A Field Study of Computer-Security Perceptions Using Anti-Virus Customer-Support Chats

Mahmood Sharif [*,†] **Kevin A. Roundy** [‡] **Matteo Dell'Amico** [‡] **Christopher Gates** [‡]
Daniel Kats [‡] **Lujo Bauer** [†] **Nicolas Christin** [†]

ABSTRACT

Understanding users' perceptions of suspected computer-security problems can help us tailor technology to better protect users. To this end, we conducted a field study of users' perceptions using 189,272 problem descriptions sent to the customer-support desk of a large anti-virus vendor from 2015 to 2018. Using qualitative methods, we analyzed 650 problem descriptions to study the security issues users faced and the symptoms that led users to their own diagnoses. Subsequently, we investigated to what extent and for what types of issues user diagnoses matched those of experts. We found, for example, that users and experts were likely to agree for most issues, but not for attacks (e.g., malware infections), for which they agreed only in 44% of the cases. Our findings inform several user-security improvements, including how to automate interactions with users to resolve issues and to better communicate issues to users.

CCS CONCEPTS

• **Security and privacy** → **Usability in security and privacy**; Intrusion detection systems; • **Human-centered computing** → *Field studies*;

KEYWORDS

Computer security; user (mis)perceptions; customer support

ACM Reference Format:
Mahmood Sharif, Kevin A. Roundy, Matteo Dell'Amico, Christopher Gates, Daniel Kats, Lujo Bauer, and Nicolas Christin. 2019. A Field Study of Computer-Security Perceptions, Using Anti-Virus Customer-Support Chats. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300308

*Work partially done as an intern at Symantec Research Labs.
†Carnegie Mellon University, Pittsburgh, PA, USA.
{mahmoods,lbauer,nicolasc}@cmu.edu.
‡Symantec Research Labs, Culver City, CA, USA.
{<firstname>_<lastname>}@symantec.com.

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05.
https://doi.org/10.1145/3290605.3300308

1 INTRODUCTION

Over the past two decades, usable security researchers have explored users' perceptions of computer security, uncovered users' struggles, and suggested interventions to help users become more secure. For example, researchers found users to be vulnerable to phishing and proposed forms of education to help them become more secure [39, 40]. As another example, researchers found users ignore or bypass security warnings, and explored designs to increase compliance [5, 14, 30, 32].

However, except for a few field studies that were either restricted to small populations [16, 21], or explored specific security behavior such as warning compliance [32], the majority of prior studies took place in controlled environments (e.g., online studies, in-lab interviews, etc.). Controlled environments, including those used in experiments that explored home users' perceptions of computer-security problems (e.g., [7, 8, 19, 44]), lend themselves better to analysis, but limit the ecological validity and generalizability of the findings, as the participants' context is unlike what it would be when encountering real security problems. For example, participants may describe security problems and their consequences in abstract terms [3], impacting research findings.

We attempt to fill this gap by exploring users' perceptions in the context of the security problems that they faced in practice. Specifically, we carried out a field study of users' perceptions of suspected computer-security problems using problem descriptions that were submitted to the customer-support desk of a large anti-virus (AV) vendor. Using 189,272 problem descriptions submitted from 2015 to 2018, we set to answer the following questions: *1)* How do users conceptualize computer-security problems?; *2)* Do users accurately diagnose security problems?; *3)* What are the most pressing security problems that users need help with?; and *4)* In what ways should we improve technology to better protect users?

Using qualitative methods, we manually analyzed a total of 650 problem descriptions. We first analyzed 150 problem descriptions to understand how users conceptualize and diagnose security problems. We found that users perceived a variety of symptoms (e.g., AV and non-AV warnings) that led them to suspect various security issues (e.g., infections and

scams). In the diagnosis process, users sometimes blamed the issues on certain parties or events (e.g., hackers or operating-system upgrades).

After discovering the themes in the data, we coded additional 500 problem descriptions according to users' perceived symptoms and issues, as well as according to experts' diagnoses of the issues. We then measured the prevalence of and relationships between symptoms and issues, and the agreement between users and experts when diagnosing issues. Among other findings, we were surprised to discover that users' diagnoses generally agreed with experts' for most issue types. The only exception is for attacks (e.g., malware infections)—users often misdiagnosed scams and potentially unwanted applications [20], attributing them to malware. Another example of an unexpected finding is that users who were uncertain about their diagnoses were more likely to agree with experts than those who were certain.

Building on our findings, we recommend several interventions to improve users' security (e.g., methods to automate interactions with users to provide better support), and conduct a preliminary evaluation of their feasibility and impact.

Next we discuss the related work (Sec. 2) and our methodology (Sec. 3). Then, we present our results (Sec. 4) and recommendations (Sec. 5), before concluding (Sec. 6).

2 RELATED WORK

Through online studies, surveys, and interviews, researchers explored users' perceptions of computer security (e.g., [7, 8, 19, 24, 44]). For example, Wash interviewed home users to study folk models of computer security [44]. He identified different ways in which users conceptualize malware and hackers, and concluded that certain security threats are difficult to eliminate since they leverage users' misperceptions. While we identified some of the models that Wash found, differently from Wash and others, we studied users' perceptions of security problems using field data provided to us by the customer-support desk of a large AV vendor.

Researchers also compared experts' and non-experts' security practices and perceptions [2, 3, 8, 19], and observed several differences. For example, experts mostly reported using password managers and installing updates to remain secure, while non-experts reported using AVs and visiting known websites [19]. In contrast, we found that users and experts often agree when diagnosing security issues. Nonetheless, disagreements might lead to severe consequences (e.g., users might be susceptible to scams).

Research studying the usability of security tools explored users' perceptions of firewalls [31]. It found that users were unaware of firewalls' functionality and role at protecting their devices. In contrast to firewalls, users acknowledge the role of AVs, with 86%–93% of study participants reporting to have AVs to protect their devices [17, 27]. Nevertheless, users who have AV may behave less securely than others—as they have higher likelihood to visit malicious websites or have infected devices [9, 16, 23, 38]. Our investigation led us to conjecture this happens because users often believe that AVs are foolproof (see Sec. 4).

Prior work found that security warnings often lead to habituation, and explored ways to increase user compliance and help users respond in the safest manner [5, 14, 30, 32]. We found that users contacting customer support often suspect problems due to warnings. The warnings, however, sometimes ignored recommended design guidelines (e.g., [4]).

