Why do Internet Devices Remain Vulnerable? A Survey with System Administrators

Tamara Bondar
Carleton University
tamara.bondar@cmail.carleton.ca

Hala Assal
Carleton University
assal@sce.carleton.ca

AbdelRahman Abdou
Carleton University
abdou@scs.carleton.ca

***Abstract*—In efforts to understand the reasons behind Internet-connected devices remaining vulnerable for a long time, previous literature analyzed the effectiveness of large-scale vulnerability notifications on remediation rates. Herein we focus on the perspective of system administrators. Through an online survey study with 89 system administrators worldwide, we investigate factors affecting their decisions to remediate or ignore a security vulnerability. We use Censys to find servers with vulnerable public-facing services, extract the abuse contact information from WHOIS, and email an invitation to fill out the survey. We found no evidence that awareness of the existence of a vulnerability affects remediation plans, which explains the consistently small remediation rates following notification campaigns conducted in previous research. More interestingly, participants did not agree on a specific factor as the primary cause for lack of remediation. Many factors appeared roughly equally important, including backwards compatibility, technical knowledge, available resources, and motive to remediate.**

I. Introduction

Software vulnerabilities are discovered daily. A vulnerability is more impactful when it is found in a popular software. The situation becomes worse when such a software runs a publicly accessible service (*e.g.,* not hidden behind a firewall) on an Internet-connected system. When a vulnerability is discovered in an actively maintained software, the software vendor typically issues an updated version. Globally, system administrators should then update their local software to remediate the vulnerability.

In reality, however, there is almost always a long tail of systems that remain on older software versions, thus vulnerable. This is usually the case, not just for software bugs, but for other types of flaws that expose a system to attacks, including configuration vulnerabilities (*e.g.,* conflicting firewall rules [36], accounts left enabled with default passwords [60]), and design flaws (*e.g.,* in network protocols, network topologies).

Compromised systems on the Internet are becoming increasingly dreadful for everyone, not just the system owner, for the following reasons. First, while online content in the early days of the web belonged mostly to the organization hosting it, such content today often involves user data; a compromised system can leak a new *data dimension*, such as a user's music interest, which improves illicit user profiling activities. Second, attackers learn and improve as they compromise more devices. For example, due to common habits of web password reuse [29], a compromised system improves password-guessing success on other systems [7], [2]. This is not limited to web passwords, and not to password reuse—attackers generally learn about password selection habits by leaking passwords on different systems [1]. Third, the reliance on libraries and frameworks across different web domains (*e.g.,* Javascript libraries, open-source web applications like WordPress) means a vulnerability in any would have a ripple effect [56]. Fourth, if the compromised device remains under attacker control, it means its computational and bandwidth resources can be used to damage other systems. It is common for such a device to be part of a botnet that contributes to DDoS attacks or malware distribution, or be part of a computational pool that is involved in illicit cryptomining, bruteforcing weak cryptographic keys/passwords, or populating misinformation through generating content from multiple sources (*e.g.,* creating fake twitter trends [23]).

Previous literature tested the effectiveness of large-scale vulnerability notification campaigns on remediation rates. A common approach in such studies is to identify a specific set of remotely testable vulnerabilities, scan the Internet for vulnerable devices, select a subset of the devices found and notify the entities registered (*e.g.,* in WHOIS [15]) as the contacts responsible for them. A subset is often left unnotified to serve as a control group. A common takeaway is that such notification campaigns do improve remediation rates, but a significant proportion of systems remain vulnerable even after notifications are sent.

Motivated by the low remediation rates found in previous literature, we were curious to understand the importance of an administrator's awareness that a vulnerability exists in one of their systems on the remediation rates. This has generalized to *RQ1. what are the main factors that prevent an administrator from remediating a vulnerability?* Consequently, we analyze *RQ2. how such factors change with variables such as, severity, company size, administrator's team size?*

Our focus is exclusively from the administrators' perspective, thus herein we solicit their thoughts on vulnerability remediation. We replicate the above methodology to identify the administrators to contact. Specifically, we extracted the abuse contacts in the WHOIS database for systems that we found are vulnerable to 9 selected vulnerabilities with varying severity. We then emailed notifying them about the identified vulnerability, along with an invitation to fill a survey. The

Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb) 2023
3 March 2023, San Diego, CA, USA
ISBN 1-891562-87-8
https://dx.doi.org/10.14722/madweb.2023.23043
www.ndss-symposium.org

survey mainly asked participants to rate the importance of 9 factors in the decision to remediate a vulnerability. Examples of such factors include cost of remediation, limited knowledge of the vulnerability or the remediation process, and compatibility issues. We sent over 13K emails; 89 participants completed the survey from 18 countries.

We did not find evidence that awareness of the existence of a vulnerability significantly affects remediation plans. This is perhaps the most compelling finding that helps explain the lack of remediation reported in previous papers (above). In fact, it appears that no factor can be considered the panacea for remediation. We did however find that, amongst the 9 factors, compatibility issues are more important than limited knowledge of the remediation process. This suggests that creating backward-compatible software patches and decommissioning old systems appear more important for higher remediation rates than, *e.g.,* training administrators.

II. Background

We provide a brief background on software vulnerabilities, device search engines (including Censys [8]), and WHOIS.

A. Vulnerabilities and Software Security

A vulnerability is typically a misconfiguration or deployment issue (including weak passwords), a design flaw (in a system or protocol), or a software/hardware weakness [60]. For software, the National Institute of Standards and Technology (NIST) tracks vulnerabilities in popular software, and makes them publicly accessible through the National Vulnerability Database (NVD). Each vulnerability gets a unique identifier, called the Common Vulnerabilities and Exposures (CVE) number, along with a severity rating based on an industry standard—the Common Vulnerability Scoring System (CVSS) [43]. When a vulnerability is discovered in a software, *responsible disclosure* is a security ethic that urges the discoverer to notify the vendor and allow time to release an update before publicly announcing the vulnerability. Vendors then issue an updated software version. However, in some cases, to honour legacy software that may still be running on some systems for compatibility, vendors can also apply the patch on old versions thus fixing the vulnerability in the old versions while maintaining the old version number—a process dubbed *backporting*. While relatively infrequent [17], and despite the advantage of keeping legacy software secure, backporting creates a challenge for vulnerability trackers that rely on software version number while scanning from the outside of an organization.

B. Devices Search Engines and Censys

In 2013, Durumeric *et al.* [22] designed ZMap, a fast scanner that can iterate over the entire IPv4 address space in 45 minutes by running independent scanning threads. A suite of open-source tools were later built on top of ZMap, including ZGrab (an application-layer scanner supporting popular protocols like HTTP and SSH) and ZCertificate (an X.509 certificate parser) [57]. Censys is an organization that uses ZMap-based tools to frequently scan the Internet (minus networks whose operators opted out of being scanned) [20]. It provides a web-based search interface and an API to query results, thus abstracting data collection for researchers. Companies offering comparable services exist, notably Shodan [51]. An empirical analysis was conducted for both engines, and it was found that both are able to reflect a status-change in an Internet-connected device within 24 hours [6].

For supported application-layer protocols, Censys extracts header data from initial exchanges—a process called *banner grabbing*. This allows Censys to store and index such data, which notably includes the software version number. For example, the following query finds all devices running OpenSSH version 7.9:

```
services.ssh.endpoint_id.software_version:
                    7.9.0.0*
```

Additional criteria can be combined using classical boolean connectors: `AND`, `OR`, and `NOT`. So the above can be modified to specifically search the standard SSH port, 22, as follows:

```
services.ssh.endpoint_id.software_version:
      7.9.0.0* AND services.port: 22
```

While this process shows how easy it can be to find all devices running a specific software version, using Censys to identify vulnerable devices by searching for unpatched software versions is challenged by the occasional backporting activities mentioned above. This methodology can result in false positives: devices mistakenly identified as vulnerable.

C. WHOIS

WHOIS is an overloaded term that typically refers to one of the following: a database, a network protocol, or a software. Elaboration follows.

A registrar is an organization that *sells* domain names, which involves maintaining a list of owned domains and coordinating with other registrars globally before registering a new domain. There are over 1000 ICANN-accredited registrars around the world [28], with `GoDaddy` being one of the largest. To conform with ICANN policies, a registrar must collect information about a prospective domain owner (registrant) before registering a domain. Such information primarily includes the registrant's name and contact address, but other information can also be collected—notably an abuse contact email address used to report domain-related abuse. Such information is stored in the registrar's database, called the WHOIS database, and is thus often referred to as WHOIS information. Other organizations that maintain WHOIS databases (typically for IP addresses) notably include Regional Internet Registries (RIRs), such as ARIN, APNIC, and AFRINIC.

