Be the Phisher – Understanding Users' Perception of Malicious Domains

Florian Quinkert
Ruhr-University Bochum
florian.quinkert@rub.de

Martin Degeling
Ruhr-University Bochum
martin.degeling@rub.de

Jim Blythe
University of Southern California
blythe@isi.edu

Thorsten Holz
Ruhr-University Bochum
thorsten.holz@rub.de

ABSTRACT

Attackers use various *domain squatting* techniques to convince users that their services are legitimate. Previous work has shown that methods like *typosquatting*, where single characters are removed or duplicated, can successfully deceive users.

In this paper, we present a study that evaluates how well participants distinguish malicious from benign domains before and after they learned and applied domain squatting techniques themselves. In a multi-part survey, 288 participants create 2,880 malicious domains based on common domain squatting techniques and rate both domains created by other participants and real-world phishing domains in terms of how convincing they are. Our key results show that participants have problems to identify legitimate domains as benign if they include unusual top-level domains, additional terms, or use subdomains. Moreover, participants rated domains created by other participants higher than real-world phishing domains. Overall, we find that participants are more sceptic of domains, and flag more benign domains as malicious, if they contain domain squatting characteristics after they gained practical experience creating phishing domains themselves. In particular, the number of falsely classified domains that were actually benign increased from 33.7% to 46.6% after our training. Our results show that training users to act as an adversary can help to increase the effectiveness of security trainings. In addition, we recommend that online services do not create domains that make use of common domain squatting techniques, to reduce confusion for users.

ACM Reference Format:
Florian Quinkert, Martin Degeling, Jim Blythe, and Thorsten Holz. 2020. Be the Phisher – Understanding Users' Perception of Malicious Domains. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security (ASIA CCS '20), October 5–9, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3320269.3384765

ASIA CCS '20, October 5–9, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-6750-9/20/06…$15.00
https://doi.org/10.1145/3320269.3384765

1 INTRODUCTION

Phishing is one of the most prevalent attack vectors threatening individual and organizational privacy and security. Kaspersky reported to have blocked almost 130 million attempts to open scam websites in the second quarter of 2019 [30]. In a phishing attack, a victim receives an e-mail appearing to be from a legitimate sender, e.g., a well-known company [12]. The e-mail motivates the victim to login to the company's website via a provided link. However, this link does not lead to the company's actual website, but to a website controlled by the attacker, who receives the victim's credentials if she enters them. In some cases, it is possible to recognize a phishing e-mail by examining the design, structure, or text of the e-mail. But phishing e-mails are nowadays often designed to mimic legitimate e-mails so that it is hard to identify them as phishing. Double-checking the provided link is a more reliable way because the attacker has to use a domain under her control, i.e., it has to be different from the original one. Attackers use a variety of techniques to create unobtrusive domains which are similar to the original ones and easily deceive the user. For example, they use the original domain as template and remove one character or swap two neighboring characters, which is called *typosquatting* [7] (e.g., *paypl.com* or *payapl.com* instead of *paypal.com*). When using *combosquatting*, attackers add suitable terms to a well-known domain to obtain a domain which looks like it could belong to the company of the well-known domain (e.g., *paypal-login.com*) [17].

Multiple studies analyzed why phishing is successful [8, 12] and how it can be prevented [19, 20]. However, a closer examination of how users understand malicious domains created with different techniques and how well they can tell them apart from benign domains is missing. A deeper understanding of user's behavior can help to prevent phishing more effectively, e.g., by avoiding certain domain structures or customizing training efforts.

In this paper, we present the design of a multi-part survey in which users classify domains into benign and malicious, create their own phishing domains, and rate domains created by other users, before they again finally classify domains into benign and malicious. We use this approach to answer the following four research questions:

(1) How well do different types of domain squatting techniques deceive users?
(2) In what ways do domains created by untrained users differ from real-world phishing domains?
(3) How do users rate the trustworthiness of user generated and actual phishing domains?

(4) Does creating phishing domains impact the ability of participants to classify domains?

We conducted a study with 288 participants who completed the previously mentioned survey to understand how users perceive and can execute domain squatting techniques.

We found that many participants are confused by legitimate domains that are close to a company's original domain (e.g. *paypal-prepaid.com* vs. *paypal.com*). In addition, participants rated domains created by other participants on average higher than real-world phishing domains (average of 2.51 vs 2.85 on a scale between 1 and 5). While the overall accuracy of correctly classified domains increased only little after participants created their own phishing domains, we observed that participants got more cautious and were more likely to classify domains as malicious when they considered them suspicious.

In summary, we make the following key contributions:

- We present the design and procedure of a study to understand users' perception of malicious domains.
- We analyze the results provided by 288 participants who performed the study.

The remaining of this paper is structured as follows: first, we introduce background information and review related work in Section 2. Afterwards, we describe the structure of the experiments in Section 3. In Section 4, we analyze the collected data, followed by a discussion and description of limitations in Section 5. Finally, we conclude the paper in Section 6.

2 BACKGROUND AND RELATED WORK

In this section, we provide an overview of domain squatting techniques and discuss previous work on phishing surveys.

2.1 Domain Squatting

Domains are an important part of the Internet because they prevent users from having to remember plain IP addresses. A domain consists of multiple labels which are separated by dots [23]. Commonly, the last label is referred to as *top-level domain* (e.g., *com* or *org*), and together with the first label to the left as *second-level domain* (e.g., *paypal.com*). Furthermore, it is possible to prepend additional labels. Such domains are called *subdomains* of a second-level domain (e.g., *developer.paypal.com*).

Attackers utilize a variety of techniques to create domains that are hard to distinguish from well-known ones or appear to be legitimate domains. The literature refers to these techniques as *domain squatting* or *cybersquatting* [14]. In the following, we introduce five common domain squatting techniques the participants of our study were offered to create squatted domains and discuss relevant literature.

Typosquatting is a technique in which an attacker uses a well-known domain as template and creates a domain which can result from a typical typing error. According to Wang et al., an attacker can remove a character, duplicate one, swap two neighboring characters, replace a character with a neighboring character on a qwerty keyboard layout, or remove the dot between *www* and the second-level domain [31]. Agten et al. created typosquatting domains according to these rules for 500 popular domains [7]. Over the course of seven months, they found typosquatted domains for 95% of the popular domains. In addition to popular domains, Szurdi et al. discovered that a majority of typosquatting domains targets less popular domains [29].

Moreover, an attacker can add suitable terms to a well-known domain to obtain a domain which is easy to tell apart from the original one but could still be a legitimate domain, e.g., *login-facebook.com*. Kintis et al. referred to this technique as *combosquatting* and detected more than one million such domains among 400 billion DNS records [17].

Nowadays, domains cannot only contain Latin letters, numbers, and hyphens, but also characters from other alphabets, such as Cyrillic. In some cases, a Latin letter is almost not distinguishable from a letter in another alphabet, which enables attackers to replace a Latin letter in a well-known domain with the corresponding letter from the other alphabet. The resulting domain is called *homograph domain*. In 2006, Holgers et al. identified only a small number of malicious homograph domains [15]. More recently, multiple studies found a growing number of homograph domains [11, 21, 26, 28], even though the number remains small compared to typosquatting and combosquatting.

An attacker can get access to a registered domain, e.g., by stealing the login credentials of the domain registration used by the victim company via a phishing attack. Afterwards, she can create subdomains of the registered domain and use them for their own, often malicious, purposes. Liu et al. referred to the created subdomains as *shadow domains* because the legitimate domain owner is not aware of their existence [22]. In our study, participants can create subdomains containing well-known domains. However, the participants did not actually get access to the second-level domain. Therefore, we refer to these domain squatting technique as *low-level domain impersonation* because the impersonated domain is in the subdomain, i.e., the lower, part of the created domain.

A simple, yet effective technique is registering a well-known domain in a different top-level domain, e.g., *paypal.top* instead of *paypal.com*. We refer to this technique as *wrong top-level domain* [3].

Additional domain squatting techniques which we do not consider because they are difficult to represent in our survey include the usage of bit errors [25] and domain names which sound similar to well-known domain names [24].

