Lessons From a Real World Evaluation of Anti-Phishing Training

Ponnurangam Kumaraguru, Steve Sheng
Carnegie Mellon University
ponguru@cs.cmu.edu, shengx@cmu.edu

Alessandro Acquisti, Lorrie Faith Cranor,
Jason Hong
Carnegie Mellon University
acquisti@andrew.cmu.edu, lorrie@cs.cmu.edu,
jasonh@cs.cmu.edu

Abstract— **Prior laboratory studies have shown that PhishGuru, an embedded training system, is an effective way to teach users to identify phishing scams. PhishGuru users are sent simulated phishing attacks and trained after they fall for the attacks. In this current study, we extend the PhishGuru methodology to train users about spear phishing and test it in a real world setting with employees of a Portuguese company. Our results demonstrate that the findings of PhishGuru laboratory studies do indeed hold up in a real world deployment. Specifically, the results from the field study showed that a large percentage of people who clicked on links in simulated emails proceeded to give some form of personal information to fake phishing websites, and that participants who received PhishGuru training were significantly less likely to fall for subsequent simulated phishing attacks one week later.**

This paper also presents some additional new findings. First, people trained with spear phishing training material did not make better decisions in identifying spear phishing emails compared to people trained with generic training material. Second, we observed that PhishGuru training could be effective in training other people in the organization who did not receive training messages directly from the system. Third, we also observed that employees in technical jobs were not different from employees with non-technical jobs in identifying phishing emails before and after the training. We conclude with some lessons that we learned in conducting the real world study.

Keywords- Design, Experimentation, Security, Human factors, Embedded training, Real world studies

I. Introduction

User education is a frequently-recommended and widely-used approach to countering phishing attacks [1, 12, 33], but few studies have evaluated the effectiveness of this approach in the real world. Researchers have demonstrated the effectiveness of PhishGuru, an embedded training system [20, 21]; and Anti-Phishing Phil, an online game [31] in laboratory studies. However, laboratory studies are unable to fully replicate real world conditions: they may lack ecological validity and do not sufficiently approximate real-world situations, which in turn may impact external validity — that is, the ability to make generalized inferences from the results [3]. The focus of this paper is to build on the earlier PhishGuru laboratory studies by conducting a similar study in a real world setting.

PhishGuru motivates users to pay attention to anti-phishing training materials by taking advantage of teachable moments. PhishGuru users are sent simulated phishing attacks via email and are presented training materials when they fall for the attacks. These emails might be sent by a corporate system administrator, ISP, or training company. The training materials present the following concepts in the form of a comic script: the definition of phishing, steps to follow to avoid falling for phishing attacks, and how criminals conduct phishing attacks easily.

Our goal is to evaluate the effectiveness of PhishGuru training in field trials and to study the effect of variations in the content of the PhishGuru training messages. To evaluate PhishGuru in the real world, we conducted a study with employees in a Portuguese company. The simulated phishing emails were all spear phishing emails targeted at the employees of the company. To investigate the effect of different training messages, we used one that had instructions on how to protect against regular phishing scams (generic training) and one that had instructions for protecting against spear phishing scams (spear training).

Our results demonstrate that the findings of PhishGuru laboratory studies do, indeed, hold up in the real world. As with the laboratory studies, our field study results showed that a large percentage of people who clicked on links in simulated emails proceeded to give some form of personal information to fake phishing websites, and that participants who received PhishGuru training were significantly less likely to fall for subsequent simulated phishing attacks one week later. In addition, we found the people trained with the spear phishing training material did not make better decisions in identifying spear phishing emails compared to people trained with the generic training material.

The remainder of the paper is organized as follows: In the next section we describe related work, including several training methods, and some relevant experimental studies. In Section 3, we present the study setup, participant demographics, and hypotheses that guided our study. In Section 4, we present the results of our evaluation, demonstrating that PhishGuru is effective in educating people in the real world. We discuss the effect of training people in the real world in Section 5. In Section 6, we present some limitations along with lessons learned. Finally, we present our conclusions and future work in Section 7.

978-1-4244-2969-1/08/$25.00 © 2008 IEEE

II. BACKGROUND

In this section we present an overview of security training methods, describe several methods for studying users' behavior in the context of phishing, and describe other experimental studies that have been conducted to evaluate the effectiveness of phishing training.