The WHOIS network protocol is defined in RFC 3912 [15], which specifies the syntax and semantics for the message exchange between a client querying a WHOIS database, and a server (listening on TCP port 43) with access to the database. Such standardization is helpful as it facilities querying the database of any compliant registrar.

Finally, a WHOIS software is a client-side implementation of the WHOIS protocol. Many such implementations exist, and are made available through a web interface (*e.g.,* https://who.is/), a GUI application, or a command-line tool (*e.g.,* `whois` command in Linux).

III. Related Work

We cover two categories of related work: (i) research efforts towards Internet-wide vulnerability notification campaigns and their effectiveness, and (ii) literature on software updates.

A. Internet-Scale Vulnerability Notification

It is believed that not knowing a vulnerability exists is amongst the primary reasons for admins' lack of remediation. We shed light on previous research targeting large-scale vulnerability notifications over the Internet.

Vulnerability notification effectiveness. Li *et al.* [36] ran a wide-scale campaign to test various notification methods, including who to contact (*e.g.,* CERTs versus WHOIS abuse contacts), email language and format (*e.g.,* verbose versus terse, English versus the recipients' language). They identified three vulnerabilities related to industrial control systems, a misconfigured IPv6 firewall (one that is more permissive than IPv4), and hosts vulnerable to UDP-based DDoS amplification attacks. The authors found that a verbose English email to the WHOIS abuse contact is the most effective. However, this method resulted in only 18% remediation rate (within 2 weeks of emailing). Other studies reached comparable results. Durumeric *et al.* [21] reported that their notifications increased Heartbleed remediation from 26.8% (unnotified) to 39.5%. Stock *et al.* [56] notifications increased remediation rates of website XSS vulnerabilities from 2% (unnotified group) to 12%. Zeng *et al.* [64] notifications dropped the proportion of websites with outdated ciphersuites to 90% (versus 93% unnotified) and with outdated TLS versions to 95% (versus 97% unnotified), 2 weeks after notifications.

Factors affecting notification success. Beyond the above efforts, Stock *et al.* [55] designed a study specifically to identify reasons behind the lack of remediation after notification. They concluded that ensuring that a notification is delivered to the affected party is not enough (likewise did Durumeric *et al.* [21]), and that several parameters come into play in convincing the recipient to invest the required time/effort to investigate and remediate as applicable. Other factors affecting remediation rates were also studied. While Cetin *et al.* [12] reported that the sender's (notifier) reputation does not affect remediation rates (for 480 notifications), Stock *et al.* [55] found that the sender's trust was an important factor in the response rates (for 24K domains). The latter also found that other means of communication, including social-media-based contacts and phones, were not significantly more successful than emails. However, Maass *et al.* [40] found that snail mail notification increased remediation rates (after 2 weeks) to 42%, compared to 33% for email-based notification and 2% unnotified. Cetin *et al.* [11] found that providing a proof-of-concept of the vulnerability did not significantly affect remediation rates.

Notification of compromised (*vs.* vulnerable) systems. In a 2016 study [37], using the Google Webmaster Console resulted in one of the highest remediation rates compared to other studies. This can likely be attributed to the severity of the situation: Li *et al.* [37] notified *hijacked* websites, which is more worrisome than a vulnerability that has not necessarily been exploited yet. Likewise, Vasek and Moore [61] focused on malware-distributing (compromised) websites, and found that 62% of such sites were cleaned after notification, versus 45% for the unnotified. Recently, Woods and Böhme [63] devised a model to explain security outcomes as a function of, *e.g.,* exposure and threat, and the effectiveness of *security intervention*, *e.g.,* direct notification. They highlight how notifications about an already compromised system, *e.g.,* [62], [37], had a more successful remediation rate than notifications about an observed security level. A slightly different, and stricter, approach to remediate compromised machines was tested in 2019, which is quarantining infected machines. This appears to be amongst the most effective remediation strategies, with 92% remediation rates for infected ISP machines [9], and 87% for end systems [10].

Beyond notification: active engagement. One noticeably different result was reported in 2014, where Kührer *et al.* [34] reported a 90% drop in the number of servers vulnerable to DoS amplification attacks over a 2-months period after an aggressive notification campaign. It is unclear whether the drop would have happened without the authors' efforts as the study lacked a control group (though a correlated remediation that followed their campaign can be observed). A stark difference between this campaign and the ones above is the campaign nature. Kührer *et al.* were in active engagement with involved parties, including collaboration with security organizations, creating technical advisories explaining how to remediate (eventually leading to public advisories on US-CERT), engaging with key industry players like Cisco (also resulting in a public Cisco advisory), and sharing their data about vulnerable devices with hundreds of institutions worldwide.

Relationship to the work herein. As questions remain regarding the reason behind too many systems remaining vulnerable for a long time after updates are issued and after admins get notified, we set out to approach this question by directly soliciting the admins' input on the matter. A proportion of our methodology is analogous to the above efforts: finding vulnerable systems, extracting contact information, and emailing them. However, as our objective is not to evaluate the effectiveness of a notification strategy, we do not test whether vulnerable systems were remediated after emailing admins. To the best of our knowledge, the work herein is the first to explicitly focus on the admins' perspective in understanding the reasons behind the lack of remediation. While several of the above papers surveyed [21], [36], [11], [55], [64] (or interviewed [9]) admins, these surveys were mainly targeted to solicit input on the research methodology, and were thus not designed to answer the research questions herein.

B. Software Updates

As discussed in Sec. II, vulnerabilities in Internet-connected systems go beyond software weaknesses; they include other aspects, *e.g.,* misconfiguration, deployment issues, design flaws, and hardware weaknesses. However, we shed light on software updates, a rich research area.

Admins approaches to software updates. Li *et al.* [38] surveyed 102 admins, and found that admins sometimes face difficulties in the update process, including in determining an update is available and applying it properly. Martius and Tiefenau [42] found that for software that is not configured to receive automatic updates, admins seek knowledge of the purpose, dependencies, and known issues with a new version

before installing it. In general, admins appear reluctant to update a software when the consequences of an update are not quite clear [58]. Focusing specifically on patch application, Jenkins *et al.* [30] analyzed the `patchmanagement.org` mailing list, and found that discussions revolved around several themes, including patch prioritization and errors/troubleshooting of new patches. The authors [30] report that the latter theme was found to be the most popular.

Software update challenges. Dissanayake *et al.* [19] conducted a literature review on applying software patches, and summarized 14 socio-technical challenges. Those include coordination overhead between stakeholders, and general limitations of software security patch-management tools (*e.g.,* usability and correctness). Combining the security and functionality perspectives, Beattie *et al.* [5] analyzed the best time to apply a patch—too early can lead to applying an immature release, too late leads to security exploits. The authors [5] highlight that the admins' challenges are exacerbated by poorly-developed patches that lead to system failures.

Relationship to the work herein. Our focus herein is on the general problem of vulnerability remediation. Keeping software up-to-date is one way of remediation. While our participant-recruitment methodology was based on searching the Internet for devices running outdated and vulnerable software (Sec. IV), our survey is designed to study the admins' views on vulnerability remediation, rather than their management of software updates (*i.e.,* motives, methods and strategies for software updates).

IV. Study Design and Methodology

We conducted an IRB-approved online survey study to answer our research questions using Qualtrics [48]. Data collection was conducted between 2020 and 2021.

A. Participant Recruitment

Our objective is to survey administrators who have a known vulnerability in their systems, as we expect their input will be more relevant to our study than others with no vulnerabilities. We focus specifically on software vulnerabilities. We randomly selected 9 vulnerabilities found in popular software over the past 8 years, with varying severity levels. Patches for these vulnerabilities were available between 4 days (POODLE) and 47 days (Exim) from the date when the vulnerability was made public. Table I shows the selected vulnerabilities and their CVE numbers. For each vulnerability, we used Censys (see Sec. II-B) to find the IP addresses of all devices running the specific software versions that are reported as vulnerable in the NVD. We then queried the RIRs WHOIS databases for contact information (see Sec. II-C), prioritizing abuse contact if found, and falling back to regular email contact otherwise.

To email survey invitations, we refrained from using our institution's email server to avoid adverse effect from wide-scale email campaigns (*e.g.,* blocklisting our domain, classification as spam). Instead, we set-up a new mail server, and configured the conventional suite of mail server authentication, namely SPF, DKIM and DMARC, to enable proper email delivery and reduce the likelihood of spam classification. We verified that our email domain is not listed as spam, *e.g.,* in spamhaus.org and www.spamcop.net. We also verified that emails sent from our mail server are properly delivered to the mailboxes of Gmail and Outlook, rather than the Spam/Junk folders, as well as the mail servers of several organizations employing friends and family, include government and private organizations.

