Can’t Stop The Scan:
An Empirical Study on Opting out of Internet-wide Scanning

Anonymous authors
Anonymous institutions

Abstract

Internet-wide scanning is prevalent. This is due to the availability and widespread adoption of high-speed scanning tools such as ZMap and Masscan, adopted by researchers and security organizations to perform Internet-wide scanning for internet censuses. However, benign scanning traffic can create undesirable noise for network administrators or researchers monitoring network traffic for security-related events. To mitigate the negative effects, prior research has proposed best practices for conducting ethical, well-regulated, Internet-wide scans. In this paper, we are the first to shed light on the practicality of these best practices, mainly focusing on opting-out of the scans. By analyzing large-scale darknet traffic, we identify 46 scan organizations, including some that had not been reported in previous studies. Still, nearly 70% of the scanners we considered to be for survey purposes do not reveal their identity. We show that even among scanners with identifiable identities, about half do not implement opt-out measures, suggesting that the effectiveness of opting out is limited. Furthermore, only seven scanners confirmed that an opt-out request was sent from a legitimate administrator, indicating a challenge in verifying the authenticity of opt-out requests. Based on these findings, we deepen the discussion on the practical coping of scan organizations and recipients to make Internet-wide scanning sustainable.

1 Introduction

Internet scanning, which communicates indiscriminately with a wide range of global IP addresses to search for responsive devices, has primarily been used by worm-type malware (hereinafter referred to as "worm") that spreads infection via networks. Due to the proliferation of IoT devices, a wide variety of devices have been connected to the Internet, which in turn makes various services run on these devices accessible, either intentionally or unintentionally, from the Internet. Such Internet-accessible devices and services can become potential targets for attacks. Mirai worm, which emerged in 2016, spread infections to IoT devices that were operating with easily guessable ID/Password combinations accessible via Telnet and subsequently carried out large-scale DDoS attacks [2]. Particularly, IoT devices often lack sufficient security measures and maintenance. As a result, vulnerabilities, if discovered in these devices, tend to remain unaddressed without proper updates or countermeasures. Consequently, researchers and security companies are making efforts to conduct wide-scale Internet scans to identify devices and services that are accessible from the Internet and are susceptible to such threats. In particular, numerous Internet scans have been reported since the development and open-source release of high-speed scanners such as Zmap [14] and Masscan [31]. By analyzing the information obtained from these Internet scans, it is possible to identify unintentionally exposed services, end-of-life (EoL) devices still in operation, and vulnerabilities in operating systems and services, among other potential risks, providing valuable insights for enhancing security measures.

On the other hand, such benign scans potentially disrupt normal network operations and security processes. Specifically, a large volume of scan packets may trigger alerts from an organization’s security devices such as IDS and IPS, increasing the monitoring workload for security operators and potentially causing delays in responding to more critical alerts. In addition, for organizations operating large-scale networks, such as universities or large corporations, the concentration of scan packets on gateway routers or firewalls can overload the devices and trigger network problems. Security researchers also engage in activities aimed at tracking the ongoing attack activities on the Internet through darknet monitoring and honeypots. However, Internet scanning can introduce undesirable noise into their observation data, thereby obstructing accurate analysis.

Internet scanning can have negative impacts. To mitigate these, prior research [14] has proposed best practices for conducting ethical scans. The best practices include disclosing the scanner’s identity, explaining the scan’s purpose without exceeding its scope, and providing an opt-out method and appropriate responses, etc. Among these best practices, the perspective of organizations affected by the negative effects of internet scanning is that processing opt-out requests appropriately is important. However, to the best of our knowledge, there is no research that has demonstrated how scan organizations respond to opt-out requests, or in other words, whether opt-out mechanisms are functioning effectively.

In this paper, we shed light on the practicality of implementing these best practices, primarily focusing on opting-out of the scans. Through an analysis of large-scale darknet traffic, we identify scan organizations that conduct internet-wide scans for survey purposes. To determine if these organizations adhere to recommended practices, we examine DNS and Whois information and details about their websites, among other factors. Regarding the opt-out process, we verify their response and assess whether they stop scanning upon receipt of an opt-out request.

Our contributions are summarized as follows:

- We find 46 scan organizations, including some that had not been reported in prior research. However, nearly 70% of the scan source IP addresses we believe to be associated with survey purposes do not reveal their identity.
- We show that many scan organizations do not disclose their source IP addresses and target ports/protocols. This lack of transparency implies that organizations being scanned lack the necessary information to preemptively filter out scan traffic.
- We are the first to empirically investigate the opt-out response of Internet-wide scanners and find that even among scanners that maintain clear identities, 44% of scan organizations fail to respond appropriately to opt-out requests and continue scanning. This finding suggests that under the current circumstances, opting-out of Internet-wide scanning offers limited effectiveness. We also discuss the difficulty in verifying the authenticity of requests, as only seven scan organizations confirm that opt-out requests come from legitimate network administrators.
- Based on our experimental results, we make additional recommendations for both the scanners and the receivers to mitigate the negative effects of Internet-wide scanning.

The rest of the paper is organized as follows. Section 2 describes the stakeholders of Internet-wide scanning and introduces the related work. In Section 3 we introduce our dataset and describe our methodology for detecting Internet-wide scans and verifying compliance with the best practices. Section 4 shows the results obtained from our methodology. In Section 5 we discuss some recommendations for scanners and receivers based on our results, as well as limitations of this study. Section 6 summarizes our results and offers some conclusions.

2 Background

In this section, we first describe the stakeholders involved in Internet-wide scans: Internet scanners, receivers, and Internet infrastructure providers. Then, we summarize related work.

2.1 Stakeholders

Scanners: The Morris worm [35], the world's first Internet worm, appeared in 1988 at the dawn of the Internet, and since then worm-infected devices have been the primary Internet-wide scanners [2, 4, 6, 26, 27, 35]. Worms search for the next targeted devices by performing Internet-wide scans from infected devices and then aggressively spread the infection by exploiting vulnerabilities found in the devices that respond. In the early 2000s, the main target of worm infection was devices with Microsoft Windows OS, and Internet-wide scans were conducted from many infected Windows devices [4, 6, 27]. However, the past decade has seen a shift in worm infections from Windows devices to IoT devices due to improved Windows OS security and the rise of IoT devices. Since its emergence in 2016, the IoT malware Mirai and its variants have infected hundreds of thousands of IoT devices worldwide, including routers and web cameras, and conducted extensive Internet-wide scans [2, 17].

Other major scanners are security researchers and security companies. The development of stateless high-speed network scanners, Zmap [14] and Masscan [31], significantly reduced the time required to scan the entire IPv4 address space. As a result, security researchers and companies started utilizing these high-speed scanners for research and survey purposes to conduct internet-wide scans [2, 11, 13, 24, 32, 36]. Several services have also emerged that allow users to search for information on IoT devices collected through Internet-wide scanning [7, 34, 40].

Receivers: Internet-wide scanning is performed over the entire IP address space, making all Internet-connected device users potential receivers. In particular, companies, universities, and other large networks are susceptible to Internet-wide scanning. Organizations with large networks receive a substantial volume of scan traffic, which places a high load on gateway routers and firewalls and can potentially lead to network disruptions. Furthermore, IDS and other security alerts triggered by these scans can increase the cost of managing security operations within an organization.

