did not detect scanning in a firewall, but rather in their web logs (primarily Apache or nginx), and in 16% of cases, administrators noticed our scanning as our HTTPS handshake appeared to be a malformed handshake in SSH or OpenVPN logs. We show a breakdown of detection mechanisms in Table 11.

5.7 Revised Recommendations

We further emphasize the importance of researchers serving an informational webpage given the high percentage of users who used this to find contact information and the high number of views by network operators. We also recommend that researchers notify the owners of

Detection Mechanism	Organizations	
Firewall logs	22	(34%)
Web server logs	14	(22%)
Intrusion detection system (IDS) logs	10	(16%)
Invalid SSH or OpenVPN handshake	10	(16%)
Public blacklists	2	(3%)
Other	6	(9%)

Table 11: **Scan detection methods** — We classify the type of evidence included in email requests to be excluded in order to understand how organizations detect scanning.

various other email accounts at the institution including postmaster and administrator, in addition to institutional help desks, departmental administrators, and IT officials.

We add the additional recommendation that researchers publish the subnet being used for their research. This allows organizations that decide to drop traffic a mechanism to blacklist the correct subnet instead of dropping traffic from the entire institution.

6 Future Work

While we shed light on the broad landscape of large horizontal scans, there remain several open questions surrounding scan detection and defensive mechanisms.

Correlating distributed scanners It remains an open research problem to detect and correlate distributed scanning events. While we are able to estimate broad patterns in scanning behavior, we excluded scanners that operate at under 10 pps or targeted fewer than 100 hosts in our darknet. This likely excludes slow, massively distributed scans [6, 15]. While there has been previous research on detecting distributed scanning, little work has applied these to darknet data, in order to understand the slow scans that are taking place. Similarly, our darknet is primarily composed of contiguous address space, which may be avoided by some operations. It reamains an open issue to analyze distributed network telescopes to determine whether attackers are avoiding large blocks of consistently unresponsive address space.

IPv6 scanning In this work, we focused on scanning within the IPv4 address space. Scanning the IPv6 address space efficiently remains an open problem, as does analyzing existing IPv6 scanning behavior.

Vertical scanning Our study focused on horizontal scanning — scanning a single port across a large number of hosts. We note that during this investigation, we also stumbled upon several cases of large vertical scanning operations, which deserve further attention.

Exclusion standards Blacklisting by external organizations indicates a lack of communication between researchers and network operators. This misalignment has lead to organizations dropping all traffic from institutional ASes, which may have other adverse impacts. There currently exists no standard for system operators to request exclusion. Further work is needed to develop a standard similar to HTTP's robots.txt to facilitate this communication.

Determining intent Given that the majority of scanning takes place from large hosting providers, it is oftentimes difficult to discern the intent of the scanner beyond scanned protocol. Follow-up work is necessary to determine the follow-up actions of these scanners. Given that these large scans are happening from a small number of hosts, it may be possible to determine owners and track from where these attacks are originating. Automated mechanisms for signaling benign intent (such as centrally maintained whitelists) could help network operators distinguish between harmful and beneficial instances of wide-scale scanning.

Understanding defensive reactions We find that a minuscule number of organizations are dropping scan traffic. However, it is unclear whether other organizations are aware of and deliberately permit this research-focused traffic, or whether they are entirely unaware of it. More investigation is needed to understand the attentiveness of these organizations.

7 Conclusion

In this work, we analyzed the current practice of Internet-wide scanning, finding that large horizontal scanning is common and is responsible for almost 80% of non-Conficker scan traffic. We analyzed who is scanning and what services they are targeting noting differences from previously reported results. Ultimately, we find that researchers and attackers are both taking advantage of new scanning tools and hosting options — adapting to new advances in technology in order to further reduce the burden for finding vulnerabilities. While the landscape of scanning is evolving, defenders have remained sluggish in detecting and responding to even the most obvious scans.

Acknowledgments

We thank Paul Royal, Adam Allred, and their team at the Georgia Institute of Technology, as well as Thorsten Holz, Christian Rossow, and Marc Kührer at Ruhr-Universität Bochum for facilitating scans from their institutions. We similarly thank the exceptional sysadmins at the University of Michigan for their help and support throughout this project. This research would not have been possible without Kevin Cheek, Chris Brenner, Laura Fink, Paul

Howell, Don Winsor, and others from ITS, CAEN, and DCO. The authors thank Michael Kallitsis and Manish Karir of Merit Network for helping facilitate our darknet analysis. We additionally thank David Adrian, Brad Campbell, Jakub Czyz, Jack Miner III, Pat Pannuto, Eric Wustrow, and Jing Zhang. This work was supported in part by the Department of Homeland Security Science and Technology Directorate under contracts D08PC75388, FA8750-12-2-0235, and FA8750-12-2-0314; the National Science Foundation under contracts CNS-0751116, CNS-08311174, CNS-091639, CNS-1111699, CNS-1255153, and CNS-1330142; and the Department of the Navy under contract N000.14-09-1-1042.