Using Censys, and aggregating data of all 9 vulnerabilities, we found $\sim$4 million potentially vulnerable devices (IP addresses) and $\sim$1.2 million associated emails in WHOIS. Of these, 13,191 email addresses were unique, to which we sent survey invitations. Table I shows this data, broken down by each of the 9 vulnerabilities. The email invitation sent to each participant lists the specific vulnerability identified in the participant's server, and an explanation of how we have obtained their contact information. Appendix B shows an example of the email we sent for the POODLE vulnerability (CVE-2014-3566). We took ethics and good Internet citizenship practices into account when sending these emails. See Sec. IV-C.

B. The Survey

Recruiting experts for research studies is challenging [4], [45]. Thus, to encourage participation, we created a relatively short survey (22 questions) to gain insights into factors influencing remediation decisions. The survey (Appendix A) included questions to provide context (*e.g.,* organization size, team size), questions relating to remediation efforts, and semantic-scale questions where participants were asked to rate different factors that could influence their decision to remediate the specific vulnerability identified in their servers. These factors were suggested by previous work as potential barriers to vulnerability remediation [53], [64], [56], [36]. All survey questions discussed herein were presented in the context of the vulnerability identified in the participant's server. Semantic scale questions were shown to participants in random order to avoid potential ordering effects.

C. Ethical Considerations

We followed our institution's ethics research board recommendations and best practices of good Internet citizenship [22] when setting up our study and sending recruitment emails.

We ensured that our recruitment email and survey consent form elucidate that participation is completely voluntary. During participant recruitment, we were transparent about our activities, setting up a web page explaining who we are, what we do, and how we can be contacted. This page was set-up on the same domain from which the emails were sent. The page also linked to our research lab's website (with our institution's domain) for further information about lab activities, and our lab's website linked to this web page, which helps further establish authenticity in our study. We provided an opt-out option, *e.g.,* from future studies, both in our email and in the web page. Finally, we limited the emails sent to each email address to one, to avoid spamming mailboxes. That is, if an email address was listed as the (abuse) contact for a vulnerable device, and we have already emailed this address for a previous vulnerability, we will not email it again.

Before starting the survey, participants were presented with a consent form explaining the purpose of the study, expected risks and benefits to participants, and data storage plans. The

Table I. VULNERABILITIES AND EMAILS IDENTIFIED. OVERALL, $n = 89$ PARTICIPANTS COMPLETED THE SURVEY, 3 OF WHOM WERE NOT LINKABLE TO A VULNERABILITY (SEE SEC. IV-D FOR DETAILS).

Vulnerability	AKA	Severity	#vuln devices	#emails found in WHOIS	#unique emails	#participants (%)
CVE-2019-6111	OpenSSH	5.9 (M)	671	587	117	-
CVE-2014-3566	POODLE	3.4 (L)	91,413	38,900	2,739	4.5 ($n = 4$)
CVE-2018-3110	Java VM	9.9 (C)	1,382	1,047	89	-
CVE-2014-0160	Heartbleed	7.5 (H)	64,187	37,824	1,802	2.2 ($n = 2$)
CVE-2019-15846	Exim vuln	9.8 (C)	618,866	401,722	1,835	30.3 ($n = 27$)
CVE-2020-6287	SAP NetWeaver	10 (C)	9,684	2,574	262	2.2 ($n = 2$)
CVE-2018-16845	Nginx	6.1 (M)	24,045	7,509	147	2.2 ($n = 2$)
CVE-2017-3169	Apache vuln	9.8 (C)	1,048,405	440,305	4,143	50.6 ($n = 45$)
CVE-2018-15599	Dropbear	5.3 (M)	2,040,824	668,513	1,464	4.5 ($n = 4$)
	n/a					3.4 ($n = 3$)

Severity levels: C - Critical, H - High, M - Medium, L - Low.

form also lists the researchers' and the IRB's contact information for questions or concerns. Participants must consent to the form before starting the survey.

Our survey is completely anonymous; we do not collect any participant-identifiable information (*e.g.,* name, age, organization name), and any identifiable information entered by participants in the open-ended questions were anonymized before data analysis. Participants were able to skip any questions they were uncomfortable answering. Survey responses were deleted from Qualtrics after data collection concluded. Anonymous data is stored on password-protected local machines, only accessible to the research team.

D. Data set

In total, we received 92 responses to our online survey. To ensure data quality, we discarded responses from three participants who selected "*I prefer not to answer*" for more than 90% of the questions. The results discussed herein are based on the remaining 89 responses. Participants took on average 12.99 ($Md = 5.2$) minutes to complete the survey.

We asked participants to indicate the IP address that we had identified in our email to help us link their responses to the identified vulnerability. All but 8 reported the IP address, but the CVE for 5 of these 8 was indicated. We were, therefore, unable to link three participants to a vulnerability. In hindsight, we should have created a separate survey link per CVE rather than asking participants to report the IP address. While our recruitment invitations spanned nine unique vulnerabilities (Table I), our participant responses were relating to seven unique vulnerabilities. As shown in Table I, the vast majority of participants in our dataset ($83\%, n = 74$) were running software versions reported as having *critical* severity [43].

Participant demographics. Table II summarizes our participants demographics. Most participants ($\approx 62\%$) work for organizations with fewer than 500 employees (considered Small-Medium Enterprises (SMEs) [16], [54]), and almost 27% of participants are employed in Large Enterprises (LEs). Moreover, as shown in Table II, most participants (64%) stated that the size of the remediation team in their organization is between 2 and 10 people. However, almost a quarter of our participants reported being the sole responsible for remediating vulnerabilities in their organization. For the 81 participants who reported back the IP addresses in the survey, we conducted an IP geolocation lookup using `ipinfo.io`. The results are shown in Table III. The devices were located in 18 countries spanning 4 continents; the majority (50) came from North America, and the rest were mostly distributed over Europe (14), Oceania (12), and Asia (5).

Table II. PARTICIPANTS' DEMOGRAPHIC ($n = 89$)

Criteria	Percentage (%)
Size of organization	
At most 500 employees	61.8
501 to 5000 employees	20.2
5001+ employees	6.7
Prefer not to answer	11.2
Size of The Remediation Team	
Just me	23.6
2 to 10 people	64.0
11 to 20 people	4.5
21+	3.4
Prefer not to answer	4.5

Table III. A COUNTRY LOOK-UP BASED ON IP ADDRESS GEOLOCATION FOR $n = 81$ PARTICIPANTS WHO REPORTED THE IP ADDRESS IN THEIR SURVEY.

Country	N	Country	N
United States	34	Japan	1
Canada	16	Netherlands	1
Australia	12	Norway	1
Austria	3	Philippines	1
France	2	Romania	1
United Kingdom	2	Switzerland	1
Belgium	1	Thailand	1
Germany	1	Turkey	1
India	1	Uzbekistan	1

E. Data Analysis

Quantitative data was analyzed using IBM SPSS v.28. Table IV summarizes the between-subject statistical tests we performed. For example, we use Pearson's Chi-Square test [44], [24] when testing the effect of the size of remediation team on participants' awareness. In this case, we are comparing more than two groups (*i.e.,* team size=1, team size=2–10, and team size=11+). However, we use Fisher's Exact test for accuracy when comparing exactly two groups [26], [24], *e.g.,* when exploring the effect of participants' organization size on their awareness (*i.e.,* the two groups being participants from SMEs and those from LEs). We use Kruskal-Wallis H test [33], [24] to determine the effect of two or more groups of independent variables on an ordinal dependent variable. For example, we test the effect of organization size (groups being SME and LEs) on each of the remediation barriers (ordinal dependent variable measured on a 5-point semantic scale)—*how important are Compatibility issues for SMEs compared to LEs?* As indicated in Table IV, post-hoc analysis with Bonferroni correction was done only when the test result was significant. We also explored the differences between the

various remediation barriers through within-subject tests using Friedman Rank Sum Test, and Wilcoxon Signed-rank Test with Bonferroni corrections when applicable. We indicate the test performed when presenting its results in Section V.