Numerous organizations, engaged in monitoring and analyzing cyber-attacks through observation of Internet traffic, are significantly impacted by the prevalence of Internet-wide scanning. For example, darknet monitoring (a.k.a network telescope) [3, 6, 10, 12, 15, 25, 26, 28, 30, 32, 39] analyzes cyber-attack activities such as scanning from worm-infected devices by observing traffic reaching unused IP addresses. However, if traffic from the Internet-wide scans for survey purposes

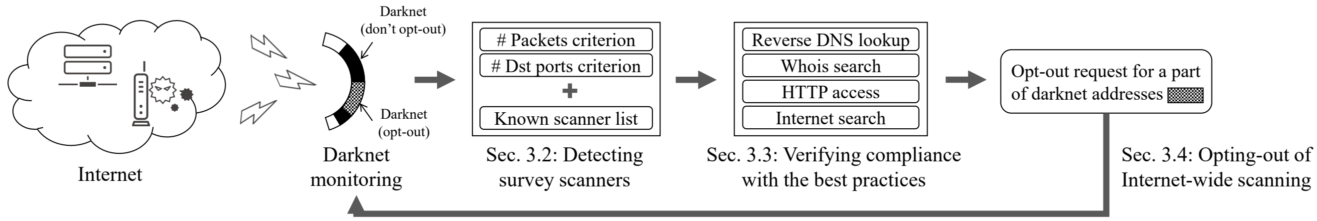


Figure 1: Overview of the methodology

is included in the darknet monitoring data because it also reaches unused IP addresses, it becomes noise in the data analysis.

Internet infrastructure providers: Internet infrastructure providers, such as ISPs positioned in the path between scanners and receivers of Internet-wide scans, may experience direct or indirect impacts due to these scans. Internet-wide scan packets are mainly TCP SYN or UDP packets to confirm port openings and are small in data size, so the traffic is unlikely to fill the ISP's network bandwidth. However, some survey scanners have been reported to perform scans at very high speeds, which may cause problems at the ISP scale [29]

Apart from direct impacts like increased traffic volume, there are also indirect ramifications to consider. For instance, the receivers of the Internet-wide scans might report network abuse to the Internet service provider or hosting service provider of the scanner based on the source IP address. Such action may impose upon those Internet infrastructure providers the cost of handling these abuse reports and could raise concerns about potential damage to their reputation.

2.2 Related Work

Darknet Monitoring: Darknet monitoring (a.k.a network telescope) is one of the major methods of observing malicious activities on the Internet [3, 25, 28]. A darknet is a set of routed but unused IP address spaces. Given the absence of real PCs and servers, darknet traffic includes only abnormal traffic and reflects malicious activities such as scanning by malware-infected hosts, sending shellcode with UDP packets, and backscatters of DDoS attacks. Pang et al. conducted the initial study exploring the broad characteristics of darknet traffic [28]. Their analysis addressed various aspects such as prevalence and variability of darknet traffic, frequently targeted ports and protocols, and the behavioral characteristics of source hosts. In subsequent work, Wustrow et al. revisited the characterization of darknet traffic [35]. Their analysis, which spanned a five-year period from 2006 to 2010, examined the changes within darknet traffic. Durumeric et al. analyzed who is scanning, what services they are targeting, and the impact of fast scanning tools using large-scale darknet monitoring data [12]. Over the past two decades, studies have been conducted analyzing the spread of various worms from Windows worms such as Code-Red [27], SQL Slammer [26], Blaster [4], Sality [10], and Conficker [6] to the IoT worms such as Carna [20] and Mirai [2] through darknet monitoring. Darknet monitoring studies have mainly focused on the IPv4 address space, but there are some previous studies on darknet monitoring in the IPv6 address space [8, 15].

Darknet traffic data can be used for Internet measurement in addition to cyber-attack analysis. Dainotti et.al proposed a method for estimating IP address space usage using darknet observation data and information from active scanning [9]. Guillot et.al proposed a system to detect remote connectivity loss by detecting deviations from periodic forecasts of darknet observation data [19].

Internet-wide scanning: The introduction of Zmap [14] and Masscan [31], high-speed scanners that appeared in 2013, changed the Internet-wide scanning landscape. Durumeric et.al demonstrated that one machine could scan the entire IPv4 address space for a single port in less than 45 minutes. These powerful scanning tools were released as open-source software, allowing many security researchers and security companies to perform their own Internet-wide scans.

Durumeric et al. surveyed the OpenSSL Heartbleed vulnerability across popular HTTPS websites and the entire IPv4 address space through extensive active scanning [13]. Mirian et al. conducted Internet-wide scans against the SCADA protocols and reported many vulnerable ICS devices accessible from the Internet [24]. Several prior studies have focused on Internet scanners and profiled scanning characteristics and scan organizations [12, 22, 36]. Most relevant to our research, Mazel et al. identified 18 scan organizations for survey purposes and investigated how each scan organization published its identity [22]. As the first search engine for IoT devices, Shodan [34] was launched in 2009. Shodan is a search engine that conducts regular Internet-wide scans and allows users to search for diverse IoT devices (webcams, routers, etc.) using a variety of filters. Censys [7], a search engine similar to Shodan, was launched in 2013 by the researchers who developed Zmap. Another search engine, ZoomEye [40], is operated by a Chinese company.

Observation period Darknet size	3/1/2022–5/31/2022 Approx. 300,000 IPs
# Packets # Packets (TCP SYN) # Packets (UDP)	127,199,253,047 115,952,109,275 11,247,143,772
# unique source (scanner) IPs	21,648,734

Table 1: Statistics of darknet traffic data

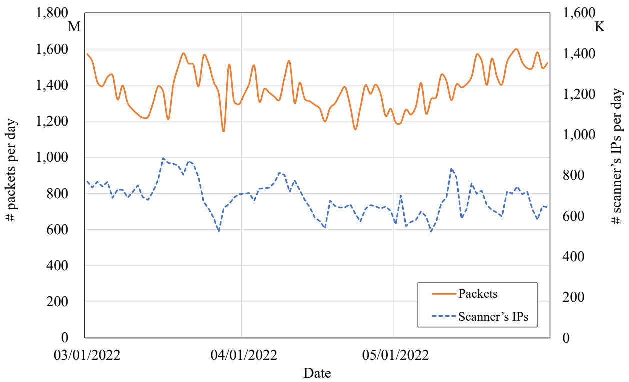


Figure 2: Daily trends in darknet traffic

3 Methodology

Figure 1 outlines the overall process of our survey. In this section, we present our dataset, criteria we employed to identify scanners intended for survey purposes, and the procedures we followed to verify adherence to best practices.

3.1 Dataset

We analyze darknet traffic data to identify Internet-wide scanners for survey purposes operated by security researchers and companies (hereinafter referred to as "survey scanners"). Our darknet monitoring system consists of about 300,000 IPv4 addresses, with sensors distributed across the address range of over 40 organizations in 10 countries. Our darknet sensors are installed across a variety of organizations, including universities, research institutions, and private companies, and each organization's darknet address block ranges from /16 to less than /24. This diverse and broad range makes it suitable for observing Internet-wide scans. Table 1 provides the statistics of our darknet traffic data, specifically limited to TCP SYN packets and UDP packets, used in this study. The observation period was three months, from March to May 2022, during which 127 billion packets targeting the darknet were observed, and 21 million IPv4 addresses were identified as scan source IP addresses. Figure 2 shows the changes during the observation period, with about 1,400 million packets observed per day from 690,000 scanner IPs.

1. Coordinate closely with local network admins to reduce risks and handle inquiries.
2. Verify that scans will not overwhelm the local network or upstream provider.
3. Signal the benign nature of the scans in web pages and DNS entries of the source addresses.
4. Clearly explain the purpose and scope of the scans in all communications.
5. Provide a simple means of opting out, and honor requests promptly.
6. Conduct scans no larger or more frequent than is necessary for research objectives.
7. Spread scan traffic over time or source addresses when feasible.