Prior work also identified that users are often misled by online and telephony scams (e.g., [10, 28, 34, 39]). Technical support scam [28]—scammers deceiving users to believe that they need support to resolve security problems for monetary gain—is particularly common in our dataset. Our unique exposure to customer-support data helped us estimate the effectiveness of various defenses.

Several research groups studied help-desk data, as well as the potential effect of various security interventions on the workloads of help desks [1, 6, 11, 26, 35]. For example, Colnago et al. used help-desk tickets to explore the usability issues of a two-factor authentication system [11]. As another example, Mercuri et al. studied the impacts of integrating password managers into Boston's Children Hospital and found that it reduced help-desk calls by 80% [26]. Differently from prior work that explored specific security-related questions, we studied users' general security perceptions, and characterized the symptoms and issues that users perceived and the connection between them.

3 METHODOLOGY

In this section we describe the dataset and our analyses methods, as well as the limitations of our study.

Dataset

For the purpose of this study, we worked with Symantec, the vendor of the Norton Security endpoint-protection software, to analyze its customer-support data. Symantec provides users of its AV with several support mechanisms—users can chat with agents online, call the support desk via phone, view frequently asked questions (FAQs) and tutorials, or discuss problems on an online forum. We studied chat data, in accordance with Symantec's privacy policy [42]. Unlike FAQs and tutorials, chat data contains statements volunteered by users, which helped us learn about their perceptions; and unlike phone and forum data, it is already transcribed or is conveniently formatted in a database to simplify our analysis. Chat is also a relatively popular means for users to contact support: ~40% of users who contact support rely on chat, which is slightly less than the percentage of users who contact support by phone.

More specifically, we analyzed the problem descriptions users provided as their opening statements when initiating chat sessions. While the complete chat transcripts contain richer information about the users' problems, it would have been infeasible to study a large number of transcripts and get a broad picture of the range of problems that the users faced. Furthermore, users volunteered these problem descriptions solely in response to the request "please describe your issue," which makes the descriptions less susceptible to priming by customer-support agents.

The dataset we analyzed contains problem descriptions submitted between July 2015 and June 2018. When problem descriptions are first provided, new cases open. In future chat sessions, users can provide case IDs to discuss old cases (e.g., with different agents). We only used data from the first chat sessions of cases. In addition to problem descriptions, which are limited to 255 characters, cases in the dataset are accompanied by information about when chat sessions began or ended, AV versions owned by the users, users' account IDs, chats' languages, and two-level categorizations of the issues introduced by agents. The categorizations' first level is general, indicating whether the problem is related to account management, security, etc. The second level of the categorizations is more specific. For example, for security problems it indicates whether the problem is related to scam, ransomware infection, Trojan infection, etc. We found the second-level categorizations highly unreliable, potentially because they are introduced in an ad-hoc manner by customer-support agents, instead of being rigorously developed. A researcher from our group manually analyzed 400 problem descriptions and estimated that as many as 37% of the chats were assigned inaccurate second-level categories, and so we decided to ignore them. Using the high-level categories and the language field, we pre-filtered the data to keep computer-security related problem descriptions written in English. After filtering, 189,272 cases remained, representing ~5% of the overall chat volume (the rest of the cases were mostly related to purchases and account management).

Since we selected problem descriptions in English, most of the customers were from the US (77%), with some from the UK (9%) and Canada (4%). The customers contacted the support desk using devices equipped with Windows (87%), Mac OS (6%), Android (5%), and iOS (2%) operating systems, all of which are supported by the AV product. One caveat when counting operating systems, however, is that customers may report issues on devices other the ones used to contact the support desk.

Data Analysis

We borrowed techniques from grounded theory [13, 41] to understand users' conceptual models of security issues. In particular, three researchers, whose computer-security experience ranged from eight to 16 years, worked jointly to open-code randomly sampled problem descriptions, and met periodically to consolidate codebooks and develop a theory that describes users' conceptual models. While coding, the researchers continuously identified and refined themes, and maintained codebooks and memos to define codes and themes. After coding 150 problem descriptions, the theory took shape (i.e., themes and relationships between them ceased to change), and the open-coding process converged (i.e., new codes stopped being added).

With the help of the codebooks that they developed, the three researchers coded a total of 500 randomly selected problem descriptions along five dimensions, such as the perceived symptoms and issues, and the expert (i.e., researcher) diagnosis of issues (see Sec. 4 for more details). We selected 500 problem descriptions because the open-coding process suggested that that many were sufficient to encounter each code multiple times, while the analysis remained feasible. Each coder coded 150 distinct problem descriptions, in addition to 50 that were common to all coders to estimate inter-coder agreement. Fleiss' Kappa [15] statistic for the different dimensions ranged from 0.59 to 0.63—indicating substantial agreement [25].

To estimate what issue types users were likely to suspect given perceived symptoms, as well as the level of agreement between users and experts when diagnosing issues, we used the χ^2-test of independence (to measure statistical significance) and the odds-ratio statistic (to measure likelihood) [37]. Additionally, we tested the feasibility of applying machine learning to automate interactions with users and detect emerging issues that affect several users at a time (see Sec. 5). To this end, we used the `Doc2Vec` representation [22]—a representation of text as a multi-dimensional vector that is learnt in an unsupervised manner—and a recently proposed clustering algorithm, `TaxoGen` [45]. We used all 189,272 problem descriptions to train the machine-learning algorithms.

Limitations

Our findings should be interpreted in the light of a few limitations. First, the dataset we analyzed was collected by a single AV vendor. Thus, some findings may be specific to the vendor. For instance, differences in the clarity of vendors' warnings may be more or less apt to confuse users. The same is true with respect to the actionability, specificity, and the nomenclature of alerts, as well as how prone the AV is to incur false positives or negatives (i.e., classifying benign or malicious software as malicious or benign, respectively).

Second, the three researchers diagnosed security issues from problem descriptions only. Their assessment may not always be correct, as they did not have access to users' devices.

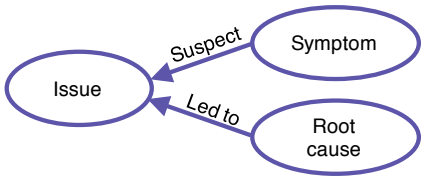

Figure 1: When describing security issues, users usually express the perceived symptoms, which lead them to suspect the (ostensible) issues. Sometimes, users also mention what they believe to be the root cause behind issue.

