Forcing Johnny to Login Safely

Long-Term User Study of Forcing and Training Login Mechanisms

Amir Herzberg and Ronen Margulies

Dept. of Computer Science, Bar Ilan University
{herzbea,margolr}@cs.biu.ac.il

Abstract. We present the results of the first long-term user study of site-based login mechanisms which force and train users to login safely. We found that interactive site-identifying images received 70% detection rates, which is *significantly better* than passive indicators' results [15, 8, 12]. We also found that login bookmarks, when used together with 'non-working' links, doubled the prevention rates of reaching spoofed login pages in the first place. Combining these mechanism provides *effective prevention and detection* of phishing attacks, and when several images are displayed in the login page, the best detection rates (82%) and overall resistance rates (93%) are achieved. We also introduce the notion of *negative training functions*, which train users not to take dangerous actions by experiencing failure when taking them.

1 Introduction

Phishing, i.e., password theft via fake websites, is an extremely worrying, wide spread and worldwide phenomenon. With billions of dollars lost and dozens of percents increase in the amount of attacks over the years [1, 19, 20], it appears that there is still a need to improve the defenses against phishing.

Psychology can provide important insights that can help improve anti-phishing defenses, by understanding what makes users so susceptible to phishing. In [11], Karlof et al. described how humans tend to develop automatic responses to situations repeating themselves. In familiar situations, the human brain makes us respond *mindlessly* with the action that is usually most appropriate; for such responses, the psychologist Robert Cialdini coined the term *click whirr* responses [4]. Cialdini says that these responses are like pre-recorded tapes in our heads, and when we encounter a familiar situation we automatically "click the play button" (which makes a whirr sound, hence click whirr). Karlof et al. explained that users developed a click whirr response to login forms, and will *automatically* submit their credentials to a login form residing on an interface they have seen before. Our main focus is on influencing users' responses to spoofed emails and web pages, and their decision-making process.

In addition, as was clearly seen on our user study (section 5), two other click whirr responses are shared by most users. The first is to follow email links from

V. Atluri and C. Diaz (Eds.): ESORICS 2011, LNCS 6879, pp. 452–471, 2011.
© Springer-Verlag Berlin Heidelberg 2011

a familiar sender, as clicking a link is a natural thing to do when meeting one (as was also noticed by Karlof et al.). The second is to put trust in a site's home page that looks familiar (even if not protected by SSL), and move to the site's login page when wanting to login (by clicking an "Enter Your Account" button/link). The three mentioned click whirr responses make the Internet a fertile ground for phishing attacks.

1.1 Current Mechanisms: Passive Indicators

Early web browsers include three main indicators to help users identify the websites they visit: the address bar (indicating the site's URL), the *https* prefix and the padlock/key image (both indicating SSL usage). Several experiments [5, 9] showed that users often enter their passwords *without validating them*. This is not surprising: the indicators are not very visible, and there is no mechanism forcing or training users to inspect them, so it is easier to just skip those checks.

Several proposals and implementations for enhanced indicators, most involving change to the browser, and few that require only support by the website, were introduced. Those indicators display warnings [21], a user-custom image/text to help the user identify the site [9, 17, 3], SSL Certificate information [9, 16], and emphasis of the domain name and protocol in the URL bar.

Both the 'classical' and 'enhanced' indicators are *passive*, i.e., they are *only displayed* to the user and *require no action* in a regular login. Several experiments measuring users' ability to detect fake sites using different (enhanced) indicators, resulted in mostly disappointing results [8, 15, 12].

1.2 Interactive Custom Indicators

We found that the use of *interactive custom indicators* can significantly improve the ability of users to avoid phishing. Interactive custom indicators force the user to interact with her customized (personal) indicator (image/text) in order to login. An example is Passpet [18] – a firefox extension which acts both as a password manager and an interactive custom indicator.

One important advantage of interactive custom indicators is that they can be implemented by changes to only the website, without requiring any change on the client side; this is significant as changes in the client side are often complex for users, requiring expensive support, and require support of multiple machines and browsers. With a site-based interactive custom indicator, the login page can hide the password text field until the user clicks her custom indicator.

The interactive nature of these indicators, which *conditions* users to find and click their custom image/text, makes users more alert to the indicator's absence on a spoofed login page. There was almost no study of the effectiveness of interactive indicators in detecting spoofed login pages [1].

[1] An exception are the encouraging results of the preliminary phase of this research (see in Dvorkin's thesis [6]): ∼20% – ∼40% detection rates for passive indicators, ∼85% for interactive.

1.3 Secure Login Using a Bookmark

The initial stage of a phishing attack is to get the user to enter a spoofed login page. A common scenario for the initial stage is to send a spoofed email containing a link to a spoofed login page. Users might follow email links, which could be risky since emails are easily spoofed. In addition, most sites' home pages are not protected by SSL and can be spoofed by a MITM; users might put their trust in a spoofed page looking similar to their target site's home page, and follow its "Enter Your Account" button which leads to the (spoofed) site's login page. Search engine results are also not protected by SSL and can be spoofed by a MITM, thus leading the user to a spoofed site.

A good habit for accessing high-value sites is to create a browser bookmark ('favorite') for a sensitive site's (https) login page, and *always surf to the site's login page via that bookmark*. Adida has presented BeamAuth [2], a *two-factor authentication* mechanism based on a login bookmark. In this mechanism, users receive a special login-bookmark from the website, containing a secret token, which identifies them to the site. To ensure that users will *always* login via their bookmark, the login page looks for the secret token and prohibits the login if a valid token is not supplied (an error message is displayed on the login page, which trains the user not to enter the login page in any way but the bookmark). A login bookmark ensures reaching the correct URL, and by containing an *https* prefix, ensures a secure channel is established.