2.2 Phishing Surveys

Multiple studies explored user's behavior when receiving phishing. Dhamija et al. asked 22 participants to classify 20 websites into phishing and non-phishing ones [12]. 23% of the participants did not include technical vectors in their decisions, e.g., the address or browser bar, so that on average 40% of the decisions were wrong. However, the authors did not focus explicitly on the used domains but rather on the appearance of the phishing website as a whole. Sheng et al. presented an online game to educate users on phishing [27]. Similar to our study, the game focuses on teaching users to identify malicious domains. While our approach uses less elements from games, we use a different perspective and let the user think like an attacker. Kumaraguru et al. proposed an education system and an online game to teach users how to avoid falling for phishing [19]. The online game focused on the detection of malicious domains, and their results show that especially inexperienced users

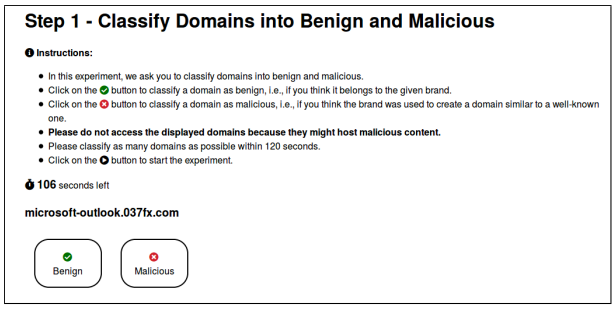

Figure 1: Screenshot of the web application showing how participants can classify domains in step 1.

benefit from such training. Blythe et al. conducted multiple studies to understand how difficult the detection of phishing is [8]. They noted that the 224 participants detected phishing in 80% of the cases, but had problems when logos of the impersonated brands were present. Domain highlighting is a technique in which the domain of a URL is highlighted while the remaining part of the URL is, for example, grayed out to focus user's attention on the domain. Lin et al. performed a study with 22 participants and found that domain highlighting is effective to a certain degree for some users but not for all so that it should be only one out of multiple measures [20]. Canova et al. proposed an Android app to educate users how to detect malicious URLs [9]. In addition, they performed a lab study and found an improved detection rate of malicious URLs after using the Android app [10]. However, small changes in the URL (e.g., swapping two letters) remain difficult to detect by users, even after the learning process.

3 SURVEY

In the following, we describe the design of the survey we use to collect information about how users understand domains. The survey was implemented as a web application written in Node.js [5] using jQuery [2] for the client-side interaction and storing results in a MySQL [4] database.[1] The survey starts with an introduction which explains the purpose of the study, displays data privacy and contact information, and details about how the survey works. Afterwards, we perform four distinct experiments, followed by a questionnaire. Based on this technical setup, we used Amazon Mechanical Turk to conduct an actual study with the survey. Therefore, we introduce Amazon Mechanical Turk at the end of this section and explain how we used it.

3.1 Step 1: Classifying Domains

In the first step, we ask the participants to classify domains into benign and malicious to understand how well participants can distinguish them. Figure 1 shows a screenshot of the web application which shows how participants can classify domains. For that purpose, we display a domain which is either a valid one used by a company or a malicious one listed in the phishing feed Phishtank [6]. Additionally, we display one button for the classification as benign domain and one for the classification as malicious.

[1]The source code is available at https://github.com/RUB-SysSec/be-the-phisher

We use a pool of 78 valid and 225 malicious domains, which belong to or target 20 well-known companies (e.g., *Amazon*, *Apple*, or *PayPal*). We choose well-known companies which we expect most participants to be familiar with. The valid domains contain both obvious examples (e.g., *microsoft.com*) and more difficult to identify examples (e.g., *paypal-prepaid.com*). Similarly, we use both easy-to-spot malicious domains (e.g., *myprofile2001.id3-440-wellsfargo.com*) and harder-to-detect ones, such as *docusignn.com*. Please see Appendix A for a full list of all used domains.

We set a time limit of 120 seconds and request the participants to classify as many domains as possible to ensure they do not spend a very long time on every domain. Besides the domain and the classification result, we measured the time a participant spent classifying a particular domain to understand which domains are more difficult to classify. Furthermore, the timing was relevant to identify a participant that just randomly presses buttons without evaluating the domain.

3.2 Step 2: Creating Domains

In the second step participants were asked to take the role of a malicious actor and create ten domains they considered to be convincing phishing domains. Figure 2 shows a screenshot of the web application which shows how participants can create domains. We wanted to understand whether phishing domains generated by untrained participants look different from actual malicious domains. Furthermore, this experiment enables us to compare the domains created by a participant with her results from step 1 to understand whether users that are more security-aware create different domains. The process is guided so that the survey script first randomly selects one out of 20 well-known domains also used in the first step the participant should impersonate. Second, the participant selects one of the five domain squatting techniques described in Section 2: typosquatting, combosquatting, wrong top-level domain, homograph domain, and low-level domain impersonation. We randomize the displayed order of the domain squatting techniques for each attempt to prevent users from always selecting the same entry.

In case of typosquatting, the participant can, based on Wang et al. [31], further select between prepending *www* without a dot to the well-known domain, omit a character, swap two neighboring characters, replace a character with its neighbor on a qwerty keyboard layout, and duplicate a character. It is easy to provide a clickable solution for most domain squatting techniques, e.g., the participant can click on a character to remove it. In contrast, combosquatting (addition of random characters or suitable terms to a domain) and low-level domain impersonation (well-known domain is a subdomain of a unsuspiciously looking domain) require the input of characters. In a pre-study, participants could input characters or terms in case of combosquatting and input unsuspiciously looking domains in case of low-level domain impersonation. Domains follow a certain format (e.g., they cannot start with a hyphen or cannot contain special characters) so that we had to explain the format, check the created domains, and provide feedback.

Participants of the pre-study said that the solution was time-consuming to understand and error-prone to use. Therefore, we developed a different approach in which participants select via buttons whether they want to add random characters or suitable

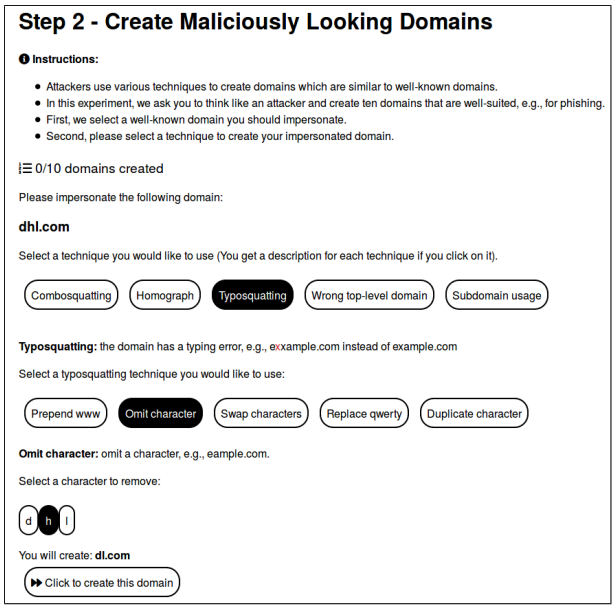

Figure 2: Screenshot of the web application showing how participants can create domains in step 2.

terms before and/or behind the well-known domain in case of combosquatting. For low-level domain impersonation, participants can choose whether they want to add additional levels with random characters or suitable terms before and/or behind the well-known domain. The system then randomly selects one of eleven suitable terms (e.g., *login* or *verify*) or creates a string containing between one and five random characters. While this approach is less accurate, it is clearly more usable and still provides valuable insights (e.g., whether terms or characters are preferred and at which position).

3.3 Step 3: Rating Domains

In the step 3 experiment, we show participants either a domain created by another participant or one of the 225 domains detected by the phishing feed Phishtank [6], which we already used in step 1. Afterwards, we ask the participant to rate the domain in terms of how likely it is that she would click on the domain if it was in an e-mail sent to her. The participant can choose between five ratings (one star through five stars) with one star meaning the domain is not convincing at all and five stars meaning the domain is very convincing.

Figure 3 shows a screenshot of the web application which shows how participants can rate domains. The likelihood that we display a domain being created from another participant is a bit higher (60%) than for domains detected by Phishtank (40%) because we are more interested in the quality of the generated domains. If a participant rated all available domains from other participants (e.g., because she is the first participant), we display a domain detected by Phishtank. In total, we ask a participant to rate ten domains and store the rated domain, the rating, and the time spent on the domain. Note that there is no time limit in this step because we are interested in a good rating rather than a fast rating. We use

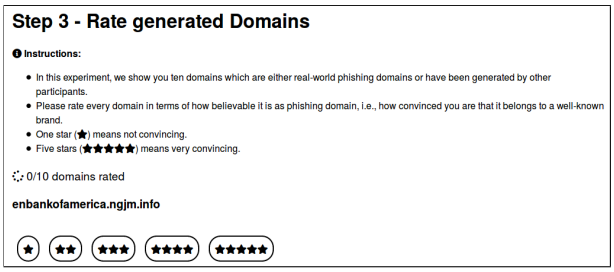

Figure 3: Screenshot of the web application showing how participants can rate domains in step 3.

this experiment to assess the quality of the domains created in step 2 as well as to compare the ratings with real-world phishing domains from Phishtank. Additionally, we can correlate the ratings of Phishtank domains with wrongly classified domains in step 1, e.g., to understand whether participants rather classify high rated domains wrongly.