Table 2: Best practices for Internet–wide scanning [14]

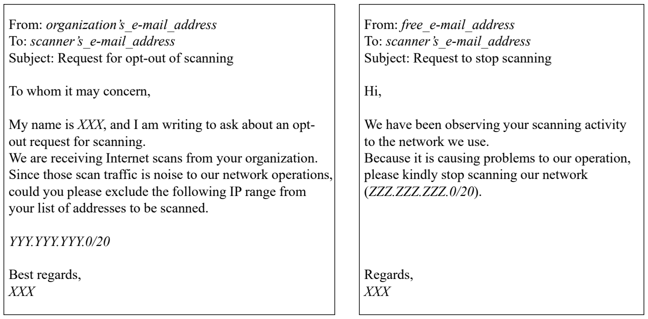
3.2 Detecting Survey Scanners

We first collect a list of IP addresses marked with the "research" tag in the threat category provided by SANS [33] and a list of IP addresses tagged with the "Actor" tag in Greynoise [16]. We then examine whether these IP addresses were active in conducting scans to the darknet to identify currently operative survey scanners. Additionally, to detect survey scanners not included in the known lists, we implement specific criteria to identify aggressive survey scanners, excluding malware-infected devices, among the scanners observed in the darknet.

Criteria: A source IP address is deemed a survey scanner if the following conditions are met within a day for observed packets (either TCP SYN or UDP).

1. the number of packets is larger than the size of our darknet, i.e., more than 300,000 packets.
2. the number of targeted destination ports exceeds 30 ports.

The criterion regarding packet count is founded on the observation that survey scanners generally operate at high speeds, leveraging high-performance servers and high-bandwidth connections. On the other hand, malware-infected devices typically initiate scanning activities from relatively lower-performance devices, such as IoT devices, utilizing narrow bandwidth connections. The criterion involving the number of ports is based on the understanding that survey scanners typically scan a wide range of ports, whereas malware-infected devices tend to scan only specific ports to facilitate the spread of infection.



From: *organization's_e-mail_address*
To: *scanner's_e-mail_address*
Subject: Request for opt-out of scanning

To whom it may concern,

My name is *XXX*, and I am writing to ask about an opt-out request for scanning.
We are receiving Internet scans from your organization.
Since those scan traffic is noise to our network operations,
could you please exclude the following IP range from
your list of addresses to be scanned.

YYY.YYY.YYY.0/20

Best regards,
XXX

(A) Opt-out request from organization's e-mail address

From: *free_e-mail_address*
To: *scanner's_e-mail_address*
Subject: Request to stop scanning

Hi,

We have been observing your scanning activity
to the network we use.
Because it is causing problems to our operation,
please kindly stop scanning our network
(*ZZZ.ZZZ.ZZZ.0/20*).

Regards,
XXX

(B) Opt-out request from a free e-mail address

Figure 3: Two types of opt-out requests

3.3 Verifying Compliance with The Best Practices

Prior research [14] has suggested seven best practices for conducting ethical Internet scans, as listed in Table 2. Among these seven items, since it is difficult to externally ascertain the level of coordination with local network administrators and Internet providers, in this study, we focus on confirming compliance with items 3, 4, and 5, as these can be assessed through external observation. For item 3, related to the scan organization's identity disclosure, we perform reverse DNS lookups, WHOIS searches, and HTTP accesses. These methods allowed us to identify the scan organization associated with the IP addresses identified as survey scanners. Regarding item 4, which relates to clarifying the purpose and scope of the scans, we conduct HTTP access to the scanner's IP address on port 80/TCP to check for the presence of a website. If accessible, we assess whether it explained the purpose and scope of the scan, including information about the targeted ports and protocols, along with the publicly available IP address range used for the survey scan. If no website is accessible via the scanner's IP address, we search for any relevant descriptions on the scan organization's official website. Regarding item 5, which concerns offering a straightforward opting-out process, similar to item 4, we examine the websites to determine if they include instructions on the opt-out procedures and details on how to submit an opt-out request. For conducting the opt-out experiment described later, we confirm contact information such as email addresses even in the absence of clear instructions regarding the opt-out process.

3.4 Opting-Out of Internet-wide Scanning

For those survey scanners for which we managed to collect contact details, we send an opt-out request to determine if these scanners would appropriately respond and stop their scanning. We submit two types of opt-out requests for each scanner: one verifiable and one anonymous, as illustrated in Figure 3. This method allows us to assess whether the response to opt-out requests varies based on the ability to verify the request's legitimacy. From the perspective of the survey scanners, verifying the validity of the received opt-out request, which confirms that the requester manages the IP address for opting out, is crucial. Accepting unverified opt-out requests may negatively influence the survey scan results due to potential fraudulent requests.

In opt-out request (A), we request the exclusion of our organization's IP address range (/20 address block) using an official organizational email address. The survey scanner operator can verify the legitimacy of the requestor's email through email communication, and affirm the IP address range ownership through a Whois search or reverse DNS lookup. Conversely, in opt-out request (B), we request the exclusion of an IP address range belonging to our partnering organization (/20 address block), using a free email address. Unlike request (A), the IP address range specified in request (B) does not offer organizational information via a Whois search or reverse DNS lookup, making it challenging for the survey scanner operator to verify the request and confirm if the requestor manages the specified IP address range. Since the timing of a survey scanner's activity is typically unknown, it is challenging to ascertain whether the scanning has ceased or if the scanner is simply inactive. However, in this experiment, both organizations have our darknet sensors installed, and the IP addresses requested for exclusion fall within the darknet sensor's observation scope. Thus, we can confirm if only the survey scans targeting the requested IP address range have stopped via darknet monitoring. If no cessation is confirmed two weeks after the opt-out request submission, the request will be resubmitted. If the survey scan continues two months post-resubmission, we deem the opt-out request unfulfilled.

4 Results

4.1 Survey Scannners Dynamics

Table 3 summarizes the results of the survey scanner detection. Using the criteria described in section 3.2, we detected 1,705 survey scanner addresses during the observation period. While these addresses represent a mere 0.08% of all scanner addresses, the packet count from this subset constitutes 48% of all darknet traffic. This highlights the substantial influence these survey scanners wield on Internet monitoring.

Out of the 1,705 addresses, 1,196 could not be associated with any organization, signifying that approximately 70% of survey scanners operate anonymously. We managed to associate the remaining 509 addresses with specific organizations by utilizing methods such as reverse DNS lookups, Whois searches, and standard Internet searches These IP addresses were then grouped by scanning organization, which allowed us to identify 13 distinct organizations. Furthermore, we detected scan packets from 3,331 and 4,244 survey scanner addresses which were obtained from SANS and Greynoise, re-

	# IPs		# Pkts [$\times 10^9$]	
Dataset	21,648,734		127	
Survey scanner (identified)	4,821	(0.022%)	35	(28%)
SANS	3,331	(0.015%)	18	(14%)
Greynoise	4,244	(0.020%)	28	(22%)
Our criteria	509	(0.002%)	26	(21%)
Survey scanner (anonymous)	1,196	(0.006%)	35	(27%)

Table 3: Acitive survey scanners

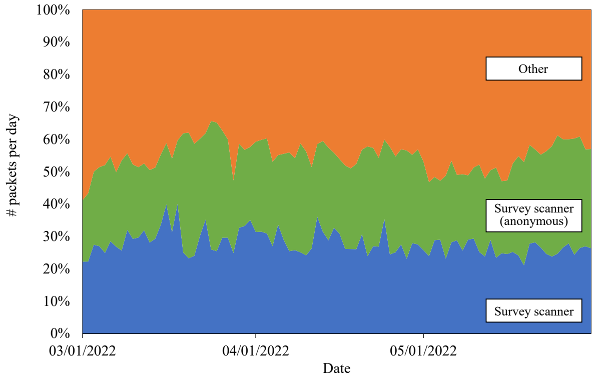


Figure 4: Packet rate of survey scannners

spectively. Combining these results, we identified a total of 46 scan organizations, operating from 4,821 scanner addresses. Table 9 shows a list of 46 organizations, of which 13 are universities or other academic institutions, 20 are companies, and the rest cannot be determined whether they are academic institutions or companies. 24 out of the 46 organizations were not reported in previous studies [22, 36], underlining the increasing prevalence of survey scans. The fact that companies outweigh academic organizations in number suggests that Internet-wide scanning has become commoditized.