Apart from initial and non-reliable results of this research [6], there was no study of the effectiveness of login bookmarks in preventing users from reaching spoofed sites. There was also no study of other mechanisms that aim to prevent users from following email links or entering the site's login page via its (non SSL protected) home page.

1.4 Challenges and Requirements for User Studies

User studies should try to emulate users' real-life activities. Most previous studies in the field [15, 8, 12] were short term (few hours) lab studies. Such studies experience problems in making participants act as in real life: if they know the true intention of the study, they will be more cautious than in real-life, and if a false intention is introduced they will be less cautious; even if a sense of risk is added (for . using their own bank accounts), the study's environment, which is not in the users' regular environment, might influence their behavior [14].

It is important to complement short-term lab studies with long-term real-life studies, where participants use an online system, with a different purpose than security, regularly for several months, and login from anywhere they want. This kind of study is closer to real-life, and even if the study's purpose is introduced, users' motivation to detect attacks is not expected to be higher than usual due to the constant usage and other purposes of the system. To the best of our knowledge, no long-term user study which examined users' response to emails and the detection rates of spoofed pages was previously conducted.

Table 1. Detection rates and overall resistance rates for a classic phishing attack. Cells are merged when results were combined for higher confidence or when it does not make sense to split (e.g. 'non-working' links don't affect the detection rates, only the prevention rates).

mechanisms	detection rate	resistance rate
none	19.61±4.95	40.22±10.24
bookmark only	42.56±5.61	49.84±14.77
bookmark+'non working' links		81.08±12.88
image only	59.84±6.24	76.12±8.44
bookmark+image	72.71±6.31	80±10.03
bookmark+'non working' links+image		
bookmark+4 images	81.94±5.17	93.24±7.8
bookmark+'non working' links+4 images		

1.5 Our Contributions

We have conducted an extensive long-term user study of real-life web and email activities, which included different kinds of simulated phishing attacks. We examined the effectiveness of different phishing defense mechanisms, including a login bookmark, interactive custom images and their combination.

We also tested mechanism sites can use to prevent users from reaching spoofed login pages – intentionally including 'non-working' links/buttons in the site's home page and email announcements. From the results we concluded that users need to *experience failure* in order to avoid dangerous actions, and introduced the notion of *negative training functions*.

We found significant differences between the mechanisms' detection rates and overall resistance rates (see table 1). From all the results of our study we derived a set of conclusions and guidelines that (high-value) sites can use. We present an important conclusion derived from table 1 in advance:

Conclusion 1. *By combining a login bookmark with 'non-working' email links and an interactive custom image, and displaying multiple images in the login page, the detection rates and overall resistance rates are higher than any other mechanism previously tested in a real-life experiment (82% and 93% respectively).*

Detailed discussion of the table and the study's results and conclusion is given in section 5. In our study we also measured the effectiveness of browsers' SSL certificate warnings and conducted a usability survey.

Another contribution is WAPP (Web Application Phishing-Protection), a site-based anti-phishing solution we designed and implemented, which combines a login bookmark with multiple interactive images to constitute a *conditioned-safe* login ceremony (see section 2.1).

1.6 Paper Organization

Section 2 describes the principles for effective anti-phishing mechanisms, which was used as a basis for WAPP (section 2.4) and for our user study (section 3). In section 4 we present a comprehensive threat analysis and users' expected behavior for simulated attacks, and section 5 presents the results and conclusions from our user study. Finally, section 6 presents the results of our usability survey.

2 Principles for Effective Anti-Phishing Mechanisms

2.1 Conditioned-Safe Ceremonies

Karlof et al. [11] introduced the notion of a *conditioned-safe ceremony*, which is a ceremony that "deliberately conditions users to reflexively act in ways that protect them from attacks", i.e., forces users to take actions that are safe.

Forcing functions, which were also mentioned by Karlof et al., are behavior-shaping constraints, used in the human reliability field aiming to prevent human errors. Forcing functions usually work by *preventing a user from progressing in her task* until she performs a specific action whose omission results in failure. Because users must take this action during every instance of the task, the forcing function trains them to *always* perform this action, and after a short experience will become a *click whirr response*.

To defeat conditioned-safe ceremonies, attackers will try to make users perform a dangerous action instead of the forcing function, thus bypassing its protection. For example, convincing users to follow a link to a spoofed login page instead of clicking the login bookmark which leads to the correct login page. Since such actions are indeed possible (as our study's results show), we introduce the notion of *negative training functions*.

Unlike forcing functions, negative training functions are *not* a part of the ceremony, and train users to *never* perform dangerous actions by making them *experience failure* when performing those actions. Negative training functions can come together with forcing functions to better train users of what should and what should not be done.

2.2 Design Goals for a Conditioned-Safe Login Ceremony

Since humans tend to make routine actions mindlessly, and in particular skip any voluntary action during a login ceremony, we should not fight this tendency. A conditioned-safe login ceremony should consist of *several forcing and negative training functions* – at least one function against each click whirr response that puts users in danger. In particular, there should be forcing and training functions against:

1. Automatic following of links.
2. Automatic submission of credentials.
3. Automatic entrance to a site's login page by clicking an "Enter Your Account" button in its home page (which could be spoofed).

