

Phishing IQ Tests Measure Fear, not Ability

Vivek Anandpara Andrew Dingman
Markus Jakobsson Debin Liu Heather Roinestad

January 15, 2007

Abstract

We argue that phishing IQ tests fail to measure susceptibility to phishing attacks. We conducted a study where 40 subjects were asked to answer a selection of questions from existing phishing IQ tests in which we varied the portion (from 25% to 100%) of the questions that corresponded to phishing emails. We did not find any correlation between the *actual* number of phishing emails and the number of emails that the subjects indicated were phishing. Therefore, the tests did not measure the ability of the subjects. To further confirm this, we exposed all the subjects to existing phishing education after they had taken the test, after which each subject was asked to take a second phishing test, with the same design as the first one, but with different questions. The number of stimuli that were indicated as being phishing in the second test was, again, independent of the *actual* number of phishing stimuli in the test. However, a substantially larger portion of stimuli was indicated as being phishing in the second test, suggesting that the only measurable effect of the phishing education (from the point of view of the phishing IQ test) was an increased concern—not an increased ability.

Keywords: phishing, phishing education, phishing IQ test

1 Introduction

Popular media routinely covers the mounting problem of phishing. Financial institutions frequently alert clients of the risks of identity theft, and many provide detailed descriptions of common attacks and how to avoid falling victim to these. With this popular focus on the problem, we must ask ourselves why the recent trends show an increase in the number of people that fall victim to phishing. Furthermore, we must pose the question whether current educational efforts are meaningful and whether the ways in which vulnerabilities are assessed work.

To be able to ask these questions, it is important first to understand why phishing works. This question has been asked by several researchers recently [5, 6, 7, 21], and a collection of insightful conclusions have been found. One reason that phishing works is that most people do not have a detailed understanding

of all the guises a given threat might take, but only react to situations that he or she has already identified as being dangerous. Another reason is that many users do not possess technical sophistication sufficient to verify whether a given email or webpage corresponds to an attempt to defraud them. The most important reason of all, though, might be that to most people, security is a *secondary* goal. In other words, the average person may very well ignore signs of risk, since he or she is not actively looking for these.

There are publicly available ‘phishing IQ tests’ published to help individuals assess their likely vulnerability to phishing scams. Examples of these can be found at Mailfrontier/Sonicwall [14, 17]; Mailfrontier also has UK and German versions [16, 15]. These tests typically take the form of a sequence of e-mail screen shots depicting messages of the sort phishers tend to emulate. Users identify the depicted message as either a legitimate message or a phishing scam, and receive a score based on the percentage of correct answers.

We argue that such phishing IQ tests are flawed on several levels. Due to their delivery format, a static file, many actual security indicators are not available. By their nature these tests also lack context present in real attacks and which can aid in making accurate decisions. In addition, and more importantly to our study, while the natural context is lacking, an artificial context may skew the test taker’s judgment. This may create a false sense of security among test takers who receive a high score on the tests. As we will show, *obtaining a high score is not an indication of ability to recognize phishing attempts*.

Because it is in the context of a test regarding the taker’s ability to correctly identify phishing, test takers may be suspicious beyond the level they normally would upon seeing the original email in their inbox. This could simultaneously mean that they correctly label some examples as phishing that they might normally be susceptible to, and/or would incorrectly identify legitimate emails as phishing. This is a well understood fact, but does not in any way affect our methodology. Quite to the contrary, we are to some extent able to quantify the effects of this type of bias; this is done by comparing the average ratings given by subjects before and after the educational step of the experiment.

We are also able to show that traditional forms of education (e.g. [8]) increase the level of fear or concern among users, but do not help them become any better at passing the phishing IQ tests.

2 Previous Work

Much of the existing literature related to phishing deals with people’s perception of website credibility and not with how people judge the emails which lure them to the fake websites. There has been a substantial amount of work on the matter of what makes a website credible or phishy to the people.

Fogg et al [6, 7] conducted a study with over 2500 subjects and investigated how different elements of websites affect people’s perception of credibility and laid down guidelines for the credible perception of websites. On similar lines, Dhamija et al [5] worked towards establishing what makes phishing websites

credible and why phishing works.

There have been user studies which focused specifically on phishing. Jagatic et al. [10] worked on improving the throughput of phishing attacks by the use of the social network of the victim to increase the credibility of the phishing emails. Publicly available databases in the form of social networks were mined to find out a victim's friend map. Their study showed with empirical data that 72% of people responded to a mail spoofed to be coming from a friend's email address. as opposed to 16% to mails coming from an unknown address. In another user study which dealt with the effectiveness of anti-phishing measures such as phishing toolbars, Wu et al [21]. examined the impact of anti-phishing toolbars in preventing phishing attacks. Their results show that even when toolbars were used to notify users of security concerns, users were tricked into providing information 34% of the time.

Several other studies have recently shed light on the problem of phishing [1, 5, 10, 11, 12, 19, 22], and several have proposed countermeasures [2, 4, 9, 12, 13, 18, 19, 20, 21]. Moreover, researchers have become increasingly interested in the role of malware in the context of phishing [3].

