# "I believe it's incredibly difficult to fight against this flood of spam": Towards Enhancing Strategies for Creating Effective Vulnerability Notifications

ANNE HENNIG, MAXIME VEIT, LEONI SCHMIDT-ENKE, FABIAN NEUSSER, DOMINIK HERRMANN, and PETER MAYER

Identifying the most effective and scalable methods for notifying website owners about compromises or vulnerabilities remains an enduring challenge. Although some success factors have been identified, results regarding effective senders and notification framing are often inconsistent, and the understanding of how recipients perceive vulnerability notifications is still limited. Heading towards a better understanding, we conducted a $3 \times 3$ randomized controlled notification experiment, examining the impact of three distinct senders and three variations of notification framings for $n = 581$ compromised German websites. Our findings revealed a promising trend: receiving any notification significantly increased remediation compared to the absence of one. Remarkably, the choice of sender and framing played only a minor role in our notification experiment, which underscores the importance of notifying compromised websites and should motivate those who find vulnerabilities to take action. Yet, despite these encouraging results, a staggering 58% of the notified websites failed to remediate. To delve deeper into this phenomenon, we conducted follow-up interviews with 42 website owners who did not remediate their websites. The insights were revealing: while our notifications were delivered, many interviewees admitted they either overlooked or dismissed them as spam. This pattern persisted across different senders and framings, highlighting a critical challenge for future notification campaigns. Moving forward, future research should focus on finding ways to cut through the overwhelming amount of daily "spam" and explore strategies for how notifications can effectively convey their importance in recipients' inboxes. Exploring strategies to raise the general awareness for cybersecurity, encouraging website owners to provide a security.txt, or providing additional assistance in the form of a self-service tool, are some proposals to increase remediation rates. We further recommend that future work should consider theories from communication science or psychology, e.g., Protection Motivation Theory (PMT) or the Elaboration-Likelihood Model, when designing notification campaigns.

Additional Key Words and Phrases: website hacking, website compromise, notification experiment, website vulnerabilities, web security, redirect hack, awareness, vulnerability notification, interview study

**ACM Reference Format:**
Anne Hennig, Maxime Veit, Leoni Schmidt-Enke, Fabian Neusser, Dominik Herrmann, and Peter Mayer. 2025. "I believe it's incredibly difficult to fight against this flood of spam": Towards Enhancing Strategies for Creating Effective Vulnerability Notifications. 1, 1 (September 2025), 41 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

It was reported that in 2022, each website had to endure, on average, 172 attacks per day [61]. While some types of website attacks, e.g., defacement or Denial of Service (DoS) attacks, are immediately noticeable by website owners, other

Authors' address: Anne Hennig, anne.hennig@kit.edu; Maxime Veit, maxime.veit@kit.edu; Leoni Schmidt-Enke, leoni.schmidt-enke@kit.edu; Fabian Neusser, fabian.neusser@stud.uni-bamberg.de; Dominik Herrmann, dominik.herrmann@uni-bamberg.de; Peter Mayer, mayer@imada.sdu.dk.


Manuscript submitted to ACM

types, like cloaking[1] or unauthorized third-party redirect hacks[2], are designed to remain hidden [58]. Attackers may have compromised a system long before the attack is finally detected. Even experienced developers might struggle to identify and remediate such unauthorized third-party redirect hacks, as described in [7]. In a preliminary investigation, a wide variety of websites were found to be affected by these redirect hacks [46], with some cases remaining undetected for several months. This highlights the necessity to notify the website owners about the compromises. The primary goal of our study was to notify affected website owners and make them aware of their websites being compromised.

While searching for the most effective, yet feasible-to-scale method of notifying website owners about vulnerabilities, we found that previous research sometimes provides conflicting results, e.g., regarding the effect of the sender [42, 81]. Furthermore, most studies have focused on specific settings, such as using only Computer Emergency Response Teams (CERTs) [39] or hosting providers [80] as senders. The influence of incentives is also still unclear (see Section 2).

To address these limitations, our study is the first to combine factors from related work that positively affect remediation into a single study design. We are the first study that partnered with two German hosting providers and the CERT of the German Federal Office for Information Security (Federal CERT) to sent out notifications directly to website owners. We conducted a mixed-methods investigation, combining a notification experiment to examine the effects of sender and framing of a notification, with qualitative interviews to explore reasons for non-remediation. Based on the results of our experiment and the feedback from affected website owners, we provide design recommendations for practitioners and researchers involved in notification campaigns. Additionally, we outline a comprehensive path for future work in the area of vulnerability notifications.

Our work was guided by the four research questions described below. We motivate these in detail from the context of the related work in Section 2.

**RQ 1** [Framing] *Which framing has what impact on the remediation of compromised websites?*

We analyzed the effects of three incentives that previous work either proposed (reputational incentives [70, 77]) or investigated (technical incentives [79, 80], no incentives). Our goal, thereby, was to identify whether providing consequences, other than legal incentives which are only applicable in certain circumstances [42], has a significant effect on remediation. To the best of our knowledge, reputational incentives as well as a direct comparison of different incentives within a single study have not been investigated before. We did not find any significant differences in the remediation rates between the three framings, indicating that for vulnerability notifications, senders can choose any of the framings in our study, whichever they feel most comfortable with (Section 4.1).

**RQ 2** [Sender] *Which sender has what impact on the remediation of compromised websites?*

We investigated the effect of three senders (university, hosting provider, Computer Emergency Response Team (CERT) of the German Federal Office for Information Security (BSI)) in comparison to one another, which have only been investigated separately in distinct studies in previous work. Further, we directly partnered with contact persons in the respective entities to make sure that notifications are sent. Our goal, thereby, was to provide clear evidence on the effects of different senders. We found that all notified websites exhibited significantly higher remediation rates and fewer days to remediation than the unnotified control group. However, we did not find a significant difference between the three senders university, hosting provider, Federal CERT (Section 4.1), indicating that the sender appears to have no impact on remediation.

[1]Cloaking describes the technique of returning different versions of a website for search engine crawlers and users.
[2]Unauthorized third-party redirect hacks, as described in Section 2.1, are redirects placed on benign websites without authorization of the website owner to redirect users to malicious websites.

**RQ 3** [Framing × Sender] *Can we identify an interaction effect between sender and framing with respect to the remediation of compromised websites?*

We aimed to identify whether we can observe interaction effects between any of the sender and framing conditions in our study, e.g., whether a technical framing was more effective when sent out by the hosting provider. Previous work indicated such interaction effects, albeit for a combination of legal sender and legal framing [42, 69]. All senders sent out notifications for all framings. However, we did not find significant differences between the different framings when correlated with a corresponding sender, further supporting the notion that sending notifications is the crucial part, while framing and sender matter less (Section 4.1).

**RQ 4** [Reasons] *What are reasons for website owners to not remediate their websites even after being notified twice?*

As related work has revealed surprisingly low remediation rates, we additionally conducted qualitative interviews with website owners, rather than using static surveys, as in [18, 35, 39, 42, 63, 65, 73, 77] to gain in-depth insights regarding *reasons for non-remediation behavior* (in contrast to *remediation strategies* as analyzed in [55]). We identified a variety of reasons why website owners did not remediate the unauthorized third-party redirect hacks. Most prominent are the visibility of vulnerability notifications among other emails and the website owners' perceived relevancy of the compromise for their website (Section 4.2).

## 2 RELATED WORK

A wide range of work has been conducted in the area of communicating security risks, particularly notifying users about potential security issues (e.g., password breaches [45], password reuse [4, 25], data breaches [32, 75], identity theft [48], or IoT hacking [57], etc.). While these findings are relevant to our work, we will focus on notification experiments that explicitly notified *website owners* about potential attacks on their websites in the following. In particular, we focused on unauthorized third-party redirect hacks, as this compromise has not been extensively investigated, and we are unaware of any studies that have raised awareness for this compromise among website owners. In our experimental design, we focused on the effects of *sender* and *framing*, as well as *interaction effects* between the two. This approach was motivated by contradictory results or research gaps identified in the existing literature. Additionally, we collaborated with specific contact persons in the Federal CERT and the hosting companies to address the limitations of previous work [34, 36, 39, 50]. We agreed on a common cover letter that each sender used, and ensured that all framings were used by each sender, and all notifications were sent.

### 2.1 Threat Model

Unauthorized third-party redirect hacks, as investigated in this paper, involve hijacking legitimate websites to lead users to malicious sites, such as fake online pharmacies or fraudulent casinos [47]. These unauthorized third-party redirect hacks are stealth attacks, meaning they are designed to hide inside an otherwise benign system [12]. The redirects are only active when users access the website through search engine results that advertise the malicious websites. The original website can still be accessed normally, i.e., no redirect takes place. At first glance, the website may appear uncompromised, but when searched using the "site:" - operator, unusual entries are visible in the search results, redirecting the visitor to an illegitimate online shop (see Figure 1 for an example). This behavior is often referred to as (malicious) search engine Spam or SEO Spam [43], website hijacking [36], Pharma Hack [7], Japanese Hack [26], or WordPress Hack [67]. All these attacks describe different variations of the same underlying problem: the ranking of a legitimate website is maliciously used to redirect to other websites. However, in this paper, we specifically investigated

a variation in which the legitimate website is compromised. To avoid using a term with an unclear or misleading definition, we refer to the compromise as “unauthorized third-party redirect hacks”.

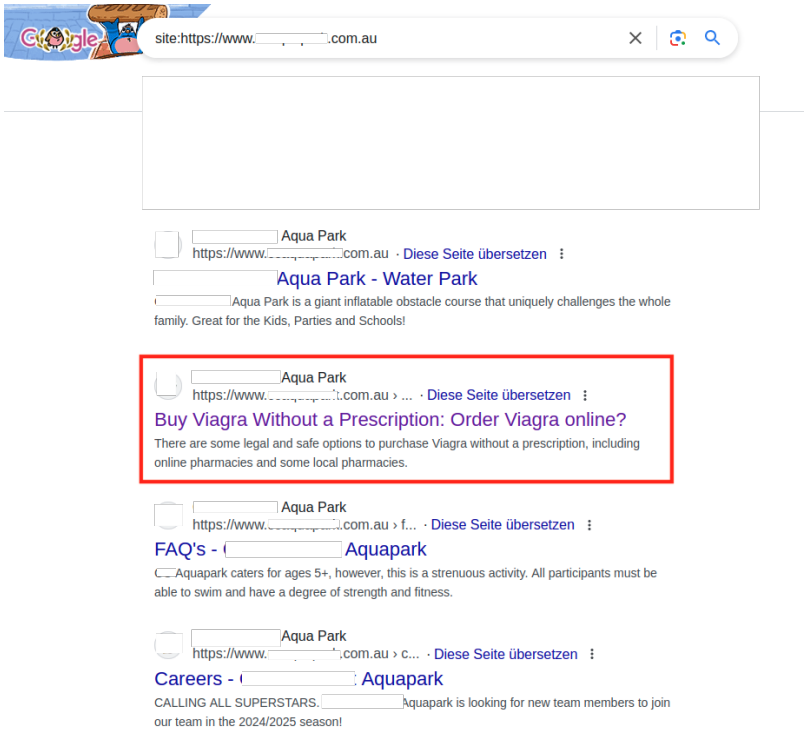



Fig. 1. Search results of a compromised website. While there are mostly legitimate results, the highlighted link would redirect the user to a fake pharmacy selling Viagra.

While seemingly “harmless”, these redirects can be seen as indicators of compromise, as mentioned in [33] and described in [27], since an attacker needs write access to the website data to place the unauthorized third-party redirect hack. Soska and Christin [62] describe risk factors for websites *before* they become malicious (e.g., for phishing or distributing malware), and based on our findings, we would add irregular search engine entries as such risk factors. As this specific unauthorized third-party redirect hack has not been analyzed on a large scale, there is only case-by-case evidence of how attackers infiltrated the systems and the extent of control they have (e.g., [7, 26, 27, 43, 67]).

### 2.2 Motivating Research Questions

*Notification Content and Framing.* There are several design factors, such as general notification design [63], translation of notification texts into native languages [36, 73], or variations in subject lines [73] that were studied in previous work, but showed no significant effects on remediation rates. On the other hand, prior research has shown that providing detailed information in the notifications [35, 36, 63, 70, 73, 81] as well as describing the security issue and highlighting its importance [73] increases the effectiveness of a notification.

Regarding the framing of the notification, Zeng et al. [73] found that pointing out consequences, i.e., framing the attack as more or less severe, has no impact on remediation rates. However, other authors suggested the opposite. Cetin

et al. [79, 80] could show that technical incentives, such as quarantining infected websites, can increase remediation rates. Maass et al. [42] found that providing legal incentives, in the form of fines, can significantly increase remediation rates. Reputational incentives, such as declining search rankings [70] or publicly naming compromised websites [77], have been proposed to increase remediation rates; however, their effectiveness has not been researched yet.

Due to conflicting results or missing research about incentives and message framing in the literature, we wanted to focus on this aspect in more detail. Our goal is to compare the effects of technical, reputational, and no incentives for the same problem within a single study, providing further insight into how to frame effective notifications. Thus arose our research question **RQ 1**: "*Which framing has what impact on the remediation of compromised websites?*"

*Sender of the Notification.* Cetin et al. [81] notified website owners via email from three different senders with varying reputations (an individual researcher, a university, and an anti-malware organization). Although the authors found a significant difference between the control group and the treatment groups, they were unable to find a significant difference between the three senders. The authors concluded that the remediation rate could neither be improved by choosing a certain sender, nor was the willingness to remediate affected by the sender's reputation [81]. Later studies could also not identify statistically significant differences between the respective sender groups (e.g., individual researcher vs. research group [63], researcher vs. Google Search Console [73], (inter)national CERTs [22, 34, 39], or hosting provider [9, 34, 55, 74, 78–80]).

In contrast, Maass et al. [42] showed that notifications with a legal framing sent from the university's law group led to significantly higher remediation rates, compared to a university computer science group, indicating that either sender or framing (or a combination of both) might make a difference. Since a vulnerability notification always requires a sender, who potentially influences the perception of the notification, our goal is to complement existing results by comparing different senders with high reputation that were used in previous studies but not yet evaluated in comparison to each other. This motivated our research question **RQ 2**: "*Which sender has what impact on the remediation of compromised websites?*"

Maass et al. [42] and Utz et al. [69] found that a framing referencing fines, ideally in combination with a sender that has the authority to impose such fines [42] leads to higher remediation. Thus, we also wanted to investigate whether we can transfer their results to other framing–sender combinations, identifying whether a specific sender is more successful with a specific framing, which could then be a further design factor for future notifications. Thus arose our research question **RQ 3**: "*Can we identify an interaction effect between sender and framing with respect to the remediation of compromised websites?*"

*Remediation Rates.* Previous research has shown that notifying website owners increases remediation rates compared to non-notification scenarios [18, 34–37, 40, 42, 63, 64, 70, 73, 77, 80, 81]. However, remediation rates, in general, turned out to be low. For website-related vulnerability notifications, Stock et al. [63] achieved a remediation rate of 24% when notifying website owners about publicly accessible Git repositories, and an even lower remediation rate of 17% when notifying about cross-site scripting vulnerabilities in WordPress. Maass et al. achieved a medium remediation rate of 56.6% ranging from 33.9% to 76.3% [42] depending on the condition[3] when informing website owners about missing IP anonymization while using Google Analytics. Zeng et al. reported a remediation rate between 7% and 34% when notifying website owners about different HTTPS misconfigurations.

[3]The authors compared twelve different conditions: university law group, university computer science group, and an individual researcher as senders; privacy, GDPR, and GDPR with fine as framings; as well as email and letter as notification channels.

Additionally, Zeng et al. [73] also observed and notified website owners about soon-to-be distrusted Symantec certificates. While in this case, notifications had no statistically significant effect compared to not notifying, remediation reached an outstanding rate of 90% after 40 days across all groups. Durumeric et al. [18] also reached a relatively high remediation rate of around 57% when notifying website owners about the infamous Heartbleed vulnerability. For both experiments, the authors acknowledged that the prominence of the respective issues had a huge impact on the high remediation rate [18, 73].