2.3 Mechanisms of Interest

On our user study we focused on mechanisms which are either forcing or negative training functions, since we believed they will be the most effective mechanisms against phishing. In addition, such mechanisms were never tested before, apart from initial encouraging results by Dvorin [6]. The forcing functions mechanisms we used were:

1. A login bookmark, which forces users to login via their bookmark only, since a login attempt not via the bookmark results in failure.
2. An interactive custom image, which forces users to find and click their custom image in order to submit their password and login.

Since the bookmark and image clicks are necessary for each login, they provide *constant training*; the more a user is trained, lower are the chances for her to omit the forcing functions. The interactive custom image defends against automatic submission of credentials and the login bookmark defends against reaching a spoofed login page (for example by following email links or entering a site's login page via a non-SSL protected home page). The combination of the two mechanisms provides both *prevention and detection* of spoofed login pages, thus achieves *defense-in-depth* and guarantees that omitting only one of them will not suffice. Therefore, a combined method is also worth testing.

To login to a site implementing the combined method, the user has to first click her bookmark which leads her to the login page and sends the secret token to the site for initial identification. The site then recognizes the user and sends her custom image to be displayed in the login page. After the user clicks her custom image she will be able to submit her password. If her username, password and secret token match, the site allows her to login.

The login ceremony of the combined method is simpler and faster than typical login ceremonies, as it requires only two mouse clicks (for the bookmark and interactive image), and saves the need for typing the site's URL (or looking for it on a search engine) and moving from the site's home page to its login page. In addition it saves the need for typing the username, as it can be automatically filled-in (see [2]). The ceremony is even faster with browser auto-completion.

A variation of interactive custom images is to choose (and click) the custom image from a small set of random images on the login page. This variation makes the user even more aware of her custom image (as seen on our study), and improves the detection rates. With only one image displayed, the user's click whirr response could cause her to immediately click on any image displayed to her, especially if she forgot her image. Making her choose the correct image, reduces the chance for an immediate mindless click. In addition, with several images being used the site can notice when a user has forgotten her image (after several mistaken clicks), and can refresh her memory. Another idea is to present animated images which attracts the human eye more than static images and could increase memorability.

In addition to the login bookmark, we also tested two negative training functions aiming to prevent users from reaching spoofed login pages:

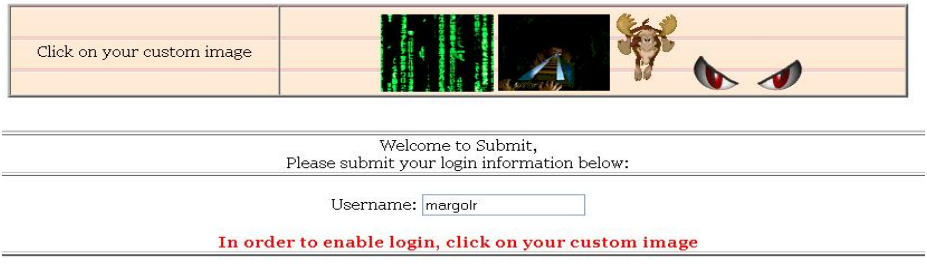

Fig. 1. The login page requires the user to click her correct custom image

1. Intentionally including "non working" links to the login page in the site's email announcements; when clicked, the user reaches the login page which displays an error telling her to login only via her bookmark. This experience trains the user to never follow links.
2. Intentionally including a "non working" account-entrance button in the site's home page. When clicked, the user reaches the login page and displayed with a similar error page. This experience trains the user to never enter the site's home page in the first place when wanting to login.

2.4 WAPP

We designed and implemented WAPP (Web Application Phishing Protection), a server side solution which implements all the above mentioned mechanisms. We refer to a site implementing those mechanisms as a WAPP-protected site. A demo of WAPP is available at `http://submit2.cs.biu.ac.il/WAPP/`.

3 Long-Term User Study

3.1 Study's Framework System

For our long-term study we used an online exercise submission system called 'submit' (`submit.cs.biu.ac.il`), which is used by most courses at the computer science department of Bar-Ilan University. With the submit system, students submit their exercises and receive emails announcing new grades. Due to its wide usage, most users logged-in to the system dozens and even hundreds of times throughout the study. We have used the system for two years (four semesters), among a population of ∼400 students, and present the results of the first three semesters (initial analysis of the results of the fourth semester correlate with the earlier results). In addition to its primary usage as an exercise submission system, we simulated several phishing attacks, and collected the attacks' results.

3.2 Introducing the Experiment

When users experience a sense of risk or are aware of the security purpose of the system, they become more concerned about security. Yet, its long-term usage

and the fact that its primary usage isn't security, but rather exercise submission, should not cause more concern than user's real life concern for high-value sites. We had two variations of our experiment:

First Experiment – Weak Motivation. In the beginning of the study we announced an up to 5 points bonus in one of Herzberg's courses of their choice for correctly detecting attacks to provide the users with an incentive to cooperate. After analyzing the results of the first two semesters and users' answers to an online survey we applied at the end of the second semester, we found that 26% of the students did not cooperate with our experiment and did not try to detect attacks. We removed the results of those users (see appendix B) and concluded that further motivation for cooperation and a higher sense of risk are needed.

Second Experiment – Extra Motivation. In the third semester we introduced our study in a more informative way: on the first login, each user, including users from previous semesters, was introduced with an instructions page which shortly described: a) what phishing is and the extent of phishing attacks; b) who we are and what are our goals; and c) the experiment details. We asked the students to cooperate and promised our gratitude. We also announced the 5 point bonus and told users they will lose bonus points upon classification mistakes. Users that used the system in the previous year knew that the bonus points were indeed granted, and that they depend on the correct classification rates.