3.4 Step 4: Classifying Domains

As step 4 experiment, we repeat the step 1 experiment to measure whether participants classify domains differently after they generated domains themselves (step 2) and rated both real-world phishing domains and other participants' domains (step 3). As specified in step 1, we store the classified domain, the classification result, and the time spent. We can use both the rate of falsely classified domains and the time a participant needed to classify a domain as metric to compare the step 4 results with the step 1 results. While the rate of falsely classified domains is an obvious metric, we argue that more time spent to classify a domain is an indicator that domains are examined more closely. Furthermore, it is interesting to see whether the number of benign domains which are rated as malicious increases because participants are more cautious in step 4.

3.5 Questionnaire

Finally, participants were asked to complete a questionnaire to collect information about the participants and their previous knowledge. The questionnaire is divided into three parts. The first part contains demographic questions, i.e., we ask for the participant's age (18 - 25; 26 - 35; 36 - 45; 46 - 55; over 55), the participant's gender (male; female; non-binary/third gender; other), the participant's education (less than high school diploma; high school diploma or equivalent; no degree; bachelor's degree; master's degree; higher than master's degree), and the participant's country of residency (north america; south america; europe; africa; asia; australia; antarctica). In addition to the mentioned answer options, we always provide an option not to answer a particular question.

The second part consists of questions about the participants' security background. For that purpose, we use the 16 questions provided by the Security Behavior Intentions Scale (SeBIS) [13]. Each question consists of a statement, e.g., *I manually lock my computer when I step away from it.* The participant is asked to select the term which applies best: never, rarely, sometimes, often, or always. Furthermore, we added two questions to verify that the

participants paid attention and understood the questions. The first attention-based question asks to select the word *never* from the list of terms and the second one not to select the word *never*.

The third part contains questions related to our survey. In particular, we ask whether the participants have been the victim of a phishing attack before (yes; no; do not know), whether they received some kind of formal phishing training (yes; no; do not know), whether they already knew some of the techniques in step 2 (yes; no; not all but at least one), and whether they believe the generation of phishing domains is going to help them in the identification of phishing domains in future (yes; no; do not know). If a participant declares she has been the victim of a phishing attack, we display an input field and ask for an optional description. Furthermore, we display a free text field so that participants can provide general feedback.

3.6 Amazon Mechanical Turk

Computer science requires manual interaction in some cases. For example, it is necessary to label data manually to use it in machine learning models or to answer survey questions to gain insights into human behavior. Amazon Mechanical Turk (MTurk) [1] is a crowdworking marketplace which enables companies or individuals to offer small tasks which are subsequently performed by others for a certain amount of money. MTurk refers to a task as *Human Intelligence Task* (HIT), to the provider of a HIT as *requester*, and to the person who performs the HIT as *worker*. A requester can choose from predefined templates for a HIT or provide a link to an external website. In the latter case, workers get a code after finishing the HIT to redeem their payment at MTurk. Furthermore, a requester can use different properties (e.g., demographic ones or having an account at a specific social media platform) to make sure the workers fit to the HIT.

We created a HIT at MTurk with a link to the survey hosted on one of our servers. Before we recruited numerous workers, we performed a pre-study with 20 participants to learn how much time they needed to perform the survey and adapt the payment accordingly. We did not require participants to have certain properties but requested only *master worker*, who have a lot of experience and proved to provide very good results. We paid every worker who completed the survey $3.00 and learned that they spent about 15 minutes on the survey so that we kept this amount because it leads to reasonable hourly earnings. Furthermore, we developed a more usable solution to create *combosquatting* and *low level domain impersonation* domains based on their feedback and made some minor changes to the survey, e.g., we started to measure the time for additional parts of the survey. Afterwards, we performed a larger study with 288 participants in November 2019.

As described previously, we added two attention-check questions in the questionnaire. However, we do not rely solely on these questions to determine whether or not a participant paid attention but additionally checked whether participants, who failed to answer both attention questions, classified the legitimate company websites (e.g. *paypal.com*) in step 1 correctly. If a participant answered both attention-based questions wrongly and classified at least one legitimate company website as malicious, we removed the participant from our evaluation.

While we had multiple participants who gave incorrect answers to one of the two attention-based questions, only one participant failed both questions. However, the participant correctly classified both company websites in step 1 so that we include all 288 participants in our analysis.

4 RESULTS

In this section, we describe the results obtained from the 288 participants who successfully completed the previously introduced survey in November 2019. We start by characterizing the participants and continue with an in-depth analysis of each survey step.

4.1 Participants

Table 1 provides an overview of demographic and technical information of the participants. A majority of participants identified themselves as male (60.4%). Most participants were between 26 and 35 (47.2%) or 36 and 45 (27.4%) years old. Older people (over 55: 6.3%) are better represented than very young people (18 to 25 years old: 2.4%). The majority of participants has at least a Bachelor's degree (more than 75%) and 3.5% a degree even higher than a Master's degree, e.g., a PhD. Most participants reported that they live in North America (70.5%) or Asia (28.1%), while other continents are underrepresented (1.4%). The vast majority of participants used a desktop system (98.3%), while few completed the survey on a mobile device (1.7%). The majority also favored Windows (86.7%) over MacOS (6.4%), Linux (3.1%), and Chrome OS (3.8%). The most often used browser is Chrome (86.5%), while Firefox (9.4%) and other browsers (4.1%) are used to a far lesser degree.

13.9% of the participants reported they have been the victim of a phishing attack before. Another 10.4% declared they do not know, and 75.7% said they have never been the victim of a phishing attack. We found no statistically significant relations between any demographic factor and the phishing experience. We will in particular analyze whether phishing victims behave differently in our experiments. A majority of 78.5% of the participants said that the creation of phishing domains in step 2 will be helpful to identify phishing in future. In contrast, 10.1% answered that it is not helpful and 11.4% that they do not know.

For the SeBIS questions, we calculated an average score for each participant, i.e., answering a question with *never* led to one point, *rarely* two points up to five points for *always*. Afterwards, we divided the sum by the number of questions. Figure 4 shows the number of participants as a function of their average SeBIS scores for intervals of size 0.5. The average SeBIS scores are normally distributed around the interval between 2.0 and 2.5. We did not observe very high or very low average SeBIS scores. We conclude that most of our participants are neither beginners nor experts, but rather average Internet users in terms of their security knowledge.

4.2 Step 1: Classify Domains

In the step 1 experiment, we asked participants to classify as many domains as possible as either benign or malicious in 120 seconds. In total, the 288 participants classified 14,178 domains in the first baseline step of the study with an average of 49.23 domains per participant (standard deviation of 21.53). Figure 5 shows the number of participants as a function of the number of classified domains

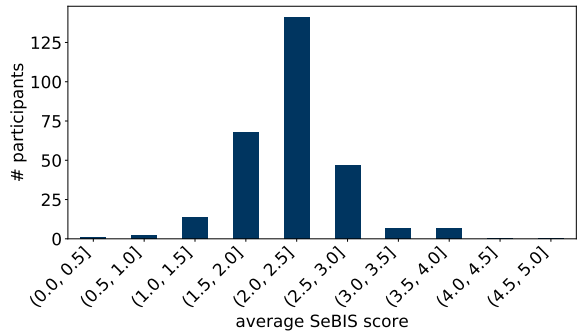

Figure 4: Number of participants as a function of their average SeBIS score for intervals of size 0.5, showing that our participants are average computer users in terms of their security knowledge.

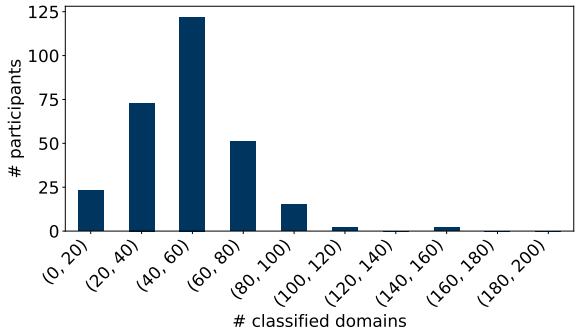

Figure 5: Number of participants as a function of the number of classified domains for intervals of size 20. It shows that the participants classified a reasonable number of domains within 120 seconds

Table 1: Demographic and technical information about the 288 survey participants in absolute numbers (column #) and percentage (column %).