Stöver et al. [66] identified further reasons why website owners would not remediate the misconfiguration by analyzing the email and survey responses that Maass et al. [42] got during their notification experiment. Among the most prominent reasons were lack of awareness for the problem, lack of technical knowledge, or lack of resources, such as time. Other reasons included deliberate lack of maintenance, ambiguous responsibilities, and complex organizational structures, which slow down or hinder remediation processes. As a result, we recognized the need to not only measure remediation rates as an indicator for the success of our notification campaign, but also to identify reasons why website owners notified about unauthorized third-party redirect hacks remain inactive. In doing so, we aim to clarify the considerable variation in the remediation outcomes reported in the literature. This motivated our research question **RQ 4**: "*What are reasons for website owners to not remediate their websites even after being notified twice?*"

### 2.3 Related Work Informing Design Decisions

*Notification Channel.* While most studies used email notifications [18, 29, 34–36, 40, 42, 63, 64, 70, 73, 77, 80, 81], some also tried other channels, such as Google Search Console messages, which were not more effective than email [73]. In addition, letters and phone calls, as well as social media, performed better than email [63]. Letters have also proven to be more effective than emails in [42]. This results in the alternative channels having a slightly higher remediation rate, at the cost of manual effort to retrieve the addresses [63], print and envelope the letters, as well as the postage fees [42]. Although sending letters proved to be less efficient, manually retrieving email addresses seems to pay off. Previous studies have shown that email bounces can be decreased [63] and deliveries can be improved by manually retrieving [29, 42, 63] or automatically crawling [69] for email addresses compared to using WHOIS [63, 64, 73] or generic emails [64, 69, 81]. *Thus, we decided for email as the notification channel and manually collected email addresses from the websites' imprints.*

*Notification Content and Study Design.* We also derived specific design decisions for our notification text and study design from interviews with $n = 25$ website owners affected by unauthorized third-party redirect hacks, that were conducted prior to this study. The methodology is not described in this paper, as further information can be found in the original publication [30]. As both studies are based on the same type of unauthorized third-party redirect hacks, we found it necessary to draw our design decisions from our pre-study specifically.

During the interviews, most participants expressed a general distrust in vulnerability notifications. Asked to identify a suitable sender, only a few interviewees could intuitively name a sender they considered appropriate. Among specific senders that were regarded suitable, the police was named 12 times, hosting provider was named nine times, research facilities were named six times, and the Federal Office for Information Security (BSI) was named four times (multiple answers permitted). A summary of the results is provided in Appendix A.2, Table 3.

Regarding suitable notification channels, 17 interviewees identified email as the most appropriate. Email notifications are valued for speed, ease of use, and practicality. Two interviewees also stated that email would be the most logical notification channel for a digital problem. On the other hand, five interviewees explained they would be somewhat

suspicious of notifications via email, but no one explicitly refused to be notified via email. Phone calls were deemed most suitable by 11 interviewees, and two indicated that they would prefer either a letter or a dedicated web portal.

Concerning the content of a notification, the results supported previous research [35, 36, 42, 63, 70, 73, 81]. In our pre-study, interviewees especially emphasized that a clear description of the attack, a clear motivation for the notification, and, if applicable, instructions on how to solve the malicious redirect should be included in a vulnerability notification. Furthermore, providing contact information – such as a phone number or email address, a signature, a letterhead, or an imprint – helps the recipients verify the sender. Also, four interviewees said that they would appreciate a personalized salutation. Two interviewees stated that they pay attention to correct orthography and spelling, while two others requested a meaningful subject. All these factors would make a vulnerability notification credible and comprehensible, thereby raising awareness for the legitimacy of the described attack. But awareness does not automatically lead to remediation: Two interviewees stated that they deemed the described compromise negligible, even if the notification was credible.

*Implications for our Study Design.* We considered all the results from our pre-study and related work, as described above, in designing our main study (see Section 3.2 for details). Firstly, we defined suitable senders and notification channels for our notification experiment based on our pre-study [30]. Since the police was named as a suitable sender most frequently, we pursued avenues to cooperate with cybercrime divisions of the German police as senders for our notification experiment. However, this ultimately proved impossible (see Section 3.3) and we had to settle for the next three most frequently named options: university/research institution, hosting provider, and BSI. Secondly, we chose email as notification channel. Most of the interviewees either named email as a suitable notification channel or were not strictly opposed to email. We found this to be the most cost-effective solution and, therefore, the most likely to be relevant in practice. We also decided to retrieve email addresses from the imprint of the websites manually. Thirdly, we made clear note of the individual features that the interviewees emphasized as necessary for the notification text, e.g., a clear description of the attack, and used them in designing the notification text. Finally, as it was found that website owners did not take the unauthorized third-party redirect hacks seriously, we deemed it necessary to point out consequences.

## 3 METHODOLOGY

To run our experiment, we utilized data from a web crawling service that identified websites affected by unauthorized third-party redirect hacks (Section 3.1). We then implemented a technique to continuously monitor the remediation status of the identified websites independently of the crawling service to find the compromised domains (Section 3.1). We used the results of our pre-study [30], which complemented the best practices for vulnerability notifications from the literature, to make informed design decisions for our study design (Section 3.2).

Based on the findings of previous work, we designed a quantitative 3×3 randomized controlled notification experiment to investigate **RQ 1** – **RQ 3**, analyzing the effects of three different senders (university, hosting provider, Federal CERT) and three different framings (neutral, technical, reputation) (Section 3.2.1). To answer **RQ 4**, we designed qualitative follow-up interviews with website owners who had not remediated the unauthorized third-party redirect hack within 56 days after our initial notification (Section 3.2.2). The overall structure of our study design is shown in Figure 2.

We end this section by discussing our ethical considerations (Section 3.3), the limitations of our work (Section 3.4), and our data analysis methods (Section 3.5). As will be described in Section 3.3, all parts of our experiment, including notification texts, interview guideline, and notification process, have been approved by the ethics board of our university.

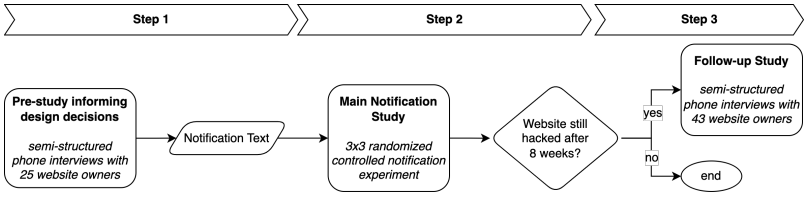



Fig. 2. General structure with all steps of our study design. Italic indicates the method used in each step.

### 3.1 Background

*Sampling Compromised Websites.* We obtained a dataset of websites affected by unauthorized third-party redirect hacks through a service specializing in large-scale web crawling [46]. No specific datasets (e.g., Tranco lists) were used to identify affected websites. Instead, the crawling identified affected websites based on public search engine results using keywords for the search. The service employed a multi-step process. First, they conducted a web search using a broad range of keywords known to be used to attract victims to malicious target websites. The keywords range from the shopping context, e.g., well-known brands, to topics such as Bitcoin or casino-related terms, and were defined by the crawling service (see Appendix A.1, Table 2 for a translated list of keywords of the websites in our sample). The URLs that were found, using the keywords-based search, were analyzed automatically to check known indicators of compromise, e.g., variations in website topics (like pharmaceutical content on a car seller's website), unusual phrases (like brand names of luxury watches on a school's website), page behavior with and without search engine identification, and other specific behaviors which the crawling service found typical for unauthorized third-party redirect hacks. Based on these indicators, the service compiled a list of potentially compromised websites every month throughout the notification experiment (i.e., 20 months). We received these monthly lists, which contained a minimum of 47 and a maximum of 926 worldwide domain names, along with the respective Google search results URLs. We used our monitoring system to check the compromised websites. We received no further demographic data for the websites (e.g., popularity, sector, or business size). We did not consider it useful to classify the websites ourselves (e.g., by industry branch or sales revenue), as these information were not available for all websites in our sample, and would, thus, mainly be drawn from the authors' assumptions based on publicly available information (e.g., the website or, if applicable, trade registers). Overall, the websites in the monthly lists were very diverse, comprising small and medium-sized enterprises, entrepreneurs, associations, projects, schools, or universities. For each month and each website, the presence of unauthorized third-party redirects was then verified by our monitoring system before the website was added to our sample.

*Monitoring Websites during Experiment.* We used a monitoring system to regularly check if the unauthorized third-party redirects remained active during our 56-day observation period for each website[4]. The monitoring system was implemented in Python, checking the presence of the redirects every six hours. It accessed each website by simulating a click from a Mozilla Firefox browser on Google search results, which were also provided by the crawling service, using the HTTP headers *Referer* and *User-Agent* respectively. We implemented several measures to counteract cloaking: a user agent string from a standard browser (Mozilla Firefox 97 on Windows 10, the most common variant found at the time of our study), an IP from a university client pool instead of a datacenter server IP, HTTP referrer header from search

[4]Note that since we obtained a new list every month, the start time for the observation period varied for all websites. However, each website was monitored for at least 56 days.

engines, and manual checks. We manually checked that our detection of the redirect worked as intended for each of the websites at the beginning of each website's 56-day observation interval. Our monitoring system then recorded whether a website's existing redirect behavior changed (e.g., a website's redirect disappeared due to the unauthorized third-party redirect hack being fixed).

If the website could not be accessed or access took more than 30 seconds, the access was retried four more times in quick succession before it was assumed that the website was taken down. The following rules were applied:

(1) It is assumed that the website is *still compromised*
   - if it contains the respective keyword, or
   - if there is a redirect to another website using HTTP redirect, JavaScript window location, or meta refresh.

(2) It is assumed that the website is *no longer compromised*
   - if an HTTP client or server error response indicates that the page has been removed, or
   - if the website has no page content, indicating that the malicious content has been removed, or
   - if the domain can no longer be resolved through the Domain Name System (DNS), indicating that the website has been taken down.

(3) If none of the above applies, no automatic decision can be made, and the monitoring system flags the website status as *unknown*, which triggers manual review.

### 3.2 Study Design Main Study

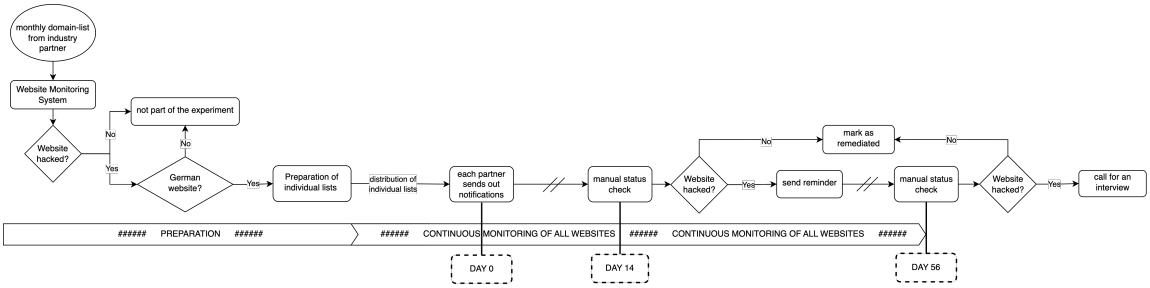



Fig. 3. Procedure RQ 1 - RQ 3.

*Notification Channel.* As described in Section 2.3, we decided to use email as the channel for our notification experiment. We purposely excluded other notification channels from our experiment, despite possible alternatives such as postal letters, social media, or phone calls [40, 63] potentially being more effective. None of these were practically possible, since none of the external senders identified as relevant in our pre-study (hosting provider and the Federal CERT) were able to integrate them into their processes for large-scale notification campaigns. Furthermore, as discussed by Stock et al. [63], the success rate of alternative communication channels was low compared to the effort and costs. It cannot, therefore, be considered cost-efficient, especially when notifying an even larger number of websites – Stock et al. [63] notified a total of 264 websites, 173 of them via alternative channels, including postal letters (67 websites), web forms (69 websites), social media (91 websites), and phone (46 websites). The results of our pre-study also supported this design decision. Thus, using email was the best choice, considering both the application of our research results at scale in practice and the notification channels accepted by website owners. To ensure that our email notifications are

delivered, we followed recommendations from previous work [29, 41, 63] and manually collected contact information from the imprint of each website. Specifically, we visited each website in our sample manually, without the aid of any automated script, and collected email addresses and the names of the persons responsible for the website from the imprint. Furthermore, we ensured that our email servers are configured according to state-of-the-art best practices, with valid SPF, DKIM, and DMARC records set to minimize the risk of our notifications being marked as spam. We purposively did not include any tracking mechanism (as Stock et al. [63] did to measure reachability) to decrease the likelihood that our emails would be filtered out.

*Notification Text.* We developed the text for our vulnerability notification based on the results of our pre-study and related work (see Section 2.3). We started with the most personal salutation possible, i.e. "Dear Mr./ Ms. [Lastname]". Next, we provided information on our motivation and on how to verify the unauthorized third-party redirect hack via the "site:"-operator in the search engine. We also explained that this operator will list all search engine entries of their domain, revealing the malicious redirects. Then, we either closed with a request to remediate the unauthorized third-party redirect hack and provided a link to our project website for further information, followed by our name, and a footer (neutral framing). Alternatively, we asked website owners to remediate the unauthorized third-party redirect hack, provided a link to our project website for further information, and subsequently included a technical or reputational incentive (technical or reputational framing). We also closed with our name and a footer that provided the sender's contact information. Since Maass et al. [42] proposed reminder notifications to increase awareness, we decided to send one reminder notification to those website owners who did not respond or remediate within two weeks after our initial notification. The reminder email had the same content as the original email, but the subject line was changed to "Reminder: Important information on your website *'domain'*". All notifications were sent only in German and addressed to German domains (see Section 3.2.1). Both the text and the framings were reiterated with the hosting providers and the Federal CERT to ensure that we use a wording that all senders can identify with. We provide the translated text and the framings in Appendix B.1 and B.2.

*Framings.* Regarding consequences, previous studies suggested to either provide technical incentives (like quarantining a vulnerable website at least temporarily [78, 80]) or pointing out negative consequences (like legal prosecution [42] or reputational damage [70, 77]) to raise awareness. However, the effectiveness of reputational framing, as well as the different framings in comparison to each other, has not been analyzed yet. To test whether – and if so which – incentives can increase remediation rates, we compared three different framings that could either be used by all senders or were phrased to fit a particular sender (see Table 1 for an overview of the different groups, and Appendix B.2 for the wording of the framings): (1) a neutral framing with no incentives (`neutral`); (2) a technical framing stating that the website can be blocked by the hosting provider (`technical-generic`), the website will be suspended after a specific date (`technical-hoster1`), or further actions will be taken by the hosting provider (`technical-hoster2`); (3) a reputational framing stating that the website suffers reputational damage (`reputation-generic`), or that web reputation services might suspend the website based on negative reputation scores (`reputation-CERT`). While we did not explicitly consider behavioral theories when designing the framings, the wording is guided by Protection Motivation Theory (PMT). All framings are inspired by related work, although none of them is exactly like previous ones, as the content needed to be adapted to our specific type of compromise. Furthermore, the technical and reputational consequences had to be iterated with the hosting providers and the Federal CERT in order to find a phrasing that all senders could approve of on behalf of their institution. In particular, the decision to announce the take-down of websites was suggested by the respective hosting provider, as this is part of their regular notification process. Note that

| Sender | Framing | # notifications |
|---|---|---|
| Federal CERT | neutral | 73 |
| | technical-generic | 70 |
| | reputation-CERT | 69 |
| | **sum** | **212** |
| hosting provider | neutral | 54 |
| | technical-hoster 1 & 2 | 54 |
| | reputation-generic | 55 |
| | **sum** | **163** |
| university | neutral | 68 |
| | technical-generic | 68 |
| | reputation-generic | 69 |
| | **sum** | **206** |
| control | – | **205** |

Table 1. Final number of notifications sent per sender and framing incl. control group.

all senders sent out notifications using each of the framings, contrary to Lone et al. [39]. Based on the remediation for each of these framings, we can answer **RQ 1**.