3.3 Users' Login Methods

Upon registration, each user was randomly assigned a login method from the following five methods:

image only an interactive custom image only
bookmark only a login bookmark only
bookmark+image a login bookmark combined with an interactive custom image
bookmark+4 images like the latter, where the login page displays 4 images
none no site-based indicator assigned, used as a control group

Users could only login using the method assigned to them upon registration. In normal usage of the system, `non-bookmark` users (`none` and `image only`) have reached the login page, which was running over https, via the system's home page, which was running over http, by providing their username. `Bookmark` users (`bookmark only`, `bookmark+image` and `bookmark+4 images`) that entered the login page the same way received an error message (except when they were attacked) telling them to login via their bookmark, and could not login. This simulated our second negative training function.

3.4 Users' Email Methods

Each user was also assigned an email method, which determined how she will get her new grade announcement emails from the system. We used three types of

emails – 1) emails that contain a link to the system's login page telling the users they have to login to view their grade and submission details; 2) emails that contain the grade and submission details within the mail body and contain no link; 3) emails like the latter containing no link and also containing a warning at the end of the email body, saying that the system never includes links in its emails since clicking links in emails is dangerous. We refer to these email methods as `link`, `no link`, and `warning` respectively.

Users always got the same type of email except when the system had sent a spoofed email. `Bookmark` users from the link group received 'non-working' links regularly (except when they were attacked), as the links directed them to the system's login page where they were shown an error message. This simulated our first negative training function.

3.5 Attacks

When users tried to enter the system's login page via its home page or via their bookmark, there was a low probability for them to be randomly directed to one of the system's spoofed login pages. We uses low attack probabilities to prevent increased awareness due to frequent attacks. Spoofed emails were also sent to users with rather low probabilities; these emails contained similar content to the system's genuine emails sent to link users, apart from the fact that the links contained the URL of a spoofed page. See the full version of our paper [10] for a detailed description of the attacks and logging of the results.

4 Threat Analysis

In this section we attempt to analyze all realistic phishing attacks against a site implementing the mentioned forcing and negative training functions mechanisms (which we refer to as a WAPP-protected site). We hypothesize users' expected behavior in the different attack scenarios, and describe how we simulated those attacks in our user study.

Classic Phishing Attack. In a classic phishing attack a spoofed email containing a link to a spoofed login page is sent to the user. WAPP defends against this attack by training the user not to follow the link and always login via her bookmark. If the user makes an error of omission and follows the link, her interactive custom image won't be displayed and the user will most likely understand she reached a spoofed page and resist the attack. We executed this kind of attack in our user study and measured whether `bookmark` users and in particular users with 'non-working' links, are less likely to follow links, and if they did, whether the interactive custom image is effective in detecting the spoofed page.

Malicious Bookmark Replacement. An attacker might trick the user into replacing her WAPP bookmark, using for example, a spoofed email that mimics the site's registration email, or overriding the bookmark by creating a bookmark with the same name when the user visits the attacker's site.

Even if the bookmark has been replaced, the secret token is not revealed and the user will most likely identify she reached a spoofed site since her custom image does not appear, thus not provide her password. This way she could suspect the new installed bookmark, delete it, and replace it with the original bookmark she received from the site. We simulated the second part of the attack by redirecting `bookmark` users to a spoofed login page after clicking their bookmark, as if it was previously replaced, and measured the detection rates.

Pharming Attack. A MITM attacker who hijacks a DNS entry will direct the user's traffic for the legitimate website towards the attacker's machine. Since the bookmark link start with https, the user's browser will try to establish an SSL connection with the attacker's machine and notice an invalid certificate. When encounter an invalid certificate, modern browsers don't display the site's content, and display active certificate warning pages, with alarming text and colors, instead. Users are required to manually add an exception or approve in order to forward to the site.

Egelman et al. [7] found high percentage (79%) of users resisting phishing attacks with active browser warnings, which makes sense as it is an interactive forcing function which prevents users from progressing with the login. For a user entering the spoofed site despite the warnings, the attacker can gain the user's secret token and custom image, and will most likely gain the user's password as well. In our user study we simulated a pharming attack by using a spoofed login page in the same domain as the submit system, which used an invalid certificate. We measured the percent of users entering the spoofed page despite the browser warnings. For a user that did enter, though a MITM could present the user with her real custom image, we did not do that in our study, as no added value could be achieved from such an attack. Instead, we did not display the image and measured the detection rates as in the other spoofed pages. We wanted to find out if the detection rates are better when a preventive forcing function is combined with a detective one.

Spoofing the Home Page. Since most sites don't apply SSL at their home page (mostly for performance reasons), the site's home page could be spoofed. A MITM can change the site's home page or the results returned by a search engine in case the user looked for the site's URL. Another scenario is that an attacker sends a spoofed email with a link to a spoofed home page instead of a spoofed login page. After reaching the spoofed home page, the user might put trust in the site and forward to the (spoofed) login page by clicking an "Enter Your Account" button/link.

WAPP users would most likely not enter the site's home page in the first place, since by previously doing so they have experienced a login failure due to the 'non-working' account entrance button. Even if the user performs an error of omission, enters the spoofed home page and forwards to the spoofed login page, she won't see her custom image and will most likely notice the attack. We have simulated this attack for `non-bookmark` users (which always logged-in from the home page) by directing them to spoofed pages. `Bookmark` users trying to enter the login page via the system's home

page mostly reached the genuine login page and received an error message, but some of their attempts also led them to a spoofed login page.