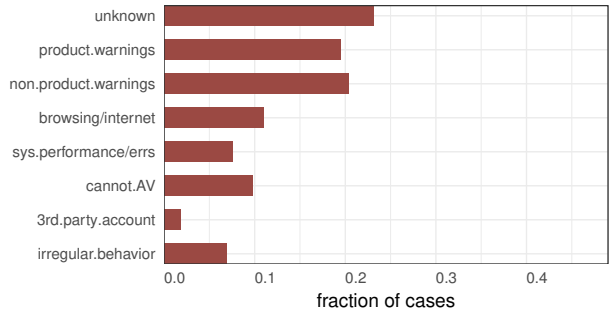

Figure 2: Fraction of cases in which perceived symptoms were encountered.

Yet, given the high inter-coder agreement among the researchers, it is likely that their assessment is accurate. Stated differently, it would have been unlikely for them to achieve high agreement while misdiagnosing cases often.

To further assess the accuracy of coding problem descriptions, we used a sample of complete chat transcripts. Specifically, two researchers coded and reconciled 10 problem descriptions, and used the corresponding chat transcripts to validate the codes. Four transcripts provided information supporting the codes–e.g., the customer-support agent's assessment in one transcript supported the coders' diagnosis of potentially unwanted software [20] (see Sec. 4), as opposed to the customer's perception of malware infection. The other six transcripts provided no additional information to support or refute the codes (mostly due to costumer-support agents not eliciting additional information, but rather moving quickly to connect remotely to customers' machines).

Third, our dataset is not perfectly filtered. Some cases in the dataset were mislabeled: we found some not to be security related (~12% were account related), and others (~1.6%) to be non-applicable for coding, as the problem descriptions were not in English, or were unintelligible. As we were dealing with field data, such imperfections are not unusual.

4 USERS' PERCEPTIONS

Fig. 1 presents the theory that describes the users' conceptual models, as found by analyzing the initial set of 150 problem descriptions. The theory consists of three themes: *issues, symptoms, and root causes*. Issues are users' perceived security problems, as expressed in problem descriptions. Symptoms are the tangible manifestations of the (ostensible) issues—they lead users to suspect issues. Often, the issues that users suspect are a result of misconceptions, potentially due to lack of technical expertise. We discuss such misconceptions later in the section. Root causes describe the entities or actions perceived as being responsible for causing the issues.

The users described issues with varied levels of certainty. Some users were certain about their suspicions, others were uncertain, while the rest simply inquired about the agents' opinion. More concretely, uncertain users often expressed doubt about their diagnoses, using phrases such as "I believe" and "I think," or explicitly asked agents for help in diagnosing security issues. Users who inquired about the agents' opinion usually presented the symptoms without hypothesizing about issues and let the agents diagnose the issues, or simply asked general questions unrelated to active security issues. In contrast, users who were certain provided their diagnoses assuming they were correct and sought help from the agents to resolve the issues they diagnosed.

As mentioned in Sec. 3, the three researchers who developed the codebooks coded 500 randomly sampled problem descriptions along five dimensions: *1)* perceived symptoms, as expressed by users; *2)* perceived issues, as suspected by users; *3)* how certain users were about the issues (certain, uncertain, or inquiring); *4)* experts' (i.e., the researchers') diagnoses of what actual issues (if any) were most likely; and *5)* how certain the experts were about their diagnoses (certain or uncertain). We highlight again that the experts' diagnoses did not result from inspecting users' machines, but rather the experts' opinions of what actual issues were most likely, based on the problem descriptions. As the agreement levels between the experts was substantial and for the majority of cases (~73%) they indicated being certain about their diagnoses, we believe that their diagnoses reflect the actual issues that the users faced. This is further corroborated by the additional validation using complete chat transcripts (see Sec. 3). We elected not to code problem descriptions along the root-cause dimension, as root cause was rarely mentioned (only ~37% of the users mentioned a root cause).

After removing problem descriptions that were not applicable for coding (due to being empty, unintelligible, or written in languages other than English), 492 problem descriptions remained. We used the codes to estimate the prevalence of perceived symptoms and issues, as well as the actual issues (from the experts' diagnoses)—see Figs. 2–3. Further, we analyzed which issues users were likely to diagnose when facing different symptoms. Last, we measured for which

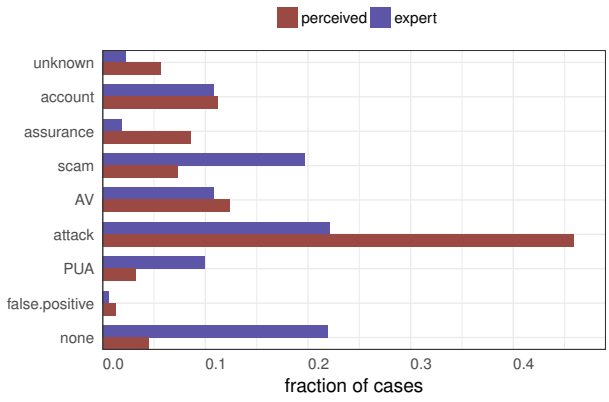

Figure 3: Fraction of cases in which issues were perceived by users (red) or assigned by experts (blue).

types of issues users' diagnoses were most likely to agree with experts'.

Next, we present the concepts (i.e., codes) that we encountered for each of the symptoms, issues, and the root causes. Afterward, we present the relationship between perceived symptoms and users' and experts' diagnoses, and discuss the agreement between users and experts.

Symptoms

Unknown. In ~23% of problem descriptions, users did not describe any symptoms. We marked the symptoms in such cases as "unknown."

Product warnings. These are warnings originating from the AV software or the vendor. The warnings may be delivered in various forms, including alerts on users' devices, emails to the addresses registered with the vendor, or alerts on the vendor's website. The warnings may carry different messages, including latest scan results, possible infection alerts that may require users to take action (e.g., selecting whether to remove or quarantine a malware), and account renewal information. Product warnings account for ~20% of the perceived symptoms, making them one of two most tangible symptoms indicating issues to users.