A. Security training methods

ISO and NIST security standards, which many companies are contractually obligated to follow, include security training as an important component of security compliance [13, 26]. These standards describe a three-level framework that includes awareness, training, and education. Security awareness activities are intended for all employees of a company and often include videos, newsletters, and posters. Training is generally intended only for employees who are involved with IT systems, mainly to provide basic computer security knowledge. Training is delivered primarily through classroom lectures, e-learning materials, and workshops. Education, intended for IT security specialists, is usually delivered via seminars or reading groups [25]. Our research offers some new approaches to delivering security awareness and training effectively.

There are many approaches to training users about phishing, including: articles about phishing on websites [8, 9, 10, 24], online cartoons about security [32], web-based phishing IQ tests [23], classroom training [28], security notices sent via email. These approaches vary in their cost as well as their effectiveness. For example, classroom training may be more effective than other training approaches because employees are required to spend dedicated time for training, but this approach is time-consuming for employees and expensive for companies that have to train a large number of employees. Online training materials are often an inexpensive approach, but it can be difficult to get people to read these materials and they are not always effective. The PhishGuru approach is to present training materials when people fall for phishing emails. This approach is effective because it motivates people to learn.

B. User study methods

To develop effective anti-phishing training materials it is essential to understand why users fall for phishing attacks and how anti-phishing tools and training materials impact their behavior. Researchers have used a variety of methods in user studies designed to gain insights into these issues. Interview studies have been conducted to gain insights into users' mental models and decision processes [7, 18]. Laboratory experimental studies where participants played a fictitious role and used personal information associated with that role have been used to test users' susceptibility to phishing attacks and evaluate the effectiveness of anti-phishing toolbars and training materials [2, 6, 14, 19, 20, 21, 31]. Laboratory experimental studies where participants used their own credentials have been used to evaluate the effectiveness of mutual authentication tools [30]. Real world studies have been used to evaluate participants' susceptibility to phishing, but not to evaluate the effectiveness of training [11, 15, 27].

Laboratory studies are very helpful in understanding user behavior in a given situation. However, each of these study methods have tradeoffs and face validity challenges: most of these studies are challenged with ecological (whether the methods, materials, and settings are similar to real life) and external (whether the results are generalizable) validity issues [3]. Laboratory studies in the context of phishing are also challenged with ethical issues of how much the researcher should inform the participant about the study and how much deception is acceptable [16, 17]. In one laboratory experimental setup, researchers showed that people who role-play behave differently from people who use their own credentials [30].

Understanding users' behavior in real world settings is critical to developing effective counter measures for phishing. Even though real world studies provide richer data, it may be difficult to control the study setup (due to many sources of variability) in the real world [29]. It can also be difficult to make the arrangements for a real world study, especially when it requires the cooperation of a company to gain access to employees or customers. Companies may not grant desired access or permit publication of study data or results. Real world studies also pose ethical challenges as they must often be conducted without obtaining prior consent from individual participants [16, 17].

C. Experimental evaluation of anti-phishing training

Few real world studies of users' behavior in the context of phishing have been conducted, and even fewer real world studies have been conducted to evaluate the effectiveness of anti-phishing training. Real world evaluations of anti-phishing training involve classroom and office training as well as training delivered via an online game. Researchers have evaluated the effectiveness of security notices and embedded training in laboratory studies.

The idea of sending fake phishing emails to test users' vulnerability has been explored by several groups. Jagatic et al. conducted a study in which they obtained information about friend relationships from social networking web sites and used it to send phishing emails to Indiana University students that appeared to come from one of their friends. A large percentage of students fell for these phishing attacks [15]. Ferguson did a two-part study among West Point cadets. In the first phase, cadets were tested for their ability to detect phishing attacks. In the second phase, cadets were given classroom training and lectures about phishing and then tested. Ferguson showed an improvement in the cadets' ability to identify phishing emails after the training [11]. Similar to the West Point cadet study, the New York state office of Cyber Security & Critical Infrastructure Coordination conducted a two-part study among their employees. In this study, participants who fell for simulated phishing attacks were presented with online educational materials on how to protect themselves from phishing. This study also showed anti-phishing training improved participants' ability to identify phishing emails [27].

Sheng et al. have shown that people can be trained about phishing URLs through an online game called Anti-Phishing Phil. In a laboratory study, they found that users made better decisions when trained with the game than with existing online training materials [31]. They found similar results while testing the game in the real world [22].