An attacker can try additional attack scenarios (which are discussed in appendix A) as a preliminary phase to a phishing attack, in order to get a hold of the user's secret token. We did not execute these attacks due to previous studies' results, complication, and ethical reasons. Our study's results empirically proved that WAPP is well protected against all kinds of realistic phishing attacks. Only two-phased attacks that try to gain the user's secret token prior to a phishing attack could effectively defeat WAPP. These complicated attacks require more effort and resources from phishers, and are likely to significantly reduce phishers' motivation to attack users of WAPP-protected sites in the first place.

5 Study Results and Conclusions

We already presented table 1, which described the detection rates and overall resistance rates of the different mechanisms against a classic phishing attack, in the introduction of this paper. We found that by combining all mechanisms the best detection and prevention rates (and hence overall resistance rates) are achieved. In this section we deal with detection and prevention rates and with the different attacks individually, and present the conclusions from the study's results.

5.1 Detection Rates Summary

In this subsection we focus on the detection rates of spoofed pages that have been entered, i.e not including invalid certificate pages which users were transferred to but not entered nor emails with spoofed links which were not clicked. Figure 2 shows a summary of the detection rates for the different defense mechanisms. Results are divided to the two variations of our study – 'weak' stands for students' weak cooperation and sense of risk in the first experiment and 'strong' stands for students' stronger cooperation and sense of risk in the second experiment. 'Combined' stands for a wighted average of the results of all three semesters, where we gave the 'strong' variation a weight of 0.7, since we believe it better suits real-life and had shorter evaluation time. The results show that:

Conclusion 2. *There is a* significant gap *between the detection rates of the different methods ranging from 20% for* `none` *users (in the combined results), up to 42% for* `bookmark only`*, 60% for* `image only`*, 72% for* `bookmark+image` *and 82% for* `bookmark+4 images`*. In the strong variation of our study, this method achieved even better and outstanding detection rates of 87%.*

In particular, `image` users (`image only`, `bookmark+image` and `bookmark+4 images`) detected *more than twice* the percent of attacks detected by `non-image` users (`none` and `bookmark`). Therefore:

Conclusion 3. *The interactive custom image is a highly effective forcing function against automatic submission of credentials.*

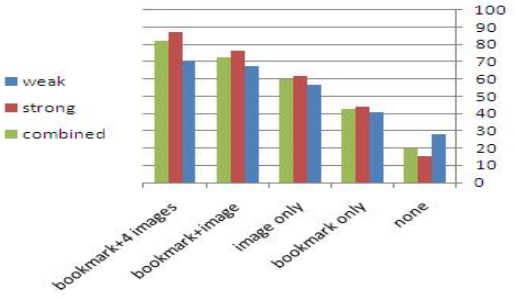

Fig. 2. Detection rates found for the different defense mechanisms

Another observation is that `bookmark only` users received better detection rates than `none` users, and `bookmark+image(s)` users received better detection rates than `bookmark only` users. We can conclude that:

Conclusion 4. *The login bookmark increases the detection rates, and its advantage is not limited to prevention.*

Finally, we noticed that the more effective a mechanism is, larger the gap in detection rates for that mechanism between the two versions of the study. In addition, the strong version of our study shows larger significance in the detection rates between the different mechanisms than in the weak version (see appendix C for further details). This implies that whether users of non-effective detection mechanisms (bookmark only and none) were cooperative did not affect their detection rates. On the contrary, when users of effective detection mechanisms were cooperative, their detection rates significantly increased. Therefore, we conclude that the strong version of our study shows the true potential of the mechanisms and better suits user's real-life behavior for sensitive sites.

5.2 Users' Response to Emails

In this subsection we deal with users' response to emails sent, both spoofed and non-spoofed. Table 2 shows the spoofed links following rates for `bookmark` and `non-bookmark` users, w.r.t. the different email methods. First we can see that there is no significant difference between `no link` and `warning` users, i.e:

Conclusion 5. *Warnings against following links in legitimate emails don't contribute in preventing users from following links in spoofed emails.*

Our results for email warnings correlate with the results achieved by Karlof et al.'s study [11]. Now, let's focus on `non-bookmark` users. The legitimate emails of `non-bookmark` users which normally receive no links (and receive their grades within the email body) look entirely different than the spoofed emails (not containing the grade and asking them to follow a link to watch it). Despite the difference, the percent of links followed by those users is as high as `non-bookmark link` users, which their legitimate and spoofed emails are similar (both contain working links). From this we conclude that:

Table 2. Links following rates for `bookmark` and `non-bookmark` users

method	email method	followed	sent	following rate
`non-bookmark`	`no link`	50	69	72.46 ± 8.84
`non-bookmark`	`warning`	68	100	68 ± 7.67
`non-bookmark`	`'working' link`	95	139	68.34 ± 6.49
`bookmark`	`no link`	53	81	65.43 ± 8.69
`bookmark`	`warning`	74	97	76.29 ± 7.1
`bookmark`	`'non-working' link`	36	116	31.03 ± 7.06

Conclusion 6. *Users don't distinguish between spoofed and non-spoofed emails, even if the email's structure is non-familiar.*

This can be explained by the fact that clicking a link is a click whirr response and the natural thing to do, and users follow email links regularly. Users might also put too much trust in the emails' 'From' header, which can be spoofed easily, and/or think that the system has changed its email announcements' structure.