	#	%
Gender		
Male	174	60.4
Female	108	37.5
Other	6	2.1
Age		
18 - 25	7	2.4
26 - 35	136	47.2
36 - 45	79	27.4
46 - 55	44	15.2
over 55	18	6.3
Not answer	4	1,5
Education		
Over Master	10	3.5
Master	51	17.7
Bachelor	161	55.9
High School	34	11.8
No degree	26	9.0
Other	6	2,1
Residency		
North America	203	70.5
Asia	81	28.1
Other	4	1.4

	#	%
System		
Desktop	283	98.3
Mobile	5	1.7
OS		
Windows	216	86.7
MacOS	19	6.4
Linux	9	3.1
Chrome OS	10	3.8
Browser		
Chrome	249	86.5
Firefox	27	9.4
Other	12	4.1
Phishing Victim		
Yes	40	13.9
No	218	75.7
Not known	30	10.4
Survey Helpful		
Yes	226	78.5
No	29	10.1
Not known	33	11.4

for intervals of size 20. We had one participant who classified only one domain and another participant who classified 156 domains. However, 75% of the participants classified less than 60 domains, i.e., the vast majority paid enough attention to make a decision but did not spend multiple seconds on a single domain. Participants spent 2.34 seconds per domain on average (standard deviation of 2.15) with a minimum of 0.27 seconds and a maximum of 93.15 seconds.

We call malicious domains *true positives* if they were classified as malicious and *false negatives* if they were classified as benign. In the same way, we refer to benign domains as *true negatives* if they were classified as benign and *false positives* if they were classified as malicious. We use two metrics to measure the quality of the participants' decisions:

- **Accuracy** is the number of true positives and true negatives divided by the number of all classified domains, i.e., (TP+TN)/(TP+TN+FP+FN). It describes the fraction of correctly classified domains.
- **Precision** is the number of true positives divided by the number of true positives and false positives, i.e., (TP)/(TP+FP). It shows the fraction of as malicious classified domains which are actually malicious.

The participants achieved an overall accuracy of 0.75 and a precision of 0.88, showing that participants classify a considerable amount of domains wrongly. In particular, we observed that 66.34% of the wrongly classified domains were false negatives, i.e., actually malicious domains which were classified as benign. Lowering this number is crucial because a misclassified malicious domain leads to a successful attack. In contrast, 33.66% of the wrongly classified domains were false positives. The results indicate that participants classified a domain only as malicious when they were confident it is, in fact, a malicious domain.

Furthermore, we examined the wrongly classified domains more closely and found every provided domain being wrongly classified at least once, including even obvious examples, such as *paypal.com*. However, in case of benign domains, we found especially three groups confusing participants: subdomains of well-known domains (e.g., *mtouch.facebook.com*), addition of terms to well-known domains (e.g., *paypal-prepaid.com*), and unusual top-level domains (e.g., *paypal.me*). While the domains make sense when their purpose is known, they confuse users who do not know it. Additionally, it simplifies attackers the use of domain squatting because it is harder for users to identify domains generated by an attacker as malicious when they know that similar domains exist. Therefore, we recommend companies to avoid such domains or at least avoid the usage

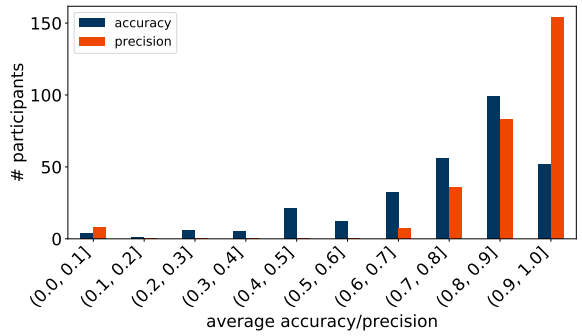

Figure 6: Number of participants as a function of accuracy and precision for intervals of size 0.1.

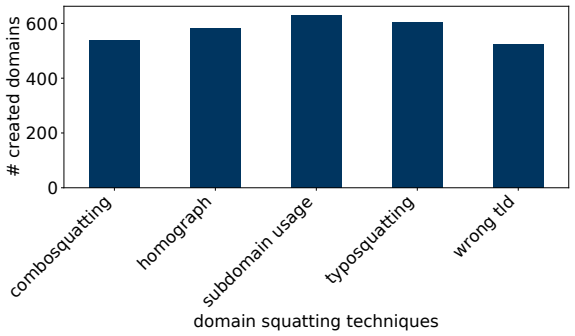

Figure 7: Number of created domains in step 2 as a function of the five squatting techniques, showing that participants do not have a favorite technique.

when contacting customers, e.g., via e-mail. On the contrary, we found participants being able to classify long and obvious malicious domains usually correctly. However, shorter domains which added suitable terms were among the malicious domains, which were most often classified as benign. Taking into consideration that, as previously explained, well-known domains sometimes use similar techniques, it is less surprising that participants had problems classifying the malicious counterparts correctly.

In addition, we were interested in the distribution of accuracy and precision among the participants to understand whether the average values are a good representation. Figure 6 shows the number of participants as a function of accuracy and precision for intervals of size 0.1. The accuracy and precision for the majority of participants are close to the average values, i.e., we did not find many participants who misclassified an exceptional high number of domains. Moreover, we analyzed whether properties like origin, gender, age or phishing experience influence the results in the step 1 experiment. A Chi-Square tests showed now statistical significance for the relation between the correct classification ratio and any demographic variable.

In summary, our results show that participants in general are able to identify malicious domains while they have problems to classify hard to identify ones correctly. Especially the high number of wrongly classified malicious domains is noteworthy.

4.3 Step 2: Create Domains

4.3.1 Overview. We asked participants to create their own phishing domains in the step 2 experiment and were interested whether the created domains differ from real-world phishing domains and whether creating phishing domains leads later on to a better detection rate. We requested participants to create ten domains so that the 288 participants created 2880 domains. Figure 7 shows the number of created domains as a function of the five domain squatting techniques. The number of created domains is very similar for each technique so that we conclude that the participants do not have a preferred squatting technique. Recent research revealed that, for example, homograph domains are used less frequently compared to other domain squatting techniques [26, 28]. However, they look very convincing so that it is logical that participants use them comparatively often. As described in Section 3.2, we randomized the order in which the domain squatting techniques are displayed for each created domain. We found that the participants on average choose the domain squatting technique at position 2.82 (standard deviation of 1.37) but focus rather on techniques displayed in the beginning: position 1 had 618 created domains, pos. 2 693, pos. 3 612, pos. 4 510, pos. 5 447.

We assumed participants would choose an appropriate domain squatting technique based on the length of the domain. For example, removing one character from a domain which is already very short does not lead to a convincing phishing domain. Nevertheless, for well-known domains with a length of at most five characters (not including the top-level domain), we found 760 created domains out of which 141 used typosquatting. That is, the ratio of typosquatting domains for short well-known domains is not considerably different than for all well-known domains. We infer that participants did not pay much attention to the well-known domain before choosing a particular domain squatting technique. In contrast, an attacker would take into account the structure of the well-known domain and select an appropriate domain squatting technique accordingly.

Furthermore, we were interested in the decisions the participants made after they chose a particular domain squatting technique. In case of the 525 domains using the domain squatting technique wrong top-level domain, we found the top-level domain *net* being used most often. In contrast, the top-level domains *de* (74) and *md* (84) were used less often, most likely because most participants were from North America and Asia and did not necessarily know that these are the country code top-level domains of Germany and Moldova, respectively. Additionally, the top-level domain *top* (94) was selected less frequently, too. However, Korczynski et al. showed that new top-level domains, such as *top*, are used more frequently by attackers [18]. The participants of our study did not have this background knowledge and chose preferably a top-level domain they know. Moreover, participants rather do not trust top-level domains they do not know as we explained in Section 4.2.

4.3.2 Typosquatting. The 604 domains which were generated with the domain squatting technique *typosquatting* distribute between

the available techniques as follows (see Section 3.2 for more details): *prepend www* (164 instances), *omit character* (153), *swap characters* (112), *duplicate character* (109), and *replace qwerty* (66). The order of their frequency is equal to the order they appear in the survey except for *duplicate character* and *replace qwerty* which switched positions. Hence, an explanation we have to take into consideration is that participants preferred a *typosquatting* technique because it was in one of the first positions. We suppose that participants used *replace qwerty* to a lesser degree because it is the most complex technique so that a participant has to think about it for some time. On the contrary, Agten et al. found a very similar distribution in their long term study of typosquatting [7]. For example, *prepend www* was used most often, *swap characters* and *duplicate character* were used with almost the same frequency, and *replace qwerty* was used to a far lesser degree. Therefore, an equally valid explanation is that the *typosquatting* domains generated by our participants represent the actual distribution pretty well.

Four of the five typosquatting techniques change a character in the well-known domain. Changing a character in the middle or back part of the well-known domain is less noticeable. Thus, we were interested at which position the participants changed a character if they used one of these techniques. Since the 20 well-known domains we provide have a different length, we introduce the concept of a *position score*: The *position score* is the position of the first changed character divided by the length of the well-known domain (without the top-level domain).