Previous research results provide strong evidence that people make better decisions when they are trained through embedded training versus the current practice of sending security notices [20]. Research also suggests that people retain and transfer more knowledge when trained with embedded training than with non-embedded training [21]. The focus of this paper is on testing embedded training in a real world setting.

III. EVALUATION

In this section we present participant demographics and study methodology along with the hypotheses that we tested in this study.

A. *Participants and demographics*

This study was conducted at a large Portuguese company. All emails and training materials were translated into Portuguese. All participants in the study worked in the same floor of an office building. Participants were from different areas of work in the company: administration, business, design, editorial, management, technical, and others.

The study included three conditions: "control," "generic training," and "spear training." Participants in the control condition did not receive any training. Participants in the generic training condition received a simulated spear phishing email and saw generic phish training material (Figure 1) when they clicked on a link in the email. Participants in the spear training condition received a simulated spear phishing email and saw spear phish training material (Figure 2) when they clicked on a link in the email. We assigned 111 employees to the control condition, 100 to the generic training condition, and 100 to the spear training condition. Table 1 presents the demographics of the study participants.

B. *Study setup*

The company we worked with was primarily interested in studying the vulnerability of their employees towards spear phishing emails, so we used spear phishing emails for all simulated phishing emails in this study. Targeted spear phishing attacks have been more successful than generic phishing attacks in coning people and causing damages to companies and individuals.

In total, participants received four emails during the study: three simulated spear phishing emails and one legitimate email containing a link. All the spear phishing emails and the legitimate email were based on actual emails that the company had received or the kind of emails that the system administrators were worried about.

The first email that employees received was a training email (Train) and was delivered on Day 0. This email was sent only to employees in the generic and spear conditions. This email was a spear phishing email that asked employees to click on a link to enter their user name and password in order to use the corporate network. When employees clicked on the link in this email, they were taken to the training material corresponding to the condition they were in. Participants in the generic training condition saw the generic phish training message shown in Figure 1, while participants in the spear training condition saw the spear phish training message shown in Figure 2.

The second email (Test 1) was designed to measure the knowledge that employees acquired through our training materials. In order to compare trained and untrained employees, this email was sent to employees in all conditions. To measure immediate effectiveness this email was sent on Day 2 of the study. This simulated spear phishing email said that the recipient's internal network password has expired and asked them to click on a link and change their password. When employees clicked on link in this email, they were taken to a fake phishing website that looked the same as the real website and was hosted on a similar-looking domain name.

Learning science literature defines retention as the ability of learners to retain or recall the concepts and procedures taught when tested under the same or similar situations after a time period δ from the time of knowledge acquisition [5]. The third email (Test 2), which was designed to measure retention, was sent on Day 7. As in Test 1, to compare the trained and untrained employees, this email was sent to participants in all conditions. This email asked employees to click on a link and update their communication information for internal corporate communication purposes. When employees clicked on the link they were taken to a phishing website that looked the same as the real website and was hosted on a similar looking domain name.

To test whether training increases participants' concern level such that they stop clicking on any links, even legitimate ones, we sent a legitimate email with a link (Test 3) on Day 10 to all participants in the study. To compare the trained and untrained employees, this email was sent to participants in all conditions. This email asked employees to click on a link to read the company's updated security policy. When employees clicked on the link, they were taken to a legitimate webpage with the updated security policy. Table 2 summarizes all emails, email types, days on which the email was sent, the conditions to which the emails were delivered, and relevant features of the email.

Phishing websites that were linked to the spear phishing emails were exact replica of real company websites but hosted on a domain that looked similar but not the same as the company's domain. All replicated websites were completely functional and allowed employees to submit information. We wanted only the employees of the company to access the training materials and fake phishing websites, so, these websites were hosted in a way that only IP addresses coming from the company's domain were able to access the websites. This also helped us in identifying the IP address and thereby the user from whose machine the request had come. The company tracked all these information and for privacy reasons, we did not receive the specific details like the IP address, etc. from the company. We tracked the clicks to the phishing websites and the training materials, as well as the information that was submitted to the phishing websites.