Finally, and most important, the results show that `bookmark` users that normally receive no links follow a similar percent of spoofed links as `non-bookmark` users. Only `bookmark` users that normally receive ('non-working') links follow *less than half* the percent of spoofed links followed by `non-bookmark` users. We found similar significant difference for non-spoofed emails. Therefore:

Conclusion 7. *The login bookmark is only effective against automatic following of links when users receive "non working" links in genuine emails and experience failure in reaching the login page by following a link.*

From this conclusion we derive that:

Conclusion 8. *Sites should intentionally include "non working" links in their email announcements to (constantly) train their users not to follow email links*[2]*.*

5.3 Spoofed Home Page Attacks Summary

In this subsection we focus on the spoofed home page attack (which lead the users to a spoofed login page). For all methods, we noticed that:

Conclusion 9. *Detection rates are lower when users enter spoofed login pages from the system's home page, in comparison to reaching them via email links or bookmark clicks*[3]*.*

A possible explanation is that:

[2] Even if the site does not send email announcements at all, it is advised to send announcements from time to time just for the training's sake.

[3] 17% vs. 26% for `none` users, 18% vs. 46% for `bookmark only`, 58% vs. 63% for `image only`, 57% vs. 74% for `image+bookmark` and 74% vs. 83% for `bookmark+4 images`.

Table 3. Entrance rates (amount entered from amount transferred) and detection rates (amount noticed from amount entered) for invalid certificate spoofed pages

method	entered	transferred	entrance rate	noticed	detection rate
none	14	31	45.16 ± 14.71	1	7.14 ± 11.32
bookmark only	7	46	15.22 ± 8.71	4	57.14 ± 30.77
image only	7	23	30.43 ± 15.78	5	71.43 ± 28.08
bookmark+image	4	33	12.12 ± 9.34	4	100
bookmark+4 images	15	34	44.12 ± 14	13	86.67 ± 14.44
total	47	167	28.14 ± 5.72	27	57.45 ± 11.86

Conjecture 1. Users put high trust in the home page of a familiar-looking site, even if it does not provide an SSL certificate.

With reduced detection rates in this attack, prevention gets higher importance. The vast majority of all `bookmark` users tried to enter the site's login page via its home page, despite their bookmark. To prevent users from doing so, the site can choose not to include an account-entrance button in its home page at all, or to include a 'non-working' button which leads the user to an error page.

We only tested the second option, and a closer look at the attacks log showed that in spite the rather high attack probability, only two `bookmark` users were attacked more than once. Therefore:

Conclusion 10. *Combining the login bookmark with a "non working" account-entrance button in the site's home page achieves effective prevention.*

Both options experience the user with failure when wanting to enter the login page via the home page, and train the user not to enter the home page in the first place when wanting to login. Yet, when a user is triggered to enter a spoofed page looking similar to her target site's home page (for example by following a link), which includes an account-entrance button, she might be tempted to click it, as she did not experience failure in this specific action. Since the vast majority of `bookmark` users did click the account-entrance button on our study, it is important to experience failure with the button click itself. Therefore:

Conclusion 11. *Sites should intentionally include a "non working" account-entrance button/link in their home page.*

5.4 Effectiveness of Active Browser Warnings

In this subsection we examine how well modern browsers' active warnings prevent users from entering spoofed pages with invalid certificates, and for the users that did enter those pages, what are the detection rates. Table 3 shows the percent of users from each method entering the spoofed site, and the detection rates for those who entered.

From the table it seems that the active browser warnings prevented 72% of the users from entering the spoofed page, which correlate with the results achieved by Egelman et al.'s study [7]. Therefore:

Conclusion 12. *Active browser warnings are an effective forcing function against spoofed sites with invalid certificates (and in particular pharming attacks).*

It Is also noticeable that `image` users have better detection rates when entered the invalid certificate page. In addition, as seen before, `bookmark+image` and `bookmark+4 images` users had better detection rates than `image only` users. Therefore, we assume the following generalization:

Conjecture 2. Combining a preventing forcing function with a detecting forcing function increases the detection rates.

5.5 False Negatives

We've seen that throughout the study only one eighth of all users have falsely reported a spoofed page, mostly once or twice within a few minutes. Since a reasonable amount of false negatives is accepted and better than false positives (better safe than sorry), we conclude that:

Conclusion 13. *All the tested mechanisms does not confuse users to falsely classify the legitimate site as spoofed.*

6 Usability Survey

Since users are forced to set the bookmark (or an authentication cookie as alternative) on each computer or browser they use, this process should be secure and usable. If the bookmark setup involves manually surfing to the website and providing some identifying information, users are susceptible to phishing attacks each time they set the bookmark. To avoid this, bookmark setup has to be *via a secondary channel*. Many of today's websites send a registration email containing a verification link which the user has to click on to complete her registration. This process can also be used to create the bookmark, and we used it on our study.

Keeping the registration email around enables a bookmark setup on other computers. If the user loses the registration email, she can request it to be sent again (to the same email address). A second email address and an SMS can also be used for fallback bookmark recovery. On emergencies where the user has no option to recover her bookmark link, social authentication techniques [13], which were originally suggested for password recovery, can be used for bookmark recovery.

Though bookmark setup is not as immediate as the login ceremony, we believe most users usually access high-value sites only from few computers. We have tested this assumption and the overall usability of the mechanisms in our

usability survey, which included the results of 136 `bookmark` and/or `image(s)` users.

The results confirmed our assumption, as users enter high-value sites from only 1.75 computers in average; medium-values sites such as social networks or webmail are entered from 3.26 computers on average.