Non-product warnings. Unlike product warnings, non-product warnings do not originate from the AV or the vendor. The warnings may be delivered via email, phone calls, or other system components (e.g., via a browser or other software). Some of non-product warnings are legitimate (e.g., messages from Internet service providers or online social networks indicating bot activity originating from the users' accounts [12, 18]), but most are illegitimate (e.g., persistent pop-ups accompanied with sound, which are indicative of technical-support scam [28]). In certain cases, the legitimate non-product warnings instructed users to download AVs from competing vendors, which led to confusion. Non-product warnings were one of the two most common symptoms, accounting for ~20% of the problem descriptions.

Browsing or Internet related. Users reporting these symptoms were unable to connect to the Internet, or suffered from repeated events that negatively affected their browsing experience. Such events consisted of repeated pop-ups showing ads, frequent redirections, or inability to change the home page.

System performance or errors. Here, users reported slow systems, software or operating systems that froze, or complete system failures (e.g., black or blue screen upon booting).

Cannot perform an AV action. This concept identifies cases in which users were unable to perform certain actions with the AV. In particular, some users were unable to install or activate the AV, some were unable to scan their device, while others were unable to find the AV's icon on the notification bar or the desktop and so were not sure whether they were being protected.

Third-party account. Some users reported inability to log into their online accounts on third-party services (e.g., email or social media) or spotting suspicious behavior on these accounts (e.g., malicious emails or messages sent to contacts).

Irregular behavior. This concept covers problem descriptions with highly unusual symptoms not covered above. For example, in one case a user reported that the character '2' was being typed indefinitely, and in another case a user reported a suspicious message written on her screen.

Issues

Unknown. In ~6% of the cases, users did not express perceived issues. We marked the issues in these cases as "unknown."

Account. Users reporting account-related issues were having trouble with subscriptions, logging-in to download the AV, or activating it. While we attempted to pre-filter the dataset to avoid cases with such issues (see Sec. 3), still ~11% of the remaining cases were account related due to analysts erroneously marking the cases as security related.

Assurance. In some cases, users felt that their devices were not being protected (e.g., because the AV's icon in the notification bar was hidden). Thus, the users contacted the support desk to get reassured about their protection status. In most cases, the users' devices did not suffer from security issues.

Technical support and fake AV scam. Users contacting the support desk about scams reported being scammed or suspected potential scams. The scammers contacted the users

via persistent pop-ups that were hard to close (often accompanied with audio) [28], email, or phone to convince users to remove purported malware from their machines. The scammers would usually instruct the users to download fake AVs, or to contact the scammers' alleged customer-support desks on toll-free numbers, and would ask for payment. Users reporting or suspecting scams were consistently accurate—i.e., they identified real scams. Nevertheless, there is a large discrepancy between the number of cases that users identified as scams (~7%) and those that experts identified as scams (~20%), indicating that scams often misled users.

AV. When dealing with AV-related issues, users mainly contacted the support desk to inquire about how to perform certain actions, or to report potential bugs in the AV software. Of the users who were unable to perform actions with the AV (e.g., scan, update, or resolve warnings), ~55% contacted the support to inquire about how to perform the actions, while ~15% attributed their inability to perform the actions to bugs. Of the users who perceived irregular behavior (e.g., flashing screen), or system performance issues (frozen or slow machines), ~14% attributed the symptoms to bugs.

Attack. Users suspecting attacks accounted for the majority of problem descriptions (~46%). Most users reporting attacks blamed malware (~44%), while some blamed hackers for remotely accessing their devices (~2%). Interestingly, users suspected attacks in nearly all situations in which email or online social network accounts behaved abnormally, such as by sending malware or spam to contacts, not considering the possibility that their passwords could have been guessed, brute forced, or (most likely) revealed in a breach [43], without unauthorized access to their devices. Some users reported receiving spam emails in a belief that the AV should have blocked such emails and that malware was likely to blame. While most purported infections were reported to have occurred after the installation of the AV, in a few cases, the users installed the AV after having been compromised. In contrast to users, experts diagnosed only ~22% of all problem descriptions as actual attacks. We explain the reason for the discrepancy in the last part of this section.

Interestingly, some users expressed surprise that their devices might get infected despite having AVs, thus expressing belief that AVs are foolproof. In reality, users who ignore AV warnings or perform risky behavior may still infect their devices. The users' belief that AVs are foolproof might explain why users who have AVs often have higher rates of infection and exposure to malicious content than others [9, 16, 23, 38].

To better understand what causes users to contact the support desk about attacks, a single researcher analyzed 100 problem descriptions that were coded as attacks. The largest category of cases involved situations in which the AV was unable to seamlessly remove infections. A closer look at the drivers behind these 29 cases reveals that in mid-October of 2016 the `Kovter` malware family used new persistence techniques to hide itself in the system registry [33], and while the AV successfully detected `Kovter`'s presence, it was unable to successfully remove it for a time. The result was a large spike in customer-support calls, which dropped quickly once the AV adapted.

In addition to these 29 calls, 24 users called to ask how to clean up infections that were detected on their systems. In all but three, the AV took preventative actions, but failed to reassure the users that all was well on their systems. Additional 15 cases of the ones marked as attacks by experts involved direct indications of successful attacks, which included ransomware, system anomalies resulting directly from attack sources described by the users, and the presence of specific malware families mentioned by the users. Finally, 18 users generically reported attacks while providing negligible additional detail, with statements like "I have a virus" or "someone has tried to hack my computer," while the final 14 users reported suspicious symptoms (e.g., Internet failure, software and software-update failures, and other irregular behavior) that convinced experts that an attack was the most probable issue, although the experts were generally uncertain about their diagnoses in these instances.

Potentially unwanted application (PUA). Differently from malware, which exhibits clear malicious actions (e.g., encrypting files on the machine or participating in denial-of-service attacks), PUAs exhibit undesirable actions (e.g., showing intrusive ads) that may not be considered objectively malicious [20]. Certain toolbars and extensions that show ads, as well as repackaged open-source software are often considered PUAs. As AVs attempt to avoid false positives (i.e., marking non-malicious applications as malicious) to avoid harming user experience, some PUAs are not blocked. This is exacerbated by the fact that what one user may consider to be a PUA, another user may consider to be a good application. The end-user license agreements for many PUAs also explain what they do, making it legally tricky to outright block PUAs or their components.