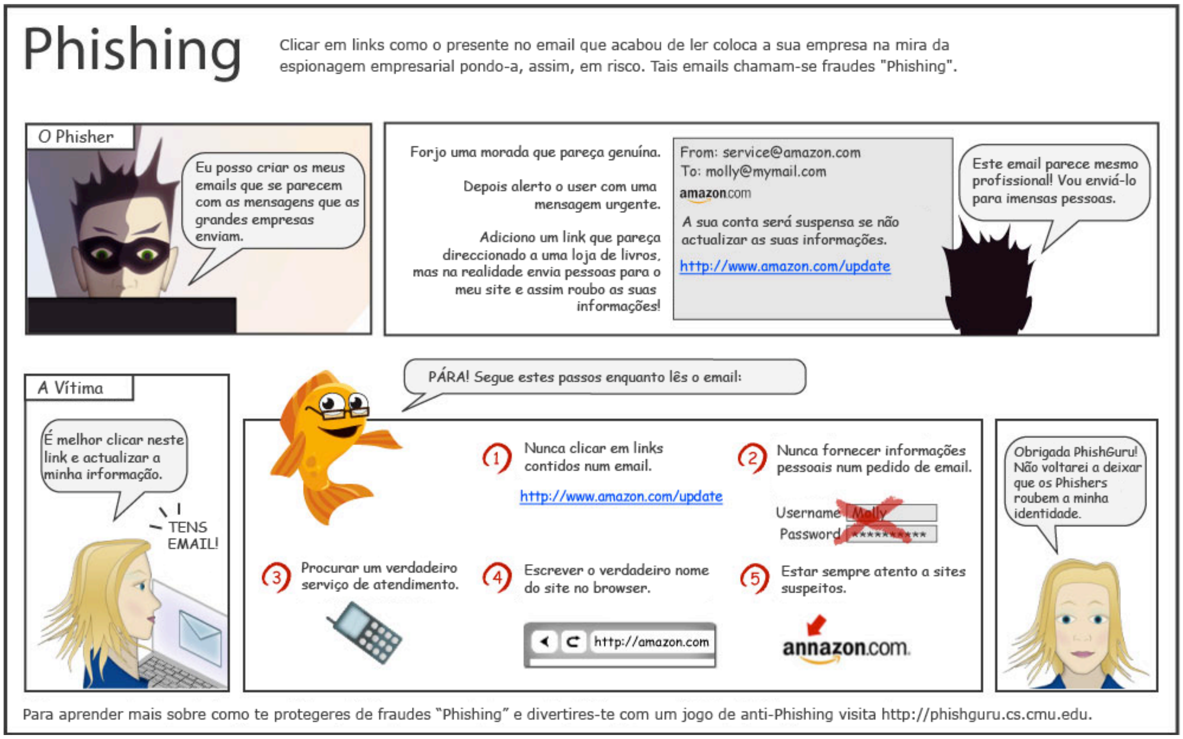
Phishing
Clicar em links como o presente no email que acabou de ler coloca a sua empresa na mira da espionagem empresarial pondo-a, assim, em risco. Tais emails chamam-se fraudes "Phishing".
O Phisher
Eu posso criar os meus emails que se parecem com as mensagens que as grandes empresas enviam.
Forjo uma morada que pareça genuína.
Depois alerto o user com uma mensagem urgente.
Adiciono um link que pareça direccionado a uma loja de livros, mas na realidade envia pessoas para o meu site e assim roubo as suas informações!
From: service@amazon.com
To: molly@mymail.com
amazon.com
A sua conta será suspensa se não actualizar as suas informações.
http://www.amazon.com/update
Este email parece mesmo profissional! Vou enviá-lo para imensas pessoas.
A Vítima
É melhor clicar neste link e actualizar a minha informação.
TENS EMAIL!
PÁRA! Segue estes passos enquanto lês o email:
1 Nunca clicar em links contidos num email.
http://www.amazon.com/update
2 Nunca fornecer informações pessoais num pedido de email.
Username Molly
Password **********
3 Procurar um verdadeiro serviço de atendimento.
4 Escrever o verdadeiro nome do site no browser.
http://amazon.com
5 Estar sempre atento a sites suspeitos.
annazon.com
Obrigada PhishGuru! Não voltarei a deixar que os Phishers roubem a minha identidade.
Para aprender mais sobre como te protegeres de fraudes "Phishing" e divertires-te com um jogo de anti-Phishing visita http://phishguru.cs.cmu.edu.

Figure 1. People in the Generic condition saw this comic strip.
An English version of this comic strip is given in the Appendix (Figure 6).