In the survey's introduction we mentioned the high prevention and detection rates achieved by the login bookmarks and interactive custom images and asked users to say if the would want to login with those mechanisms[4] to medium-value and high-value sites. The results showed that most users (72% ± 7.42) want to use a login bookmark to access high-value sites, and half of them (51% ± 8.31) also want to use it on medium-value sites. In addition, only 36% of the users who did not want to use the bookmark picked the bookmark setup as the reason why they wouln't want to use this method. We can conclude that:

Conclusion 14. *When considering their security benefits, users are willing to use login bookmarks and other mechanisms that require registration of the computer via a secondary channel.*

We also found that 64% of the users are willing to use interactive custom images on high-value sites and 41% of them also want to use them for medium-value sites. The results are encouraging, and we expect them to be much higher due to two main reasons that biased users' answers in the survey and are described in appenidx D.

7 Conclusions

We have conducted a realistic long-term user study which tested the effectiveness of different defense mechanisms that use forcing and negative training functions, against different phishing attacks. The study's results showed that forcing and negative training functions are very effective in both prevention and detection due to their constant trainings. A combined method, which uses a login bookmark with interactive custom images, displays several images in the login page, and intentionally includes "non working" links in the site's email announcements and a "non working" account-entrance button in its home page, received outstanding prevention and detection rates.

Acknowledgments. The authors would like to thank Ben Adida for his feedback and helpful suggestions and Alex Dvorkin for the initial results of this study. This work was supported by Israeli Science Foundation grant ISF1014/07.

References

[1] Aaron, G., Rasmussen, R.: Global Phishing Survey: Trends and Domain Name Use in 2H2009. Anti-Phishing Working Group (May 2010), `http://www.antiphishing.org/reports/APWG_GlobalPhishingSurvey_2H2009.pdf`

[4] Bookmark only and `image only` users were only asked a yes/no question about their method and `bookmark+image(s)` users were given four options – bookmark only, image only, both or none.

[2] Adida, B.: Beamauth: two-factor web authentication with a bookmark. In: CCS 2007: Proceedings of the 14th ACM Conference on Computer and Communications Security, pp. 48–57. ACM, New York (2007)
[3] Sitekey Bank of America, http://www.bankofamerica.com/privacy/index.cfm?template=sitekey
[4] Cialdini, R.: Influence: Science and Practice, 5th edn. Allyn and Bacon, Boston (2008)
[5] Dhamija, R., Tygar, J.D.: Why phishing works. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 581–590. ACM Press, New York (2006)
[6] Dvorkin, A.: Evaluation of the Tools for User Protection against Web Site and Electronic Mail Based Attacks. Master's thesis, Bar-Ilan University (December 2008)
[7] Egelman, S., Cranor, L.F., Hong, J.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: CHI 2008: Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 1065–1074. ACM, New York (2008)
[8] Herzberg, A.: Why Johnny can't surf (safely)? Attacks and defenses for web users. Computers & Security (2008)
[9] Herzberg, A., Jbara, A.: Security and identification indicators for browsers against spoofing and phishing attacks. ACM Trans. Internet Techn. 8(4) (2008)
[10] Herzberg, A., Margulies, R.: Long-term user study of forcing and training login mechanisms against phishing. Tech. rep., Bar Ilan University (March 2011), http://submit2.cs.biu.ac.il/WAPP/WAPP_primary.pdf
[11] Karlof, C., Tygar, J.D., Wagner, D.: Conditioned-safe ceremonies and a user study of an application to web authentication. In: SOUPS 2009: Proceedings of the 5th Symposium on Usable Privacy and Security (2009)
[12] Schechter, S., Dhamija, R., Ozment, A., Fischer, I.: The emperor's new security indicators. In: SP 2007: Proceedings of the 2007 IEEE Symposium on Security and Privacy, pp. 51–65. IEEE Computer Society, Washington, DC, USA (2007)
[13] Schechter, S., Egelman, S., Reeder, R.W.: It's not what you know, but who you know: a social approach to last-resort authentication. In: CHI 2009: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 1983–1992. ACM, New York (2009)
[14] Sotirakopoulos, A., Hawkey, K., Beznosov, K.: "i did it because i trusted you": Challenges with the study environment biasing participant behaviours. In: SOUPS User Workshop (2010)
[15] Wu, M., Miller, R.C., Garfinkel, S.L.: Do security toolbars actually prevent phishing attacks? In: CHI 2006: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 601–610. ACM, New York (2006)
[16] Better website identification and extended validation certificates in ie7 and other browsers (November 2005), published in Microsoft Developer Network's IEBlog http://blogs.msdn.com/b/ie/archive/2005/11/21/495507.aspx
[17] Yahoo: What is a sign-in seal?, http://security.yahoo.com/article.html?aid=2006102507
[18] Yee, K.-P., Sitaker, K.: Passpet: convenient password management and phishing protection. In: SOUPS 2006: Proceedings of the Second Symposium on Usable Privacy and Security, pp. 32–43. ACM, New York (2006)
[19] Gartner survey shows phishing attacks escalated in 2007 more than $3 billion lost to these attacks (2007), http://www.gartner.com/it/page.jsp?id=565125

[20] Gartner says number of phishing attacks on u.s. consumers increased 40 percent in 2008 (2008), http://www.gartner.com/it/page.jsp?id=565125
[21] Mcafee siteadvisor (2009), http://www.siteadvisor.com/

A Additional Attack Scenarios

1. ***Tricking the User to Give Away the Secret Token.*** An attacker could send a spoofed email on behalf of the target site's admin announcing some sort of technical problem and asking the user to copy&send the bookmark URL to the attacker, or to forward the registration email/SMS to the attacker. Even if this attack works, and the attacker finds the user's custom image, a second stage of the attack is still necessary. The attacker has to succeed in a phishing attack convincing the user to follow a link in another spoofed email. For this phase, the custom image won't help, but the constant trainings of the bookmark and non-working links should assist in preventing the user from following the spoofed link.