References

[1] Backdoor found in Linksys, Netgear routers. https://news.ycombinator.com/item?id=6997159.

[2] Hackers spend Christmas break launching large scale NTP-reflection attacks. http://www.symantec.com/connect/blogs/hackers-spend-christmas-break-launching-large-scale-ntp-reflection-attacks.

[3] Open resolver project. http://openresolverproject.org/.

[4] Shadowserver open resolver scanning project. https://dnsscan.shadowserver.org/.

[5] M. Allman, V. Paxson, and J. Terrell. A brief history of scanning. In *Proc. 7th ACM SIGCOMM conference on Internet measurement*, pages 77–82, 2007.

[6] Anonymous. Internet census 2012. http://census2012.sourceforge.net/paper.html, Mar. 2013.

[7] L. Bello. DSA-1571-1 OpenSSL — Predictable random number generator, 2008. Debian Security Advisory. http://www.debian.org/security/2008/dsa-1571.

[8] A. J. Bonkoski, R. Bielawski, and J. A. Halderman. Illuminating the security issues surrounding lights-out server management. In *Proc. 7th USENIX Workshop on Offensive Technologies*. USENIX, 2013.

[9] J. W. Bos, J. A. Halderman, N. Heninger, J. Moore, M. Naehrig, and E. Wustrow. Elliptic curve cryptography in practice. In *Proc. 18th International Conference on Financial Cryptography and Data Security (FC)*, 2014.

[10] S. Checkoway, M. Fredrikson, R. Niederhagen, M. Green, T. Lange, T. Ristenpart, D. J. Bernstein, J. Maskiewicz, and H. Shacham. On the practical exploitability of dual EC in TLS implementations.

[11] R. Chirgwin. Hacker backdoors Linksys, Netgear, Cisco and other routers. http://www.theregister.co.uk/2014/01/06/hacker_backdoors_linksys_netgear_cisco_and_other_routers/.

[12] J. Conway. Ecatel's harboring of spambots and malware causes BGP peers to stop peering with them. http://www.sudosecure.com/ecatels-harboring-of-spambots-and-malware-causes-bgp-peers-to-stop-peering-with-them/.

[13] E. Cooke, M. Bailey, Z. M. Mao, D. Watson, F. Jahanian, and D. McPherson. Toward understanding distributed blackhole placement. In *Proc. ACM workshop on Rapid malcode*, pages 54–64, 2004.

[14] J. Czyz, K. Lady, S. G. Miller, M. Bailey, M. Kallitsis, and M. Karir. Understanding IPv6 Internet background radiation. In *Proc. 13th ACM SIGCOMM Conference on Internet Measurement*, 2013.

[15] A. Dainotti, A. King, F. Papale, A. Pescape, et al. Analysis of a /0 stealth scan from a botnet. In *Proc. 12th ACM SIGCOMM Conference on Internet Measurement*.

[16] Z. Durumeric, D. Adrian, M. Bailey, and J. A. Halderman. Heartbleed bug health report. https://zmap.io/heartbleed/.

[17] Z. Durumeric and J. A. Halderman. Internet-wide scan data repository. https://scans.io.

[18] Z. Durumeric, J. Kasten, M. Bailey, and J. A. Halderman. Analysis of the HTTPS certificate ecosystem. In *Proc. 13th ACM SIGCOMM Internet Measurement Conference*, Oct. 2013.

[19] Z. Durumeric, E. Wustrow, and J. A. Halderman. ZMap: Fast Internet-wide scanning and its security applications. In *Proc. 22nd USENIX Security Symposium*, Aug. 2013.

[20] P. Eckersley and J. Burns. An observatory for the SSLiverse. Talk at Defcon 18 (2010). https://www.eff.org/files/DefconSSLiverse.pdf.

[21] S. Gallagher. Backdoor in wireless DSL routers lets attacker reset router, get admin. http://arstechnica.com/security/2014/01/backdoor-in-wireless-dsl-routers-lets-attacker-reset-router-get-admin/.