Figure 4 shows the percentage of packets from survey scanners in darknet traffic on a daily basis. We find that about 20% to 30% of the darknet packets consistently originate from these 46 scan organizations. Additionally, roughly 20% of the darknet packets stem from anonymous survey scanners. The results indicate that these scan organizations routinely conduct Internet-wide scans.

Table 4 shows the top organizations performing aggressive survey scans. Censys and Recyber generated particularly high volume of scan traffic, performing nearly 10 times more scans compared to other organizations. Recyber claims to be a project that assists researchers, universities, and other educational institutions, but lacks transparency regarding its operating entity. Recyber is known for performing aggressive scans, and several organizations have complained about the intensity of the scans [29]. Scans during the observation period from Shodan and Recyber were performed on all TCP ports, and it appears that they are not focused on specific services or vulnerabilities in their scans, but rather on visibility into the Internet as a whole.

Scan organization	Pkts [$\times 10^9$]		IPs	
Censys	11.8	(33.4%)	279	(5.8%)
Recyber	10.5	(29.8%)	171	(3.5%)
Shadowserver	1.9	(5.4%)	529	(10.9%)
cyber.casa	1.6	(4.5%)	252	(5.2%)
Open Port Statics	1.2	(3.3%)	19	(0.4%)
Shodan	1.1	(3.0%)	47	(1.0%)
CriminalIP	0.98	(2.8%)	63	(1.3%)
BitSight	0.87	(2.5%)	311	(6.4%)
Abor	0.80	(2.3%)	3	(0.1%)
Other	4.6	(13.2%)	3,161	(65.4%)

Table 4: Top organizations performing aggressive scans

4.2 Targeted Services

Survey scanners and other scanners have different characteristics of targeted services. Table 5 shows the distribution of target ports for scan packets across the 46 scan organizations. The target ports of survey scanners are widely distributed, even for HTTPS (TCP/443), the most frequently scanned port, which accounts for a mere 0.78% of packets from the survey scanners. Survey scanners target numerous ports related to HTTP and HTTPS. We observed many scans directed not only at default ports but also at high ports used by specific devices and HTTP servers (e.g., TCP/8080, TCP/8443, TCP/8081, TCP/9000). In recent years, the WebUI for managing webcams, routers, and other IoT devices has been running on a high port instead of the default port and is accessible from the Internet side, and survey scanners are interested in scanning these devices as well. We also observed many scans against RDP (TCP/3389) and VNC (TCP/5900) from survey scanners. This is likely due to the increase in users using remote access services such as RDP and VNC for working from home due to the COVID-19 pandemic, and many scan organizations showing interest in those accessible devices. Additionally, other scans often target UDP-based devices that can be leveraged in amplification attacks (e.g., NTP. SNMP, DNS, CoAP) and exposed databases (e.g., Elasticsearch, PostgreSQL, Redis).

Table 6 presents the distribution of target ports for scanners, excluding those attributed to survey scanners. Unlike the survey scan, the other scans are more concentrated on ports associated with the propagation of malware infections. Telnet, which is the most frequently targeted service, along with SSH, the second most targeted, are exploited by Mirai and other IoT malware to proliferate infections. Combined, these two account for over 20% of all scan packets. As for the sixth and seventh most targeted services, Docker REST API,

Service	%	Service	%	Service	%
HTTPS (TCP/443)	0.78%	Alt-HTTPC (TCP/8081)	0.31%	Elasticsearch (TCP/9200)	0.23%
HTTP (TCP/80)	0.74%	DNS (UDP/53)	0.31%	Alt-HTTP (TCP/8090)	0.23%
Alt-HTTP (TCP/8080)	0.52%	CoAP (UDP/5683)	0.30%	Alt-HTTP (TCP/8888)	0.23%
Alt-HTTPS (TCP/8443)	0.42%	SSH (TCP/22)	0.27%	SSDP (UDP/1900)	0.22%
Microsoft RDP (TCP/3389)	0.40%	Alt-HTTP (TCP/9000)	0.26%	PostgreSQL (TCP/5432)	0.21%
FTP (TCP/21)	0.38%	VNC (TCP/5900)	0.26%	Redis (TCP/6379)	0.21%
NTP (UDP/123)	0.35%	Telnet (TCP/23)	0.25%	NetBIOS Name Svc (UDP/137)	0.21%
SNMP (UDP/161)	0.35%	Alt-HTTP (TCP/8000)	0.25%	Other	92.3%

Table 5: Commonly targeted services by survey scanners (identified)

Service	%	Service	%	Service	%
Telnet (TCP/23)	16.0%	SMB over IP (TCP/445)	1.7%	NTP (UDP/123)	0.73%
SSH (TCP/22)	5.0%	Alt-HTTP (TCP/8080)	1.4%	RPC (TCP/111)	0.73%
HTTP (TCP/80)	2.7%	HTTPS (TCP/443)	1.4%	SQL Server (TCP/1433)	0.67%
Redis (TCP/6379)	2.4%	SIP (UDP/5060)	1.2%	Alt-Telnet (TCP/2323)	0.59%
ADB/Alt-HTTP (TCP/5555)	2.3%	Unknown (UDP/60138)	1.0%	Huawei Vuln. (TCP/37215)	0.56%
Docker REST API (TCP/2375)	2.0%	Microsoft RDP (TCP/3389)	0.97%	PBX (TCP/5038)	0.52%
Docker REST API (TCP/2376)	1.8%	Realtek Vuln. (TCP/52869)	0.81%	PBX (TCP/50802)	0.47%
Alt-HTTP (TCP/81)	1.8%	Misc. (TCP/4200)	0.75%	Other	52.4%

Table 6: Commonly targeted services (excluded survey scanners)

Country	Pkts [$\times 10^9$]		IPs	
United States	18.1	(51.3%)	2,056	(42.5%)
Netherlands	13.3	(37.6%)	440	(9.1%)
United Kingdom	1.78	(5.1%)	511	(10.6%)
Germany	0.37	(1.0%)	1,103	(22.8%)
Portugal	0.27	(0.8%)	94	(1.9%)
Other	1.49	(4.2%)	631	(13.1%)

Table 7: Top countries originating survey scans

AS	Pkts [$\times 10^9$]		IPs	
IP Volume Inc	12.8	(36.2%)	271	(5.6%)
CENSYS-ARIN-01	6.7	(18.9%)	155	(3.2%)
CENSYS-ARIN-03	2.9	(8.1%)	72	(1.5%)
CENSYS-ARIN-02	2.2	(6.4%)	52	(1.1%)
HURRICANE	1.9	(5.4%)	529	(10.9%)
Constantine Cybersecurity Ltd.	1.6	(4.5%)	252	(5.2%)
ARBOR	0.80	(2.3%)	3	(0.06%)
CARINET	0.67	(1.9%)	272	(5.6%)
GOOGLE-CLOUD-PLATFORM	0.65	(1.9%)	230	(4.8%)
Akamai Connected Cloud	0.58	(1.7%)	408	(8.4%)
Other	4.5	(12.9%)	2,591	(53.6%)

Table 8: Top providers originating survey scans

malware has been reported to spread infection through those services [23]. Furthermore, we observed other scans targeting vulnerabilities unique to certain devices (e.g., Realtek vulnerability, Huawei router vulnerability). On the other hand, only 1.7% of scans targeted SMB over IP service, which Windows malware such as Conficker and Wannacry exploited for infection, indicating that the worm's primary infection targets have shifted from Windows devices to IoT devices.