*Sender.* To answer **RQ 2**, we partnered with dedicated contact persons from IONOS and GoDaddy, two major hosting providers in Germany, as well as the BSI (German Federal CERT), serving as senders. Additionally, we sent notifications ourselves for the university condition. In total, we had four senders, i.e., one person in each of the four entities. The same person always sent the notifications to achieve homogeneity within the groups. All senders can also be related to one of the framings, similar to [42], where the legal framing was related to the law group. Hosting providers can be associated with technical framing due to their oversight of the infrastructure on which the compromised websites are hosted, and their resulting technical authority. The Federal CERT can be related to a reputational framing, as it maintains public lists of malicious websites and could potentially add affected websites to these lists. The university relates to the neutral framing, with neither apparent technical nor reputational authority. Each of the three senders sent out email notifications using each of the three framings (see Table 1 for an overview of the different groups), allowing us to answer **RQ 3**. Some websites were not notified as a control condition to allow a baseline comparison of the notifications (see Section 3.2.1).

*3.2.1 Procedure **RQ 1** – **RQ 3**: Framing & Sender.* The notifications were sent out monthly over 20 months between April 2022 and November 2023. Each month, we processed a completely new list obtained from the crawling service, totaling 20 lists with between 47 and a maximum of 926 domains from around the world per month. Websites that were already detected in any of the previous months or that had already been called (see Section 3.2.2) were excluded from the sample. Websites that were already remediated, non-German, or lacked valid contact information were also excluded from the sample. The procedure each month was as follows (see also Figure 3).

(1) We obtained a domain list with vulnerable websites detected in the previous month from the crawling service. This domain list was then uploaded to the monitoring system to test all websites for their most recent status (see Section 3.1).
(2) All websites showing indicators of unauthorized third-party redirect hacks were manually checked for language and headquarters by the authors. For German websites, email addresses and recipients' names were then manually collected from the imprint. Websites from other countries were not included in the sample.
(3) Websites with valid contact information were then allocated semi-purposively to individual lists for each of our partners. The two hosting providers were assigned their own customers based on their Autonomous System Number (ASN). The remaining websites were then assigned to the Federal CERT, the university, or the control

group. Note that websites, where it seemed critical to inform them about the vulnerability (i.e., kindergartens, schools, hospitals, doctors, lawyers) were never allocated to the control group.

(4) Finally, each website from the treatment groups was randomly assigned to one of the three framings. The framings were distributed evenly to all senders throughout the study. Meaning that if we could not evenly distribute the three framings to the number of domains we found for each partner in one month (i.e., when we had a number of domains that could not be divided by three), we then balanced it out in the following month.

(5) The individual lists, as well as further instructions, were sent to our contact persons at the hosting companies and the Federal CERT, respectively. Email notifications were sent by all senders every Wednesday in the third week of each month (Day 0). Two weeks later (Day 14), a reminder email was sent to websites that were still compromised.

*3.2.2 Procedure **RQ 4**: Reasons.* To answer **RQ 4**, we, as researchers from the university, called website owners who were still compromised eight weeks after the first notification email was sent out (Day 56). In the initial phone call, we invited them to a follow-up interview to get more in-depth information on why they had not remediated. Again, we processed each month's list individually to ensure that we do not contact website owners from any previous list. The phone number was retrieved from the imprint or the contact information given on the website. Website owners were contacted a total of no more than three times by one of the authors if we were unable to reach them on the first call. In the call, we specifically asked for the person who received and processed our notification. If the person did not agree to an interview, we briefly noted down what they told us regarding our notification (see Appendix C.3 for the categories we used to code the phone calls).

If the person agreed to participate in an interview, a return call was scheduled. We then conducted semi-structured telephone interviews. Note that all interviews were conducted in German as it was the first language for both participants and the interviewer. The ethics committee of our institution approved the interview guideline. Informed consent was obtained before the interviews. The participants were not compensated, and the interviews took 15 minutes on average, depending on the amount of information the interviewees recalled (see Appendix C.1 for the translated interview guideline, including informed consent). All interviews were recorded and manually transcribed verbatim without using any transcription AI, while anonymizing personal data in the process. Only the anonymized transcripts were analyzed. For coding the interviews, we applied open coding with three coders. The lead researcher developed a first draft of the codebook based on the interview guideline. After that, three of the authors independently coded three interviews and then met to discuss new codes and disagreements within existing codes. Thus, while the development of the initial codebook followed a deductive approach, further codes were added in an inductive approach. This process was repeated twice. After 23% of the interviews had been independently coded by all three coders, and an IRR of $\kappa = 0.82$ was reached in the last iteration, the remaining interviews were coded independently by only two coders. However, all three coders met regularly to discuss any questions that arose. In the end, all codings were checked by the lead author.

### 3.3 Ethical Considerations

All parts of our study were approved by the ethics committee of our institution. There were no ethical concerns about the permissibility of our research. In designing our study, we carefully considered any critical factors reported in related work. Firstly, neither we nor the crawling service we used exposed the websites or web servers to any risk, e.g., through performing any form of attack (in contrast to, e.g., Wu et al. [72]). We only sampled websites that showed indicators of being compromised as explained in Section 3.1.

Secondly, we carefully discussed the necessity and permissibility of a legal framing, i.e., a reference to legal obligations to remediate the hacking. As mentioned in the context of the Princeton study [49, 68], legal framings can cause fear and anxiety among recipients, and may lead to anger about the notification and mistrust in the senders. Since legal framings, i.e., referring to the General Data Protection Regulation (GDPR) and potential fines, had proven most effective in other notification studies [42, 69], we discussed with several legal entities and law enforcement agencies (e.g., several cybercrime divisions of the polices in different states, a university law group, and contact persons at the Federal CERT) the feasibility of a legal framing similar to the one used by Maass et al. [42]. However, due to German legislation, this was impossible. There were only vague grounds for holding affected websites liable in case of unauthorized third-party redirect hacks, and prosecution varied significantly across the federal states. Therefore, we deliberately abandoned this condition – even if it limits the generality of our results – so as not to put legal pressure on the recipients or cause them unnecessary stress.

Another potential issue was the processing of personal data, including names, phone numbers, and email addresses. It was not possible to obtain consent for processing this personal data in advance without contacting participants at least once. Doing so would have compromised the integrity of our experiment. For contacting the affected websites, we used only publicly available data. In our email notification, we linked to our project website to connect the notification with the research project. A proper debriefing took place when we were able to reach the website owners via phone. Informed consent was obtained from all interviewees (see Appendix C.1). During the interviews, the participants were not exposed to any physical or mental risk at any time. While all personal data were deleted at the end of the experiment at the latest (e.g., personal data in the transcripts was removed during transcription), the domain names of the websites are retained for ten years as part of our sample documentation, in accordance with guidelines for proper scientific practice.

### 3.4 Limitations

*Internal Validity.* The internal validity of our results may be compromised by sampling bias. As described in Section 3.1, compromised websites were identified by searching for specific keywords. This approach disregards websites that might be compromised by unauthorized third-party redirect hacks, but cannot be identified via the selected keywords. To mitigate this bias, the crawling service we used updated its crawling model to include new keywords over time. We must acknowledge that we did not have control over the keywords used for the crawling, which means that there are likely affected websites that were not found. However, this would only affect the potential sample size and would not impact our results. As described in Section 3.5, we required a sample size of 400 websites for our statistical analysis, which we successfully achieved.

We took countermeasures against cloaking before we added the websites to our sample (see Section 3.1). Nevertheless, there is a potential risk that attackers could have enhanced their cloaking technique during the 56-day observation period, enabling them to evade our countermeasures. This would result in false positives, where "compromised" websites are erroneously assumed to be "not compromised" by our monitoring system. However, random manual checks indicated that such behavior is unlikely to have occurred; therefore, we are confident that this has not affected our results.

Another limitation concerns our follow-up interviews. Our sample only consists of people who answered our calls *and* agreed to conduct an interview. Website owners who chose to participate in our interviews might systematically differ from those who did not, potentially leading to self-selection bias. Thus, our interview data likely represent a small and non-representative sample of website owners. Furthermore, our data might be affected by recall biases. However, we reached code saturation and gained a wide range of answers, allowing us to identify trends. Thus, we are confident

in the value of our data. Furthermore, we only contacted website owners when the unauthorized third-party redirect hack was *not* fixed. In turn, we did not interview website owners who fixed the unauthorized third-party redirect hack. While interviewing website owners who had fixed the unauthorized third-party redirect hack could have provided interesting results, it was out of scope for this experiment, and we leave it to future work to address this.

By ensuring that our email servers had valid SPF, DKIM, and DMARC records, we addressed potential issues related to reachability. With our sample being diverse in terms of receiving mail servers we are confident that we did not encounter the problems with one service, like Google, filtering most of our notifications, as described in Stock et al. [63]. However, there is still the general risk that recipients email spam filters rejected our notifications. Reputation monitoring systems or spam blocking lists are helpful to check the reputation of the sender's email address [41]. As we wanted the experiment to be integrated into the regular notification process of the Federal CERT and the hosting providers, we had no control over their mail server configurations and could not subscribe to any spam reporting service. We acknowledge that potential differences could have an impact on deliveries itself as well as delivery rates and time (also affecting the appearance of our notifications in recipients' inboxes), which might have influenced whether the notification was read.

However, we can be sure that these institutions regularly check the reputation of their email addresses. Our notifications were treated the same as any other of their usual communications. We recorded if one of our emails bounced and also asked our partners to report emails that bounced. We are aware of a total of two emails that could not be delivered, indicating a negligible number of emails experiencing technical delivery issues. We only checked whether our university address and the CERT's address are listed on Spamcop.net and Spamhaus.org after completing our experiment, but no problems were found with either address. There is still a high probability that our notifications did not pass the "human" spam filter, as our results in Section 4.2 show. Unfortunately, we cannot quantify how many notifications might have been discarded as spam by the recipients.

*External Validity.* The non-representativeness of our sample limits the external validity of our findings. We could only include websites that were discovered in the web crawling. Furthermore, our notification experiment is exclusively focused on German websites, constraining its external validity. The findings might not be generalizable to other geographical locations, as websites may operate under different legal, cultural, or technological environments. However, we wanted to focus on only one country first to avoid introducing additional variables to our experiment (e.g., legal requirements in different countries, policies of CERTs or hosting providers, or translation of the notification). Furthermore, as Maass et al. [42] noted, restricting notification campaigns to a specific country has the advantage that notifications are better understood and the names of senders are somewhat familiar, which increases trust in the organization sending the notification. Nevertheless, replicating our study in other countries and for other regulatory and cultural contexts represents an important direction for future work to confirm that our results are valid for a global audience as well.

Lastly, regarding the interaction effect between sender and framing, we cannot be certain that our participants share our view of relating the Federal CERT to the reputational framing, hosting providers to the technical framing, and universities to the neutral framing, which might have affected the effectiveness of our framings.

### 3.5 Data Analysis

For our notification experiment, we measured and controlled two independent variables (framing and sender) with three nominal characteristics each (neutral, technical, reputation, and university, hosting provider, Federal CERT). The

existence of third-party redirects was continuously measured by our monitoring system four times a day throughout the observation period, as defined in Section 3.1 (see also Figure 3). Thus, we measured remediation as our dependent variable as days until remediation (continuous).

We analyzed the differences between the framings (**RQ 1**), and the differences between the senders, and between the senders and the control group (**RQ 2**), using a single-factor ANOVA. To identify a possible correlation between sender and framing, we used a two-factor ANOVA (**RQ 3**). Additionally, we used survival analysis to compare the time until remediation between our treatment groups. For feasibility reasons, we only interpreted the results of the automatically determined status once a day. Therefore, we decided to use right-censoring and Kaplan-Meier estimator, as in previous studies where survival analysis was used [39, 42, 73, 78, 81], instead of, e.g., interval censoring.

We used an alpha level of .05 for all statistical tests, and applied post-hoc Holm-Bonferroni correction to counter alpha error cumulation. We also used a priori power analysis to determine the estimated sample size for our statistical analysis. We found that by assuming a medium effect size, we would need 280 websites to run a single-factor ANOVA with four groups (university, hosting provider, Federal CERT, control), and 400 websites to run a two-factor ANOVA with six groups (three senders, three framings). In the absence of similar rules for survival analysis, we aimed for a minimum sample size of 280 websites and attempted to collect up to 400 websites. With a final sample size of $n = 467$ websites, we surpassed that threshold.

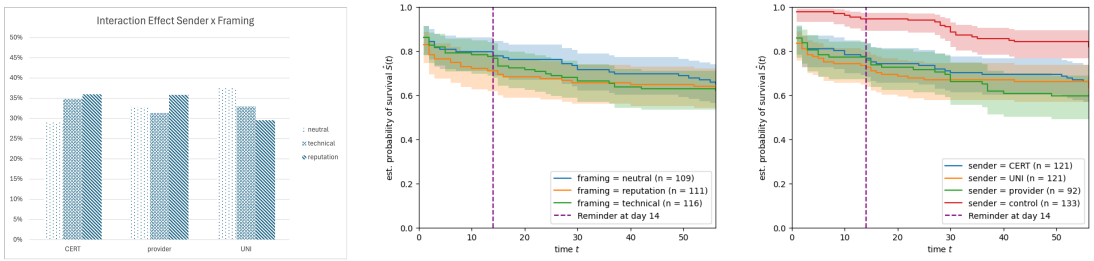
## 4 RESULTS

Between April 2022 and November 2023, we notified 581 website owners about unauthorized third-party redirect hacks. We did not notify 205 website owners who were in our control group. In total, two emails bounced and were not delivered. These were added to the control group. We are not aware of any other delivery failures and, therefore, assume that all our notifications reached their intended destinations. In total, our sample included 786 websites. Table 1 provides the number of notifications sent out in each group. The overall remediation rate for our notification experiment was 42.0%, which means that 58.0% of websites were still compromised eight weeks after two notifications.

Our sample also included websites with an unknown status, for which our monitoring system could not determine a status (see Section 3.1). If the monitoring system returned "unknown", we manually checked the websites periodically, but not daily. Thus, for these websites, we only have a few manually verified data points within the 56-day observation period. To measure the remediation in days as continuous variable we excluded the websites for which we could not clearly define the status or recorded a change in status over the 56-day period, and created a sub-sample for our longitudinal analysis containing $n = 467$ websites.

### 4.1 RQ 1 – RQ 3: Effect of Framing and Sender

To determine the effects of **sender** on the time to remediation as a continuous variable, we first used one-way ANOVA with the sub-sample ($n = 467$). University, hosting provider, and Federal CERT did not show any outliers in the box-plot-diagram, but the control group had 16 extreme outliers, which seemed reasonable to us: Contrary to the otherwise long time to remediation – or non-remediation until the end of the observation period – these cases were indeed remediated and sometimes even relatively quickly. Data were not normally distributed (Shapiro-Wilk test, $p < .001$), but ANOVA has proven robust against violations of the assumption of normality [8, 24, 28, 38, 59]. There was no homogeneity of variance (Levene's test, $p < .001$), and we, thus, interpreted the results of Welch's test. For **sender** we had defined four categories: control group, university, hosting provider, and Federal CERT. The mean time to remediation was 51.61 days (95%-CI[49.30, 53.91]) for the control group, 39.83 days (95%-CI[35.41, 44.26]) for university,




(a) Percentage of remediated websites. (b) Kaplan Meier curve for framing. (c) Kaplan Meier curve for sender.

Fig. 4. RQ1 - RQ3: Effect of Framing and Sender.