   Karlof et al. [11] included a similar attack in their long-term, real-life study. Each time a user had to register a new computer (to receive a cookie), the website sent her an email with a URL containing a single-use secret token. After the link click, the user reached a page which set a persistent authentication cookie, and deleted the token from its data base, to prevent further usage. Karlof et al. executed two kinds of attacks – asking the users to copy&paste the link to a spoofed site or forward the registration email to the attacker. In both attacks users got a message announcing "technical problems". The researchers stated that email-based registration is a conditioned-safe ceremony, which makes users click on the link they see rather than making one of the actions they have been asked to do by the attacker, which are of course more complicated to perform than a link click. Results have shown ∼40% success for those attacks.
2. ***Attacking the Email Account.*** If a registration email is used and the attacker breaks into a user's email account, the WAPP token is compromised, and the attacker will gain the user's custom image. Like the previous attack, the attacker needs to perform a phishing attack hoping the user will make an error of omission and follow a spoofed link sent to her.
3. ***Temporarily Using the User's Computer.*** An attacker might temporarily gain access to the user's computer and immediately gain the user's secret token and custom image. Fortunately, the attacker will not be able to immediately login as the user, as he still needs her password. This reduces to the last two attacks, requiring the attacker to phish the user hoping she will make an error of omission and follow a spoofed link.
4. ***Malware on the user's computer.*** An attacker who injects untrusted code into the user's computer can completely control the system, read the browser's bookmark content, and keylog the user's password. WAPP is completely vulnerable to this kind of attack. On the other hand, WAPP can assist in protection against malware, since attackers might phish downloads

sites in purpose of convincing users to download and install malware. Even though downloads sites usually don't maintain users' information and does not require a login, they could also integrate WAPP to prevent phishing attacks and allow only users who clicked the bookmark and their custom image to reach the downloads page.

B Removing Outliers

Prior to analyzing the results, we had to remove the results of users that did not cooperate with our experiment. Our methodology in finding outliers was to find the amount of attacks and amount of detections for each user. Then, for each method, we looked at the average amount of attacks and detections for all users from that method. By filtering users that had fewer detections than the average of their method and more attacks, than non-cooperating users were easy to find. For instance, if `bookmark+4 images` users had 4.48 detections of 6.77 attacks in average, than we filtered users with 4 hits or less which were attacked 7 times or more. For all methods, our filtering technique had left only users that were *noticeably non-cooperating*, for example having 0 detected attacks of 7 attacks or 1 of 23 and so on, without any 4 of 7 or other values which are just a bit lower than the average.

With the mentioned filtering technique we found 108/411 (26.27%) non-cooperative users in the first two semesters, and 75/402 (18.65%) in the third semester, 39 of whom (52%) were also non-cooperative in the previous semesters. In addition, several more non-cooperative users were hand-picked and removed. Theses included also users with many hits from many attacks. For instance, one user had 47 hits of 47 attacks; by closely looking at his attack-log entries, we saw that he have reached a spoofed page and did not enter his password. The same page was visited over and over, with exactly one hour between each visit, suggesting that the user had left his browser open and the browser is being refreshed constantly every hour.

To support our assumptions that the filtered users were indeed non cooperative, we have sent them an email asking them if they did not try to find spoofed pages (telling them that it is OK if they did not). 68% non-cooperating users of all three semesters had confirmed by email, none rejected, and the others did not respond.

C Different Versions of Our Study – Conclusions

Let's examine the difference between the two versions of our study: weak vs. strong cooperation and sense of risk. Assuming that users of all methods felt similar will to cooperate and similar sense of risk in each version of the study, one would expect to find a similar increase in detection rates for each method between the two versions, and similar gaps in detection rates between the methods in each version (for instance, a 5% increase in detection rates for all methods, keeping the gaps between different methods the same for the weak and strong versions).

Table 4. Increase in detection rates between the two version of the study

method	weak	strong	increase
`none`	27.88	15	-46.21
`bookmark only`	40.5	43.65	7.8
`image only`	56.54	61.9	9.48
`bookmark+image`	67.6	76.07	12.53
`bookmark+4 images`	70.7	87.07	23.16

Yet, the results turned out to be different than expected; it is noticed (see table 4) that the more effective a method is, larger the gap in detection rates for that method between the two versions of the study and in the strong version of our study there was a larger gap in the detection rates between the different methods. In particular, there was only a minor increase in detection rates for `bookmark only` users and a noticeable *decrease* for `none` users, suggesting that whether these users cooperated and felt a sense of risk or not *had no affect on the results*. On the other hand, for `bookmark+4 images` users the increase is the largest (23%), suggesting that lack of cooperation and sense of risk highly affected their results. From this we conclude that lack of cooperation and sense of risk inserts noise to the statistical data, and the true potential of each method's effectiveness is evidenced in the strong version of the study.

D Interactive Images Usability

There were two reasons that biased the survey answers of `image` users. First, we performed an additional experiment before the survey, where users were asked to remember few custom images instead of just one, and one of the custom images was randomly chosen and displayed on the login page on each login, along three non-custom images. This caused approximately half of the users that did not wish to use the interactive custom image to complain about the method, due to mistaken clicks which increased the authentication times. Most mentioned that a single custom image would be better. Due to the bias caused by the multiple-images experiment, we believe many of the users would want to use a single custom image.

Second, most of the other users that did not wish to use the interactive custom image stated that they don't need its protection, or because they did not understand its purpose (many thought of it as a user-to-site identification instead of the opposite). Those users did not say it was a bother to use the image, so we believe most of them won't be unhappy if it was applied on a site they use.