[22] C. Gates. Coordinated scan detection, 2009.

[23] R. Graham. MASSCAN: Mass IP port scanner. https://github.com/robertdavidgraham/masscan.

[24] J. Green, D. J. Marchette, S. Northcutt, and B. Ralph. Analysis techniques for detecting coordinated attacks and probes. In *Workshop on Intrusion Detection and Network Monitoring*, pages 1–9, 1999.

[25] N. Heninger, Z. Durumeric, E. Wustrow, and J. A. Halderman. Mining your Ps and Qs: Detection of widespread weak keys in network devices. In *Proc. 21st USENIX Security Symposium*, Aug. 2012.

[26] R. Holz, L. Braun, N. Kammenhuber, and G. Carle. The SSL landscape: A thorough analysis of the X.509 PKI using active and passive measurements. In *11th ACM SIGCOMM conference on Internet measurement (IMC)*, 2011.

[27] V. Jacobson, C. Leres, and S. McCanne. libpcap. Lawrence Berkeley National Laboratory, Berkeley, CA. Initial release June 1994.

[28] M. Javed and V. Paxson. Detecting stealthy, distributed SSH brute-forcing. In *Proc. ACM SIGSAC conference on Computer & communications security*, pages 85–96, 2013.

[29] J. Jung, V. Paxson, A. W. Berger, and H. Balakrishnan. Fast portscan detection using sequential hypothesis testing. In *Proc. IEEE Symposium on Security and Privacy*, pages 211–225, 2004.

[30] J. Kasten, E. Wustrow, and J. A. Halderman. CAge: Taming certificate authorities by inferring restricted scopes. In *17th International Conference on Financial Cryptography and Data Security (FC)*, 2013.

[31] C. Leckie and R. Kotagiri. A probabilistic approach to detecting network scans. In *Network Operations and Management Symposium (NOMS)*, pages 359–372, 2002.

[32] D. Leonard and D. Loguinov. Demystifying service discovery: Implementing an Internet-wide scanner. In *Proc. 10th ACM SIGCOMM conference on Internet measurement (IMC)*, pages 109–122, 2010.

[33] G. F. Lyon. *Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning*. Insecure, USA, 2009.

[34] J. C. Matherly. SHODAN the computer search engine, Jan 2009. http://shodanhq.com.

[35] MaxMind, LLC. GeoIP, 2013. http://www.maxmind.com/en/city.

[36] N. Mehta and Codenomicon. The heartbleed bug. http://heartbleed.com/.

[37] D. Moore, C. Shannon, G. M. Voelker, and S. Savage. *Network telescopes: Technical report*. Department of Computer Science and Engineering, University of California, San Diego, 2004.

[38] H. Moore. Security flaws in universal plug and play. Unplug. Don't Play, Jan. 2013. http://community.rapid7.com/servlet/JiveServlet/download/2150-1-16596/SecurityFlawsUPnP.pdf.

[39] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of Internet background radiation. In *Proc. 4th ACM SIGCOMM Conference on Internet Measurement*, pages 27–40, 2004.

[40] P. Porras, H. Saidi, and V. Yegneswaran. Conficker C analysis. *SRI International*, 2009.

[41] M. Prince. Technical details behind a 400Gbps NTP amplification DDoS attack. http://blog.cloudflare.com/technical-details-behind-a-400gbps-ntp-amplification-ddos-attack.

[42] C. Rossow. Amplification hell: Revisiting network protocols for DDoS abuse. In *Proc. Network and Distributed System Security Symposium*, Feb. 2014.

[43] S. E. Schechter, J. Jung, and A. W. Berger. Fast detection of scanning worm infections. In *Recent Advances in Intrusion Detection*, pages 59–81. Springer, 2004.

[44] E. Vanderbeken. TCP-32764: some codes and notes about the backdoor listening on tcp-32764 in linksys WAG200G. https://github.com/elvanderb/TCP-32764.

[45] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston. Internet background radiation revisited. In *Proc. 10th ACM SIGCOMM Conference on Internet Measurement*, pages 62–74. ACM, 2010.

[46] V. Yegneswaran, P. Barford, and D. Plonka. On the design and use of internet sinks for network abuse monitoring. In *Recent Advances in Intrusion Detection*, pages 146–165. Springer, 2004.

[47] J. Zhang, Z. Durumeric, M. Bailey, M. Karir, , and M. Liu. On the mismanagement and maliciousness of networks. In *Proc. Network and Distributed System Security Symposium (NDSS)*, Feb. 2014.