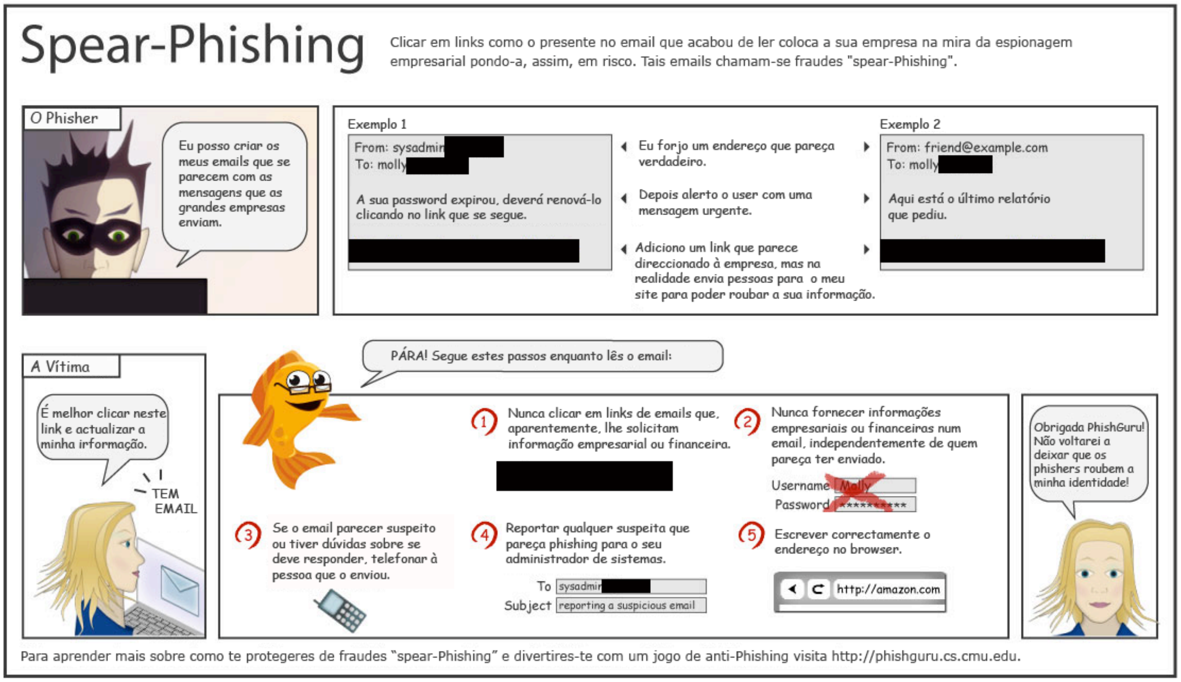
Spear-Phishing
Clicar em links como o presente no email que acabou de ler coloca a sua empresa na mira da espionagem empresarial pondo-a, assim, em risco. Tais emails chamam-se fraudes "spear-Phishing".
O Phisher
Eu posso criar os meus emails que se parecem com as mensagens que as grandes empresas enviam.
Exemplo 1
From: sysadmi[redacted]
To: molly[redacted]
A sua password expirou, deverá renová-lo clicando no link que se segue.
[redacted]
Eu forjo um endereço que pareça verdadeiro.
Depois alerto o user com uma mensagem urgente.
Adiciono um link que parece direccionado à empresa, mas na realidade envia pessoas para o meu site para poder roubar a sua informação.
Exemplo 2
From: friend@example.com
To: molly[redacted]
Aqui está o último relatório que pediu.
[redacted]
A Vítima
É melhor clicar neste link e actualizar a minha informação.
TEM EMAIL
PÁRA! Segue estes passos enquanto lês o email:
1 Nunca clicar em links de emails que, aparentemente, lhe solicitam informação empresarial ou financeira.
[redacted]
2 Nunca fornecer informações empresariais ou financeiras num email, independentemente de quem pareça ter enviado.
Username Molly
Password **********
3 Se o email parecer suspeito ou tiver dúvidas sobre se deve responder, telefonar à pessoa que o enviou.
4 Reportar qualquer suspeita que pareça phishing para o seu administrador de sistemas.
To sysadmi[redacted]
Subject reporting a suspicious email
5 Escrever correctamente o endereço no browser.
http://amazon.com
Obrigada PhishGuru! Não voltarei a deixar que os phishers roubem a minha identidade!
Para aprender mais sobre como te protegeres de fraudes "spear-Phishing" e divertires-te com um jogo de anti-Phishing visita http://phishguru.cs.cmu.edu.