In ~3% of the 492 problem descriptions, users contacted the support desk to report PUAs on their system, and to inquire about ways to remove them. In contrast, the experts identified ~10% of the problem descriptions as PUA related. The discrepancy is because users often identified PUAs as malware, as we discuss below.

False positive. In just above 1% of the problem descriptions, users believed that the AV was erroneously preventing them from installing benign software—i.e., they believed the AV incurred false positives. Experts agreed with the users in half of those cases, judging that the AV might have been overly aggressive in marking PUAs as malicious.

None. In ~4% of the cases, users did not contact the support desk to report active issues, but rather to inquire about the AV (e.g., whether it can be downloaded from Apple's App Store), or to ask for help with issues unrelated to security (e.g., retrieving missing files). In contrast, the experts deemed that ~22% of the cases that users called about did not involve active security issues. In most of these cases users perceived attacks (rather than one of the other issues mentioned above) but the experts did not find security issues. Often, we found, confusing warnings misled users to believe they suffered from an attack (~5% of cases overall).

Root Causes

In ~37% of the cases, users attributed issues to entities or actions that they perceived caused the issues. Some users, especially those who perceived unauthorized remote access, believed that hackers were responsible. Such users viewed hackers as "burglars who break into computers for criminal purposes," as described by Wash [44]. Users who thought they were being scammed sometimes identified illegitimate firms (e.g., "Global Technical Support") as responsible for the scams. Some users who suspected attacks or PUAs attributed the issues to malicious software (e.g., Fragilepottery), toolbars (e.g., Delta search), malicious websites they visited, or emails they received. In certain cases, users attributed their inability to use the AV to recent operating-system upgrades, license renewals, or the purchase of new computers.

Relationship Between Symptoms and Issues

To better understand users' thought process, we analyzed what issues they were likely to suspect when observing certain symptoms. We found a strong dependency between user-perceived symptoms and user-perceived issues (χ^2=250, p<0.01, dof=42). To characterize the nature of this dependency, we tested which pairs of perceived symptoms and issues are dependent (correcting for multiple comparisons using Bonferroni correction). For each pair that is statistically significantly dependent, we used odds ratio to measure the likelihood of the issue being suspected by users when the symptom is perceived. As a baseline to compare with, we selected the odds of the issue being suspected when users could not perform actions with the AV.

The results are shown in Fig. 4a. For all symptoms users' odds of suspecting an attack were significantly higher than the baseline. This finding indicates that users may have concluded too quickly that attacks were taking place. This is especially dangerous when users faced non-product warnings, potentially from scammers. In those cases, users' odds of suspecting attacks were roughly ten times higher than the baseline. In fact, ~66% of users facing non-product warnings suspected attacks.

Experts' diagnoses were even more deterministically linked to the symptoms than users'—as one would expect. The experts' judgments exhibited strong dependencies between symptoms and issues (χ^2=476, p<0.01, dof=42; see Fig. 4b). Attacks had higher odds than the baseline only in the cases of product warnings, symptoms related to third-party accounts, and irregular behavior. Non-product warnings had higher odds of leading experts to diagnose scams than the baseline (~73% of non-product warnings were diagnosed by experts as resulting from scams), while browsing and Internet-related symptoms led to experts' diagnoses of PUAs (mostly due to the intrusive ads that PUAs introduce). Experts exhibited high odds to suspect that no security issues existed in cases of low system performance and errors, symptoms related to third-party accounts, and irregular behavior. In fact, experts judged that in ~69% of cases in which users perceived low system performance and errors they did not suffer from security issues. In those cases, the experts believed, the symptoms could be explained by hardware (e.g., failing disks) or resource-exhaustion issues (e.g., too many processing were running).

When Are Users and Experts Likely to Agree?

Next, we studied whether users' and experts' diagnoses were likely to agree and in what situations. Initially, we studied the dependencies between issues identified by experts and issues suspected by users. The dependency between the two is strong (χ^2=1,360, p<0.01, dof=49). Then, to characterize the dependency, we tested for pairs of expert-diagnosed issues and user-perceived one whether they are dependent (again, using Bonferroni correction to correct for multiple tests). For each pair that is statistically significantly dependent, we used odds ratio to measure the likelihood of the expert-diagnosed issue given the user-perceived issue. As a baseline to compare with, we selected the odds of the expert-diagnosed issue when users perceived an AV-related issue.

The results are shown in Fig. 5. One can see that the diagonal values are the highest for most issues. This indicates that the odds of experts agreeing with users were high. For instance, when users perceived scams, they were likely to be in agreement with experts—the odds that experts would diagnose scams when users did so were 5×10^8 times higher than the baseline (i.e., experts' odds to diagnose scams when users perceived AV-related issues).

The agreement between experts' and users' diagnoses was high, but nevertheless imperfect. Differently from other issues, when users perceived attacks, experts were often likely to disagree. The experts diagnosed attacks only in ~44% of the cases in which users perceived attacks. In ~18% of cases in which users perceived attacks, experts diagnosed no active security issues, with roughly ~25% of the disagreements

Figure 4: Odds of issues being suspected by users (a) or experts (b) when certain symptoms were perceived compared to when the user could not perform actions with the AV (i.e., "cannot.AV" is the baseline). For example, users were seven times more likely than not to perceive a attacks when they perceived browser and Internet-related symptoms compared to when they could not perform actions with the AV. White tiles indicate statistically insignificant dependencies between symptoms and issues ($p \geq$0.05).

resulting from confusing warnings. Warnings for network-based threats where the nature and source of the attack appeared to be unknown to users were especially confusing. Similarly, repetitive warnings that did not clearly suggest possible actions and warnings instructing users to download additional removal software also confused users.

In ~10% of cases in which users perceived attacks, experts diagnosed PUAs on the users' devices. Conflating PUAs with attacks indicates that some users may need AVs to handle PUAs more aggressively (e.g., by putting them in quarantine, limiting their access, or removing them completely).