We use this value as a metric for the changed position. On average, the changed character was at position score 0.57 (standard deviation 0.25). 25% of the created domains had a position score less or equal than 0.40 and 75% a position score less or equal than 0.78. That is, most participants changed rather a character in the middle or back of the well-known domain. This indicates that they were aware that changing a character in the beginning is easier to spot and leads to a suboptimal phishing domain.

4.3.3 Homograph Domain. The domain squatting technique *homograph domain* had 581 created domains. Quinkert et al. found not necessarily the most equally looking character pairs being used and concluded that often a less similar character is sufficient to delude a victim [26]. In contrast, the participants of our study commonly utilized character pairs which were almost not distinguishable, which can be expected when the participants are asked to create a most convincing phishing domain. In terms of which characters were replaced, the participants' results are in line with Quinkert et al. because in both cases, vocals and easy to spoof characters, such as *l*, were used.

Compared to *typosquatting*, it is less important to change a character in the middle or back part of a domain because in most cases a *homograph domain* is almost not distinguishable from the well-known domain. We calculated the same position score as in case of typosquatting and found slightly smaller values (average 0.45, standard deviation 0.29, 25% of the domains had a value less or equal than 0.20, and 75% had a value less or equal than 0.67). Comparing these values with the values for *typosquatting*, we conclude that participants were aware that a *typosquatting* domain benefits from a changed character in the middle or back part while it is sufficient to change any character in a *homograph domain*.

4.3.4 Combosquatting. In case of the 539 domains created with the domain squatting technique *combosquatting*, the participants had eight different options to create the domains (characters, terms or none in front and/or behind with none not being allowed for both front and behind because the result would be the well-known domain itself). The participants chose to use none in the front and characters (20 instances) and terms (36) behind at the least. This is surprising because having the well-known domain in the beginning guarantees that a victim recognizes the well-known part almost immediately. On the contrary, the participants chose to have a term in the front and nothing behind of the well-known domain most often (141 instances). Such a domain is very convincing if the selected term fits to the well-known domain (note that participants could not choose a term but it was chosen by the system due to usability). A *combosquatting* domain with characters in the front and nothing behind was created less often (82), while combinations of terms and characters in front or behind of the well-known domains were created between 44 and 89 times.

4.3.5 Low-Level Domain Impersonation. The domain squatting technique *low-level domain impersonation* was used to create 631 domains. Similarly to *combosquatting*, participants could choose to add characters, terms or none in front and/or behind the well-known domain. In contrast to *combosquatting*, it was allowed to choose none for both front and behind, which would result in a valid subdomain *well-known-domain.test-domain.com*. Participants did not opt for *low-level domain impersonation* domains starting immediately with the well-known domain, which is comparable to the results for *combosquatting* (chars behind: 13 instances, none behind: 28, terms behind: 36).

The most often selected variant was terms in front and none behind (128), which is in line with the *combosquatting* results, too. The second most often used variant is to create random characters in front and behind the well-known domain (117). To a certain degree, it is explainable because the well-known domain stands out in the random characters in front and behind it. Nevertheless, the resulting domain typically does not look very believable so that it remains unclear why participants chose this combination. The remaining variants were used with a similar frequency and a slightly tendency to characters in front (with none behind: 86, with terms behind: 78) instead of terms in front (with chars behind: 56, with terms behind: 89).

4.3.6 Summary. In summary, the results show that the participants basically understood how phishing domains work, even though the generated domains differ in certain aspects from actual phishing domains. *Combosquatting* and *low-level domain impersonation* are, to a certain degree, special cases because the created domains are never as close to the well-known domains as it is the case for the other domain squatting techniques. Therefore, we suppose that participants also wanted to try these techniques but lacked a clear understanding of how to use them.

4.4 Step 3: Rate Domains

In the step 3 experiment, we asked participants to rate both real-world phishing domains and domains generated by other participants in step 2 in terms of how convincing they are as phishing

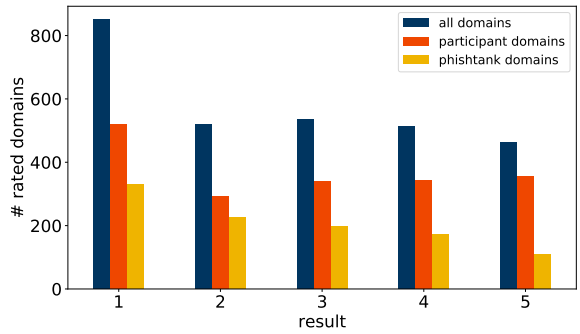

Figure 8: Number of domains as a function of the ratings one through five for all domains (blue), domains generated in step 2 (orange), and phishtank domains (yellow).

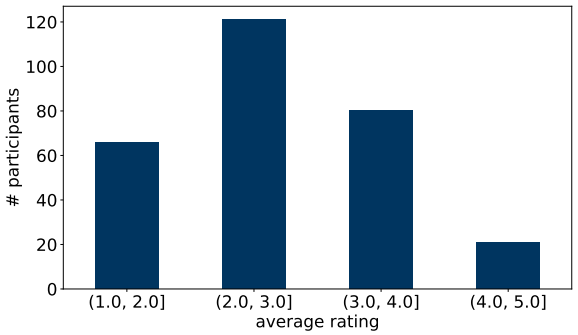

Figure 9: Number of participants as a function of the average rating they provided for intervals of size 1.0, showing that lower ratings are preferred.

domains. Participants could choose between five ratings from one (not convincing) through five (very convincing). We were interested both whether real-world phishing domains are more convincing than the generated ones and how convinced participants were of the domains created by other participants. We asked the participants to rate ten domains and got 2,879 ratings (due to technical problems, one result got lost). Overall, the participants rated 1,307 distinct domains out of which 1,084 were created in step 2 (1,846 ratings) and 223 from phishtank (1,033 ratings).

Figure 8 shows the number of domains as a function of the ratings one through five for all domains (blue), domains generated in step 2 (orange), and phishtank domains (yellow). Considering all domains, the rating one was used most often while the ratings two through five were used almost equally. The ratings for domains created by participants show a similar distribution, while the phishtank domains were rated with a descending tendency. First, we thought the high number of domains rated with one appeared because some participants rated all domains with one. However, we found only four participants who rated every domain with one. When an attacker creates a phishing domain, it does not necessarily have to be perfect to deceive a victim because it is not unlikely that she does not pay attention in particular to the domain. Therefore, the real-world phishing domains from phishtank are less convincing and consequentally lower rated. In contrast, the participants were asked in step 2 explicitly to create well-suited phishing domains so that the ratings are overall slightly better. Nevertheless, we discovered in Section 4.3 that the created domains can differ from actual phishing domains which explains the prevalence of ratings as one.

Figure 9 shows the number of participants as a function of the average rating they provided for intervals of size 1.0. Most participants had an average rating between 2.0 and 3.0 or between 3.0 and 4.0. That is, they actually tried to use the full scale between one and five and did not rate the domains only in terms of convincing or not convincing. Low average ratings were more common than high average ratings which is expected to some degree because we already analyzed that the participants preferred lower ratings.

Furthermore, we analyzed how similar ratings of the same domain are. Similar ratings suggest a similar understanding of how convincing a phishing domain looks like, while different ratings

Table 2: Number of domains with a standard deviation in the particular interval for ratings of all domains, domains created in step 2, and phishtank domains which have at least two ratings, indicating that participants agreed more on domains created in step 2.

Standard deviation interval	All domains	Step 2 domains	Phishtank domains
0.0 - 1.0	285	208	77
1.0 - 2.0	245	122	123
2.0 - 3.0	86	75	11

indicate that the participants have a different understanding of the particular domain. Analyzing the similarity of ratings is only useful for domains which had at least two ratings. In case of 616 domains, we received at least two ratings and calculated the standard deviation between the ratings for each domain. 285 domains have a standard deviation between 0.0 and 1.0, while 245 domains have a standard deviation between 1.0 and 2.0 and 86 have one between 2.0 and 3.0. That is, in most cases, the ratings are similar. Table 2 summarizes the number of domains with standard deviations in the intervals 0.0 through 1.0, 1.0 and 2.0, and 2.0 and 3.0 for all domains, the ones created in step 2, and the phishtank domains. Interestingly, the participants agree more on the ratings of domains created in step 2 than they do on the ratings of phishtank domains.

4.5 Step 4: Classify Domains

As step 4 experiment, we asked participants again to classify as many domains as possible in 120 seconds. Compared to the 14,178 domains the participants classified in step 1, they now classified 16,274 domains, which is an increase by 14.78%. The average number of classified domains slightly increased to 56.70 (standard deviation 22.41), and 75% of the participants classified less than 70 domains. These numbers already indicate that participants classified domains faster than in step 1. Indeed, they spent 2.05 seconds on a domain (compared to 2.34 seconds in step 1).