Qualitative data (*e.g.,* responses to survey open-ended questions and email responses) was analyzed using thematic analysis [3]. We performed *open coding* to abstract and conceptualize the data [14] by assigning *codes* to describe the main themes or ideas discussed in each excerpt [35]. Through open coding we look for interesting themes and common patterns in the data that are relevant to our research questions. For example, we looked at participants' reasoning to remediate or forgo remediation, additional remediation barriers, remediation plans, as well as any other relevant comments. Affinity diagrams [32], [31], through a virtual collaboration tool, were used to discuss the codes and themes, and identify patterns in the data. Codes were written on (virtual) sticky notes, and similar notes were grouped together to represent higher abstractions or themes [31]. These groupings were not based on preconceived ideas, rather they are grounded in the data. As recommended by previous work [13], open coding was done by a single researcher with extensive experience in qualitative data analysis, and the research team regularly met to discuss codes, themes, and affinity diagrams to ensure consistency and that they reflect the sentiment and information in the quotations.

F. Limitations

In this paper, we focus on the view of specific admins whose systems remain vulnerable despite the technical availability of a remediation mechanism (a "software patch" in our case herein). While our findings may not generalize to *all* admins, they provide insights into the issues facing admins with vulnerable systems to help identify ways to support them.

Because the survey is anonymous, we cannot be entirely sure that participants are in fact employed as system admins. However, we have done our due diligence to reach out to the proper demographic by emailing the *abuse contact* in WHOIS, typically a technical network operator [50].

We sent thousands of recruitment emails, hoping to reach a demographic with a wide range of vulnerability severity. Despite our efforts, our sample is skewed towards highly severe vulnerabilities . A candidate research follow-up can examine if a more balanced sample would provide further insights.

Finally, we recognize that our sample size ($n = 89$) could have been larger. While we do not claim that it is fully representative, we believe the findings herein are useful to the community and provide insights into a real problem shaping the path to support system administrators.

V. Results

All our survey questions include a *Prefer not to answer* option, so participants were able to skip questions. We report the actual number of data points when reporting the results.

A. Remediation and Awareness

a) Remediation response: Overall, our participants were almost equally split in their remediation response; $45\%(n = 40)$ of participants indicated that they have remediated or attempted to remediate the vulnerability, and $51\%(n = 45)$ have not (see Table V). Encouragingly, the majority of participants ($67\%, n = 30$) who have not remediated the vulnerability indicated that they plan to do so (see Fig. 1). Whereas, $27\%(n = 12)$ indicated that they do not plan to remediate the identified vulnerability.

b) Awareness of the vulnerability: As a first step to determine the role of awareness in remediation, we asked participants if they were aware of the identified vulnerability prior to our contact. We again found an almost equal split, $51\%(n = 45)$ of participants were previously unaware of the vulnerability, and $46\%(n = 41)$ were aware of it (see Table V).

Interestingly, not all participants who did not remediate the identified vulnerability were unaware of it. As shown in Fig. 1, of the participants who were aware of the vulnerability, $61\%(n = 25)$ indicated they have remediated or attempted to remediate it, yet more than third ($37\%, n = 15$) have not attempted remediation despite their awareness. On the other hand, $29\%(n = 13)$ of participants who were previously unaware of the vulnerability indicated that they had already fixed it by the time they filled our survey. It is unclear though if participants who have remediated (or attempted to remediate) the vulnerability have done so after our contact and before responding to the survey, or if remediation was a byproduct of another task.

Regardless of prior awareness of the vulnerability, some participants indicated that they plan to remediate ($n = 21$: previously unaware, and $n = 9$: previously aware). Although not our main research objective, our recruitment email may have served as a notification of the vulnerability and impetus for taking action towards remediation. On the other hand, twelve participants indicated that they do not intend to fix the vulnerability ($n = 5$: previously aware, and $n = 7$: previously unaware). We analyzed qualitative responses from these participants to understand the reasoning behind such concerning outcome. We found that the reasons for foregoing remediation centered mainly around the lack of benefits from remediation. For example, some participants explained that the vulnerability is a false positive;[1] the vulnerable server is old, its shutdown is planned and *"no one is interested in fixing unused systems on that server"*(P54); or that the server acts as a honeypot and does not include vulnerable data. We also found cases where participants indicated they are unable to, or not responsible to, remediate the vulnerability (*e.g.,* participants from Internet Service Providers (ISPs) and the identified vulnerability belongs to their customers).

We conducted statistical analysis to explore association between participants' awareness and remediation of the identified vulnerability. We assumed an optimistic outlook, considering participants who plan to remediate will be successful in following through their plan, and conducted between subject analysis with those, who remediated or plan to remediate as one group, and those who did not remediate and do not plan to do so as the other group. Fisher's Exact Test did not show statistically significant association between awareness and vulnerability remediation response ($p = 0.757, N = 80$).

[1] It is unclear if participants verified this claim

Table IV. SUMMARY OF BETWEEN-SUBJECT STATISTICAL TESTS PERFORMED. POST-HOC ANALYSIS WITH BONFERRONI-CORRECTION WAS PERFORMED ONLY IN CASE OF STATISTICAL SIGNIFICANCE.

	Organization size	*Severity level*	*Size of remediation team*	*Awareness*
Awareness	Fisher's Exact	Fisher's Exact	Pearson Chi-Square	—
Remediation response	Fisher's Exact	Fisher's Exact	Pearson Chi-Square	Fisher's Exact
Remediation barriers	Kruskal-Wallis	Kruskal-Wallis	Kruskal-Wallis & Dunn's	—

An entry in row i column j shows the Test performed to test the effect of j (independent variable) on i (dependent variable).

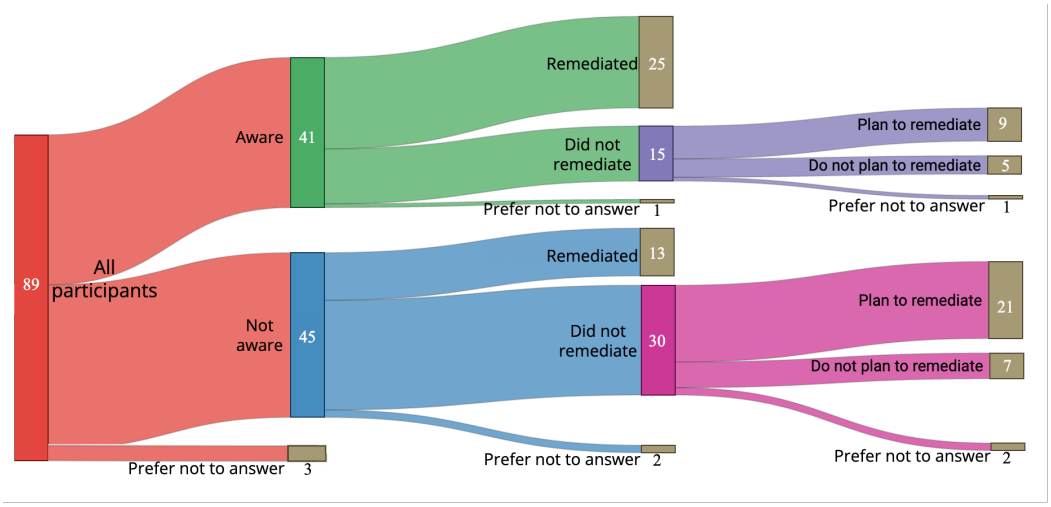

Figure 1. Awareness, remediation response and remediation plans ($n = 89$)

Table V. SURVEY RESPONSES RELATING TO AWARENESS, REMEDIATION RESPONSE, AND PLAN TO REMEDIATE THE IDENTIFIED VULNERABILITY

Criteria	Percentage (%)
Aware $(n = 89)$	
Yes	46% $(n = 41)$
No	51% $(n = 45)$
Prefer not to answer	3% $(n = 3)$
Remediated $(n = 89)$	
Yes	45% $(n = 40)$
No	51% $(n = 45)$
Prefer not to answer	4% $(n = 4)$
Plan to remediate $(n = 45)$	
Yes	67% $(n = 30)$
No	27% $(n = 12)$
Prefer not to answer	6% $(n = 3)$

B. Remediation Decisions Beyond Awareness

Using 5-point semantic-scale questions, participants rated the different factors commonly believed to be barriers to remediation [53], [64], [56], [36]. These questions were presented to participants in random order to avoid ordering effects. The factors are:

- Third-party dependencies (*e.g.,* hosting provider, certificate authority) - *[3rd party dependency]*
- Compatibility issues (*e.g.,* backwards compatibility, legacy code, library compatibility) - *[Compatibility issues]*
- Limited access to relevant resources that are not controlled by the remediation team (*e.g.,* data from other teams) - *[Limited access to rel resources]*
- Issues impending the collaboration within the remediation team or with other stakeholders - *[Collaboration issues]*
- Limited vulnerability tracking tools - *[Limited vuln tracking]*
- Limited remediation tools - *[Limited rem tools]*
- Limited knowledge of vulnerability - *[Limited knowledge of vuln]*
- Cost of remediation outweighs risk - *[Cost outweighs risk]*
- Limited knowledge of remediation process - *[Lim knowledge of rem process]*

Participants were also able to indicate additional factors in an open-ended question (See Section V-C). As shown in Fig. 2, the top two of the barriers presented to participants are compatibility issues and third-party dependencies.