4.3 Distiburion of Scanners

To analyze the geographical distribution of scanner IP addresses by country and AS, we use the MaxMind GeoIP2 databases [21]. Tables 7 and 8 list the leading countries and AS, respectively, engaged in high-volume survey scans. We observed 25 countries as source countries for survey scans, with 88.9% originating from hosts in the United States and the Netherlands. In the United States, Censys was responsible for 65% of survey scan traffic. They operate their own ASes, CENSYS-ARIN-01, CENSYS-ARIN-02, and CENSYS-ARIN-03, which are in the top two to four survey scan source ASes. Arbor also conducts scans only from the United States and operated its own ASes. HURRICANE, which is in the top five survey scan sources, is operated by HURRICANE Electric LLC, which is headquartered in the United States. All observed Shadowserver scanners are in the HURRICANE network. As for the Netherlands, most of the scans originated from a single AS, IP Volume Inc. IP Volume Inc, also known as Ecatel Network and Quasi Network,

Scan Organization	# IPs	identifiable through			information disclosure			opt-out		
	(darknet)	PTR	Whois	Website	# IPs	purpose	port/proto	desc.	contact	result
Adscore	30	✓							Web form	
Alpha Strike Labs	1021		✓	*		✓		✓	email	
Arbor	3	✓	✓		256	✓		✓	email	(A)/(B)
BinaryEdge	727	✓				✓			email	(A)/(B)
BitSight	311	✓	*			✓		✓	email	(A)/(B)
Univ. of Cambridge	1	✓	✓	✓	1	✓	*	✓	email	(A)/(B)
Censys	279	✓	✓	*	1280	✓	✓		email	
Cloud System Networks	47	✓						✓	email	
CriminalIP	63	✓	✓	*		✓		✓	email	
cyber.casa	252	✓	✓		256	✓		✓	email	(A)
CyberGreen	1	✓		✓		✓	✓	✓	Web form	(A)/(B)
Cymru	3	✓		✓	3	✓		✓	email	
DIVD CSIRT	2		✓	✓	258	✓	*		email	
ESET	1	✓		✓		✓			email	(A)
FH Muenster	1	✓	✓	✓		✓			email	(A)/(B)
GDNP	47		✓			✓		✓	email	
INTERNET-MEASUREMENT	44	✓			50	✓			email	
InterneTTL	93	✓			19	✓				
Intrinsec	12	✓	✓	✓		✓		✓	email	(A)/(B)
IOStation	2	✓				✓		✓	email	(A)
ipip.net	40	✓	✓			✓			email	
LeakIX	7	✓				✓			email	(A)/(B)
MPI-INF	8	✓	*	✓	18	✓		✓	email	(A)
Net Systems Research	47	✓	✓			✓		✓	email	(A)/(B)
netsecscan	16	✓	✓	✓		✓		✓	email	(A)/(B)
ONYPHE	165	✓		✓	327	✓	*	✓	email	(A)
Open Port Statics	19	✓						✓	Web form	
Palo Alto	243		✓					✓	email	(A)/(B)
Project Sonar	359	✓	✓	✓	384	✓	*	✓	email	(A)
Project25499	88			✓	2	✓		✓	email	
QuadMetrics	6	✓							email	(A)/(B)
Recyber	171	✓	✓					✓	Web form	(A)/(B)
research-scanner.com	2	✓				✓	*	✓	email	
research.knoq.nl	1	✓		✓						
RWTH Aachen Univ.	2	✓	✓	✓	64	✓		✓	email	(A)/(B)
ScanOpticon	3	✓							Web form	
SecurityTrails	3		✓			✓			email	
Shadowserver	529	✓				✓	*	✓	email	(A)
Shodan	47	✓				✓			email	
Stanford Univ.	11	✓	✓	✓	512	✓			email	(A)
Stretchoid	68	✓				✓		✓	Web form	(A)/(B)
TUM	2	✓	✓	✓	258	✓	*	✓	email	(A)
Univ. of Colorado	1	✓	✓	✓		✓		✓	email	
Univ. of Michigan	41	✓			1536	✓			email	
Univ. of Sydney	1	✓	✓							
Winnti Scan Host	1	✓				✓		✓	Web form	
46 organizations		40 (87%)	24 (52%)	20 (43%)	16 (35%)	37 (80%)	9 (20%)	28 (61%)	43 (93%)	24/15 (52%/33%)

In "Whois" and "Website" columns, ∗ indicates that only some scanner IP addresses are identifiable. In "port/proto" column, ∗ indicates that only partial information is disclosed. In "result" column, (A) and (B) indicate that the scan was stopped by opt-out requests (A) and (B), respectively.

Table 9: Compliance with the best practices

is a well-known bulletproof hosting provider, whose servers have previously been linked to a source of Spam emails, malware hosting, and child pornography storage. Recyber, Open port statics, and CriminalIP use the network provided by IP Volume Inc to perform Internet-wide scans.

While some scan organizations located their scanners in specific countries or ASes, Shodan and Bitsight distributed their scanners widely. For example, Shodan scanned from eight networks in six countries, including the United States and the Netherlands, while Bitsight scanned from five countries. Wan et al. reported that geographical and topological differences in scan origin affect Internet-wide scan results [38].

Therefore, distributing scanners increases the visibility of Internet-wide scans.

4.4 Compliance with The Best Practices

Table 9 presents a summary of the compliance with the best practices by the 46 scan organizations identified in Section 4.1. This compliance includes identity disclosure, sharing scanner IP addresses, explanation of survey purpose and scope, and outlining the opt-out process.

Identification of scanner:

We could identify 40 out of the 46 scan organizations through performing a reverse DNS lookup of the scanner IP addresses. This result indicates that scanners use the value of PTR record as the primary communication channel to disclose their identity to receivers. On the other hand, only 24 organizations, or about half of the scan organizations, could be identified through Whois search. There could be several reasons for this. One reason might be that the network used for scanning is not owned by the scan organization but by hosting provider or cloud service provider thus Whois data is not associated with the scan organizations. Other reason may be due to privacy concerns and regulations like GDPR, some Whois data is redacted thus IP addresses cannot be associated with the network user. For example, two of those organizations, BitSight and MPI-INF, use multiple networks and for a certain IP address, although a reverse DNS lookup will reveal their identity through PTR record, Whois data only provide general information about the network operators.

Running a web server on the scanner host to advertise the benign intention to receivers is a known practice. Thus, we additionally performed HTTP accesses to the scanner IP addresses. We could successfully reach the websites of 20 scan organizations. However, in two of those scan organizations, we were not able to access their websites of some scanner IP addresses, showing some inconsistency in the adaption of the practice. Providing access to a Website through the scanner's IP address is an effective method of information disclosure because it allows the scan receivers to easily access detailed information about the scan details, such as target of the scan and opt-out procedures.

Transparency: Prior research [12] recommends publishing the source IP addresses used for Internet-wide scans and the target ports and protocols so that scan receivers can properly filter the scan traffic. However, we find that only 16 of the 46 scan organizations published their scanner IP addresses on their Website. In addition, some scan organizations had discrepancies in the published scan source IP addresses and the actual source IP addresses they used. For instance, InternetTTL discloses 19 scanner IP addresses, yet we identified 93 scanner IP addresses being used by InternetTTL during our observation period. Regarding the purpose and target of the scan, 37 organizations clearly stated on their Websites that they conduct Internet-wide scans, but we were unable to find any such declaration from the remaining 9 organizations. Moreover, we managed to find detailed descriptions of the target port and protocol in only 9 organizations. Overall, the task of locating comprehensive scan information among numerous webpages proved to be challenging, particularly for large-scale entities such as universities and major corporations. Many scan organizations, even those that do disclose their identities, do not properly disclose the IP addresses of their scanners and the ports and protocols they are scanning, making it difficult for scan receivers to properly filter those scan packets.

	w/ reply		w/o reply		All	
Stopped	20	(47%)	4	(9%)	24	(56%)
Continued	5	(12%)	10	(23%)	15	(35%)
Undetermined	1	(2%)	3	(7%)	4	(9%)
All	26	(60%)	17	(40%)		

Table 10: Results of opt-out request (A)

	w/ reply		w/o reply		All	
Stopped	11	(26%)	4	(9%)	15	(35%)
Continued	13	(30%)	12	(28%)	25	(58%)
Undetermined	2	(5%)	1	(2%)	3	(7%)
All	26	(60%)	17	(40%)		

In Tables 10 and 11, scan organizations classified in the "Undetermined" line are those for which scan cessation could not be determined because a sufficient amount of scans were not observed during the observation period following the opt-out request.