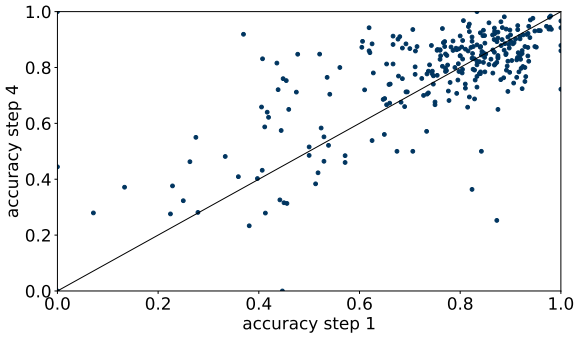

Figure 10: The scatterplot shows the accuracy in step 1 (x-axis) and step 4 (y-axis). Dots in the upper half represent participants who improved their accuracy in step 4.

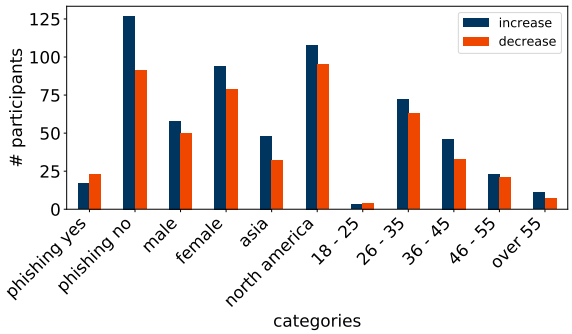

Figure 11: Number of participants who increased (blue) and decreased (orange) their accuracy between step 1 and step 4 as a function of different categories.

The participants reached an overall accuracy of 0.77 (0.75 in step 1) and a precision of 0.86 (0.88). At first glance these results seem to be not very encouraging. However, we analyzed the wrongly classified domains in more details and found a considerable lower number of false negatives (malicious domains which were classified as benign). In particular, 53.44% of the wrongly classified domains were false negatives (66.34% in step 1) and 46.56% false positives (33.66%). That is, participants got more cautious and classified domains rather as malicious if they were uncertain. In general, it is better to be cautious and do not click on a link if in doubt. Therefore, we consider the results to be promising because letting participants create their own phishing domains and rating others leads to a higher awareness.

Furthermore, we analyzed how many participants improved their accuracy compared to step 1 and whether there are determining factors for an improvement. Overall, 158 participants had an increased accuracy in step 4 compared to step 1, while 129 had a decreased one. Figure 10 shows a scatterplot which compares the participants' accuracy between step 1 and step 4. Each dot is a participant with the accuracy in step 1 on the x-axis and the accuracy in step 4 on the y-axis. Dots in the upper half represent participants who improved their accuracy in step 4. The difference between the results in step 1 and step 4 is statistically significant (two-sample t-test p<0,001).

Figure 11 shows the number of participants who increased (blue) and decreased (orange) their accuracy between step 1 and step 4 as a function of multiple categories. Note that we excluded categories with very few participants, e.g., origins other than North America and Asia or genders other than male and female. The category *phishing yes* means the participant declared that she has been the victim of a phishing attack and *phishing no* accordingly that the participant stated she has not been the victim of a phishing attack. In most groups, the accuracy increased for more participants than it decreased. Nevertheless, it is surprising that participants who have been the victim of a phishing attack rather decrease their accuracy between step 1 and step 4. Phishing victims have on average a slightly lower SeBIS average score (2.02 compared to 2.25), a difference that is significant when evaluated with a t-test (p<0,001). This result highlights the importance of general computer security trainings.

5 DISCUSSION

The results described above help us to better understand how users classify domains and showed that practical knowledge of domain squatting and phishing techniques can help raising awareness. More specifically, our results show that it is difficult for users to tell apart legitimate and malicious domains if the legitimate domains make use of common domain squatting techniques, such as additional terms (e.g., *paypal-prepaid.com* vs. *paypal.com*), unusual top-level domains (e.g., *paypal.me*), or subdomains (*mtouch.facebook.com*). Therefore, we recommend domain owners to avoid using such domains in the future.

We also showed that phishing domains created by participants were often similar to the actual malicious domains which suggests that participants understood the explanations of how malicious domains are built by attackers. Furthermore, it shows that our survey helps to teach participants.

Furthermore, we found that participants also considered domains created by other participants to be more convincing than real-world phishing domains. We asked participants explicitly to create convincing phishing domains so that they focused on this aspect, while a real-world phishing domain often does not have to be such convincing to deceive a victim.

Last, creating phishing domains and rating domains of other participants as well as phishing domains from phishtank led to a higher cautiousness when finally classifying domains. In particular, legitimate domains containing common domain squatting techniques as aforementioned were later often classified as malicious. We consider this to be promising because even though the domains are legitimate, they still look like their malicious counterparts and should be handled with care.

5.1 Limitations

We distinguish between two types of limitations. First, limitations caused by our survey design. Second, limitations originating from the usage of Amazon Mechanical Turk.

The only limiting part of our survey is the generation of phishing domains in the step 2 experiment. We had to select a reasonable number of domain squatting techniques and had to develop a usable

solution to create corresponding domains. While it certainly had been possible to add more domain squatting techniques (e.g., bitsquatting or soundsquatting as mentioned in Section 2), we focused on the most often used and in publications covered ones. Furthermore, participants could only change one character if a domain squatting technique aimed at changing characters and could not type characters or terms in case of *combosquatting* and *low-level domain impersonation*. However, it was important for us to keep the survey as easy understandable and usable as possible and do not overwhelm participants. Therefore, we consider these limitations as necessary. Moreover, they have only a small impact on the quality of the results and still allow the understanding of users' perception of malicious domains.

Other limitations arise from using Amazon Mechanical Turk. For example, we cannot guarantee that participants focused solely on our survey and did not work, for example, on another HIT at the same time. However, we analyzed multiple parameters (e.g., time spent on a domain/to create a domain or number of classified domains) to ensure an adequate performance of participants. Furthermore, we did not set a minimum number of domains to classify in steps 1 and 4 to get paid because it leads to classifications which were only performed to collect enough classifications for the payment. Additionally, we did not find an extraordinary number of participants classifying only very few domains so that we conclude the participants usually performed the task they were asked to do.

5.2 Ethical Considerations

In our study design, we followed best practices of our research institution for survey design as there is no ethics board. When we performed the study at Amazon Mechanical Turk, human workers executed the experiments in our survey and answered the questionnaire's questions. On the start page of the survey, we provide all necessary information, such as who is responsible for the study, who can be contacted in case of questions, what kind of data is collected, how is it stored and processed, and that the participation is voluntary and can be discontinued at any time. Furthermore, we always provide an option not to answer a question without giving reasons if it possibly collects personal information, e.g., the age group or gender. Neither we can nor we tried to identify users based on the provided information.

We paid every worker $3.00 after having completed the survey and provided the correct completion code at Amazon Mechanical Turk. Completing the survey takes approximately 15 minutes so that $3.00 lead to hourly earnings of $12.00, which is above the federal hourly earnings of $7.25 and above most states' hourly earnings in the United States [16]. Therefore, we consider our payment sufficient for the given task.

6 CONCLUSION

Previous work studied either why users fall for phishing e-mails or the prevalence of different domain squatting techniques. In this paper, we combined both perspectives and proposed the design of a user study in which we explain domain squatting techniques, and let participants create and rate phishing domains to measure whether the practical experience helps them to better classify domains.

Based on the results from 288 participants on Amazon Mechanical Turk, we learned that participants had problems to classify legitimate domains correctly if they use techniques similar to domain squatting, e.g., adding a common word to a well-known domain. Furthermore, comparing the domains created by participants with real-world phishing domains revealed that users created similar domains. However, techniques leading to domains which have additional characters or terms were difficult to understand for the participants. Additionally, participants were more convinced by the domains created by other participants than by real-world phishing domains. After the participants finished the creation and rating of phishing domains, they got more cautious when classifying domains into benign and malicious. They tended to classify even benign domains as malicious if they found properties often used in actual malicious domains.

In summary, we consider this study a starting point to learn more about users' perception of malicious domains.