Friedman test indicated a statistically significant difference in the importance of different factors for all participants ($\chi^2(8) = 31.172, p < 0.001, N = 76$). Post hoc analysis using Wilcoxon signed-rank tests with a Bonferroni correction showed that *Limited knowledge of remediation process* was significantly less important than *Compatibility issues (e.g. backwards compatibility, legacy code, library compatibility)* factors ($Z = -3.599, p = 0.012$).

We then divided our participants into two groups: those who remediated the vulnerability, and those who did not. We conducted Friedman test to explore if there is a significant difference in the importance of the factors within each group. We found a significant difference between factors for participants who have remediated the vulnerability ($\chi^2(8) = 20.248, p = 0.009, N = 35$), and a significant difference for those who have not ($\chi^2(8) = 16.166, p = 0.040, N = 40$). The post hoc analysis using Wilcoxon signed rank test with Bonferroni correction did not show any significant differences during the pairwise comparisons for both groups.

C. Other Barriers to Remediation

We asked participants to report other factors that they consider as barriers to remediation beyond those described in the previous section. Through analyzing participants' open-ended responses, we identified six main themes of remediation barriers. *Lack of control* over the vulnerable system was one of the main barriers discussed by participants. In these cases, participants are offering a hosting service, and it is the customer's responsibility to ensure security. Some participants mentioned their duty ends at alerting their customers of the vulnerabilities, while others would take further actions to ensure remediation as P70 explained, *"[..] once we notify the responsible party of the vulnerability we allow 24 hours to remediate before traffic to the affected IP address will be blocked until remediation is completed."* *Politics* was also one of the most prominent remediation barriers. Participants explained that they sometimes have to go through bureaucratic processes with other departments, or have to convince management and other stakeholders of the importance of remediating vulnerabilities. One participant considered *"Political/Business Infrastructure supportive of time and personnel"* as an *"extremely important"* factor. Our analysis also shows that when the *benefit does not outweigh the cost*, admins tend to forgo remediation. Under this theme, participants consider the impact of the remediation process on existing services (*e.g.,* service downtime), the perceived risk and expected losses from a vulnerability exploitation, and their plans to decommission old vulnerable servers. Participants also indicated barriers related to *limited resources*, such as the lack of time to keep servers updated, lack of personnel (especially those who are qualified), the unavailability of vendor patches, and the lack of proper documentation from previous admins (*e.g.,* P7 said, *"I am the network administrator. The administrator of this system recently died. I was not completely aware that this server had a public network exposure[...]"*. Finally, our analysis shows that having to deal with *complex remediation processes*, and *legacy systems* where remediation may not be possible (or patches are unavailable) are also influential barriers.

D. Influential Parameters

In this section, we investigate parameters that may influence awareness of the vulnerability, remediation response and remediation barriers. When discussing remediation response, we consider participants who have remediated or plan to remediate the identified vulnerability as a "positive remediation response" group, and those who have not remediated and do not plan to as a "negative remediation response" group.

1) Organization size: We grouped participants according to the size of their organization [59] into SMEs with up to 500 employees, and LEs with 500+ employees.

a) Awareness: We used the Fisher's Exact Test to explore association between participants' awareness and organization size. We found statistically significant association $(p = 0.043, N = 76)$. Participants from LEs are more likely to be aware of the vulnerability compared to those from SMEs.

b) Remediation: We conducted between subject analysis to explore the association between organization size and remediation response (positive or negative remediation response). The Fisher's Exact Test did not show significant association between them $(p = 0.715, N = 74)$.

c) Remediation barriers: The Kruskal-Wallis H test did not show significant difference in the importance of any of the barriers between SME participants and those of LEs.

Within groups, the Friedman test showed that barriers varied significantly for both SMEs $(\chi^2(8) = 18.632, p = 0.017, N = 49)$ and LEs $(\chi^2(8) = 24.980, p = 0.002, N = 21)$. However, post hoc analysis using Wilcoxon test with Bonferroni correction did not show significant difference.

2) Severity level: We grouped participants into two groups according to vulnerability severity [43]: 1) low and medium severity (L-M); 2) high and critical severity (H-C).

a) Awareness: We ran the Fisher's Exact Test to explore association between participants' awareness of the identified vulnerability and its severity level. Test results did not show a significant association $(p = 0.090, N = 83)$.

b) Remediation: The Fisher's Exact Test showed significant association between severity level and participants' remediation response $(p = 0.026, N = 79)$. We found that vulnerabilities with L-M severity levels are more likely to have a negative remediation response.

c) Remediation barriers: The Kruskal-Wallis H test did not show significant difference in the importance of any barrier between participants of L-M severity and those of H-C.

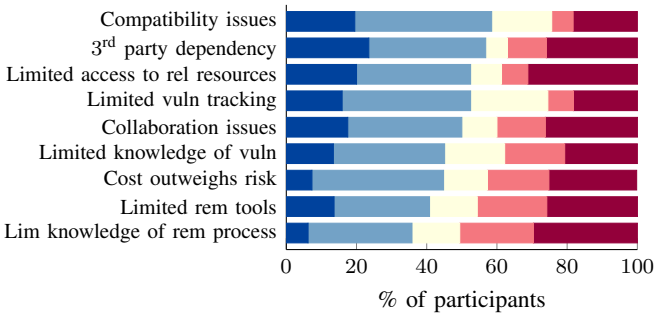

Figure 2. Barriers to remediation $(n = 80)$

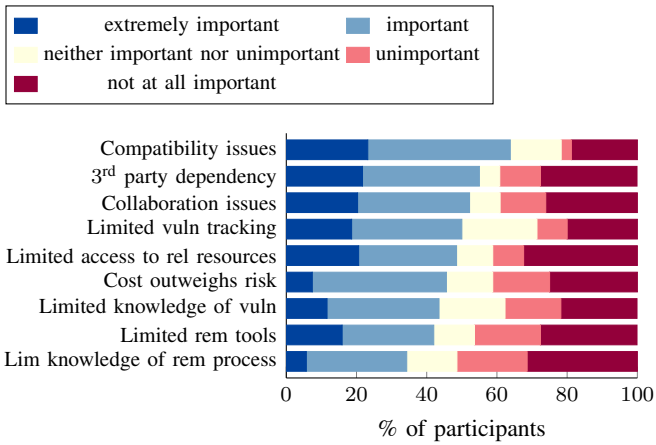

Figure 3. Barriers to remediation for participants with high and critical severity vulnerabilities $(n = 66)$

We used the Friedman test to investigate the importance of the barriers within each group. There was no significant differences between barriers for L-M severity participants ($\chi^2(8) = 9.029, p = 0.340, N = 7$). However, we found that the barriers vary significantly for group with H-C severity vulnerabilities ($\chi^2(8) = 31.056, p < 0.001, N = 66$). Post hoc analysis using Wilcoxon test with Bonferroni correction showed that for this group *Compatibility issues* was significantly more important than *Limited knowledge of remediation process* ($Z = -1.780, p = 0.007$). Figure 3 shows the importance of each barrier to the H-C severity group.

3) Size of the remediation team: We grouped participants into three groups based on the size of the remediation team: 1) a single admin; 2) 2 to 10 people; 3) 11 people or more.

a) Awareness: The Pearson Chi-Square Test was used to explore association between participants' awareness of the identified vulnerability and the size of the remediation team. We did not find statistically significant association ($\chi^2(2) = 2.916, p = 0.233, N = 82$).

b) Remediation: Similarly, the Pearson Chi-Square Test did not show statistically significant association between participants' remediation response and size of the remediation team ($\chi^2(2) = 1.612, p = 0.445, N = 79$).

c) Remediation barriers: The Kruskal-Wallis H test showed that the importance of *"Issues impeding the collaboration within the remediation team or with other stakeholders"* was significantly different between remediation teams of different sizes ($\chi^2(2) = 11.033, p = 0.004, N = 77$). The mean rank factor importance was 26.68 for a single admin, 41.7 for teams between 2–10 admins, and 54.93 for 11+ teams. Post hoc analysis, using Dunn's pairwise tests with a Bonferroni correction showed that this barrier is significantly more important for teams of 11+ admins ($p = 0.009$) and 2–10 teams ($p = 0.026$) compared to sole admins.