Most users' misdiagnoses were caused by scams, which are specifically designed to exploit misconceptions. For ~26% of the cases suspected to be attacks by users, the experts diagnosed scams. Those cases can be extremely harmful, as the users expressed no skepticism about the scammers' assertions that their systems were compromised. Indeed, we found that within a sample of 100 chats coded by experts as scams, five users made payments to scammers before contacting the support desk, and of these, three realized the deception prior to the support chats, while two were still under the illusion that they had paid legitimate parties. Out of the 100 users, 48 contacted the support desk thinking that the scams were genuine. For those, it seems likely that the availability of support desk played a meaningful role in reducing the impact of the scams. However, some users fall prey to support scams without contacting customer support (e.g., to erroneously complain about the excess charges). We observed that the dominant delivery mechanism affecting users were persistent pop-ups, which sometimes included auditory elements. AVs, and security products in general, have a potential to protect users by preventing the delivery mechanisms. Still, in ~7% of the scam cases that were diagnosed by experts, the scams were delivered through phone calls, hence limiting the ability of security products running on users' devices to protect against the scams.

Of the users contacting the support desk, ~55% were certain about their diagnoses, ~14% were uncertain, and the rest were inquiring. Surprisingly, we found that the odds of agreement between experts' diagnoses and users' were ~1.7 times higher for uncertain users (agreeing ~52% of the time), than for certain users (agreeing ~39% of the times). The dependency is marginally statistically significant (χ^2=3.76, p=0.05, dof=1). In practice, this indicates that agents should question users' diagnoses, particularly if they are conveyed with certainty.

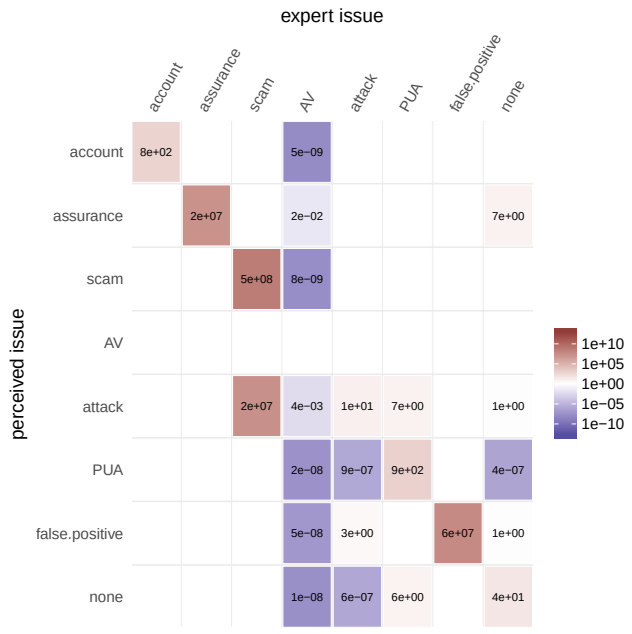

Figure 5: Odds of issues being diagnosed by experts when certain issues were perceived by users compared to the odds of the same issues being diagnosed when users perceived AV-related issues (i.e., the "AV" row is the baseline). White tiles indicate statistically insignificant dependencies ($p \geq 0.05$)

5 RECOMMENDATIONS

We build on our findings to recommend several interventions to improve user security, and report on preliminary experiments evaluating the feasibility of certain recommendations.

Making Customer Support More Effective. Our findings show that perceived symptoms and issues are two main elements of problem descriptions that are often provided, and are strongly correlated with expert issues. These findings suggest that automatically extracting symptoms and predicting expert issues from problem descriptions may be feasible. Solving these problems can enable certain mechanisms to make customer support more effective, including: *1)* assigning cases to qualified agents that are trained to give support in certain topics (e.g., assigning attack-related cases to agents trained to remove malware); *2)* retrieving the most relevant FAQs and tutorials that can help users solve issues or can educate them about certain topics (e.g., scam); *3)* automatically targeting questions to ask users in order to help agents diagnose issues. Such mechanisms may reduce the number of users in need for live support and make the support process faster.

As an early investigation to test the feasibility of our recommendation, we trained and evaluated machine-learning algorithms to automatically extract symptoms and predict experts' diagnoses from problem descriptions. Specifically, we applied `Doc2Vec` [22] on the dataset containing the 189,272 problem descriptions to convert each problem description into a feature vector of length 50. Subsequently, we trained and evaluated standard multi-class logistic regression classifiers using the 492 problem descriptions that we coded. Ten rounds of 10-fold cross-validation resulted in 60% ± 1% accuracy for expert-issue prediction, and 51% ± 1% for perceived symptom extraction. These results are promising, as they are already ~5 and ~4 times better than random assignment, respectively. We believe that one way to improve the accuracy would be through increasing inter-coder agreement and coding more problem descriptions to increase the training set's size. Furthermore, it would be useful to encourage users to divulge perceived symptoms to aid in the diagnoses. In our case, we found that a substantial fraction of users (~23%) did not provide symptoms. To mitigate this, we suggest updating the input form to ask users not only about the problem description in general, but also specifically about the symptoms they perceive.

Trend Identification. As users' diagnoses often agrees with experts', we hypothesized that there is potential in leveraging the customer-support data for detecting trends. To put this hypothesis to the test, we applied `TaxoGen` [45] on a dataset of 210,850 problem descriptions, submitted between July 2015 and August 2018 (this dataset includes additional two months of problem descriptions compared to the dataset presented in Sec. 3). `TaxoGen` creates, in an unsupervised manner, a taxonomy of keywords from a corpus of text data, resulting in a hierarchical clustering where each cluster contains a set of documents and keywords.

Examining the clusters manually, we found that certain clusters contain problem descriptions related to attacks, while others are centered around scams or AV-related issues (please refer to the Supplementary Material for a deeper overview). For example, problem descriptions in attack-related clusters contain names of malware and vulnerabilities as part of the most prominent keywords (e.g., `Kovter`, `Meltdown`, `WannaCry`, etc.). We found that `TaxoGen` provides a sensible clustering for keywords with as little as 10 mentions. Hence, analyzing the spikes for keywords together with the clustering associated with the keywords appears to be a reasonable way to detect and categorize trends. This could be used to inform the customer-support desk and other teams (e.g., the AV developers) about abnormal situations that require action. For instance, in the case of `Kovter` (which led to an increase in customer-support chats due to not being removed properly for a certain period of time, see Sec. 4) a cluster mostly of recent problem descriptions related to malware may inform malware outbreaks and/or AV issues.