REFERENCES

[1] "Amazon Mechanical Turk (MTurk)," https://www.mturk.com/, accessed: 2019/12/09.
[2] "jQuery," https://jquery.com/, accessed: 2019/12/09.
[3] "LEGO vs Cybersquatters: The burden of new gTLDs," https://news.netcraft.com/archives/2017/04/14/lego-vs-cybersquatters-the-burden-of-new-gtlds.html, accessed: 2019/12/09.
[4] "MySQL," https://www.mysql.com/, accessed: 2019/12/09.
[5] "Node.js," https://nodejs.org/en/, accessed: 2019/12/09.
[6] "Phishtank," https://www.phishtank.com/, accessed: 2019/12/09.
[7] P. Agten, W. Joosen, F. Piessens, and N. Nikiforakis, "Seven months' worth of mistakes: A longitudinal study of typosquatting abuse," in *Network and Distributed System Security Symposium (NDSS)*, 2015.
[8] M. Blythe, H. L. Petrie, and J. A. Clark, "F for fake: four studies on how we fall for phish," in *Conference on Human Factors in Computing Systems (CHI)*, 2011.
[9] G. Canova, M. Volkamer, C. Bergmann, and R. Borza, "NoPhish: An Anti-Phishing Education App," in *Security and Trust Management (STM)*, 2014.
[10] G. Canova, M. Volkamer, C. Bergmann, and B. Reinheimer, "NoPhish App Evaluation: Lab and Retention Study," in *Workshop on Usable Security and Privacy (USEC)*, 2015.
[11] D. Chiba, A. H. Akiyama, T. Koide, Y. Sawabe, S. Goto, and M. Akiyama, "DomainScouter: Understanding the Risks of Deceptive IDNs," in *Research in Attacks, Intrusions, and Defenses (RAID)*, 2019.
[12] R. Dhamija, J. D. Tygar, and M. Hearst, "Why Phishing Works," in *Conference on Human Factors in Computing Systems (CHI)*, 2006.
[13] S. Egelmann and A. Peer, "Scaling the Security Wall: Developing a Security Behavior Intentions Scale (SeBIS)," in *Conference on Human Factors in Computing Systems (CHI)*, 2015.
[14] O. G. Hatch, "The Anticybersquatting Consumer Protection Act," https://www.gpo.gov/fdsys/pkg/CRPT-106srpt140/html/CRPT-106srpt140.htm, 1999, accessed: 2019/12/09.
[15] T. Holgers, D. E. Watson, and S. D. Gribble, "Cutting through the Confusion: A Measurement Study of Homograph Attacks." in *USENIX Annual Technical Conference*, 2006.
[16] T. E. P. Institute, "Minimum Wage Tracker ," https://www.epi.org/minimum-wage-tracker, 2019, accessed: 2019/12/09.
[17] P. Kintis, N. Miramirkhani, C. Lever, Y. Chen, R. Romero-Gómez, N. Pitropakis, N. Nikiforakis, and M. Antonakakis, "Hiding in Plain Sight: A Longitudinal Study of Combosquatting Abuse," in *Conference on Computer and Communications Security (CCS)*, 2017.
[18] M. Korczynski, M. Wullink, S. Tajalizadehkhoob, G. C. M. Moura, A. Noroozian, D. Bagley, and C. Hesselman, "Cybercrime After the Sunrise: A Statistical Analysis of DNS Abuse in New gTLDs," in *ASIA Conference on Computer and Communications Security (AsiaCCS)*, 2018.
[19] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong, "Teaching Johnny Not to Fall for Phish," in *ACM Transactions on Internet Technology (TOIT)*, 2010.
[20] E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock, "Does Domain Highlighting Help People Identify Phishing Sites?" in *Conference on Human Factors in Computing Systems (CHI)*, 2011.
[21] B. Liu, C. Lu, Z. Li, Y. Liu, H. Duan, S. Hao, and Z. Zhang, "A Reexamination of Internationalized Domain Names: The Good, the Bad and the Ugly," in *International Conference on Dependable Systems and Networks (DSN)*, 2018.

[22] D. Liu, Z. Li, K. Du, H. Wang, B. Liu, and H. Duan, "Don't Let One Rotten Apple Spoil the Whole Barrel: Towards Automated Detection of Shadowed Domains," in *Conference on Computer and Communications Security (CCS)*, 2017.
[23] P. Mockapetris, "RFC 1035 - Domain Names - Implementation and Specification," https://tools.ietf.org/html/rfc1035, 1987, accessed: 2019/12/09.
[24] N. Nikiforakis, M. Balduzzi, L. Desmet, F. Piessens, and W. Joosen, "Soundsquatting: Uncovering the use of homophones in domain squatting," in *International Conference on Information Security (ISC)*, 2014.
[25] N. Nikiforakis, S. Van Acker, W. Meert, L. Desmet, F. Piessens, and W. Joosen, "Bitsquatting: Exploiting bit-flips for fun, or profit?" in *International World Wide Web conference (WWW)*, 2013.
[26] F. Quinkert, T. Lauinger, W. Robertson, E. Kirda, and T. Holz, "It's Not What It Looks Like: Measuring Attacks and Defensive Registrations of Homograph Domains," in *Conference on Communications and Network Security (CNS)*, 2019.
[27] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge, "Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish," in *Symposium On Usable Privacy and Security (SOUPS)*, 2007.
[28] H. Suzuki, D. Chiba, Y. Yoneya, T. Mori, and S. Goto, "ShamFinder: An Automated Framework for Detecting IDN Homographs," in *Internet Measurement Conference (IMC)*, 2019.
[29] J. Szurdi, B. Kocso, G. Cseh, J. Spring, M. Felegyhazi, and C. Kanich, "The Long "Taile" of Typosquatting Domain Names," in *USENIX Security Symposium*, 2014.
[30] M. Vergelis, T. Shcherbakova, and T. Sidorina, "Spam and phishing in Q2 2019," https://securelist.com/spam-and-phishing-in-q2-2019/92379/, 2019, accessed: 2019/12/09.
[31] Y.-M. Wang, D. Beck, J. Wang, C. Verbowski, and B. Daniels, "Strider Typo-Patrol: Discovery and Analysis of Systematic Typo-Squatting," in *USENIX Workshop on Steps Reducing Unwanted Traffic on the Internet (SRUTI)*, 2006.

A STEP 1/4 DOMAINS

Malicious domains (225 domains)

Amazon

- secure.amazon.com.achiverscs.com
- amazonorigin.com
- amazon.com.a4b.it
- secure.amazon.com.best-effectresources.com
- center-amazon-services.com
- www.amazon-bestsellers.net
- amazon-co-uk-ssl-mail.co.uk
- amazon.update.com.sonsofsamhorn.com
- www.amazon.com.paulowanderley.com.br
- amazon.com.update.nochelatinaclub.com
- www.amazonecard.com
- amazon-ssl-login.co.uk
- amazon.com-welcome-amazon-login.update-cgi954745-002.tkladders.ie
- amazon.com-log-service.himanshusabharwal.com

American Express

- online.americaexpress.moracom.co.uk
- www.americanexpress-sx89gox03gqn.com
- online-americaexpress.depilab.cl
- srv5-americanexpress.com
- americanexpress-tvbg.com
- global-americanexpress.com
- americanexpressbeta.com
- americanexpress-secure.com
- americanexpressverify.com
- welcomeamericanexpress.com
- americanexpress-fraud.com
- verifybyamericanexpress.com

Apple

- itunes.apple.verification.compte.id.admiral-placements.co.za
- id-apple.store.update.cmd-pool.data.bullamakanka.com.au
- apple-id.com.update.slavna-plus.com.ua
- itunes-appleid.com
- verifcation.itunees.applee.uk.id.login.divulga-se.info
- applesupdate.com
- secure.store.apple.com.next-sign-up.com
- ssl.allegro.fi-apple.com
- www.userid-apple.net
- appleid.apple.com.idapple-dk.tk
- id.apple.moodylake.com
- apple-updates.cloudapp.net

AT&T

- secureonline.att.homephoneandinternet.actlap.org
- login.att.net.issue-authentication.ml
- admin.att.mobiliariostore.com
- www.online.att.net.homephone.uverse.admasonry.ca

Bank of America

- onlinebanking-bankofamerica-com.tk
- bankofamerica-com.ml
- secure-bankofamerica.com
- enbankofamerica.ngjm.info
- www.secure.bankofamerica.spirotechkk.ml
- bankofamerica-online-reconnect.ga
- bankofamericar.miwomensconference.org
- bankofamerica.com.signinv2.es
- bankofamerica.online.dhadetailing.com
- login.bankofamerica-service.com
- email-bank0famerica.ga

Chase

- jpmorganchaase.mzf.cz
- jpmorganchaseorganizer.com
- chase.com.checkingaccounts-creditcards.online.confirm.id-find.support-user-mail.data.unitam.edu.mx
- chaseonline.chase.com.billsyst-paypal.com
- myjpmorganchaseupdate.hpage.co.in
- chasecomcredit-cards-rtbl-account-access-rtbl.000webhostapp.com
- jpmorganchaseonlinebankingverification.typeform.com
- chasebank.scotibanks.com
- chasesbanks.com
- jpmorganchaseauthe.ghaffarigroup.com