To investigate the importance of barriers within each group, we performed the Friedman test. We found a significant differences for participants who are the sole admin ($\chi^2(8) = 16.470, p = 0.036, N = 20$), those who work in teams of 2–10 ($\chi^2(8) = 23.420, p = 0.003, N = 47$) and for participants in 11+ admins teams ($\chi^2(8) = 19.304, p = 0.013, N = 6$). However, post hoc analyses using Wilcoxon signed-rank tests with a Bonferroni correction did not show significant differences.

E. Response to our Study

We received 156 email replies to our recruitment emails. Many were automated replies ($n = 50$) confirming email receipt, opening a ticket, or directing us to external links to submit our request. We also received 7 emails looking to verify our recruitment email. In this section, we discuss our analysis of the remaining 99 emails,[2] as well as feedback that was provided by participants in the survey.

Reactions towards our study varied widely, from appreciative, all the way to threatening legal actions against the research team. Four main themes emerged from our analysis.

[2]We do not include quotes from emails as we do not have the admins' explicit consent to share the content of their emails, unlike survey responses.

Appreciation. It was encouraging to find that many of the admins we emailed appreciated our research, even in cases where they believed the vulnerability was a false positive or has already been remediated. P81 said, "*Thank you for the initiative and proactive approach. Very helpful [in] preempting security vulnerabilities.*" We found that even admins who take the security of their systems seriously may miss vulnerabilities due to various reasons (*e.g.,* the lack of resources), and in such cases communications such as ours are welcomed. P85 explained, *"[...] we do monitor our abuse accounts for reports such as yours in an effort to tighten up anything that is brought to our attention. Every now and then, it seems something slips through the cracks,"* and P75 said, *"we cannot proactively scan all services hosted on our network, it would take too much time to setup/maintain. As such, receiving relevant notifications (with a low rate of false positive) from researchers is helpful."* Additionally, some admins were interested in furthering our research and provided their personal phone numbers in case we wanted to contact them for more information.

Negative feelings. We observed negative feelings towards our study, mostly from those who believed the vulnerability was a false positive. Understandably, some admins felt that the time they spent verifying the vulnerability was wasted. Others erroneously believed that we scanned their systems and were disconcerted. In an extreme case, an admin was convinced that the research team was withholding information that would be crucial to their system's security (*e.g.,* they believed that the research team hacked their system and obtained data from it). The admin threatened legal action against the research team, despite multiple communications explaining the methodology and assurances that the team has shared all relevant information. We tried to mitigate negative feelings by clearly explaining our methodology in the recruitment email, and avoiding wording that implied that we have confirmed the existence of the vulnerability. In Section VI-C, we discuss ethical implications for this line of research.

Remediation and confusion. We received multiple communications from system admins about the status of their (potentially) vulnerable systems. For example, some explained that they have investigated the issue, remediated the vulnerability, plan to remediate it, or that they are in the process of decommissioning the server. In some cases, the admins were confused as to why their systems appeared to be vulnerable despite having remediated the vulnerability before. They would explain the remediation steps previously taken and ask us to verify whether these steps were the correct ones.

Requesting further information. Aligning with previous research [11], [36], we found that system admins sought detailed information about the scans that identified the vulnerable software versions (*e.g.,* timestamp, logs, port number), and the tools and methodologies we used to aid their investigations and remediation efforts. For example, P80 said, *"Could you please share the tools and techniques of your research? What software was used; was it specially developed or modified according to your needs?"* One of the reasons for the interest in the tools used was that some admins considered finding a tool that scan their network and cross references vulnerabilities to the CVE database, a barrier to remediation. Some admins also requested a proof-of-concept of the vulnerability exploitation.

VI. Discussion

A. Answering the Research Questions

RQ1: Factors preventing vulnerability remediation. Our analysis suggests that no single factor was reported by admins to be of significant importance over others—*i.e.,* no one-size-fits-all solution. Furthermore, we did not find evidence that awareness of the existence of a vulnerability affects remediation plans, which advances our understanding of findings in previous literature on vulnerability notification.

RQ2: Variables affecting factors' importance. All factors received some attention from our participants. First, we found that compatibility issues were more important than limited knowledge of remediation process. This was also true for participants who responded to high and critical severity vulnerabilities. This suggests that it is not as necessary to spend more time training admins, versus creating software patches that are backwards compatible and decommissioning old systems. However, for low and medium severity vulnerabilities, we found that participants are more likely to have negative response to remediation, *i.e.,* to answer that they have not remediated the vulnerability, or that they are not planning to.

As for company sizes, while we found that the importance of factors does not change with company size, we found that LEs are more likely to be aware than SMEs. Additionally, we found that "issues impeding collaboration between members of the remediation team and other stake holders" was less important to sole admins compared to teams. Finally, from our qualitative analysis, we found other factors that influence remediation decisions, including politics, benefit versus cost, limited resources, and the maintenance of legacy systems.

B. Reflections on the Study Methodology

The utility of Censys in finding vulnerabilities. Some participants indicated how our identified vulnerability was a false positive. Such false positives are artifacts of our methodology, and not related to Censys. The reason is that Censys performs a banner grab to determine the software version number. If a software was backported (Sec. II), it becomes much harder to tell whether this version is vulnerable. It is unclear how this challenge can be fixed without actively interacting with the device and, *e.g.,* attempting to exploit the vulnerability. Backporting thus, while it provides the advantage of keeping old software patched [49], it creates a challenge for non-intrusive methodologies that use engines like Censys.

Alternatives to WHOIS for vulnerability notification. While previous literature has tested various contact methods [55], [40], emails remain cheapest and most scalable. It is clear, however, that WHOIS emails are not quite optimal for several reasons. First, WHOIS lacks standardized structure [39], and addresses are often difficult to extract [11]. Second, privacy-protective regulations, notably GDPR, are making it increasingly harder for contact information to be publicly available on WHOIS records [41]. Third, WHOIS contact information often belongs to an admin that has no direct control over a specific server. For example, an organization with a /16 IPv4 addresses (65K IP addresses) will likely have several admins in different administrative domains responsible for all of them. This is also a common scenario with ISPs, especially since WHOIS can be queried using either an IP address or a domain name, and the results are not necessarily identical. Alternative contact mechanisms include email addresses in the Start of Authority (SOA) of DNS records, and `security.txt` files [27]. Soussi *et al.* [52] found more promising (responsive) results with SOA addresses compared to WHOIS. The `security.txt` is a proposed standard whereby websites can make available information about the website's policy (what to do when reporting a security issue), preferred language of correspondence, security positions available for hiring, and who to contact with security-related issues [47]. Findlay and Abdou [25] queried WHOIS for the contact information of every domain in Tranco's top million [46] that had a `security.txt` file (∼5000 domains). The authors found zero matches, raising further questions about the utility of WHOIS for effective notification.

C. Ethics for Human-centric Internet Measurements Research

We use publicly-available data in the research presented herein (*i.e.,* IP addresses of potentially-vulnerable machines), thus did not anticipate adverse events from our study. Nevertheless, we found cases where the admins whom we contacted were shocked by the availability of this data and our ability to use it. As explained in Sec. V-E, some admins experienced relatively negative feelings. In our study, an Internet service provider mentioned that they would block their customer if the latter fails to remediate a vulnerability in a time period following notification. That is, notifying a service provider of a vulnerability may lead to adverse event for their customers.

Concrete considerations for ethical large scale notification-related research in security and privacy are yet to be defined [41]. We hope to raise awareness of the importance of taking adverse events into account when designing similar studies, and recognizing that these events may extend beyond study participants. For example, adverse events can occur simply due to receiving a recruitment email (or a vulnerability notification) without actually participating in the study, or to a third-party as a byproduct of an admin's study recruitment (*e.g.,* as in the case of a *non-participating* customer blocked by a service provider). Another factor is that large scale recruitment across the Internet spans different jurisdictions,[3] which potentially exacerbates the risk of adverse events.

Finally, with the growing intersection between large scale Internet measurements studies and user studies, it is becoming more important that IRBs account for established best practices for good Internet citizenship (*e.g.,* put forth by Durumeric *et al.* [22]) as part of the IRB clearance process.