39.61 days (95%-CI[34.77, 44.45]) for the hosting provider, and 41.93 days (95%-CI[37.75, 46.11]) for the Federal CERT. We observed that the time to remediation was significantly different for at least one group, but only with a small effect, Welch's F(3, 230.87) = 16.68, p < .001, $\eta^2$ = .054. There were no statistically significant differences between the three treatment groups (p > .05). Still, each sender was significantly different from the control group (Games-Howell post-hoc analysis, p < .001: university - Control: -11.77, 95%-CI[-18.31, -5.24], Federal CERT - Control: -9.68, 95%-CI[-15.93, -3.42], provider - Control: -12.0, 95%-CI[-19.02, -4.98]). Applying the Holm-Bonferroni correction confirmed these findings (see Appendix D, Table 4 for all results).

Second, we performed a two-way ANOVA to assess the effects of **framing**, **sender**, and the **interaction effect** between framing and sender on the time to remediation as continuous variable. For framing, there were no outliers, as indicated by the box-plot diagram. In none of the categories, the data were normally distributed (Shapiro-Wilk test, p < .001), but again, we can assume robustness against violations of the assumption of normality, since our sample size was larger than 15 domains for each group. We also determined homogeneity of variances using Levene's test, which showed that equal variance could be assumed (p = .054). We excluded the control group as sender from our data, as we only wanted to determine effects if a notification was sent ($n$ = 334). For **framing** we had defined three categories: neutral, technical, and reputation. The mean time to remediation was 42.59 days (95%-CI[38.24, 46.95]) for the neutral framing, 40.05 days (95%-CI[35.74, 44.36]) for the technical framing, and 39.02 days (95%-CI[34.32, 43.72]) for the reputational framing. Interestingly, as shown in Figure 4a, the reputational framing resulted in the highest remediation rates for Federal CERT as sender, and the neutral framing resulted in the highest remediation rates for university as sender. However, our assumption that there is a general interaction effect between the framing and sender was not valid for the hosting provider as sender, where both the reputational and the neutral framing resulted in higher remediation rates than the technical framing. Furthermore, our main model was not significant, F(8, 325) = 0.53, p = .834, and neither of our framings (neutral, technical, reputation), F(2, 325) = 0.09, p = .916, nor any of our senders (university, hosting provider, Federal CERT), F(2, 325) = 0.41, p = .661, or the interaction between framing and sender, F(4, 325) = 0.75, p = .559, were significant.

Third, we conducted survival analysis using the sub-sample ($n$ = 467). The websites that were notified with a reputational **framing** were estimated to remediate with a mean of 39.02 days (95%-CI[34.39, 43.65]), requiring the least time until remediation compared to a technical framing (M = 40.05, 95%-CI[35.81, 44.3]), or a neutral framing (M = 42.59, 95%-CI[38.25, 46.93]). However, the results of the log-rank test showed that survival distributions do not differ significantly, $\chi^2(2) = 0.299$, p = .861 (see Figure 4b).

We also analyzed the effect of **sender** within our survival analysis. The websites that the hosting providers notified were estimated to remediate with a mean of 39.61 days (95%-CI[34.86, 44.36]), requiring the least time until remediation, whereas websites that were notified by the university (M = 39.84 days, 95%-CI[35.48, 44.2]), or the Federal CERT (M = 41.93 days, 95%-CI[37.76, 46.12]) took longer. All treatment groups required significantly fewer time until remediation than the control group (M = 51.61 days, 95%-CI[49.33, 53.88]). The log-rank test confirms that significant differences exist between at least two of the four groups, $\chi^2(3) = 17.49$, $p < .001$. Pairwise post-hoc log-rank tests showed statistically significant differences in the survival distributions of the control group and the Federal CERT, $\chi^2(1) = 10.63$, $p = .001$, the control group and the provider, $\chi^2(1) = 15.79$, $p < .001$, as well as the control group and the university, $\chi^2(1) = 13.05$, $p < .001$. However, there were no significant differences in the survival distribution of the treatment groups (see Figure 4c).

### 4.2 RQ 4: Reasons for Non-Remediation

Approximately eight weeks after our initial notifications, we contacted the website owners who had not remediated their websites. Between October 2022 and December 2023, we called 316 website owners via phone. We managed to reach 210, of whom 42 agreed to an interview. While not all agreed to an interview, most website owners provided some information in the phone calls, e.g., whether they received our notifications or already knew about the attack, which we also noted down as codes (see Appendix C.3 for an overview of the codes we used). Thus, we were also able to analyze the responses we received during the phone calls. See Figure 5 for an illustration of the process and Figure 6 for the results.

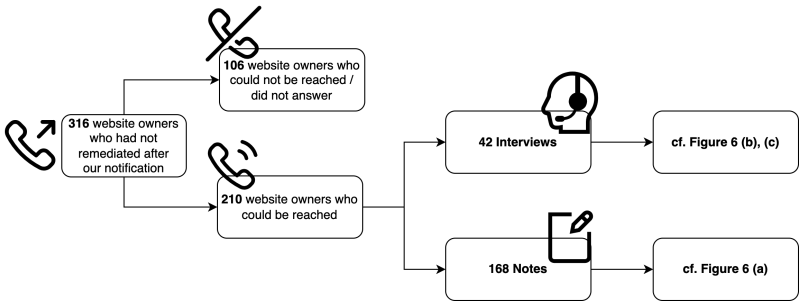



Fig. 5. Illustration of the Process of the Follow-up Interviews.

Of those who had not agreed to an interview ($n = 168$), 75 website owners said they were not aware of our email notification. In contrast, 26 website owners said they got some information, and eight already knew about the attack before our notification. We were unable to collect further information from 51 website owners, who either would not speak with us (14 website owners) or did not respond as promised during the initial call (37 website owners). Instead, in our interviews ($n = 42$), we mainly talked to website owners who had received our notification: 28 interviewees had received and read our notification, two website owners said they had not received our notification, and 12 did not remember if they received it.

All 14 interviewees who did not – knowingly – receive and, thus, read our notification stated that they overlooked our email, probably because it got buried among the daily spam. Or, as [P29] summarizes: *"Well, I have to be honest and say that I think a message like this [...] comes so unprepared and you receive such an abundance of emails every day*

*suggesting that something needs to be done in some way, um, that I think it's incredibly difficult to fight against [...] this flood of spam and it's, therefore, incredibly difficult to actually convey the seriousness that it needs to be received and read [...] at all."* P[36] also explained the email address we wrote to was too general, and it probably got lost due to vacation: *"You sent it to a rather general address. [...] Maybe someone was on vacation, I don't know, but in any case it went through."* P[29] further added that the email address we used belongs to an old website and is only infrequently checked: *"[I]t is our former company account, [which] no longer exists, and [...] is only checked very rarely."* Some participants also mentioned that the subject line we used did not help the email to stand out. P[7] explained: *"'Important information' sounds a bit more like... I get a lot of e-mails like that, where it says something like 'very important, very important'. That actually tends to get thrown out, because most of the time [...] it's just the opposite, when it somehow says 'super important' or something like that and then I click on it and then I see, ok, that's just something that's not important at all."*

Even if they had knowingly received and read the notification, 11 interviewees said they regarded the notification as spam and, thus, did not immediately react to it. P[4] said the sender was not familiar (*"I'd say that the average user, who sells his cake on the internet, won't know the CERT anyway"*). P[37] mistrusted Federal CERT as sender, questioning their motivation, and found this kind of notification unfamiliar: *"So, [...] cert@BSI didn't mean anything to me."* P[22] also said that a federal office scanning websites at no cost and informing users about vulnerabilities seemed unrealistic to them: *"Or CERT [...] this seems implausible itself. Because, I think, why should some irrelevant website be of interest [...]? You know[...] in those offices, where they usually have other things to do, [...] and then they additionally write emails to people [...] So why should they do that? [...] [A]nd then I thought, no, not really, that can only be spam."*

In addition, nine interviewees did not consider the unauthorized third-party redirect hack relevant enough to take immediate action. The main reason, given by three interviewees, was that the website has no priority for them. P[4] said they did not have time to fix the unauthorized third-party redirect hack immediately and then forgot about it: *"To be honest, I haven't had time to take care of it. It sounds a bit weird that we as a web-agency don't take care of our website, but we're so busy with work that we never really actively advertise on the website. So it wasn't a big priority for me at first."* P[7] felt that troubleshooting is more expensive than just launching a new website, so they did not react: *"[...] if I hire a web designer or a programmer who spends, I don't know, 3 days searching for something, it will cost me more than building a new website."* Two interviewees said they had not understood the problem, and four could not reproduce it. P[20] could not reproduce the problem because they did not read the advice on how to verify the unauthorized third-party redirect hack carefully enough. Two interviewees stated that they had checked their files but could not identify anything suspicious. Two interviewees admitted that they had understood the problem and found it relevant enough to act on, but did not know how to remediate, and 11 interviewees thought the attack was already remediated. We provide relevant quotes from the interviews in Appendix E, Table 5.

## 5 DISCUSSION AND FUTURE WORK

As discussed in [63] three key parameters contribute to the success of notification campaigns: (1) successful delivery of the notifications; (2) trust in the notification process; and (3) enhancing recipients' ability to remediate.

We must assume that our manual efforts in retrieving email addresses from the imprints of the websites were worthwhile, as the majority of notifications seemingly reached recipients' inboxes. This is supported by the fact that only two notifications bounced[5]. However, we observed that the second and third parameter, "reaching out" and "breaking through", were significant challenges in our notification experiment.

[5]As described in Section 3.2, we purposively did not include further tracking mechanisms, so we can only assess delivery success based on the number of bounces.

### 5.1 Reaching out: Notification Channel and Sender

We learned that a significant number of website owners, who did not remediate the unauthorized third-party redirect hack within 56 days, could not recall receiving our notification. Some interviewees mentioned that the contact information we used was too unspecific or outdated, which diminished the reliability of our email notifications. One alternative could be to reach out to those affected via alternative notification channels, such as phone calls – the second most preferred notification channel identified in our pre-study [30]. However, as highlighted in previous research [42, 63], and confirmed during our follow-up interviews, calling website owners is time-consuming and costly, making it unsuitable for large-scale notification campaigns.

Another reason why some email notifications went unnoticed or were regarded as spam was the lack of familiarity with the sender, distrust regarding the sender's intention, or simply the fact that they were notified at all. When we asked for recommendations for suitable senders, we found that opinions varied greatly among the interviewees. Website owners indicated a preference for notifications from a trustworthy and reputable sender but could not provide specific examples. We assumed that the Federal CERT, as also investigated in [39, 50], or the respective hosting providers are perceived as reputable senders. However, even when websites were notified of these trusted entities, our experiment demonstrated that remediation is not significantly higher than if the notifications are sent by the university. This, thus, confirms that the perception of a sender's trustworthiness depends highly on the individual recipient, as already stated in Hennig et al. [31]. However, since all our treatment groups performed significantly better than the control group, we can support previous research [18, 34–37, 40, 42, 63, 64, 70, 73, 77, 80, 81] in that sending out notifications does indeed encourage remediation.

We encourage future research to investigate these findings in the light of the Elaboration Likelihood Model (ELM) of persuasion, a theoretical model from communication science in the context of media effects research. We can consider our notification as a stimulus in the form of a piece of persuasive information that should motivate the recipients to change their attitude or behavior, i.e., remediate their vulnerable websites. The ELM proposes two "routes" by which recipients interact with the information (e.g., in a notification) based on the level of elaboration that is stimulated. If deemed convincing, the notification leads to a change in attitude or behavior, i.e., results in the remediation of the described issue. First, when elaboration is high, e.g., if the recipient has a high need for cognition and is motivated to interact with the notification, the information is processed on the *central route*. Here, the content and its arguments are closely scrutinized, and the information is carefully examined. However, based on what we learned in our follow-up communication with the recipients, we have to assume that vulnerability notifications via email are processed on the *peripheral route*, where elaboration is low and information is mainly processed based on heuristic principles, e.g., credibility heuristics where the sender's perceived expertise is assessed. Peripheral cues, i.e., external factors such as the perceived credibility of the sender, have a particularly strong effect on the information processing via the *peripheral route*, as they serve to assess whether a change in attitude or behavior is necessary.

In the context of vulnerability notifications, this leaves researchers and practitioners with two challenges. First, external circumstances should be improved so that the recipient is empowered and motivated to process a notification directly on the *central route*. Taking sender and notification channel as an example, elaboration motivation could be stimulated by personal relevance (e.g., choosing a sender that has a personal relationship to the participant), and elaboration ability could be stimulated by sufficient prior knowledge about the topic or a non-distracting environment (e.g., choosing a notification channel that is exclusively used to address tech-savvy contacts directly, as opposed to a universal notification channel that receives many irrelevant messages and is mainly managed by non-tech-savvy contacts

who do not expect to handle critical vulnerability notifications). Future work should use the ELM to systematically identify external factors in the context of vulnerability notifications that increase elaboration motivation and elaboration ability.

RFC 9116 [60] proposes a file format (security.txt) in which dedicated contact information for technical contacts are provided so that vulnerability notifications can be sent in a "non distracting environment". Unfortunately, related work has shown that the adoption of security.txt is alarmingly low [19, 52]. Yet, we still lack a fundamental understanding of the reasons behind this hesitance among website owners to provide such contact information. To address this gap, we recommend future work to directly engage with website owners, explore potential misconceptions or obstacles they face, and identify resources website owners need. Furthermore, it might be beneficial to raise awareness for the importance of providing such contact information while fostering a broader understanding of website security within a comprehensive awareness campaign. Future work might evaluate the impact of concepts like a "Web Security Awareness Month", which could serve as a rallying call for better practices in the community. Organizations such as chambers of commerce, industry associations, national CERTs, and external service providers – including hosting and content management platforms like WordPress – are in a unique position to lead these efforts. By coming together to offer comprehensive information, workshops, or webinars focused on website security, they can create a powerful network of support and knowledge-sharing, inspiring website owners to embrace best practices and enhance their security posture. Based on the recommendations of Gerber et al. (2025) [23], website owners may find the problem and its remediation less daunting if they feel supported and have the opportunity to connect with others, which leads to an increase in website owners' self-efficacy and consequently, increases their ability to remediate.

Future work should also investigate how theoretical foundations from the context of organizational information security awareness campaigns (see, e.g., Bada et al. [6] for an overview), could shape the organization of and the communication around such events, as has been explored in related but distinct contexts, such as Social Network Analysis to select security champions in companies [15], or Transactive Memory System Theory (TMS) to facilitate sharing IT security knowledge within a group [3]. Further factors that influence information security behavior of a website owner, such as job characteristics, personality traits, or real-life exposure [21], should also be investigated to identify different types – or personas – of website owners. This is helpful information for ultimately tailoring such events to different target groups and making them as effective as possible for participants.

Second, peripheral cues have to be researched in more detail. This is especially important as it has to be considered that, depending on the context, every cue – or factor – has different effects. A sender's reputation, for example, might affect the degree of elaboration, i.e., whether the information is processed on the central or the peripheral route. Or it may affect decision-making within each of the routes, i.e., serves as a peripheral cue within the peripheral route, to motivate the recipient to take the notification as credible based on their credibility heuristic, or influences the valence of elaboration within the central route by increasing personal relevance. In general, every factor that accompanies the communication can be a peripheral cue. For communication in (mass) media, factors such as characteristics of the speaker (e.g., voice, presentation style, appearance, prominence, etc.), characteristics of the situation (e.g., environment in which a message is presented), characteristics of the message itself (e.g., length, repetition), or personal characteristics of the recipient (e.g., mood, distraction, attitude towards a message or the sender, attitude of other persons towards the message, etc.) are described as influencing factors [11, 51, 71]. It would be interesting to know which of these factors are especially effective within risk communication, i.e., vulnerability notifications, in contrast to other persuasive communication, such as advertisements.

Our results regarding suitable channels and senders are twofold. Firstly, since none of the senders in our experiment proved to be most effective, we recommend that practitioners designate a single authority, like the Federal CERT, to handle security or privacy-related vulnerability notifications. This entity should be promoted nationally, possibly through partnerships with business associations to amplify its reach. Secondly, for future research on notification campaigns, we recommend exploring additional strategies to raise a general awareness among website owners for cybersecurity, and encourage them to provide accurate security contacts (e.g., security.txt). Although email scales well for notification campaigns, it should be accompanied by efforts to increase general awareness of security notifications to ensure they stand out amidst other emails. Within the context of the ELM, enhancing background knowledge and the relevance of a topic positively impacts elaboration ability and motivation, thereby increasing the likelihood that a notification is processed via the *central route*.