DHL

- www.dhl.info.pl
- dhltrackid.com
- dhl-tracking.bkth-bkk.com
- dhlexpress.cf
- dhlexp.cf
- dhlexpress.gq
- dhl.ogistics-files.us
- shipment-dhl-notification.000webhostapp.com
- delivery-shipment-dhl.000webhostapp.com
- dhlexpress.delivery.recordcorrection.net
- dhlship.tk
- dhl.deliveryinprogress.com
- dhlservices.deliveryinprogress.com

Docusign

- docusign-docs.tk
- docusignfile.com
- docusignn.com
- docusign.serviceim.com
- document-share-docusign-mess.classicalschoolathome.com
- docusignfolder.com
- docusign.signedcopy.online
- secureserver.docusign.completeddocs.com
- in-docusigns-files.com
- secureserver.docusign.confidentialfolder.com
- login.docusign.securedfolder.net
- login.docusign.encryptfolder.net
- www.docusigner.org

Dropbox

- dropbox.reddirtbbq.com
- www.securedropbox.file.godaddi.win
- dropbox.verification-wellsfargo.com
- dl-dropbox.github.io
- www.dropbox.com.secure.index-upload.files8h2s.com.yashmatiitc.org
- www.dropbox-en.ml
- w3dropbox.com
- dropbox.com.businesss.upload-documents78shh.com.steelmarineservice.com
- dropbox.org.mx
- webdropbox.net
- dropbox-documents.us

Facebook

- facebookbeta.ga
- vrf-facebook-account-com.cf
- www.facebook.ellom.in
- faceboksecurty.com
- facebooke.info
- facebookbr.net
- facebook1.co

- facebook.lbyts.com
- facebook-login.mywire.org
- www.facebook.getoffer.info
- login-facebook.hkdaily02.one
- www.facebook.com.bckiwanis.org

Google

- ssl-google-com-secured.spiritualscholars.com
- google.drive.chileboats.cl
- drive.google.cnc-style.de
- googlefile.comli.com
- googleuploader.com
- ssl-google-com-seured.strengthgrind.com
- drive.google.com.continue.mrstu.net
- drive.google.com.yaho.ml
- google.drive.amazonascde.com
- drive.google.com.continue.cqr4.biz
- ssl.gstatic.google.com.ventura.co.zm
- login.drive.google.avdihd.tk

Instagram

- instagram.com-accounts-login.frozenfoodsehat.com
- instagram.helpverifycation.com
- instagramn.co.nf
- www.instagramn.co.nf
- instagram.rezaee.ir
- insttagram.hol.es
- instgram.besaba.com
- instagram-verif.cf
- www-instagram.com-account-login.googlequest.gq
- www.instagram-verified.com
- instagram.com.userig.gq

Linkedin

- linkedin.n-koei-jp.com
- linkedinverification.eu
- www.login-linkedin.com
- linkediin.internetdatingstories.com
- linkedinn2.weebly.com
- linkedinvalidation.eu
- linkedin.com.marinyaki.com.au
- linkedn.altervista.org
- www.update-linkedin.com
- accounts-linkedin.gq
- linkedinsecurity.athenafinance.com.au
- refresh.linkedin.com.cgibin3302asuas.login-submit.cgibinsq435 .memberrefresh301231.delcos.org

Microsoft

- microsoftupgrade-net.ml
- microsoft-openings-security-alert-errorpage111.online
- match.microsoftexceldoc.com
- microsoft-outlook.355service.com
- microsoft-outlook.037fx.com
- outlook-microsoft.xv4567.net
- microsoftdrop.com
- login-microsoft-us.com
- login.microsoftonline.the-angel-network.com
- login.microsoftonline.com.l–6.us
- micrsoftomline.com

Netflix

- mail.netflix-pym.com
- netflix-pending.com
- wwwnetflixmobile.com
- accnetflixdate.com
- netflix-memservice.com
- netflixmaster.com
- premieraccount-netflix.com
- premium.billing.netflix.premier.login.premieraccount-netflix.com
- verify.netflix-alerts.com
- login-account-netfliix.com
- netflixptt.com
- netflixonlineverification.com

Orange

- www.orange.infoclientrsst.com
- www.orange.fr.facturation74589.site
- id.orange.fr.authuser2.bin.waltermarquez.com.ve
- www.orange.ne
- id.orange.fr.piggyhawk.net
- id.orange.service-orange-check.com
- idauthweb.orange.fr.bodrumsatilikemlak.net
- remboursement.orange.fr.indulgencehbs.com.au

Paypal

- paypal-resolve-now.sign-in-customer.info
- paypal.com.information.uaecorp.gq
- mail.paypal-account-limitied.ml
- paypal.verif-serv.info
- security-paypal.com
- paypal.support.webmpps-service.tk
- suport-secur.com-paypall.ga
- account-secure-paypql.ga
- paypal-myaccounts.com
- paypal.com-webapps.safeauth-key.me
- secure.paypal.com.serviceaccount-loginpage.privycy.info
- update-info-data-paypal.usa.cc

Wells Fargo

- myprofile2001.id3-440-wellsfargo.com
- servicewellsfargocom.ga
- wellsfargo.com.aandagruoupbd.net
- my.profile49394-wellsfargo.com
- wellsfargo.pekur.sk
- myprofile4001.9432948-wellsfargo.com
- welsfargo-fund.com
- updatebankingwells.acc-wellsfargo-info.xyz
- 1https.connect.secure.wellsfargo.com.tropfsteinhoehle.com
- wellsfargo.users.personalsecurity.findmybooks.in
- vnotice1001.wellsfargo3032030102.com
- disabled.3692832935.cipd.wellsfarg0.one-it.net

Xfinity

- xfinity.comcast.net.gupzi.com
- xfinity-redirect.hj.cx
- account-xfinity-update.hmm-wood.com
- xfinity-update.jordanstyles.gq
- login.comcast.net-xfinity-securelogin-validateaccount.x009.net
- connect.xfinity.com.free-flash-games-online.com
- upgradexfinitynow.com
- xfinity-signup.knottsapp.com
- www.xfinity-serviceupdates.com
- login.xfinity.com.hastyfreights.com
- xfinitysignin.from-va.com

Yahoo

- yahoo.update.azadari.co.uk
- mg15-yahoo.com
- account-yahoo.mooo.com
- www.yahoo.accountservices.ververoom.com
- www.access.logon.online.yahoo.ververoom.com
- yahooppn.tk
- www.deletion.account.yahoo.ververoom.com
- yahooupdatingprogram.cnc-cs.com
- login-yahoo-com.wisci.us
- yahoo.com.verifyuser.account.catpellet.com.tr
- login.mail.yahoo.verify2.enerqy.co.uk
- yahoo-mailverification.com

Benign domains (78 domains)

Amazon

- amazon.com
- account.amazon.jobs
- aws.amazon.com
- docs.aws.amazon.com

American Express

- americanexpress.com
- online.americanexpress.com
- global.americanexpress.com

Apple

- apple.com
- apps.apple.com
- support.apple.com
- secure1.store.apple.com

AT&T

- att.com
- paygonline.com

Bank of America

- bankofamerica.com

- secure.bankofamerica.com
- myhealth.bankofamerica.com
- prepaid.bankofamerica.com

Chase
- chase.com
- jpmorganchase.com
- access.jpmorgan.com
- paymentnet.jpmorgan.com

DHL
- dhl.com
- sso.dhl-usa.com
- mydhl.dhl.com
- logistics.dhl
- mydhl.express.dhl

Docusign
- docusign.com
- account.docusign.com
- support.docusign.com
- docusign.net
- docusign.utexas.edu

Dropbox
- dropbox.com
- help.dropbox.com
- dropboxforum.com

Facebook
- facebook.com
- mtouch.facebook.com
- newsroom.fb.com

Google
- google.com
- accounts.google.com
- play.google.com
- support.google.com

Instagram
- instagram.com
- help.instagram.com
- instagramers.com
- instagram-press.com

Linkedin
- linkedin.com
- mobile.linkedin.com
- business.linkedin.com

Microsoft
- microsoft.com
- onedrive.live.com
- office.live.com
- support.microsoft.com
- products.office.com
- windowscentral.com

Netflix
- netflix.com
- help.netflix.com
- dvd.netflix.com
- media.netflix.com

Orange
- orange.com
- topup.orange.com
- boutique.orange.fr

Paypal
- paypal.com
- paypal-prepaid.com
- paypal.me
- paypal-community.com

Wells Fargo
- wellsfargo.com
- connect.secure.wellsfargo.com
- wellsfargofinancialcards.com
- ebpp3.wellsfargo.com

Xfinity
- xfinity.com
- login.xfinity.com
- customer.xfinity.com
- xfinityprepaid.net

Yahoo
- yahoo.com
- mail.yahoo.com
- login.yahoo.com
- compose.mail.yahoo.com
- my.yahoo.com