VII. Conclusion

We conducted a survey with system administrators from 18 countries to understand reasons behind the persistent lack of vulnerability remediation that plagues the Internet. To recruit participants, we used Censys to find systems running vulnerable software versions, obtained the contact email addresses for these systems from WHOIS, and emailed them an invitation to our survey. We received 89 complete surveys overall. Our analysis shows that awareness of the existence of a vulnerability

[3]Of 156 surveyed countries, Diop *et al.* [18] found that only Albania and Philippines have legal articles "requiring" admins to remediate vulnerabilities.

does not affect remediation plans, which paves the way towards explaining the low remediation rates in previous literature. While most participants reported compatibility issues and third-party dependencies as the most influential factors, we found no evidence that eliminating a specific barrier is a panacea for the long tail of vulnerable systems. Many factors were discussed by participants, including backwards compatibility, lack of resources, the admin's technical knowledge of the vulnerability, importance of the at-risk assets to the owner/admin, limited knowledge of vulnerability tracking, internal company politics, benefit versus cost, limited resources, and maintenance of legacy systems. We believe these findings well complement previous literature on vulnerability remediation.

Acknowledgments

Assal and Abdou, each acknowledge funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) through their Discovery Grants. We thank Daniela Napoli and Quinn McGarry for their help in the initial stages of this research. Thanks to our participants for taking the time to participate in our study.

References

[1] A. Abdou, D. Barrera, and P. C. v. Oorschot, "What lies beneath? Analyzing automated SSH bruteforce attacks," in *International conference on PASSWORDS*. Springer, 2015, pp. 72–91.

[2] F. Alaca, A. Abdou, and P. C. van Oorschot, "Comparative Analysis and Framework Evaluating Mimicry-Resistant and Invisible Web Authentication Schemes," *IEEE Transactions on Dependable and Secure Computing (TDSC)*, vol. 18, no. 2, pp. 534–549, 2019.

[3] J. Aronson, "A pragmatic view of thematic analysis," *The qualitative report*, vol. 2, no. 1, pp. 1–3, 1995.

[4] H. Assal and S. Chiasson, "'Think secure from the beginning' A Survey with Software Developers," in *ACM CHI conference on human factors in computing systems*, 2019.

[5] S. Beattie, S. Arnold, C. Cowan, P. Wagle, C. Wright, and A. Shostack, "Timing the Application of Security Patches for Optimal Uptime," in *USENIX LISA*, 2002.

[6] C. Bennett, A. Abdou, and P. C. van Oorschot, "Empirical Scanning Analysis of Censys and Shodan," in *NDSS Measurements, Attacks, and Defenses for the Web (MADWeb) Workshop*, 2021.

[7] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano, "The quest to replace passwords: A framework for comparative evaluation of web authentication schemes," in *IEEE Symposium on Security and Privacy (S&P)*, 2012.

[8] Censys, "Company - Censys," https://censys.io/, 2022, online; accessed 15 February 2022.

[9] O. Cetin, C. Ganan, L. Altena, T. Kasama, D. Inoue, K. Tamiya, Y. Tie, K. Yoshioka, and M. van Eeten, "Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai," in *Network and Distributed System Security (NDSS)*, 2019.

[10] O. Cetin, C. Ganan, L. Altena, S. Tajalizadehkhoob, and M. Van Eeten, "Tell Me You Fixed It: Evaluating Vulnerability Notifications via Quarantine Networks," in *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2019.

[11] O. Cetin, C. Ganan, M. Korczynski, and M. van Eeten, "Make notifications great again: learning how to notify in the age of large-scale vulnerability scanning," in *Workshop on the Economics of Information Security (WEIS)*, 2017.

[12] O. Cetin, M. Hanif Jhaveri, C. Gañán, M. van Eeten, and T. Moore, "Understanding the role of sender reputation in abuse reporting and cleanup," *Journal of Cybersecurity*, vol. 2, no. 1, pp. 83–98, 2016.

[13] K. Charmaz, *Constructing grounded theory*. sage, 2014.

[14] J. Corbin and A. Strauss, *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications, 2014.

[15] L. Daigle, "WHOIS Protocol Specification," RFC 3912 (Draft Standard), Internet Engineering Task Force, Sep. 2004. [Online]. Available: http://www.ietf.org/rfc/rfc3912.txt

[16] Daniel Liberto, "Small and Mid-size Enterprise (SME)," https://www.investopedia.com/terms/s/smallandmidsizeenterprises.asp, 2011, online; accessed 14 January 2022.

[17] A. Decan, T. Mens, A. Zerouali, and C. De Roover, "Back to the Past–Analysing Backporting Practices in Package Dependency Networks," *IEEE Transactions on Software Engineering (TSE)*, 2021.

[18] S. M. Diop, J. D. Ndibwile, D. Fall, S. Kashihara, and Y. Kadobayashi, "To Coerce or Not to Coerce? A Quantitative Investigation on Cybersecurity and Cybercrime Legislations Towards Large-Scale Vulnerability Notifications," in *IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW)*, 2019.

[19] N. Dissanayake, A. Jayatilaka, M. Zahedi, and M. A. Babar, "Software security patch management - A systematic literature review of challenges, approaches, tools and practices," *Information and Software Technology*, vol. 144, p. 106771, 2022.

[20] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, "A search engine backed by Internet-wide scanning," in *ACM Conference on Computer and Communications Security (CCS)*, 2015.

[21] Z. Durumeric, F. Li, J. Kasten, J. Amann, J. Beekman, M. Payer, N. Weaver, D. Adrian, V. Paxson, M. Bailey *et al.*, "The matter of heartbleed," in *ACM Internet Measurement Conference (IMC)*, 2014.

[22] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications," in *USENIX Security Symposium*, 2013.

[23] T. Elmas, R. Overdorf, A. F. Özkalay, and K. Aberer, "Ephemeral astroturfing attacks: The case of fake twitter trends," in *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2021.

[24] A. Field, *Discovering statistics using IBM SPSS statistics*. sage, 2018.

[25] W. Findlay and A. Abdou, "Characterizing the Adoption of Security.txt Files and their Applications to Vulnerability Notification," in *NDSS Measurements, Attacks, and Defenses for the Web (MADWeb) Workshop*, 2022.

[26] R. A. Fisher, "On the interpretation of χ 2 from contingency tables, and the calculation of p," *Journal of the royal statistical society*, vol. 85, no. 1, pp. 87–94, 1922.

[27] E. Foudil and Y. Shafranovich, "A File Format to Aid in Security Vulnerability Disclosure," IETF Internet Draft (draft-foudil-securitytxt-12), Internet Engineering Task Force, May 2021. [Online]. Available: https://datatracker.ietf.org/doc/html/draft-foudil-securitytxt-12

[28] Internet Corporation for Assigned Names and Numbers—ICANN, "List of Accredited Registrars," https://www.icann.org/registrar-reports/accredited-list.html, 2022, online; accessed 15 February 2022.

[29] B. Ives, K. R. Walsh, and H. Schneider, "The domino effect of password reuse," *Communications of the ACM*, vol. 47, no. 4, pp. 75–78, 2004.

[30] A. Jenkins, P. Kalligeros, K. Vaniea, and M. K. Wolters, ""anyone else seeing this error?": Community, system administrators, and patch information," in *IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020.

[31] P. Jennifer, R. Yvonne, and S. Helen, "Interaction design: beyond human-computer interaction," *NY: Wiley*, 2002.

[32] J. Kawakita, "The original kj method," *Tokyo: Kawakita Research Institute*, vol. 5, 1991.

[33] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[34] M. Kührer, T. Hupperich, C. Rossow, and T. Holz, "Exit from hell? Reducing the impact of amplification DDoS attacks," in *USENIX Security Symposium*, 2014.

[35] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[36] F. Li, Z. Durumeric, J. Czyz, M. Karami, M. Bailey, D. McCoy, S. Savage, and V. Paxson, "You've got vulnerability: Exploring effective vulnerability notifications," in *USENIX Security Symposium*, 2016.

[37] F. Li, G. Ho, E. Kuan, Y. Niu, L. Ballard, K. Thomas, E. Bursztein, and V. Paxson, "Remedying web hijacking: Notification effectiveness and webmaster comprehension," in *Conference on World Wide Web (WWW)*, 2016.

[38] F. Li, L. Rogers, A. Mathur, N. Malkin, and M. Chetty, "Keepers of the machines: Examining how system administrators manage software updates for multiple machines," in *USENIX SOUPS*, 2019.

[39] S. Liu, I. Foster, S. Savage, G. M. Voelker, and L. K. Saul, "Who is. com? Learning to parse WHOIS records," in *ACM Internet Measurement Conference (IMC)*, 2015.

[40] M. Maass, M.-P. Clement, and M. Hollick, "Snail Mail Beats Email Any Day: On Effective Operator Security Notifications in the Internet," in *Conference on Availability, Reliability and Security (ARES)*, 2021.