Figure 2. People in the spear condition saw this comic strip.
An English version of this comic strip is given in the Appendix (Figure 7).

TABLE I. DEMOGRAPHICS OF THE PARTICIPANTS.

	Control Condition (N=111)	Generic training condition (N=100)	Spear training condition (N=100)
Gender			
Male	77%	27%	67%
Female	23%	73%	33%
Areas of work			
Administration	1%	1%	1%
Business	2.7%	5%	9%
Design	5.4%	3%	7%
Editorial	4.5%	5%	7%
Management	22.5%	19%	20%
Technical	39.6%	36%	35%
Others	24.3%	31%	21%

To make sure the employees received the emails that were part of the study, system administrators bypassed the corporate email filters and placed them in participants' inboxes.

We asked all participants to complete a post-study survey on Day 20. The survey consisted of questions regarding (1) the interest level of participants in receiving such emails in future, (2) participants' feedback on the training, and (3) participants' feedback on the instructions.

C. *Hypotheses*

In this section we introduce three hypotheses which informed the study described in the paper.

1) Replicating laboratory study results

Earlier laboratory studies have shown that a large percentage of participants who click on links in simulated emails proceed to give some form of personal information to the phishing website. This percentage was around 90% in earlier laboratory studies [20, 21]. Our goal was to investigate whether this is true in a real world setting. This result may show that people have to be trained not to click on links, otherwise, there is low probability that they will click and not give personal information to phishing websites.

Hypothesis 1: *A large percentage of people who click on links in simulated emails proceed to give some form of personal information in the real world.*

A laboratory study showed that users learn, retain, and transfer effectively when training materials are presented after they fall for a phishing attack [21]. Our goal was to investigate whether this is true in a real world setting.

Hypothesis 2: *PhishGuru (embedded training) is effective in training people in the real world.*

To evaluate the effectiveness of PhishGuru, we calculated the following: (1) percentage of participants who clicked on a link in phishing emails and gave information to fake phishing websites immediately after the training; (2) percentage of participants who clicked on a link in phishing emails and gave information to fake phishing websites after a delay of 7 days from the training; and (3) percentage of participants who clicked on a link in legitimate emails after the training.

2) Generic and spear training instructions

The content of training materials makes a difference in the way people learn and reproduce knowledge. Researchers have shown that people make better decisions if the testing situation is the same or similar to the training situation and the training materials than if the testing situation is different [5]. To investigate the effect of the difference in the instructions, we developed one set with anti-phishing instructions that were generic and another one specific to spear phishing emails. Figure 1 and Figure 2 have the same content except for the instructions in the lower pane of the material. As the training materials used in the study were in Portuguese, the translated English version of the instructions is given in Table 3. The English version of the messages is given in the Appendix (Figure 6 and Figure 7).

Hypothesis 3: *People trained with spear training material make better decisions in identifying spear phishing emails compared to people trained with generic training material.*

IV. RESULTS

In this section we present the results of our study. The results from this study support Hypotheses 1 and 2, but not Hypothesis 3. We found a large percentage of the participants who clicked on links in simulated emails gave away some form of personal information to the fake phishing websites that were part of the study. We found participants in the training conditions made significantly better decisions after the training compared to before the training. Our results suggest that users retained knowledge gained from PhishGuru for at least 7 days after the training. However, the difference in the instructions in our training materials did not have a significant effect on the participants' ability to identify phishing emails. Surprisingly, our results also suggest that PhishGuru training could be effective in training other people in the organization who did not receive training messages directly from the system. The complete decision tree for all the three conditions is given in the Appendix.

A. *Giving away personal information*

In this study we found that a large percentage of the participants who clicked on links in simulated phishing emails went ahead and gave some form of personal information to the phishing websites. The system administrators in the company who helped us conduct the study had access to the information that was entered into phishing websites. They were able to check the usernames and other details that were entered. We found that 88% of the participants who clicked on links went ahead and gave some form of personal information to the fake phishing websites. In laboratory studies, researchers have found that 90 to 93 percent of participants who clicked on links gave their personal information to fake phishing websites [20, 21]. Table 4 gives the percentage of participants in each condition who clicked on a link in phishing emails, and who clicked and gave information to fake phishing websites.

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**