Table 11: Results of opt-out request (B)

Opt-out process: 28 organizations stated on their websites that they offer the option to opt-out of scanning. Almost all of these scanning organizations provided an email address as their designated opt-out contact. 7 organizations used web forms to accept opt-out requests where receivers can enter email address and CIDR blocks that they would like to have removed. In addition, we chose to gather contact information and submit opt-out requests for those scanning organizations that did not specify any opt-out procedure. As a result, we managed to collect contact information for 43 scanning organizations in total. However, we did not include the University of Sydney because we were unable to locate the relevant department or laboratory to contact. Despite clear declarations from certain scanning organizations on their websites stating their non-compliance with opt-out requests, we still made the deliberate decision to send them opt-out requests nonetheless.

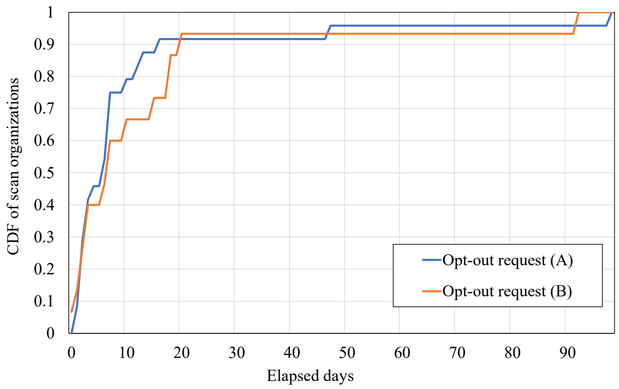


Figure 5: CDF of elapsed days for opt-out implementation

4.5 Response to Opt-out Requests

Table 10 and Table 11 summarize the results of whether the scan organizations stopped or continued their scans in response to the two types of opt-out requests shown in Figure 3. In terms of replies to opt-out requests, 26 scan organizations provided replies to both opt-out requests, including auto-response emails. IOStation and Net Systems Research responded solely to opt-out request (A), while DIVD CSIRT and the University of Colorado responded exclusively to opt-out request (B). It is unclear whether these differences in response were intentional or due to oversight. Two organizations, Project25499 and GDNP, are categorized as "w/o reply" in Table 10 and Table 11 because the opt-out request we sent to their provided contact email address was unsuccessful. The emails bounced back with "User Unknown" error suggesting that their email addresses may not exist or are no longer active.

For scanners that failed to respond to our opt-out requests, the type of opt-out request made little difference in the response rate. The majority of scanners that did not respond to our request continued their scanning activity. Among these were Adscore, CriminalIP, INTERNET-MEASUREMENT, ipip.net, Open Port Statics, ScanOpticon, and Shodan. GDNP and Project 25499, which returned the bounce-back error messages to our request, also continued scanning. Regarding the organizations that responded to only one request, IOStation halted scans solely for the addresses associated with opt-out request (A), while Net Systems Research ceased scanning for both requests (A) and (B). Conversely, DIVD CSIRT did not stop scanning for any of the addressed requests, and it remained uncertain whether the University of Colorado had discontinued its scanning activities. Shodan, an early player in the IoT search service landscape and a tool widely used by researchers and corporations, was identified as a scanner that does not actively comply with best practices. This assessment is based on the observation that Shodan does not disclose its scanner address on its website and persists in scanning despite receiving opt-out requests. Thus, it can be inferred that Shodan does not proactively adhere to best practices. In contrast, netsecscan, QuadMetrics, and Stretchoid, despite not responding to either request, stopped their scanning activity for both addresses.

The responses from scanners that acknowledged the application were varied. A total of 12 organizations that stopped scanning on opt-out request (B) (Arbor, BinaryEdge, BitSight, University of Cambnridge, CyberGreen, FH Muenster, Intrinsec, LeakIX, Net Systems Research, Palo Alto, Recyber, and RWTH Aachen University) accepted both applications without any form of validation and stopped scanning. While this aligns with best practices in terms of responding appropriately to opt-out requests, it simultaneously raises a concern that false requests might also be accepted if submitted anonymously. On the other hand, Censys, the most active scanner, continued its scanning despite the requests. They responded by stating it would not enact manual opt-out measures and advised that the receiver should block the scanner's IP address instead. According to our darknet observations, the persistence of scanning activities by Censys accounted for 33% of the traffic from the scanners we identified, using the specified criteria, including anonymous survey scanners. Consequently, 18% of scanning packets persisted, highlighting the considerable impact of non-compliance with opt-out requests by aggressive scanners.

Seven organizations (cyber.casa, ESET, MPI-INF, ONYPHE, Project Sonar, Stanford University, and TUM) exhibited different responses to the two types of requests. These organizations verified the authenticity of the anonymous requests before taking actions. They requested that applications be submitted from an official, rather than a free, email address. Additionally, several of these organizations required submission of information that would validate the ownership of the application address. Table 12 summarizes how those organizations verify the authenticity of opt-out requests. Proof of ownership requested typically included responses from the administrative email address listed in the Whois, responses from the organization's email address found in the BGP information, or information regarding the organization and the applicant's affiliation. Since no additional proofs were submitted in this experiment, scans from those organizations continued.

In response to opt-out request (A), where the validity of the request could be verified, 24 scan organizations stopped scanning during the observation period. While most of the scan organizations that replied to the request stopped scanning, five organizations (Alpha Strike Labs, Censys, Cymru, University of Michigan, and Winnti Scan Host) continued scanning. These five organizations demonstrated diverse reactions. As mentioned above, Censys does not facilitate manual opt-outs. Alpha Strike Labs responded by requesting proof of ownership for the address range we applied for. Although

Category	Verification Method	# organizations
Email address verification	send back with admin or abuse contact address provided by Whois	2
	send back with organization's email address in BGP information	1
	send back with company's email address	1
	send back with official email address	1
Request for additional information	provide organization and affiliation information	1
	proof of official responsibility for this network	1
	prove you are the owner of that network block	1
	provide verification of your ownership of the range	1

Table 12: How to verify the authenticity of opt-out requests

we answered their query, we received no further communication, and the scanning persisted. Cymru, on the other hand, responded asking for more details about the scan. After we replied, they indicated that they had added us to their exclusion list. Yet, their scan continued during our observation period. Similarly, the University of Michigan notified us that they had added our IP addresses to their exclusion list. However, we could not verify the cessation of their scanning. Winnti Scan Host sent an automatic email confirming the receipt and validity of our email address, but their scan continued even after we responded to their message.

Figure 5 shows the CDFs from the two types of opt-out requests until the scan was stopped. It is crucial to clarify that the dates presented do not correspond to the dates of addition to the exclusion list. Instead, they depict the time elapsed until we could affirm, through our monitoring, that the scanning had indeed ceased. The data illustrated in the figure indicates that 40% to 50% of organizations respond to opt-out requests within a week, and as many as 90% respond within a three-week period.