### 5.2 Breaking through: Framing and Notification Content

We discovered that website owners who did not remediate often did not consider the notification relevant enough to take (immediate) action against the unauthorized third-party redirect hacks. Interestingly, in contrast to previous research (e.g., [42]), we see that trust in the notification *content* was less of an issue, a finding supported by our remediation rate. Once the notification was considered relevant, a significant percentage (42%) of website owners remediated the unauthorized third-party redirect hacks. However, as described above, it appears the notifications did not convey sufficient relevance to motivate elaboration and, thus, to surpass the daily "flood" of unsolicited messages. This was also echoed in our interviews: Website security was not a main priority for the website owners we interviewed. Some were unable to gauge the extent of the attacks and were often reluctant to invest more resources than absolutely necessary in website maintenance, especially considering that it is typically not their bread and butter.

Providing incentives to raise awareness for the problem's severity did not significantly enhance the effectiveness of our notifications. Neither of the two framings with incentives (reputation, technical) was significantly more effective. We observed a slight advantage of the reputational framing in terms of time to remediation compared to the technical or the neutral framing, but the differences were not significant. We also examined whether a framing sent from an authoritative source – one that could impose fines for non-compliance, as suggested by [42] – might enhance the effectiveness of our vulnerability notifications. In the absence of a legal framing in our experiment, we investigated the effect of other framing–sender combinations (i.e., technical framing from a hosting provider, and reputational framing from the Federal CERT). Our study did not uncover any correlation between sender and framing, making it impossible to generalize the results from Maass et al. [42] and Utz et al. [69] to other framing–sender combinations.

While we could not generalize results from Cetin et al. [79, 80], who found that technical incentives in the form of quarantining compromised domain owners is effective, or Maass et al. [42] and Utz et al. [69] who proposed that certain framings are only effective in combination with a corresponding sender, we were able to confirm results from Zeng et al. [73] in that framings have no (additional) effect on remediation. As both of our framings with incentives were only slightly more effective than the framing without any incentive, we do not observe a significant effect of a certain type of notification being discarded as spam more often (e.g., notifications with a neutral framing being discarded more frequently than those with a reputational framing). Perhaps the wording of the technical and reputational framings were not explicit enough to motivate the recipients to remediate and protect themselves more than was the case with the neutral framing.

Possible explanations can be derived from the Protection Motivation Theory (PMT) [56]. According to PMT, reactions towards warnings are based on two appraisal processes: the *coping appraisal* and the *threat appraisal*. In the process of *threat appraisal* individuals assess their perception of a threat, based on three factors: perceived severity, which refers to the expected degree of harm caused by the threat, perceived vulnerability, which indicates the likelihood of experiencing that harm, and maladaptive rewards, meaning the aspects that reinforce insecure behaviors. For individuals to feel motivated to react to a warning, the perceived severity and perceived vulnerability must outweigh any maladaptive rewards. In our notification experiment, we aimed to increase the expected degree of harm and enhance participants' perceived severity by using technical and reputational incentives. However, it is possible that the incentives we provided were not strong enough to effectively influence participants' perceptions.

Some researchers propose that increasing the perceived risk associated with vulnerabilities, or threatening website owners with public disclosure if they do not remediate, could provide stronger stimuli to evoke action and improve remediation rates significantly [35, 63]. Based on our results, we would argue against employing threatening incentives. One interviewee proposed that notifications should be addressed to browser vendors, allowing them to directly block malicious websites in the respective browsers and, thereby, alert website owners about the severity of the unauthorized third-party redirect hack. While such warnings can effectively attract the website owner's attention and significantly increase remediation, they must be accompanied by additional information about the compromise and clear remediation instructions, as noted by Li et al. [36]. Furthermore, stimulating threat perception must not lead to fear arousal [44, 76]. Research has shown that heightened pressure and concern may lead to refusal and mental overload, as observed in contexts such as passwords [5, 16] or cybersecurity incidents [17, 53]. Related work has also shown that security and privacy topics in general are perceived as complex and frightening by users, leading to people feeling overwhelmed and frustrated [14, 23, 54]. Instead of stimulating negative emotions that might inhibit recipients to take action, recent work recommends to frame security and privacy topics as "more engaging and enjoyable" [23]. In the context of vulnerability notifications, prior results [31, 39] also propose that positive reinforcement – such as emphasizing that remediation reduces the threat to others – may be more effective in motivating website owners to remediate compromises.

In the context of the PMT, positive reinforcement can also take place in the process of *coping appraisal*, where individuals assess their ability to cope with a threat based on three factors: self-efficacy, response efficacy, and response costs. To motivate protective actions, e.g., remediation in our context, Mayer et al. [44] describe that self-efficacy and response efficacy must outweigh response costs. In the context of vulnerability notifications this means that the resources the recipients need to spent for remediation (e.g., time, knowledge, or costs for staff or external support) must be outweighed by the possibilities they have to remediate (e.g., access to the vulnerable system, technical knowledge), and a perceived high probability that the vulnerability will be removed after remediation. Thus, another reason our framings were not effective is that our notification did not stimulate the recipients' ability to cope with the threat sufficiently to offset the expected costs.

Reasons mentioned by website owners for not remediating included the belief that they had already remediated, or that they were unsure about the appropriate solution. This points to transparency issues, where individuals are not able to perceive the status of their system. Some interviewees expressed a desire for more information such as a PDF attachment or links to further resources. Previous studies also indicated that providing a self-service tool or detailed reports alongside a notification can be valuable [42, 63, 64, 77, 81]. Furthermore, some of our interviewees suggested that notifications should include offers of support, such as specific assistance for a fee or help from the Federal CERT. This indicates a distinct need to understand the systems individuals use and regain control over it. Future work in this regard should investigate creating additional materials based on cognitive and behavioral theories, such as

Protection Motivation Theory (PMT), which was, for example, used to design interventions that should encourage users to change their passwords after a data breach [76], or Theory of Planned Behavior (TBP) [1, 2], as was, for example, used by Bulgurucu et al. [10] who investigated the intention of employees to comply with security policies in their company. Specifically, factors like "attitude" (TBP), which was found to have a reliable medium effect on increasing secure information security behavior [44] and a large effect on security policy compliance [13], "self-efficacy" (PMT, TBP), which is a factor in both theories and has proven to have a reliable weak [44] to medium positive effect [13] especially in combination with "controllability" [44], as well as "subjective norms" (TBP) and "response efficacy" (PMT) (reliable weak positive effect [44], medium to large effect [13]), should be investigated to address the issues and needs identified in our interviews.

Another possible reason why our incentives had no (additional) effect might be habituation effects. As several interviewees mentioned, they often receive emails that urge them to take action. Therefore, any incentive that demands action might be dismissed as yet another spam message. However, this underlines the importance of sending notifications without any additional call-to-action, and helping recipients to distinguish legitimate vulnerability notifications from those with marketing interests.

It might be worth discussing whether a larger sample size could reduce the likelihood of underestimating potential differences between our treatment groups (type II error) in our statistical analysis. However, our findings revealed statistically significant differences between the control group and the treatment groups, and we met the required sample size determined by a priori power analysis. Therefore, simply increasing our sample size would not alter our findings in meaningful ways. Nonetheless, as discussed in Section 3.4, replicating our study with a more diverse sample that includes websites from different geographical regions and legal contexts would enhance the robustness of our results.

Rather than using threats to compel website owners to remediate compromises, we recommend supporting website owners in their coping appraisal by encouraging them, offering support, and including references to further information in a broader awareness campaign. Additionally, we suggest sharing information materials with third parties, such as CERTs, hosting providers, or browsers (e.g., Google Safe Browsing), which can provide further support or distribute warnings. Future research should develop materials based on theoretical foundations such as Protection Motivation Theory or Theory of Planned Behavior and explore whether supplementary materials enhance remediation or foster mistrust. We also recommend investigating encouraging framings for future work. Our notification will be freely available to facilitate its use in practice[6], but we advice practitioners to use it without any specific framing.

## 6 CONCLUSION

In our study, we focused on unauthorized third-party redirect hacks, which are not easily recognizable by non-experts. While the websites display normal content when accessed directly, infected websites will list unusual and eventually malicious URLs in search engine results. This is not only harmful to a website's reputation, but it also indicates that an attacker gained write access to the website. The goal of our research was to combine results of previous work and, based on qualitative interviews with 25 website owners conducted in a pre-study [30], develop an effective notification process. We then conducted a quantitative $3 \times 3$ randomized controlled notification experiment, measuring differences

[6]Before publication, we will revise our notification template based on the feedback from the interviews.

between senders that were deemed suitable by the interviewees (i.e., university, hosting provider, and Federal CERT), and different framings that should incentivize website owners to remediate. Between April 2022 and November 2023 we notified 581 website owners via email, observing an additional 205 that were in our control group. We found that only 42.0% of websites remediated within 56 days. We can confirm previous research, which also found that we could not identify a sender or framing that was more effective than others [63, 73, 81]. It may be that, for example, a legal framing or a sender with legal authority might be more effective; however, the use of these conditions should be carefully considered in light of potential ethical issues. Nevertheless, remediation rate was significantly better in the treatment groups compared to the control group, indicating that notification campaigns are effective in increasing remediation.

In addition to notifying website owners via email, we called them when the unauthorized third-party redirect hack was not remediated within 56 days. We conducted 42 qualitative follow-up interviews to find reasons for non-remediation. The main reason for non-remediation was that website owners could not recall receiving our notification or regarded it as spam. Thus, we found increasing the perceptibility of notifications to be a major issue. We recognize a general need to raise awareness about security notifications and the provision of proper security contacts by website owners, especially for those whose websites are not their primary business. We also recommend that future work investigates how email notifications can stand out from the mass of daily emails in recipients' inboxes, e.g., by embedding notifications in a broader awareness campaign, by analyzing framings that encourage recipients to open an email, or by following up with website owners who have remediated the problem to investigate reasons for remediation. We also found that providing additional support, e.g., in the form of a self-service tool or a PDF attachment, might increase recipients self-efficacy and encourage website owners to take action.

We hope that our results and the recommendations derived from them can guide researchers in finding more effective ways to notify website owners of compromised websites. We make our revised notification texts freely available online to help practitioners when notifying the victims of compromises.

## ACKNOWLEDGMENTS

This research is supported by the German Federal Ministry of Education and Research as part of the INSPECTION project (Zuwendungsnummer 16KIS1113), and by funding from the topic Engineering Secure Systems, topic 46.23.01 Methods for Engineering Secure Systems, of the Helmholtz Association (HGF) and by KASTEL Security Research Labs. Special thanks to Alexandra Pawelek, Elly Reich, Lauritz Kanyi, and Miriam Mutter who helped at different stages of this research as part of their jobs as student assistants.

## REFERENCES

[1] Icek Ajzen. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 2 (1991), 179–211. Issue 50. https://doi.org/10.1016/0749-5978(91)90020-T

[2] Icek Ajzen. 2002. Perceived Behavioral Control, Self-Efficacy, Locus of Control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology* 4 (2002), 665–683. Issue 32. https://doi.org/10.1111/j.1559-1816.2002.tb00236.x

[3] Saad Alahmari, Karen Renaud, and Inah Omoronyia. 2023. Moving beyond cyber security awareness and training to engendering security knowledge sharing. *Information Systems and e-Business Management* 21 (2023), 123–158. Issue 2023. https://doi.org/10.1007/s10257-022-00575-2

[4] Yusuf Albayram and Jaden Walker. 2024. Investigating Effectiveness of Informing Users About Breach Status of Their Email Addresses During Website Registration. *International Journal of Human–Computer Interaction* 0, 0 (2024), 1–20. https://doi.org/10.1080/10447318.2024.2404721

[5] Nora Alkaldi and Karen Renaud. 2018. Encouraging Password Manager Adoption by Meeting Adopter Self-Determination Needs (Extended Version). *SSRN Electronic Journal* (2018). https://doi.org/10.2139/ssrn.3259563

[6] Maria Bada, Angela Sasse, and Jason Nurse. 2015. Cyber Security Awareness Campaigns: Why do they fail to change behaviour? 118–131.

[7] BitofWP. 2019. WordPress Infected with the Pharma Hack? How to Detect, Clean and Secure your site from it - DEV Community. https://dev.to/bitofwp/wordpress-infected-with-the-pharma-hack-how-to-detect-clean-and-secure-your-site-from-it-4fja. [last accessed 2024-04-19].

[8] M. J. Blanca, R. Alarcón, R. Arnau, J.and Bono, and R. Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* 29, 4 (2017), 552–557. https://doi.org/10.7334/psicothema2016.383

[9] Brennen Bouwmeester, E.R. Turcios Rodriguez, Carlos Gañán, Michel van Eeten, and Simon Parkin. 2021. "The thing doesn't have a name": Learning from emergent real-world interventions in smart home security. In *Proceedings of the 17th Symposium on Usable Privacy and Security, SOUPS 2021 (Proceedings of the 17th Symposium on Usable Privacy and Security, SOUPS 2021)*. USENIX Association, 493–512.

[10] Burcu Bulgurcu, Hasan Cavusoglu, and Izak Benbasat. 2010. Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness. *MIS Quarterly* 34, 3 (2010), 523–548. http://www.jstor.org/stable/25750690

[11] Roland Burkart. 2021. *Kommunikationswissenschaft*. Boehlau Verlag.

[12] Lorena Cazorla, Cristina Alcaraz, and Javier Lopez. 2018. Cyber Stealth Attacks in Critical Information Infrastructures. *IEEE Systems Journal* 12, 2 (2018), 1778–1792. https://doi.org/10.1109/JSYST.2015.2487684

[13] W. Alec Cram, John D'Arcy, and Jeffrey G. Proudfoot. 2019. Seeing the forest and the trees: a meta-analysis of the antecedents to information security policy compliance. *MIS Q.* 43, 2 (June 2019), 525–554. https://doi.org/10.25300/MISQ/2019/15117

[14] Joseph Da Silva and Rikke Bjerg Jensen. 2022. "Cyber security is a dark art"': The CISO as Soothsayer. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 365 (Nov. 2022), 31 pages. https://doi.org/10.1145/3555090

[15] Duy Dang-Pham, Siddhi Pittayachawan, and Vince Bruno. 2017. Applications of social network analysis in behavioural information security research: Concepts and empirical analysis. *Computers & Security* 68 (2017), 1–15. https://doi.org/10.1016/j.cose.2017.03.010

[16] Marc Dupuis, Anna Jennings, and Karen Renaud. 2021. Scaring People is Not Enough: An Examination of Fear Appeals within the Context of Promoting Good Password Hygiene. In *Proceedings of the 22nd Annual Conference on Information Technology Education (SIGITE '21)*. Association for Computing Machinery, New York, NY, USA, 35–40. https://doi.org/10.1145/3450329.3476862

[17] Marc Dupuis and Karen Renaud. 2021. Scoping the ethical principles of cybersecurity fear appeals. *Ethics and Information Technology* 23, 3 (2021), 265–284. https://doi.org/10.1007/s10676-020-09560-0

[18] Zakir Durumeric, Frank Li, James Kasten, Johanna Amann, Jethro Beekman, Mathias Payer, Nicolas Weaver, David Adrian, Vern Paxson, Michael Bailey, and J Alex Halderman. 2014. The Matter of Heartbleed. In *Proceedings of the 2014 Conference on Internet Measurement Conference, IMC '14 (IMC '14)*. Association for Computing Machinery, Vancouver, BC, Canada, 475–488. https://doi.org/10.1145/2663716.2663755

[19] William Findlay and AbdelRhaman Abdou. 2022. Characterizing the Adoption of Security.txt Files and their Applications to Vulnerability Notification. *Proceedings of the 2022 Workshop on Measurements, Attacks, and Defenses for the Web* (2022). https://doi.org/10.14722/madweb.2022.23014

[20] Thomas Franke, Christiane Attig, and Daniel Wessel. 2018. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction* 35, 6 (2018), 456–467. https://doi.org/10.1080/10447318.2018.1456150

[21] Steven Furnell and Anish Rajendran. 2012. Understanding the influences on information security behaviour. *Computer Fraud & Security* 2012 (2012), 12–15. Issue 3. https://doi.org/10.1016/S1361-3723(12)70053-2

[22] Oliver Gasser, Quirin Scheitle, Carl Denis, Nadja Schricker, and Georg Carle. 2017. Security Implications of Publicly Reachable Building Automation Systems. In *2017 IEEE Security and Privacy Workshops (SPW)*. 199–204. https://doi.org/10.1109/SPW.2017.13

[23] Nina Gerber, Verena Zimmermann, Alexandra von Preuschen, and Karen Renaud. 2025. Unpacking the Social and Emotional Dimensions of Security and Privacy User Engagement. In *Proceedings of the 21th Symposium on Usable Privacy and Security, SOUPS 2025 (Proceedings of the 21th Symposium on Usable Privacy and Security, SOUPS 2025)*. USENIX Association.