[41] M. Maass, H. Pridöhl, D. Herrmann, and M. Hollick, "Best Practices for Notification Studies for Security and Privacy Issues on the Internet," in *Conference on Availability, Reliability and Security (ARES)*, 2021.

[42] F. Martius and C. Tiefenau, "What does this update do to my systems?– an analysis of the importance of update-related information to system administrators," in *SOUPS Workshop on Security Information Workers (WSIW)*, 2020.

[43] National Institute of Standards and Technology (2018), "National vulnerability database: Vulnerability metrics," https://nvd.nist.gov/vuln-metrics/cvss, 2018, online; accessed 22 January 2022.

[44] K. Pearson, "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[45] O. Pieczul, S. Foley, and M. E. Zurko, "Developer-centered Security and the Symmetry of Ignorance," in *New Security Paradigms Workshop (NSPW)*, 2017.

[46] V. L. Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Network and Distributed System Security (NDSS)*, 2019.

[47] T. Poteat and F. Li, "Who you gonna call? an empirical evaluation of website security.txt deployment," in *ACM Internet Measurement Conference (IMC)*, 2021.

[48] Qualtrics, https://www.qualtrics.com/, 2005, online; accessed 14 January 2022.

[49] Red Hat Customer Portal, "Security Backporting Practice," https://access.redhat.com/security/updates/backporting/, 2022, online; accessed 15 February 2022.

[50] RIPE NCC, https://www.ripe.net/support/abuse, 2020, online; accessed 18 January 2022.

[51] Shodan, "Shodan Search Engine," https://www.shodan.io/, 2022, online; accessed 15 February 2022.

[52] W. Soussi, M. Korczynski, S. Maroofi, and A. Duda, "Feasibility of large-scale vulnerability notifications after GDPR," in *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2020.

[53] J. Staddon and N. Easterday, ""it's a generally exhausting field" a large-scale study of security incident management workflows and pain points," in *IEEE Conference on Privacy, Security and Trust (PST)*, 2019.

[54] Statistics Canada, https://www150.statcan.gc.ca/n1/pub/11f0027m/2011069/part-partie1-eng.htm##archived, 2011, online; accessed 14 January 2022.

[55] B. Stock, G. Pellegrino, F. Li, M. Backes, and C. Rossow, "Didn't you hear me?–Towards more successful Web vulnerability notifications," in *Network and Distributed System Security (NDSS)*, 2018.

[56] B. Stock, G. Pellegrino, C. Rossow, M. Johns, and M. Backes, "Hey, you have a problem: On the feasibility of large-scale web vulnerability notification," in *USENIX Security Symposium*, 2016.

[57] The ZMap Team, "The ZMap Project," https://zmap.io/, 2022, online; accessed 15 February 2022.

[58] C. Tiefenau, M. Häring, K. Krombholz, and E. von Zezschwitz, "Security, Availability, and Multiple Information Sources: Exploring Update Behavior of System Administrators," in *USENIX SOUPS*, 2020.

[59] US Small Business Administration, "What's new with small business?" https://www.sba.gov/sites/default/files/Whats_New_With_Small_Business.pdf, 2016, online; accessed 19 January 2022.

[60] P. C. van Oorschot, *Computer Security and the Internet: Tools and Jewels*. Springer, 2020.

[61] M. Vasek and T. Moore, "Do Malware Reports Expedite Cleanup? An Experimental Study," in *USENIX Workshop on Cyber Security Experimentation and Test (CSET)*, 2012.

[62] M. Vasek, M. Weeden, and T. Moore, "Measuring the impact of sharing abuse data with web hosting providers," in *ACM Workshop on Information Sharing and Collaborative Security (WISCS)*, 2016.

[63] D. W. Woods and R. Böhme, "SoK: Quantifying cyber risk," in *IEEE Symposium on Security and Privacy (S&P)*, 2021.

[64] E. Zeng, F. Li, E. Stark, A. P. Felt, and P. Tabriz, "Fixing HTTPS misconfigurations at scale: An experiment with security notifications," in *Workshop on the Economics of Information Security (WEIS)*, 2019.

APPENDIX A
SURVEY QUESTIONS

1) Please enter the IP of a vulnerable host (hereon referred to as "The Host") we identified in our corresponding email (This IP address will only be used to connect the survey response to the network it came from, for removal data if requested, or to later check of the vulnerabilities were fixed):
   a) _____
   b) Prefer not to answer
2) Did we contact the appropriate person in your organization?
   a) Yes
   b) No (Why not? _____)
3) What is the size of your organization?
   a) At most 500 employees
   b) 501 to 5000 employees
   c) 5001+ employees
   d) Prefer not to answer
4) How many people are involved in addressing issues related to the Host's security/privacy vulnerabilities (hereon referred to as "The Remediation Team")?
   a) Just me
   b) 2 to 10 people
   c) 11 to 20 people
   d) 21+
   e) Not sure
   f) Prefer not to answer
5) Were you previously aware of the vulnerability we detected?
   a) Yes
   b) No
   c) Prefer not to answer
6) Have you already remediated, or previously attempted to remediate, the vulnerability we detected?
   a) Yes
   b) No
   c) Prefer not to answer
7) If you have not, do you plan to remediate the vulnerability we detected?
   a) Yes
   b) No
   c) Prefer not to answer
8) Have you previously received notifications from the other security researchers regarding potential vulnerabilities related to your network?
   a) Yes
   b) No
   c) Prefer not to answer
9) Please rate the importance of the following considerations that may contribute to The Remediation Team's reasoning to forgo remediating the vulnerability we have identified. (Note: Respondents rate the following statements on a 5-point Likert scale - "Not at all important", "Unimportant", "Neither important nor unimportant", "Important", "Extremely important", "Prefer not to answer". To counteract the potential ordering influence, the following statement were presented in random order per participant.)
   a) Cost of remediation outweighs risk
   b) Limited knowledge of vulnerability
   c) Limited knowledge of remediation process
   d) Issues impending the collaboration between The Remediation Team and other stakeholders
   e) Limited remediation tools
   f) Limited vulnerability tracking tools
   g) Third-party dependencies (e.g. hosting provider, certificate authority)
   h) Compatibility issues (e.g. backwards compatibility, legacy code, new libraries)
   i) Limited access to relevant resources that are not controlled by the Remediation Team (e.g. data, tools)
10) Are there any other considerations we did not mention?
   a) Yes
   b) No
11) If yes, Please describe the other considerations and provide an importance rating (e.g. "Not at all important", "Important", or "Extremely Important") for each additional consideration.
12) If we did not contact the appropriate person, please explain who this survey was forwarded to.
13) In general, how similar would your responses be if we asked these questions regarding other vulnerabilities the Remediation Team detects?
   a) Not at all similar
   b) Somewhat similar
   c) Similar
   d) Very similar
   e) Not sure
   f) Prefer not to answer
14) If you would like to elaborate on any of your responses in this survey, or have any other comments, please provide your feedback below. Please make sure not to include any personal information, or information which may give away your identity.

APPENDIX B
RECRUITMENT EMAIL EXAMPLE

Subject: "Vulnerability Notification for [IP address] and Remediation Survey"

Hello,

You are being contacted because your email address is registered as point of contact in the WHOIS data associated with [IP address]. This message is intended for this host(s) operator. Please share it with a relevant individual should you not be the appropriate contact.

We are security researchers from [Institution Name]. In our research we have used public search engines to scan the Internet for hosts with potential vulnerabilities. Our scan suggests that your host(s) may be effected by the vulnerability described at the end of this email.

The purpose of our research is to better understand reasons a network operator may be unable to address a vulnerability or choose to forgo remediation. To help us measure these factors, we invite you to participate in the following survey: [link to the survey].

We estimate our survey will take you 15 minutes to complete. Your responses will be anonymous and will shape future efforts to help operators, such as yourself, in managing the security of their network.

———————

Detected Vulnerability:

CVECODE: CVE-2014-3566

DESCRIPTION: The SSL protocol 3.0, as used in OpenSSL through 1.0.1i and other products, uses nondeterministic CBC padding, which makes it easier for man-in-the-middle attackers to obtain cleartext data via a padding-oracle attack, aka the "POODLE" issue.

CVSSSCORE: 3.4 LOW

NVDLINK: https://nvd.nist.gov/vuln/detail/CVE-2014-3566

———————

Again, we invite you to participate in our survey measuring the reasons a network operator may not remediate the vulnerability we've detected: [link to the survey]. Your contribution is greatly appreciated.

This research has been cleared by [Institution's Research Ethics Board].

If you have questions or concerns, or would like to opt-out of future studies, please respond to this email.

Sincerely,

[Researchers]