5 Discussion

5.1 Ethics Consideration

We followed the ethical principles laid out in the Menlo Report [5]. The scanner detection is conducted using passively observed data, ensuring that it has no detrimental impact on general users or the scanner itself. HTTP access on port TCP/80 for scanner identity verification was performed only once for each scanner IP address. In addition, HTTP accesses to the same /24 subnet were spaced at least 10 seconds apart to avoid overloading the network. In this experiment, we made two types of applications to verify the opt-out compliance of scan organizations. However, we did not make any false applications as both application IP addresses were managed by either our own organization or the collaborative organization. Our opt-out requests reduce the address range of Internet-wide scans but do not impact the observation results of scan organizations since we are applying for darknet addresses.

5.2 Recommendations for scanners

Some recommendations for scanners to complement the best practices derived from the experimental results are described below.

5.2.1 Improve identity disclosure and searchability

Our survey found that 70% of detected scanners were anonymous, with only a small number of scanners revealing their identities. Since it is difficult for receivers to discern the intent of a scan based solely on received packets, it is necessary for organizations conducting Internet-wide scans to disclose their identity. If an identity isn't revealed, these mass scans could be mistaken for malicious activity, leading to abuse reports and overly cautious blocking measures. We were able to identify 13 out of the 46 organizations by name, but we couldn't determine the actual operator, be it a company, university, or individual researcher. This level of anonymity is similar to that of an anonymous scanner and is insufficient in terms of identity disclosure.

Ease of finding information is also important. Of the 46 organizations, 20 had a website running on the scanner's IP address. This meant we had to first identify the organization's name through reverse DNS lookup or Whois search, find its official website through an Internet search, and then search for web pages describing the scanning information. This was a time-consuming process. Each organization should have a dedicated webpage that summarizes its scanning information and is easily accessible from the main page of the organization's website. This is especially important for organizations that have a lot of information on their official Web site, such as universities and corporations.

5.2.2 Improve transparency

Transparency of a scanning activity is vital for scan receivers to comprehend the nature of the scan and act accordingly. In our survey, most scanning organizations do not disclose enough information about their scans. For example, only 16 organizations disclosed the IP addresses of their scanners. Information such as scan duration, frequency, target ports, and protocols are rarely revealed. Given this situation, receivers struggle to filter scans preemptively before any issues occurs. It's equally important that scanning information is regularly updated and easily accessible by receivers. Scan organizations could consider providing information in accessible formats such as JSON and XML. It is also a good idea to consider providing a REST API if possible.

Furthermore, the scan organization should carefully and precisely explain the objective of the scan. As Internet-wide scanning typically entails sending large packets without the receiver's consent, a rational explanation acceptable to the receiver is mandatory. If a receiver wishes to opt-out, the survey scanner should provide a legitimate reason for denying the request. The best practices require that scans should not exceed their stated objective. However. if the objective is not adequately clarified, the receiver cannot determine whether the scan is confined to the minimum necessary scope to fulfill its intended purpose.

5.2.3 Promote information-sharing and cooperation among scan organizations

Our research has shown that Internet-wide scanning has become increasingly commoditized. Today, virtually anyone can readily perform an Internet-wide scan. However, we need to reconsider whether each organization really needs to perform its own Internet-wide scan. In our observations, each of the 48 organizations was scanning for a large number of destination ports and services. In other words, many organizations are repetitively performing their own scans for the same port/service. Nevertheless, only a select few of the 48 organizations, such as Shodan and Censys, offer their collected scan results to external parties. Most organizations use the scan data exclusively for their internal purposes. If more organizations were to undertake Internet-wide scans in the future, they would generate a large number of packets to collect the same response from the same service operating on identical devices, leading to high inefficiency. For this reason, we recommend that scan organizations promote information sharing and collaboration. Similar information-sharing frameworks exist, for example, Virustotal [37] and MalwareBazaar [1] for sharing malware samples, and Phishtank [18] for sharing phishing URLs. Increased information sharing of scan data among scan organizations will reduce the negative impact of Internet scans by encouraging them to perform their own scans only when shared data is insufficient.

5.2.4 Provide direct benefits to the receivers

As shown in related studies, data obtained from Internet-wide scanning is useful for research and development in various fields as well as in the cybersecurity field. While the benefits are recognized by researchers and security companies, the benefits are not accessible to the majority of end users. In order to ensure that end users who actually receive scans understand the effectiveness and significance of Internet-wide scans, it is vital to consider ways of delivering direct benefits to them.

A potential approach to extending direct benefits to receivers involves providing relevant information. As previously mentioned in our recommendation for inter-organizational information sharing, most scanning organizations presently utilize scan data solely for their internal use. By freely sharing this data with users, they might gain awareness of potential misconfigurations or vulnerabilities within their own devices. Moreover, scanning organizations can proactively send alerts to network administrators within companies and other organizations. For instance, Shadowserver freely shares its observations with network administrators. This approach of gradually extending benefits to receivers could enhance public acceptance of Internet-wide scanning and contribute towards its long-term sustainability.

5.3 Recommendations for receivers

Because Internet scanning is performed without the consent of the receiver, there are not many workarounds that can be taken by the receiver. However, we also discuss some recommendations for receivers to reduce the negative effects of Internet-wide scans.

5.3.1 Send opt-out requests even if the effect is limited

In our experiment, only half of the survey scanners responded to our opt-out requests. Despite this, these scanning organizations managed to exclude our address range from their scans within one to two weeks. If an organization experiences issues due to survey scans, it is recommended to initially submit an opt-out request before resorting to scan filtering. This can be particularly effective as some scanning organizations use dynamic IP addresses or regularly add new scanner IP addresses. Thus, an opt-out request can considerably mitigate the impact. When submitting an opt-out request, using the organization's email address to include affiliation details can slightly enhance the chance of acceptance.

Reporting abuse to ASs and bulletproof hosting providers is basically ineffective and not recommended. In our experiment, emails were sent to several networks' abuse contact points, particularly those with aggressive anonymous survey scanners, requesting intervention. However, responses were not received from any. As for bulletproof hosting, often associated with malicious activities, it may be better for an organization to block the entire network range.

5.3.2 Share scanner information

Our observations have revealed several survey scanners that are not included in the known list. Since it is difficult for individual organizations to keep track of all these survey scanners, we recommend that information about survey scanners be shared so that each organization can freely use them for filtering purposes. In addition to the SANS and GreyNoise lists employed in our study, scanner lists are available from individual contributors on platforms like Github, which might be beneficial. On the other hand, we find that some scan organizations use dynamic IP addresses or hosting services to conduct their scans, so we must be careful about over-blocking. The emergence of new scanners in the future necessitates a community-driven mechanism to curate and maintain a current and reliable scanner list. Crafting such a mechanism that can keep pace with emerging scanners will be a critical undertaking.

5.4 Limitations

In this study, we employed our criteria to identify aggressive survey scanners in addition to a list of known survey scanners. These criteria were developed heuristically based on the characteristics of survey scanners. However, it should be noted that certain scan organizations distribute their scanning activities among multiple scanners, with each scanner focusing on a limited set of destination ports. Consequently, such distributed survey scanners may not have been detected by using our criteria. Conversely, if an attacker is performing Internet-wide scans over a wide range of destination ports in preparation for launching an attack, we may detect attacker-operated scanners as well as survey scanners.

6 Conclusion

In this work, we analyzed the current situation of Internet-wide scanning. We identified active survey scanners through darknet monitoring and found that Internet-wide scanning is becoming more prevalent. We verified the compliance of scan organizations with the best practices in terms of identity disclosure, information disclosure, and opt-out response. We found that many of the survey scanners are anonymous, with only inadequate scan disclosure, and more than half do not implement opt-out requests. Based on our findings, we recommended some actions for the scanners and the receivers to make Internet scanning sustainable.

References

[1] abuse.ch. Malwarebazaar | malware sample exchange. https://bazaar.abuse.ch/, 2023. [Online; accessed 1-June-2023].