[24] Gene V Glass, Percy D. Peckham, and James R. Sanders. 1972. Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research* 42, 3 (1972), 237–288. https://doi.org/10.3102/00346543042003237

[25] Maximilian Golla, Miranda Wei, Juliette Hainline, Lydia Filipe, Markus Dürmuth, Elissa Redmiles, and Blase Ur. 2018. "What was that site doing with my Facebook password?": Designing Password-Reuse Notifications. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (Toronto, Canada) *(CCS '18)*. Association for Computing Machinery, New York, NY, USA, 1549–1566. https://doi.org/10.1145/3243734.3243767

[26] Paul Goodchild. 2024. Rectifying Google Rankings: A Primer on Japanese Keyword Hack Recovery. https://getshieldsecurity.com/blog/japanese-keyword-hack/. [last accessed 2024-09-05].

[27] Stephan Halder. [n. d.]. Website Redirects im Umfeld von Fake Webshops und SEO Fraud. https://www.bdo.de/de-de/insights/aktuelles/assurance/website-redirects-im-umfeld-von-fake-webshops-und-seo-fraud. [last accessed 2025-02-14].

[28] Michael R. Harwell, Elaine N. Rubinstein, William S. Hayes, and Corley C. Olds. 1992. Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics* 17, 4 (1992), 315–339. https://doi.org/10.3102/10769986017004315

[29] Anne Hennig, Heike Dietmann, Franz Lehr, Miriam Mutter, Melanie Volkamer, and Peter Mayer. 2022. "Your Cookie Disclaimer is Not in Line with the Ideas of the GDPR. Why?". In *Human Aspects of Information Security and Assurance (HAISA 2022) (IFIP Advances in Information and Communication Technology, Vol. 658)*. Springer, Cham, 218–227. https://doi.org/10.1007/978-3-031-12172-2_17

[30] Anne Hennig, Fabian Neusser, Aleksandra Alicja Pawelek, Dominik Herrmann, and Peter Mayer. 2022. Standing out among the daily spam: How to catch website owners' attention by means of vulnerability notifications. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 317, 8 pages. https://doi.org/10.1145/3491101.3519847

[31] Anne Hennig, Nhu Thi Thanh Vuong, and Peter Mayer. 2023. Vision: What the hack is going on? A first look at how website owners became aware that their website was hacked. In *Proceedings of the 2023 European Symposium on Usable Security* (Copenhagen, Denmark) *(EuroUSEC '23)*. Association for Computing Machinery, New York, NY, USA, 312–317. https://doi.org/10.1145/3617072.3617101

[32] Yue Huang, Borke Obada-Obieh, and Konstantin Beznosov. 2022. Users' Perceptions of Chrome Compromised Credential Notification. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 155–174. https://www.usenix.org/conference/soups2022/presentation/huang

[33] Patchmuthu Ravi Kumar, Perianayagam Herbert Raj, and Perianayagam Jelciana. 2019. A Framework to Detect Compromised Websites Using Link Structure Anomalies. *Advances in Intelligent Systems and Computing* (2019), 72–84. https://doi.org/10.1007/978-3-030-03302-6_7

[34] Marc Kührer, Thomas Hupperich, Christian Rossow, and Thorsten Holz. 2014. Exit from Hell? Reducing the Impact of Amplification DDoS Attacks. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 111–125. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/kuhrer

[35] Frank Li, Zakir Durumeric, Jakub Czyz, Mohammad Karami, Michael Bailey, Damon McCoy, Stefan Savage, and Vern Paxson. 2016. You've Got Vulnerability: Exploring Effective Vulnerability Notifications. In *25th USENIX Security Symposium (USENIX Security 16)*. 1033–1050. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/li

[36] Frank Li, Grant Ho, Eric Kuan, Yuan Niu, Lucas Ballard, Kurt Thomas, Elie Bursztein, and Vern Paxson. 2016. Remedying Web Hijacking: Notification Effectiveness and Webmaster Comprehension. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Montreal, Quebec, Canada, 1009–1019. https://doi.org/10.1145/2872427.2883039

[37] Frank Li, Lisa Rogers, Arunesh Mathur, Nathan Malkin, and Marshini Chetty. 2019. Keepers of the Machines: Examining How System Administrators Manage Software Updates. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 273–288. https://www.usenix.org/conference/soups2019/presentation/li

[38] Lisa M. Lix, Joanne C. Keselman, and H. J. Keselman. 1996. Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance F Test. *Review of Educational Research* 66, 4 (1996), 579–619. https://doi.org/10.3102/00346543066004579

[39] Qasim Lone, Alisa Frik, Matthew Luckie, Maciej Korczyński, Michel van Eeten, and Carlos Gañán. 2022. Deployment of Source Address Validation by Network Operators: A Randomized Control Trial. *2022 IEEE Symposium on Security and Privacy (SP)* 00 (2022), 2361–2378. https://doi.org/10.1109/sp46214.2022.9833701

[40] Max Maaß, Marc-Pascal Clement, and Matthias Hollick. 2021. Snail Mail Beats Email Any Day: On Effective Operator Security Notifications in the Internet. In *The 16th International Conference on Availability, Reliability and Security (ARES 2021)*. ACM, New York, NY, USA, Vienna, Austria, 1–13. https://dl.acm.org/doi/10.1145/3465481.3465743

[41] Max Maaß, Henning Pridöhl, Dominik Herrmann, and Matthias Hollick. 2021. Best Practices for Notification Studies for Security and Privacy Issues on the Internet. In *The 16th International Conference on Availability, Reliability and Security (The 16th International Conference on Availability, Reliability and Security)*. Association for Computing Machinery, Vienna, Austria, 1–10. https://doi.org/10.1145/3465481.3470081

[42] Max Maaß, Alina Stöver, Henning Pridöhl, Sebastian Bretthauer, Dominik Herrmann, Matthias Hollick, and Indra Spiecker. 2021. Effective notification campaigns on the web: A matter of Trust, Framing, and Support. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 2489–2506. https://www.usenix.org/conference/usenixsecurity21/presentation/maass

[43] Art Martori. 2020. Spamdexing: What is SEO Spam and How to Remove It. https://blog.sucuri.net/2020/02/spamdexing-seo-spam.html. [last accessed 2021-10-28].

[44] Peter Mayer, Alexandra Kunz, and Melanie Volkamer. 2017. Reliable Behavioural Factors in the Information Security Context. In *Proceedings of the 12th International Conference on Availability, Reliability and Security* (Reggio Calabria, Italy) *(ARES '17)*. Association for Computing Machinery, New York, NY, USA, Article 9, 10 pages. https://doi.org/10.1145/3098954.3098986

[45] Peter Mayer, Yixin Zou, Florian Schaub, and Adam J. Aviv. 2021. "Now I'm a bit angry:" Individuals' Awareness, Perception, and Responses to Data Breaches that Affected Them. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 393–410. https://www.usenix.org/conference/usenixsecurity21/presentation/mayer

[46] mindUp Web '&' Intelligence GmbH. [n. d.]. Fake-Online-Shops - Erkennung von Fake-Shops auf gehackten Webseiten. https://www.mindup.de/data-scientists/anwendungsfaelle/fake-online-shops. [last accessed 2025-02-14].

[47] 'mindUp Web & Intelligence GmbH'. [n. d.]. Gezieltes Finden gehackter Webseiten. https://www.mindup.de/nachrichten/artikel/gezieltes-finden-gehackter-webseiten. [last accessed 2025-02-14].

[48] Caitlyn N. Muniz, Taylor Fisher, Katelyn Smith, Roan Ali, C. Jordan Howell, and David Maimon. 2024. Hello, You've been hacked: a study of victim notification preferences. *Journal of Crime and Justice* 0, 0 (2024), 1–17. https://doi.org/10.1080/0735648X.2024.2340554

[49] n.a. 2021. Princeton researcher apologizes for GDPR/CCPA email study . https://news.ycombinator.com/item?id=29650719. [last accessed 2024-04-18].

[50] Yevheniya Nosyk, Maciej Korczyński, Carlos H. Gañán, Michał Król, Qasim Lone, and Andrzej Duda. 2023. Don't Get Hijacked: Prevalence, Mitigation, and Impact of Non-Secure DNS Dynamic Updates. *2023 IEEE 22nd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* 00 (2023), 1480–1489. https://doi.org/10.1109/trustcom60117.2023.00202

[51] Daniel O'Keefe. 2015. Elaboration Likelihood Model. In *The Concise Encyclopedia of Communication*.

[52] Tara Poteat and Frank Li. 2021. Who you gonna call? an empirical evaluation of website security.txt deployment. *Proceedings of the 21st ACM Internet Measurement Conference (IMC '21)* (2021), 526–532. https://doi.org/10.1145/3487552.3487841

[53] Karen Renaud, Rosalind Searle, and Marc Dupuis. 2021. Shame in Cyber Security: Effective Behavior Modification Tool or Counterproductive Foil? *New Security Paradigms Workshop* (2021), 70–87. https://doi.org/10.1145/3498891.3498896

[54] Karen Renaud, Verena Zimmermann, Tim Schürmann, and Carlos Böhm. 2021. Exploring cybersecurity-related emotions and finding that they are challenging to measure. *Humanities and Social Sciences Communications* 8, 75 (2021). https://doi.org/10.1057/s41599-021-00746-5

[55] Elsa Rodríguez, Susanne Verstegen, Arman Noroozian, Daisuke Inoue, Takahiro Kasama, Michel van Eeten, and Carlos H Gañán. 2021. User compliance and remediation success after IoT malware notifications. *Journal of Cybersecurity* 7, 1 (07 2021), tyab015. https://doi.org/10.1093/cybsec/tyab015

[56] Ronald W. Rogers. 1975. A Protection Motivation Theory of Fear Appeals and Attitude Change. *The Journal of psychology* 1 (1975), 93–114. Issue 91. https://doi.org/10.1080/00223980.1975.9915803

[57] Asreen Rostami, Minna Vigren, Shahid Raza, and Barry Brown. 2022. Being Hacked: Understanding Victims' Experiences of IoT Hacking. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*. USENIX Association, Boston, MA, 613–631. https://www.usenix.org/conference/soups2022/presentation/rostami

[58] Nayanamana Samarasinghe and Mohammad Mannan. 2021. On cloaking behaviors of malicious websites. *Computers & Security* 101 (2021), 102114. https://doi.org/10.1016/j.cose.2020.102114

[59] Emanuel Schmider, Matthias Ziegler, Erik Danay, Luzi Beyer, and Markus Bühner. 2010. Is It Really Robust? Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption. *Methodology* 6, 4 (2010), 147–151. https://doi.org/10.1027/1614-2241/a000016

[60] Y. Shafranovich and E. Foudil. 2022. RFC 9116: A File Format to Aid in Security Vulnerability Disclosure — datatracker.ietf.org. https://datatracker.ietf.org/doc/html/rfc9116 [last accessed 12-01-2025].

[61] SiteLock. 2022. Cybersecurity Statistics Report 2022. https://www.sitelock.com/resources/security-report/. [last accessed 2023-10-10].

[62] Kyle Soska and Nicolas Christin. 2014. Automatically Detecting Vulnerable Websites Before They Turn Malicious. In *23rd USENIX Security Symposium (USENIX Security 14)*. USENIX Association, San Diego, CA, 625–640. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/soska

[63] Ben Stock, Giancarlo Pellegrino, Frank Li, Michael Backes, and Christian Rossow. 2018. Didn't You Hear Me? - Towards More Successful Web Vulnerability Notifications. In *Proceedings of the 25th Annual Symposium on Network and Distributed System Security (NDSS '18)*. 1 – 15. https://doi.org/10.14722/ndss.2018.23171

[64] Ben Stock, Giancarlo Pellegrino, Christian Rossow, Martin Johns, and Michael Backes. 2016. Hey, You Have a Problem: On the Feasibility of Large-Scale Web Vulnerability Notification. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX, 1015–1032. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/stock

[65] StopBadware and Commtouch. 2012. Compromised Websites: An Owner's Perspective. https://www.stopbadware.org/files/compromised-websites-an-owners-perspective.pdf. (2012), 1 – 15.

[66] Alina Stöver, Nina Gerber, Henning Pridöhl, Max Maass, Sebastian Bretthauer, Indra Spiecker genannt Döhmann, Matthias Hollick, and Dominik Herrmann. 2023. How Website Owners Face Privacy Issues: Thematic Analysis of Responses from a Covert Notification Study Reveals Diverse Circumstances and Challenges. *Proc. Priv. Enhancing Technol.* 2023 (2023), 251–264. https://petsymposium.org/popets/2023/popets-2023-0051.php

[67] Karishma Sundaram. 2022. Fix WordPress Pharma Hack and SEO. https://www.malcare.com/blog/what-is-pharma-hack-how-to-clean-it/. [last accessed 2024-04-19].

[68] Princeton University. 2021. Princeton-Radboud Study on Privacy Law Implementation. https://privacystudy.cs.princeton.edu/. [last accessed 2024-04-18].

[69] Christine Utz, Matthias Michels, Martin Degeling, Ninja Marnau, and Ben Stock. 2023. Comparing Large-Scale Privacy and Security Notifications. *Proceedings on Privacy Enhancing Technologies* 2023, 3 (2023), 173–193. https://doi.org/10.56553/popets-2023-0076

[70] Marie Vasek and Tyler Moore. 2012. Do Malware Reports Expedite Cleanup? An Experimental Study. In *5th Workshop on Cyber Security Experimentation and Test, CSET '12, Bellevue, WA, USA, August 6, 2012*. USENIX Association, 1 – 8. https://www.usenix.org/conference/cset12/workshop-program/presentation/vasek

[71] Werner Wirth and Rinaldo Kühne. 2013. Grundlagen der Persuasionsforschung. Konzepte, Theorien und zentrale Einflussfaktoren. In *Handbuch Medienwirkungsforschung*.

[72] Qiushi Wu and Kangjie Lu. 2021. On the Feasibility of Stealthily Introducing Vulnerabilities in Open-Source Software via Hypocrite Commits. https://api.semanticscholar.org/CorpusID:233479632.