Defending Against Scams. Technical support and fake AV scams are prevalent, and pose a real threat to users. We estimated that ~20% of the users faced scams, but only ~7% of the users actually suspected scams. The scammers even managed to extort money from some users. The prevalence and risk of scams require us to take action. Miramirkhani et al. [28] proposed interventions at two levels—user education and improving browsers' design.

Further research is needed to estimate the effectiveness of user education, as it depends on various factors, such as the education method [40]. Nevertheless, as prior work proposes [28], there seems no harm in educating users not to trust webpages that claim that the users' devices are infected (as webpages cannot scan devices, by browser design), and persistent webpages that are difficult to close (as legitimate webpages are unlikely to exhibit such behavior). We found that users in our study who followed these guidelines or similar to identify scams did so successfully—experts diagnosed scams in ~92% of the problem descriptions in which users suspected scams.

As a technical solution, prior work proposed to update browsers' design to make it difficult for intrusive webpages to persist [28]. As a complementary defense, one may consider providing AVs with a functionality to detect persistent pop-ups and help users close them. In addition to technology for detecting illegitimate emails (which are sometimes used to deliver scams), such technical approaches have the potential to prevent ~93% of the scams in our dataset that did not take place over the phone. For the remaining ~7% that were delivered by directly calling potential victims, other interventions would be required (e.g., smartphone apps to flag or block such attempts).

Adapting AVs' Aggression Levels to Users. Our findings show that, on the one hand, PUAs are often considered as malware by users. On the other hand, a few users considered cases in which PUAs were flagged by the AV to be false positives. Traditionally, AVs try to avoid false positives, and so generally avoid blocking PUAs. However, as emerged from our analysis, such a one-size-fits-all approach may not satisfy all users. Instead, AVs may improve users' experience by distinguishing PUAs from benign and malicious software [20], and reacting more (e.g., by putting PUAs in quarantine) or less (e.g., by allowing PUAs to run) aggressively to PUAs depending on users' profiles (e.g., profiling users by the software they have installed [29], knowledge [31], or culture [36]).

More User-Friendly AVs. By studying users' problem descriptions, we discovered three avenues in which the AV can be improved to become more user friendly. First, warnings should convey messages more clearly to users, potentially suggesting actionable steps. For instance, in the case of `Kovter` [33], when the AV failed to remove a newer version of the malware, it started to repeatedly show the same warning to users without suggesting a course of action. Such warnings could lead to confusion, or even worse, to habituation [5]. Suggesting actionable items (e.g., pointing users to tutorials or documentation, or suggesting that they contact support), could help prevent confusion and habituation [4]. Second, the AV should be self-contained and should not request that users download an additional software to remove an infection, as these requests led to confusion. Third, the AV should always make itself visible in the notification bar to clearly show the protection status, as users who did not see the icon in the notification bar sometimes required reassurance.

Who Should Warn Users? We encountered a few cases in which users were warned by legitimate parties other than Symantec (e.g., an Internet service provider via email) that their devices were infected [12, 18]. These warnings confused users, who sometimes questioned their legitimacy. The users might have expected that only the AV or its vendor should be able to warn them about potential attacks. Further research is needed to understand users' expectations, and how multiple parties in the ecosystem should interact to effectively help users remain secure.

6 CONCLUSION

We conducted a field study of users' computer-security perceptions using 189,272 problem descriptions that were provided to the customer-support desk of a large AV vendor from 2015 to 2018. We used a mixture of qualitative and quantitative methods to identify security issues that users faced and the symptoms that led them to diagnoses. Comparing users' diagnoses with those of experts, we found that users and experts agreed for most types of issues, except for attacks (e.g., malware infections). The disagreements, however, uncovered several misconceptions that may expose users to risks (e.g., falling victim to technical support and fake AV scams). Our findings inform how we can tailor technology to better protect users. For example, by automatically extracting symptoms or predicting experts' diagnoses from problem descriptions (which, as we found, can be done relatively accurately), one could assign users to trained agents for help, or even target users with specific questions to help accurately diagnose issues. The AV vendor is using our recommendations to drive further research and improvements.

ACKNOWLEDGEMENTS

We would like to thank Maria Dossin for her help understanding the dataset's format and collection process. This work was partially supported by CyLab at Carnegie Mellon University via a CyLab Presidential Fellowship and by a Symantec Research Labs Graduate Fellowship.

REFERENCES

[1] Divyangi Anchan and Mahmoud Pegah. 2003. Regaining single sign-on taming the beast. In *Proc. SIGUCCS*.

[2] Farzaneh Asgharpour, Debin Liu, and L Jean Camp. 2007. Mental models of security risks. In *Proc. FC*.

[3] Steffen Bartsch and Melanie Volkamer. 2013. Effectively communicate risks for diverse users: A mental-models approach for individualized security interventions. In *GI-Jahrestagung*.

[4] Lujo Bauer, Cristian Bravo-Lillo, Lorrie Cranor, and Elli Fragkaki. 2013. *Warning design guidelines.* Technical Report. Carnegie Mellon University, Pittsburgh, PA.

[5] Cristian Bravo-Lillo, Lorrie Cranor, Saranga Komanduri, Stuart Schechter, and Manya Sleeper. 2014. Harder to Ignore? Revisiting Pop-Up Fatigue and Approaches to Prevent It. In *Proc. SOUPS*.

[6] Sacha Brostoff and M Angela Sasse. 2003. "Ten strikes and you're out": Increasing the number of login attempts can improve password usability. In *Proc. CHI Workshop on HCI and Security Systems*.

[7] Z Byrne, M Roberts, AE Howe, Malgorzata Urbanska, and I Ray. 2012. The psychology of security for the home computer user. In *Proc. S&P*.

[8] Jean Camp, Farzaneh Asgharpour, Debin Liu, and IN Bloomington. 2007. Experimental evaluations of expert and non-expert computer users' mental models of security risks. In *Proc. WEIS*.

[9] Nicolas Christin, Serge Egelman, Timothy Vidas, and Jens Grossklags. 2011. It's all about the Benjamins: An empirical study on incentivizing users to ignore security advice. In *Proc. FC*.

[10] Nicolas Christin, Sally Yanagihara, and Keisuke Kamataki. 2010. Dissecting One Click Frauds. In *Proc. CCS*.

[11] Jessica Colnago, Summer Devlin, Maggie Oates, Chelse Swoopes, Lujo Bauer, Lorrie Cranor, and Nicolas Christin. 2018. "It's not actually that horrible": Exploring adoption of two-factor authentication at a university. In *Proc. CHI*.