[2] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, D. Menscher, C. Seaman, N. Sullivan, K. Thomas, and Y. Zhou. Understanding the Mirai Botnet. In *Proceedings of the 26th USENIX Conference on Security Symposium*, SEC '17, pages 1093–1110, 2017.

[3] M. Bailey, E. Cooke, F. Jahanian, and J. Nazario. The internet motion sensor: A distributed blackhole monitoring system. In *Proceedings of the 12th Annual Network and Distributed System Security Symposium*, NDSS '01, 2005.

[4] M. Bailey, E. Cooke, F. Jahanian, and D. Watson. The blaster worm: then and now. *IEEE Security Privacy*, 3(4):26–31, 2005.

[5] M. Bailey, E. Kenneally, D. Maughan, and D. Dittrich. The menlo report. *IEEE Security amp; Privacy*, 10(02):71–75, mar 2012.

[6] CAIDA. Conficker/conflicker/downadup as seen from the ucsd network telescope. https://www.caida.org/archive/ms08-067/conficker/, 2009. [Online; accessed 1-June-2023].

[7] Censys. Exposure management and threat hunting solutions. https://censys.io/. [Online; accessed 1-June-2023].

[8] J. Czyz, K. Lady, S. G. Miller, M. Bailey, M. Kallitsis, and M. Karir. Understanding ipv6 internet background radiation. In *Proceedings of the 2013 Conference on Internet Measurement Conference*, IMC '13, page 105–118, 2013.

[9] A. Dainotti, K. Benson, A. King, kc claffy, M. Kallitsis, E. Glatz, and X. Dimitropoulos. Estimating internet address space usage through passive measurements. *SIGCOMM Comput. Commun. Rev.*, 44(1):42–49, 2014.

[10] A. Dainotti, A. King, kc Claffy, F. Papale, and A. Pescapè. Analysis of a "/0" stealth scan from a botnet. In *Proceedings of the 2012 Internet Measurement Conference*, IMC '12, page 1–14, 2012.

[11] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman. A search engine backed by internet-wide scanning. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, page 542–553, 2015.

[12] Z. Durumeric, M. Bailey, and J. A. Halderman. An internet-wide view of internet-wide scanning. In *Proceedings of the 23rd USENIX Conference on Security Symposium*, SEC '14, page 65–78, 2014.

[13] Z. Durumeric, F. Li, J. Kasten, J. Amann, J. Beekman, M. Payer, N. Weaver, D. Adrian, V. Paxson, M. Bailey, and J. A. Halderman. The matter of heartbleed. In *Proceedings of the 2014 Conference on Internet Measurement Conference*, IMC '14, page 475–488, 2014.

[14] Z. Durumeric, E. Wustrow, and J. A. Halderman. ZMap: Fast internet-wide scanning and its security applications. In *Proceedings of the 22nd USENIX Conference on Security Sumposium*, SEC '13, pages 605–620, 2013.

[15] M. Ford, J. Stevens, and J. Ronan. Initial results from an ipv6 darknet. In *Proceedings of the International Conference on Internet Surveillance and Protection*, ICISP '06, pages 13–13, 2006.

[16] GreyNoise. Greynoise is the source for understanding internet noise. https://www.greynoise.io/, 2023. [Online; accessed 1-June-2023].

[17] H. Griffioen and C. Doerr. Examining mirai's battle over the internet of things. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, CCS '20, page 743–756, 2020.

[18] Cisco Talos Intelligence Group. Phishtank | join the fight against phishing. https://phishtank.org/, 2023. [Online; accessed 1-June-2023].

[19] A. Guillot, R. Fontugne, P. Winter, P. Merindol, A. King, A. Dainotti, and C. Pelsser. Chocolatine: Outage detection for internet background radiation. In *Proceedings of the 2019 Network Traffic Measurement and Analysis Conference*, TMA '19, pages 1–8, 2019.

[20] E. Le Malécot and D. Inoue. The carna botnet through the lens of a network telescope. In *Proceedings of the 6th Symposium on Foundations and Practice of Security*, FPS '13, page 426–441, 2013.

[21] MaxMind. Geoip2 databases. https://www.maxmind.com/en/geoip2-databases, 2023. [Online; accessed 1-June-2023].

[22] J. Mazel, R. Fontugne, and K. Fukuda. Profiling internet scanners: Spatiotemporal structures and measurement ethics. In *Proceedings of the 2017 Network Traffic Measurement and Analysis Conference*, TMA '17, pages 1–9, 2017.

[23] Trend Micro. Xorddos, kaiji variants target exposed docker servers. https://www.trendmicro.com/en_us/research/20/f/xorddos-kaiji-botnet-malware-variants-target-exposed-docker-servers.html, 2020. [Online; accessed 1-June-2023].

[24] A. Mirian, Z. Ma, D. Adrian, M. Tischer, T. Chuenchujit, T. Yardley, R. Berthier, J. Mason, Z. Durumeric, J. A. Halderman, and M. Bailey. An internet-wide view of ics devices. In *Proceedings of the 14th Annual Conference on Privacy, Security and Trust*, PST '16, pages 96–103, 2016.

[25] D. Moore. Network telescopes: Observing small or distant security events. In *Proceedings of the 11th USENIX Security Symposium*, SEC '02, 2002.

[26] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, and N. Weaver. Inside the slammer worm. *IEEE Security Privacy*, 1(4):33–39, 2003.

[27] D. Moore, C. Shannon, and KC claffy. Code-red: A case study on the spread and victims of an internet worm. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurment*, IMW '02, page 273–284, 2002.

[28] R. Pang, V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson. Characteristics of internet background radiation. In *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, IMC '04, page 27–40, 2004.

[29] Radware. Internet noise is taxing online services and businesses. https://www.radware.com/security/ddos-threats-attacks/threat-advisories-attack-reports/internet-noise-is-taxing-online-services-and-businesses/, 2022. [Online; accessed 1-June-2023].

[30] P. Richter and A. Berger. Scanning the scanners: Sensing the internet from a massively distributed network telescope. In *Proceedings of the Internet Measurement Conference*, IMC '19, page 144–157, 2019.

[31] robertdavidgraham. Masscan: Mass ip port scanner. https://github.com/robertdavidgraham/masscan, 2013. [Online; accessed 1-June-2023].

[32] C. Rossow. Amplification hell: Revisiting network protocols for ddos abuse. In *Proceedings of the 21st Annual Network and Distributed System Security Symposium*, NDSS '14, pages 23–26, 2014.

[33] SANS. Internet storm center/dshield api. https://isc.sans.edu/api/threatcategory/research/, 2023. [Online; accessed 1-June-2023].

[34] Shodan. Shodan search engine. https://www.shodan.io/. [Online; accessed 1-June-2023].

[35] E. H. Spafford. The internet worm program: An analysis. *SIGCOMM Comput. Commun. Rev.*, 19(1):17–57, jan 1989.

[36] S. Srinivasa, J. M. Pedersen, and E. Vasilomanolakis. Open for hire: Attack trends and misconfiguration pitfalls of iot devices. In *Proceedings of the 21st ACM Internet Measurement Conference*, IMC '21, page 195–215, 2021.

[37] VirusTotal. Virustotal. https://www.virustotal.com/gui/home/upload, 2023. [Online; accessed 1-June-2023].

[38] G. Wan, L. Izhikevich, D. Adrian, K. Yoshioka, R. Holz, C. Rossow, and Z. Durumeric. On the origin of scanning: The impact of location on internet-wide scans. In *Proceedings of the ACM Internet Measurement Conference*, IMC '20, page 662–679, 2020.

[39] E. Wustrow, M. Karir, M. Bailey, F. Jahanian, and G. Huston. Internet background radiation revisited. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, page 62–74, 2010.

[40] ZoomEye. Zoomeye - cyberspace search engine. https://www.zoomeye.org/. [Online; accessed 1-June-2023].