[73] Eric Zeng, Frank Li, Emily Stark, Adrienne Porter Felt, and Parisa Tabriz. 2019. Fixing HTTPS Misconfigurations at Scale: An Experiment with Security Notifications. In *The 2019 Workshop on the Economics of Information Security (2019)*. Boston, MA, 1 – 19. https://www.semanticscholar.org/paper/Fixing-HTTPS-Misconfigurations-at-Scale%3A-An-with-Zeng-Li/b22c522c6201f8545e1626deaf6ca43db52444d7

[74] Jia Zhang, Haixin Duan, Wu Liu, and Xingkun Yao. 2017. How to Notify a Vulnerability to the Right Person? Case Study: In an ISP Scope. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*. 1–7. https://doi.org/10.1109/GLOCOM.2017.8253993

[75] Yixin Zou, Shawn Danino, Kaiwen Sun, and Florian Schaub. 2019. You 'Might' Be Affected: An Empirical Analysis of Readability and Usability Issues in Data Breach Notifications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300424

[76] Yixin Zou, Khue Le, Peter Mayer, Alessandro Acquisti, Adam J. Aviv, and Florian Schaub. 2024. Encouraging Users to Change Breached Passwords Using the Protection Motivation Theory. *ACM Trans. Comput.-Hum. Interact.* 31, 5, Article 63 (Nov. 2024), 45 pages. https://doi.org/10.1145/3689432

[77] F. O. Çetin, C. Hernandez Ganan, M. T. Korczynski, and M. J. G. van Eeten. 2017. Make notifications great again: learning how to notify in the age of large-scale vulnerability scanning *(16th Workshop on the Economics of Information Security (WEIS 2017))*. San Diego, 1–23. http://resolver.tudelft.nl/uuid:621f4a4f-e5d9-4f04-abc4-46252f9db3db

[78] Orçun Çetin, Lisette Altena, Carlos Gañán, and Michel van Eeten. 2018. Let Me Out! Evaluating the Effectiveness of Quarantining Compromised Users in Walled Gardens. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. USENIX Association, Baltimore, MD, 251–263. https://www.usenix.org/conference/soups2018/presentation/cetin

[79] Orçun Çetin, Carlos Hernandez Gañán, Lisette Altena, Takahiro Kasama, Daisuke Inoue, Kazuki Tamiya, Ying Tie, Katsunari Yoshioka, and Michel van Eeten. 2019. Cleaning Up the Internet of Evil Things: Real-World Evidence on ISP and Consumer Efforts to Remove Mirai. *Proceedings 2019 Network and Distributed System Security Symposium* (2019). https://www.ndss-symposium.org/ndss-paper/cleaning-up-the-internet-of-evil-things-real-world-evidence-on-isp-and-consumer-efforts-to-remove-mirai/

[80] Orçun Çetin, Carlos Gañán, Lisette Altena, Samaneh Tajalizadehkhoob, and Michel van Eeten. 2019. Tell Me You Fixed It: Evaluating Vulnerability Notifications via Quarantine Network. *2019 IEEE European Symposium on Security and Privacy (EuroS&P)* 00 (2019), 326–339. https://doi.org/10.1109/eurosp.2019.00032

[81] Orçun Çetin, Mohammad Hanif Jhaveri, Carlos Gañán, Michel van Eeten, and Tyler Moore. 2016. Understanding the role of sender reputation in abuse reporting and cleanup. *Journal of Cybersecurity* 2, 1 (2016), 83–98. https://doi.org/10.1093/cybsec/tyw005

# A BACKGROUND

In this section, we provide some of the keywords used to identify compromised websites (Appendix A.1) and results from our pre-study (Appendix A.2).

## A.1 Keywords for Identifying Compromised Websites

Table 2 lists selected keywords that were used to find compromised websites for our sample (translated). Note that we only report the keywords from the hacked websites in our sample. This is *not* a comprehensive list of all keywords used for the crawling.

| | |
|---|---|
| **pharmacy** | Adderall, Alprazolam, Amoxicillin, Ativan, Azithromycin, Buserelin, Canadian Pharmacy, Cefixim, Cialis, Cialis non prescriptive, Cialis super active, Clomid, Clonazepam, corona mask, Dapoxetin, Desinfectant, Dexedrin, Diazepam, diflucan, Enanthate, Fentanyl ratiopharm, Fildena super active, finasteride, Gamma-Hydroxybutyrat, Generika, Generika non prescriptive, hormone, hydrocodone, Kamagra, Kamagra non prescriptive, Lasix, Levitra, Levitra non prescriptive, Lexapro, Lorazepam, Lovegra, Magnesium, Methadon, Methyltestosteron, Midazolam, Modafinil, Modafinil non prescriptive, Novaldex, Orlistat, Oxycodon, Pantoprazole, Pantoprazole non prescriptive, Pharmacy, Phentermine, Pregabalin, Priligy, Propecia, respiratory protection mask FFP3, Ritalin, Rivotril Roche, Rohypnol, Sildenafil, Sildenafil non prescriptive, Tadalafil, Tadalafil ratiopharm, Tadalista super active, Tebonin, Testosteron, Testosteron non prescriptive, Tramadol, Tramal, Valacyclovir, Valium, Vardenafil, Ventolin, Viagra, Viagra non prescriptive, Viagra super active, Vicodin, Xanax, Xenical |
| **watches** | Breitling, Breitling 2019, Breitling 2020, Fortis, Montblanc, Patek Philippe, Patek Philippe Nautilus, Paul Hewitt Miss Ocean, Rolex, Rolex 2019, Rolex 2020 |

| | |
|---|---|
| **clothes** | Adidas by Stella McCartney, Armani Jeans, Balducci, Barbour, Barcelona Tricot, Ben Sherman, Betula licensed by Birkenstock, Birkenstock, BOSS, BOSS Green, BOSS Orange, Camel Active, Canada Goose, Chocolate Schubar, Christian Louboutin, Clarks, Clarks Originals, Dachstein, Daniel Hechter, DC Shoes, Deuter, Deuter Backpack, Dorothy Perkins, Dr. Martens, El Naturalista, Enzo Marconi, Esprit, Falke, Fjällräven, Fjällräven Jacket, Fjällräven Backpack, Fjällräven Bag, Floris van Bommel, Franceschetti, Fratelli Rossetti, French Connection, Gabor, Gianvito Rossi, Giorgio 1958, Giuseppe Zanotti, Gucci, Guess, H.I.S., Haglöfs, Head Edge Lyt 100 W, Head Graphene 360+ Radical 120 SB, Head Infinity Jacket Women, Head Kore 40, Hoffenheim Tricot, Jack Jones, KangaROOS, KARL LAGERFELD, Kenneth Cole Reaction, Cologne Tricot, LAGERFELD, Lauren Ralph Lauren, Le Coq Sportif, Levis, Lloyd 1888, Louboutin Pumps, Lumberjack, Makita, Mammut Rainjacket, Marc O Polo, MARCIANO GUESS, Moheda Toffeln, Mustang, New York Yankees, Nike, Nike Performance, Nike Sportswear, Nordisk Jacket, Original Penguin, Panama Jack, Pedro Miralles, PERLATO, Picard Portemonnaie, Picard Shopper, Picard Bag, Pierre Balmain, Polo Ralph Lauren, Prada, Prime Shoes, Puma, Puma Golf, Ricosta, Robeez, Roberto D'Angelo, S.Oliver, Sanchita Shoes, Schöffel Trousers, Schöffel Jacket, Shellys London, Sir Oliver, Skechers, Speedo, Superdry, Superfit, Tamaris, Tartine et Chocolat, Tatonka Rainponcho, Tatonka Backpack, The North Face, Timberland, Tommy Hilfiger, Tricot, Under Armour, Vans, Vaude, Vaude Arco, Vaude women's clothes, Vaude men's clothes, Vaude Jacket, Versace, Versace Collection, Versace Jeans, Versus Versace, Zign |
| **tech** | Acer Aspire, ASICS, Asus, Canon Eos, Garmin GPSmap, Garmin Smartwatch, Hewlett Packard, HP Pavillion, HP Toner, Huawei, LEDLENSER, Lenovo, Macbook Air, Nikon Z, Nintendo Switch, Panasonic Lumix, Playstation 5, ps 5, Sony Alpha, Sony Cybershot, Sony PS5, Trekstor |
| **home** | Beko Washer, Bosch AXT, Bosch Mum, Bosch Rotak, Braun IdentityCollection, broil king crown pellet 400, Cube, Cube allroad, Cube green silver, Cube stereo hybrid, De Longhi Magnifica, deuter sleeping bag, Dolmar PS, Dyson vacuum cleaner, Einhell, Gaggia Milano, Gardena, Goodyear, Grundfos, Grundig, Güde, Hancook, Head Edge Lyt 100 W, Husqvarna, Jura ENA, Jura Vollautomat, Kärcher, Kärcher K, Kawasaki, Kitchenaid Artisan, Krups, Magura, Magura MT8, Makita, Michelin Primacy, Miele, Montblanc Pen, Neff, Nordisk Tarp, Pirelli, Proficook, Robomow, Rossignol Bike, Saeco, Samsonite, Scheppach, Scott Addict, Scott Scale, Scott Spark, Siemens EQ, siemens eq 6 plus s500, siemens eq 6 plus s700, Stiga, Stihl, Thermomix, Thule, Thule ProRide, Thule VeloCompact, uvex helmet, Vorwerk, Vorwerk Kobold, Vorwerk Thermomix, WMF, Wolf Garden |
| **other** | 8800, Banner, Bounce, buy, Casino, Cristiano Ronaldo CR7, date, Disney, Eskorte, gold, gold bars, IT Invest, Kickers, Lego, Lyssna, Pokemon, Porn, Starwars, win, xxx |

Table 2. Selection of keywords from our sample (translated). Note, that we only report the keywords from the hacked websites in our sample. This is not a complete list of all keywords that were used for the crawling.

## A.2 Results Pre-Study

Table 3 provides the results from our pre-study. The table and further discussion of the results can also be found in [30].

Table 3. Overview of sender and notification channels of the initial notification

| Interview No | Number of Employees | ATI | Initial Sender | Suitable Sender | Initial Notification Channel | Suitable Notification Channel |
|---|---|---|---|---|---|---|
| 1 | – [7] | 4.22[8] | university | BSI [9], relatives with IT-Know-How, police, research facilities | email | letter |
| 2 | 10-49 | 4.33 | project partner | do not know, well-known sender, BMWI [10], BSI, data protection authorities, IT security, business association, police, hosting provider, research facilities | phone call | letter, email, web portal [11] |
| 3 | 1-9 | 3.33 | police, project partner | do not know, police, IT security companies | email, phone call | letter, email |
| 4 | 1-9 | 4 | relatives, project partner | indifferent, well-known and verifiable email address | email, phone call | suspicious of every channel, email |
| 5 | 1-9 | 5 | self | indifferent, hosting provider, police | – | email |
| 6 | – | 5.11 | project partner, police | BSI | email, phone call | email, phone call |
| 7 | 1-9 | 3.44 | hosting provider | hosting provider | email | phone call |
| 8 | 1-9 | 4 | unknown | authorities in general, accounting firm | email | phone call + email |
| 9 | 1-9 | 5.33 | self | hosting provider, Joomla | – | email |
| 10 | 1-9 | 4.78 | police | do not know, police, hosting provider | email | letter |
| 11 | 1-9 | 4.22 | police | indifferent | phone call | phone call |
| 12 | – | 3.22 | university | content more important than sender, some official body, hosting provider, police, research facilities | email | letter, phone call |
| 13 | 10-49 | 4.22 | website users | our agency | unknown | phone call |
| 14 | 1-9 | 5.44 | uninvolved third party | indifferent, police | email | email, phone call |
| 15 | 1-9 | 4.44 | police | police | phone call and email | email) |

*Continued on next page*

[7] If it was not a company website, we could not determine company size by the number of employees
[8] Determined via the Affinity for Technology Interaction (ATI) scale [20] with a 6-point Likert scale from 1.0 to 6.0, with 1.0 meaning low and 6.0 meaning a high interest in technology
[9] Federal Office for Information Security
[10] Federal Ministry for Economic Affairs and Climate Action
[11] Only for vulnerability notifications from hosting provider

**Table 3 – *continued from previous page***

| Interview No | Number of Employees | ATI | Initial Sender | Suitable Sender | Initial Notification Channel | Suitable Notification Channel |
|---|---|---|---|---|---|---|
| 16 | 1-9 | 4.44 | hosting provider | hosting provider would be fine | email | email, web portal [12] |
| 17 | – | 5.78 | university | indifferent, research facilities, hosting provider | email | email |
| 18 | 1-9 | 4 | university | do not know, authorities, business association, BMG [13], city council, police, research facilities | email | email |
| 19 | 1-9 | 4.56 | self, website users | well-known sender, research facilities | unknown | email |
| 20 | – | 4.44 | police | indifferent, has to match the signature | phone call | email, phone call |
| 21 | 10-49 | 5.67 | police | do not know, BSI, police | email | letter, email |
| 22 | 1-9 | 3.78 | police | police, official organizations | phone call | phone call + email |
| 23 | 1-9 | 5.67 | self | official organizations, hosting provider | – | phone call |
| 24 | 10-49 | 4.89 | police | do not know, official organizations, police | phone call | do not know |
| 25 | 50-249 | 3.33 | police | do not know, police | phone call | phone call |

[12] only hosting provider
[13] Federal Ministry of Health

## B NOTIFICATION TEXT

In this section, we provide the text for the notification emails we sent (Appendix B.1), including the framings that were used (Appendix B.2).

### B.1 Notification Text

To: <recipient.email>

From: <sender>

Subject: Important information on your website <domain>

<title> <lastname>,

We are contacting you with important information about your website <domain>.

Within a research project we currently analyze search engine entries and detected several subpages which redirect to fraudulent online shops. It looks like your website was manipulated by third parties.

You can easily reproduce it when you search in your browser for "site:<source.fqdn> <hacking_keyword>". The search results show all entries to your website that are known to the search engine. You can see that some entries do not match the content of your website. If you clicked on those links, you would be redirected to a fraudulent online shop.

We recommend you to restore the security of your website. Please verify our information and solve the problem as soon as possible. You can find further information at https://www.web-inspection.de/faq.

<framing>

Best regards,
<sender>
——————————
<signature>

### B.2 Framings

<neutral>: without framing

<technical-generic>: Please check this issue. If the problem is not remediated, we will have to report your case to your hosting provider. They can block your webspace or the affected files in your webspace.

<technical-hoster1>: Please check this issue and let us know what you have done. Please note that if there is no response or if the problem persists, we may be forced to block your webspace or the affected files in your webspace on

<date>.

<technical-hoster2>: Please check this and let us know what action you have taken. Please note that if there is no response or if the problem persists, we will be forced to take further action.

<reputation-generic>: Please check this issue. Please note that you may suffer reputational damage if you do not remediate the problem. This can be, for example, block listings with security providers or warning messages from the browser, so that your website is indexed as a security risk for outsiders.

<reputation-CERT>: Please check this issue and let us know what you have done. Please note that various web reputation services assess the credibility of websites according to a scoring procedure, which may also include issues such as this one. Users of such web reputation services may therefore receive a warning, or their access may be blocked completely.

## C INTERVIEW GUIDELINE AND CODEBOOK

In this section, we provide the translated Interview Guideline for our Follow-up interviews (Appendix C.1) and the translated codebook used to analyze the interviews (Appendix C.2). We also provide the codes that we used to categorize the information obtained from the phone calls (Appendix C.3).

### C.1 Interview Guideline for Main Study Interviews

**Introduction**

We will start the interview right away. First, I would like to explain our procedure to you once again. I apologize if I repeat some parts of our previous conversation in the following. For the record, I would like to ensure once again that we have provided you with comprehensive information about our project. As already mentioned, the aim of our research project is to inform website owners about manipulations by cybercriminals on their websites. We are particularly interested in why you have not yet been able to solve the problem. In the following, I will ask you a few questions about this. From our experience, the interview will take about 5-10 minutes. All your answers are voluntary, and you can withdraw your consent to participate at any time without giving any reason. You will not suffer any disadvantages if you do not take part in the interview. There are no right or wrong answers, we are interested in your *personal* experiences and your *personal* opinion. The interview will be recorded and afterwards transcribed by the research team. The audio files will then be deleted. For further scientific analysis of the interview texts, all personal information that could lead to your person or your company being identified will be changed in or removed from the text. In scientific publications, interviews are only quoted in excerpts. In this way, we ensure that your individual statements cannot lead to the identification of your person or your company. Personal contact information will be stored separately from interview data and will not be accessible to third parties. Your contact information will be automatically deleted after the end of the research project.

If you like, I can send you a detailed privacy policy by email afterwards.

[waiting for response]

If you have any questions about the project or the use of your data, you can contact me at any time. I will be happy to provide you with my contact information again after the interview.