[12] Comcast. 2018. Bot Detection and Removal. https://goo.gl/3ttpqb. Online; accessed 24 Dec 2018.

[13] Norman K Denzin and Yvonna S Lincoln. 1994. *Handbook of qualitative research.* Sage Inc.

[14] Adrienne Porter Felt, Alex Ainslie, Robert W Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL warnings: Comprehension and adherence. In *Proc. CHI*.

[15] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378.

[16] Alain Forget, Sarah Pearman, Jeremy Thomas, Alessandro Acquisti, Nicolas Christin, Lorrie Faith Cranor, Serge Egelman, Marian Harbach, and Rahul Telang. 2016. Do or do not, there is no try: User engagement may not improve security outcomes. In *Proc. SOUPS*.

[17] SM Furnell, P Bryant, and Andrew D Phippen. 2007. Assessing the security perceptions of personal Internet users. *Computers & Security* 26, 5 (2007), 410–417.

[18] Chetan Gowda. 2014. Making malware cleanup easier. https://goo.gl/9VzaWt. Online; accessed 24 Dec 2018.

[19] Iulia Ion, Rob Reeder, and Sunny Consolvo. 2015. "... no one can hack my mind": Comparing expert and non-expert security practices. In *Proc. SOUPS*.

[20] Platon Kotzias, Leyla Bilge, and Juan Caballero. 2016. Measuring PUP Prevalence and PUP Distribution through Pay-Per-Install Services. In *Proc. Usenix Security*.

[21] Fanny Lalonde Levesque, Jude Nsiempba, José M Fernandez, Sonia Chiasson, and Anil Somayaji. 2013. A clinical study of risk factors related to malware infections. In *Proc. CCS*.

[22] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proc. ICML*.

[23] Fanny Lalonde Lévesque, José M Fernandez, and Anil Somayaji. 2014. Risk prediction of malware victimization based on user behavior. In *Proc. MALWARE*.

[24] Michelle L Mazurek, JP Arsenault, Joanna Bresee, Nitin Gupta, Iulia Ion, Christina Johns, Daniel Lee, Yuan Liang, Jenny Olsen, Brandon Salmon, Richard Shay, Kami Vaniea, Lujo Bauer, Lorrie F Cranor, Gregory R Ganger, and Michael K Reiter. 2010. Access control for home data sharing: Attitudes, needs and practices. In *Proc. CHI*.

[25] Mary L McHugh. 2012. Interrater reliability: The Kappa statistic. *Biochemia medica: Biochemia medica* 22, 3 (2012), 276–282.

[26] Rebecca T Mercuri. 2004. The HIPAA-potamus in health care data security. *Commun. ACM* 47, 7 (2004), 25–28.

[27] George R Milne, Lauren I Labrecque, and Cory Cromer. 2009. Toward an understanding of the online consumer's risky behavior and protection practices. *Journal of Consumer Affairs* 43, 3 (2009), 449–473.

[28] Najmeh Miramirkhani, Oleksii Starov, and Nick Nikiforakis. 2017. Dial one for scam: A large-scale analysis of technical support scams. In *Proc. NDSS*.

[29] Michael Ovelgönne, Tudor Dumitras, B Aditya Prakash, VS Subrahmanian, and Benjamin Wang. 2017. Understanding the relationship between human behavior and susceptibility to cyber attacks: A data-driven approach. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 4 (2017), 51.

[30] Fahimeh Raja, Kirstie Hawkey, Steven Hsu, Kai-Le Wang, and Konstantin Beznosov. 2011. Promoting a physical security mental model for personal firewall warnings. In *Proc. CHI Extended Abstracts*.

[31] Fahimeh Raja, Kirstie Hawkey, Pooya Jaferian, Konstantin Beznosov, and Kellogg S Booth. 2010. It's too complicated, so I turned it off!: Expectations, perceptions, and misconceptions of personal firewalls. In *Proc. SafeConfig*.

[32] Robert W Reeder, Adrienne Porter Felt, Sunny Consolvo, Nathan Malkin, Christopher Thompson, and Serge Egelman. 2018. An experience sampling study of user reactions to browser warnings in the field. In *Proc. CHI*.

[33] "Symantec Security Response". 2015. Kovter malware learns from Poweliks with persistent fileless registry update. https://goo.gl/CzeSeV. Online; accessed 18 Sep 2018.

[34] Merve Sahin, Aurélien Francillon, Payas Gupta, and Mustaque Ahamad. 2017. SoK: Fraud in telephony networks. In *Proc. EuroS&P*.

[35] M Angela Sasse and Ivan Flechais. 2005. Usable security: Why do we need it? How do we get it? O'Reilly.

[36] Yukiko Sawaya, Mahmood Sharif, Nicolas Christin, Ayumu Kubota, Akhiro Nakarai, and Akira Yamada. 2017. Self-Confidence Trumps Knowledge: A Cross-Cultural Study of Security Behavior. In *Proc. CHI*.

[37] Howard J Seltman. 2012. *Experimental design and analysis.* Carnegie Mellon University.

[38] Mahmood Sharif, Jumpei Urakawa, Nicolas Christin, Ayumu Kubota, and Akira Yamada. 2018. Predicting impending exposure to malicious content from user behavior. In *Proc. CCS*.

[39] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proc. CHI*.

[40] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Anti-phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. In *Proc. SOUPS*.

[41] Anselm Strauss and Juliet M Corbin. 1990. *Basics of qualitative research: Grounded theory procedures and techniques.* Sage Inc.

[42] Symantec. 2018. Privacy policy. https://www.symantec.com/privacy. Online; accessed 22 Dec 2018.

[43] Kurt Thomas, Frank Li, Ali Zand, Jacob Barrett, Juri Ranieri, Luca Invernizzi, Yarik Markov, Oxana Comanescu, Vijay Eranti, Angelika Moscicki, et al. 2017. Data breaches, phishing, or malware?: Understanding the risks of stolen credentials. In *Proc. CCS*.

[44] Rick Wash. 2010. Folk models of home computer security. In *Proc. SOUPS*.

[45] Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han. 2018. TaxoGen: Constructing Topical Concept Taxonomy by Adaptive Term Embedding and Clustering. In *Proc. KDDI*.