Can you please briefly confirm for the record that you agree to the interview under the conditions just explained?

[waiting for response]

**Questions**

(1) Email not received
- Have you seen/received the email? If not: Why did you not receive the email / what could have been the reason? Do you have suggestions for improvement?
- Did you receive a notification about the problem from another person? If yes: By Whom? Why? Content?

(2) Email not read
- Did you also read the email? If not: Why did you not read the email? What could have been the reason? Do you have suggestions for improvement?

(3) Email read but did not find it relevant enough to take any actions
- Did you find the content of the email relevant enough to deal with it further? If not: Why not? What should have been different? Do you have suggestions for improvement?

(4) Email was relevant enough to deal with further, but did not understand the urgency
- Did you understand the risk described? If not: What was unclear? What should have been different? Do you have suggestions for improvement?
- Did you try to find help? If yes - where? To what extent was this helpful? Do you have suggestions for improvement?
  - If an agency or external service provider was involved: Why could the problem not be remediated?
- Did you try to find information on the website named in the email? If yes: To what extent helpful? What should have been different? Do you have suggestions for improvement? *If necessary explain the risk now*

(5) Understood the risk but could not reproduce
- Were you able to reproduce the hack? Was the procedure for reproducing the hack clear to you? If not: What was unclear? What should have been different? Do you have suggestions for improvement?
- Did you try to find help? If yes - where? To what extent was this helpful? Do you have suggestions for improvement?
  - If an agency or external service provider was involved: Why could the problem not be remediated?
- Did you try to find information on the website named in the email? If yes: To what extent helpful? What should have been different? Do you have suggestions for improvement?

*If they told that they used a malware scanner which did not find anything suspicious, ask, which software they used*

(6) Understood the risk but remediation was unclear
- Was the solution to the problem clear to you? If not: What was unclear? What should have been different? Do you have suggestions for improvement?
- Did you try to find help? If yes - where? To what extent was this helpful? Do you have suggestions for improvement?
  - If an agency or external service provider was involved: Why could the problem not be remediated?
- Did you try to find information on the website named in the email? If yes: To what extent helpful? What should have been different? Do you have suggestions for improvement?

*If they told that they used a malware scanner which did not find anything suspicious, ask, which software they used If necessary explain solution now*

(7) Hack is currently being fixed or has been fixed in the meantime
- Could you briefly describe how you plan to remediate the hack? Or how you remediated it?

**Gratitude and Final Comments**

Thank you very much! That brings us to the end of the questionnaire. On behalf of the university, I would like to thank you very much for your participation. Finally, do you have any questions or comments that you would like to share with us?

### C.2 Codebook for Main Study Interviews

**Notification Delivery / Sender:**

- Original email received [yes / no / cannot remember / n.a.]
- Suggestions for improvement of delivery [n.a. / text]
- Other related to delivery [text]

**Notification Content**

- Original e-mail read [yes / no / cannot remember / n.a.]
- if read: why not remediated(1)? [email considered spam / problem not considered relevant e.g. no time to fix, "not that bad" / problem not understood e.g. do not know what happened, what these redirects are / problem not reproducible e.g. if advice with search operator not understood / solution unclear e.g. understood that action should be taken, but no idea what to do / problem considered fixed / other (text)]
- if read: why not remediated(2)? [email considered spam / problem not considered relevant e.g. no time to fix, "not that bad" / problem not understood e.g. do not know what happened, what these redirects are / problem not reproducible e.g. if advice with search operator not understood / solution unclear e.g. understood that action should be taken, but no idea what to do / problem considered fixed / other (text)]
- if read: why not remediated(3)? [email considered spam / problem not considered relevant e.g. no time to fix, "not that bad" / problem not understood e.g. do not know what happened, what these redirects are / problem not reproducible e.g. if advice with search operator not understood / solution unclear e.g. understood that action should be taken, but no idea what to do / problem considered fixed / other (text)]
- Suggestions for improvement of content or criticism content [n.a. / no suggestions for improvement e.g. notification was actually good, or person cannot think of anything on the subject / text]
- Other related to content [text]

**Further information**

- Type of notification known [yes / no / n.a.]
- Ever had problems with the website / experienced other hacking [yes / no / n.a.]
- Sender already known [yes (positive / negative / n.a.) / no (positive / negative / n.a.) / n.a.]
- Sender of notification contacted? [yes / no / n.a.]
- if sender contacted: why contacted? [thanked for notification / sender verified / help sought to resolve problem e.g. further information requested / other (text)]
- Other external help obtained [yes / no / n.a.]
- if help obtained: how or from whom? [external service provider (temporary/permanent) / IT manager (in-house) / friends, acquaintances, relatives / other (text)]

- Other important information in general [Text]
- Other suggestions for improvement [Text]

### C.3 Codebook for Notes during Phone Calls

- not called *(e.g. no valid phone number)*
- not reached after three attempts *(e.g. no one answered the phone)*
- no interview - not interested *(e.g. person said that she is not interested in an interview without providing any further information)*
- no interview - no more response *(e.g. we were told that we should reach out again to a dedicated person for an interview, but then no further contact could be established)*
- no interview - notification not received *(e.g. person explained that they did not know about our notification but also did not want to participate in an interview)*
- no interview - know about problem *(e.g. person explained that they already knew about the problem before we called them but they also did not remember to have received our notification)*
- no interview - know about notification *(e.g. person explained that received our notification but would not want to participate in an interview)*
- interview

## D POST-HOC ANALYSIS

Table 4 provides the results of the Games-Howell post-hoc analysis for the ANOVA, including the Bonferroni corrections.

| (I) Sender | (J) Sender | difference in means(I-J) | standard error | Sig. | Sig. corr. | 95% confidence interval lower limit | 95% confidence interval upper limit |
|---|---|---|---|---|---|---|---|
| Federal CERT | control | -9.68 | 2.41 | <.001 | .006 | -15.93 | -3.42 |
| | provider | 2.33 | 3.22 | .888 | > .999 | -6.03 | 10.68 |
| | university | 2.1 | 3.07 | .904 | > .999 | -5.85 | -10.05 |
| university | control | -11.77 | 2.52 | <.001 | .006 | -18.31 | -5.24 |
| | provider | 0.23 | 3.3 | 1.000 | >.999 | -8.33 | 8.79 |
| | Federal CERT | -2.1 | 3.07 | .904 | >.999 | -10.05 | 5.85 |
| provider | control | -12.0 | 2.7 | <.001 | .006 | -19.02 | -4.98 |
| | university | -0.23 | 3.3 | 1.000 | >.999 | -8.79 | 8.33 |
| | Federal CERT | -2.33 | 3.22 | .888 | >.999 | -10.68 | 6.03 |
| control | Federal CERT | 9.68 | 2.41 | <.001 | .006 | 3.42 | 15.93 |
| | university | 11.77 | 2.52 | <.001 | .006 | 5.24 | 18.31 |
| | provider | 12.0 | 2.7 | <.001 | .006 | 3.42 | 15.93 |

Table 4. Post-hoc analysis for differences between the sender groups.

## E RESULTS FROM THE FOLLOW-UP INTERVIEWS

In this section, we provide translated quotes from our follow-up interviews (Section 3.2.2) in Table 5, and figures to describe the results of our phone calls and interviews (Figure 6).

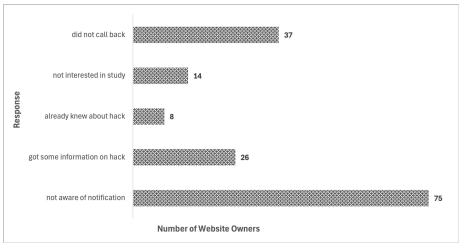
| | |
|---|---|
| sender | "Yeah so, I mean, of course if I get a warning directly through WordPress, then of course I check this differently than anything else." [P2] |
| | "But otherwise I think it's a very unfortunate communication channel, because a lot of spam comes in in the name of the Federal CERT, and I'd say that the average user, who sells his cake in the internet, won't be know the Federal CERT anyway. But they will recognize the name of their hosting provider. [...] In general, I think it would be better to receive both - an e-mail directly from the Federal CERT and perhaps also from the provider as well [...]." [P4] |
| | "Yeah, so this [the sender being *email address*] definitely sounds like it went straight to spam." [P7] |
| | "First of all, it must be clear that this is a legitimate sender. [...] because you get so many e-mails where you don't know what exactly is behind it, whether they want to sell something or something else." [P18] |
| | "Or Federal CERT, right, there... this seems implausible itself. Because, I think, why should some irrelevant website be of interest [...]? You know, I myself know that, in those offices, who they usually have other things to do, then with some s*** like that, and then they additionally write emails to people [...] So why should they do that? That was actually the question I asked myself, and then I thought, no, not really, that can only be spam." [P21] |

| | |
|---|---|
| | "So, Federal CERT dot , ,I mean… something that has , ,then you'll read it, but CERT Federal CERT didn't mean anything to me." [P37] |
| notification channel | "Definitely via phone [...] But then also in a way that you don't have the feeling that someone is trying to sell you something [...] So I would always combine it with a scientific background [...] that there is a real legitimacy behind it […]." [P8] |
| | "An e-mail everyone can write, but if someone goes the extra mile to write a letter, this has a totally different flavor. […] Letter or phone, yes. So phone call is, especially when you [...] send your number [...], this has a different liability, as when I write an e-mail and, as I told before, the people are then overwhelmed." [P13] |
| | "[…] and it was not before you approached us personally via phone that we actually became aware that this is no spam." [P41] |
| | "Phone is always great." [P35] |
| other | "[…] just the fact that it was sent three or four times, um, was, um, awkward and I didn't want to have anything to do with it." [P14] |
| improve subject line | "[…] but for me it probably was in the subject line, that I said I'm not interested, and the I'll delete that immediately" [P2] |
| | "'Important' information sounds a bit more like... I get a lot of e-mails like that, where it says something like 'very important, very important'. That actually tends to get thrown out, because most of the time [...] it's just the opposite, when it somehow says 'super important' or something like that and then I click on it and then I see, ok, that's just something that's not important at all. Um... Yes, I don't know, maybe just write in the subject line which institute you're from, simply, that seems a bit more trustworthy [...]" [P7] |
| | "Okay, then I delete something like that anyway, because I don't even look at it. Because, something like 'we have your page' and 'we would like to optimize SEO' and 'there's a security issue' and all that, this I delete, I don't even look at it, because this is for sure some spam stuff." [P8] |
| | "Hmmh, maybe there had been something, I don't know, something from your - I am not sure, are you a university? […] Maybe there had been something like this in the subject line at first, […] so that you see it as legitimate […]." [P28] |
| | "[...] there is this CERT in square brackets. Um, CERT minus , ,hashtag, and then there's a very long number with a date in it, I think that's [...] irritating, yes. [...] So you really have to make an effort to read down to the sender, because otherwise it just looks very unusual and I didn't know how to classify it." [P37] |
| provide assistance | "Yeah, [...] there really has to be a hint. Maybe also to an external website [...] or an extra attachment as a PDF or so. The e-mail itself, that is already okay lengthwise and I am also with you, if it is even longer, it will only be skimmed and maybe also half of it will be ignored. Um, but the person who really wants to deal with it, I think he's grateful for some help." [P5] |
| | "So, if you know where this vulnerability is, it would of course be great if you could tell me and what we can do now." [P18] |

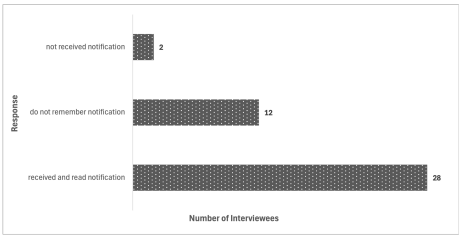
| | |
|---|---|
| | "And yes, above all, what you should or can do specifically [...]. Without the usual sentence, uh sit down with your mastermind - or how this is called [...]. But rather say explicitly what one should do, if there is really something bad." [P18] |
| | "[...] I found the email very [...], very open, so um so rather focused on the problem instead of focused on the solution [...]. Um, there is a problem, but I would definitely liked to have solutions, maybe even tips for programming or so, that you will, whether you pay for it or not, um, at least somehow find the solution, is it remediated or is it not remediated [...]." [P30] |
| | "The initial information from the provider was TOTALLY undetailed. It was only a hint that allegedly on our subdomain there are some conspicuous links that would not match the rest [...]." [P31] |
| | "So, for me, as a layperson, it would have been nice if there had been some PDF files with a step-by-step guide on what can be done to prevent this from happening again, or to actually clean up this attack. And there was actually nothing there. Um, just with this Google search and um, it just said, please check which vulnerabilities were exploited [...]." [P32] |
| improve formatting | "What I found a bit unfortunate is actually [...] the URL that you are using, the *name of website*. [...] In general, we warn for links in texts [...] and then it is of course the formatting. Um that is actually a bit university-style that you say, ok, I'm writing here a plaintext e-mail, non-formatted [...] with a strange font [...] that could be designed more appealing, then it can rather attract attention." [P13] |
| | "In particular, I didn't go to the links, um where I could get more information, um, about it. [...] So that was suspicious for me [....] that I myself had to go, um, to, uh, another page or via the link to another area." [P39] |
| | "[...] this e-mail had a red banner running across it on the top and this must somehow have looked as if it was spam [...]." [P16] |
| | "What definitely helped was the footer, [...] that this didn't come from something strange, uh what also helped was that there weren't some strange links [...] but just *name of website*, so, there you [...] can do little wrong, if you simply type that in the address bar [...]." [P27] |
| | "I would rather start immediately with Google in the first step. [...] Because there it became obvious [...] 'Look at your search results'. 'There is something wrong'. [...] And, uh, not 'you have to do something' or 'react to something here', [...] just 'look at the results'." [P33] |
| | "Well, you do a good job, but whenever a logo from a governmental authority appears - and those usually pay attention of me - then I'm already suspicious whether it is a fake. The only thing that convinced me was when you appeared as a university and mentioned several companies, tax consultants etc., that actually exist." [P35] |
| motivation | "Um, because here came just... well, ok, the message, urgent message, but then just manipulated and search engine entries blah blah blah [...] maybe talk about, who you are exactly, or what your message is about." [P26] |
| provide contact information | "Maybe it would have been better state at the beginning of the e-mail that you can call a phone number [...] to get the content and the correctness, the legitimacy of the email confirmed ..." [P34] |

| | |
|---|---|
| Blocklisting through browser | "Finally, it might also be a possibility [...] that Chrome, that Google actively maintains lists somewhere, where the pages with malware, um, are stored, so that Chrome issues a warning before you visit the page, so that you actually take care of that. That you notify the institutions, like Google, [...] where a malware page is, and this [page] will then be blocked in Chrome." [P4] |
| one responsible entity | "It said, if there are doubts or any questions, one can get in touch with them. So I called them and they told me, well, we're an association and private individuals, this number would only be for companies if they have a problem, and since I'm calling on behalf of an association, it doesn't affect them at all, so you'd have to get in touch with someone else, but they wouldn't know exactly who to get in touch with. [...] And then I thought, ok, [...] if I call there at the hotline and then I am so carelessly discarded, then the problem can't really serious." [P12] |
| | "Because actually is a topic for the Federal CERT. An then [...] this is something were you immediately recognize, this is someone you can ask [...] you can also turn directly to the Federal CERT, which is an official body, they deal with such things and they give you advice [...] then it doesn't have this sales flavor." [P33] |

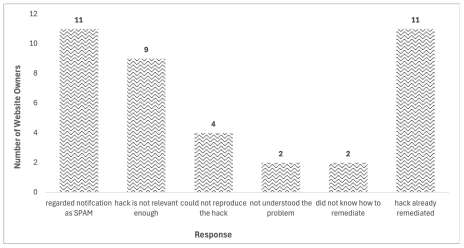
Table 5. Summary of recommendations for future vulnerability notifications




(a) Responses Notes.




(b) Responses Interviews.




(c) Responses of Website Owners who read our Notifications.

Fig. 6. Illustration of the Results of the Phone Calls.