Many of these papers fail to recognize that although it is not commonly happening today, the various indicators of security which are emphasized can be spoofed by phishers (see, e.g., [5]). Thus, sophisticated phishing attacks may be difficult to detect, even to people who are reasonably aware of what to look for. It is also worth mentioning that the multitude of studies which perform general evaluations of phishing vulnerabilities largely neglect the subject-expectancy effect. The subject-expectancy effect is a cognitive bias that occurs when a subject expects a given result and therefore unconsciously manipulates an experiment or reports the expected result, partly to avoid embarrassment. This effect is applicable to phishing IQ tests (e.g., [14]), where the cognizance of being involved in a phishing IQ test makes the subjects unusually suspicious and this can substantially skew the results of the test.

There have been no previous attempts to gather any empirical data on the effectiveness of these phishing IQ tests.

3 Experiment

We claim that phishing IQ tests do not work; phishing education makes one more paranoid and that phishing IQ tests measure this fear and not the ability to recognize phish.

In order to substantiate our hypotheses statistically, we asked subjects to take a phishing IQ test before and after being educated on phishing. Our aim was to calculate the correlation between the number of phishing examples identified and actually present. We also wanted to test the effect of phishing education on subjects in terms of their performance on the phishing IQ tests. Thus we divided the test into a pre and post test. A comparison of the results of the two tests can provide an idea of the effect of reading the educational material on the subject.

Subjects completed an initial five-question ‘phishing IQ test’ consisting of questions taken from a publicly available instance of such a test, the SonicWall IQ Test [17]. Based on random assignment, the initial set of five examples included either 1, 2, 3, or 4 phishing messages. The subject’s responses were recorded, but scores were withheld until completion of the full experiment protocol. After completing the first set of questions, subjects were given a short, publicly available text intended to educate the reader in identifying and avoiding phishing scams.

This was followed by a second five-question test. This second set contained $5 - p_1$ phishing messages, where p_1 is the number of phishing messages presented to the same subject in the first round. (E.g., if a subject’s first test had 2 phishing e-mails, the second test had 3.) Once again, the subject’s responses were recorded, but scores were withheld until completion of the full experiment protocol.

Phishing IQ Tests We took our examples from an existing phishing IQ test, as we wish to measure the effectiveness of phishing IQ tests in their current state. Our full test includes five each of phishing and legitimate emails. We then showed each subject half of these examples, selected to match the proportion corresponding to the user’s assigned group.

We then compared the number of actual phishing examples to the number they labeled as suspicious. A correlation between the number of examples labeled as phishing and the number that actually were examples of phishing would be expected, if the tests really tested the vulnerability of subjects.

If, however, the context of the test determined how subjects answer, there would be no relation between the number of phishing examples and the number of examples labeled as phishing. When conscious of being tested on their ability to identify suspicious emails, subjects may be more likely to default with the ‘bad’ label, thereby potentially overestimating the number of potential attacks they see.

4 Data and Results

We did a test on 40 subjects for this data analysis. As our study asks about how an average user performs at the task, we excluded subjects with an unusual knowledge of computer science or security, and asked that subjects be people who either use or would consider using online shopping, banking, or bill paying.

A first phishing IQ test. For the first part of the experiment, we gave subjects a short sample test with five screenshots in which we varied the number of those screenshots corresponding to actual phishing emails. An example screenshot is shown in Figure 1. Our hypothesis predicted that the portion of screenshots subjects labeled as suspicious would be roughly the same regardless of the group of the subject (i.e., independent of how many screenshots *were* “bad”). Furthermore, we hypothesized that this portion would be high, due

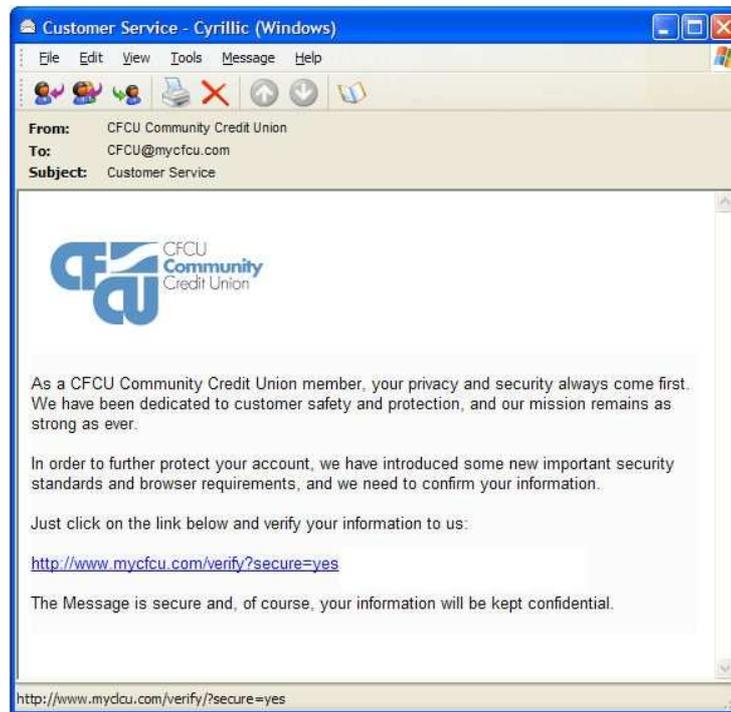


Figure 1: A sample screenshot from the Mailfrontier phishing IQ test. This stimulus corresponds to a phishing attack, as test-takers aware of domain names would recognize – see the link target at [www.myclcu.com](http://www.myclcu.com/verify?secure=yes) in the status bar, which conflicts with the link text [www.mycfcu.com](http://www.mycfcu.com/verify?secure=yes).

to the previous knowledge of the subjects that we were testing their ability to identify suspicious emails.

Phishing education. After the first sample test, we gave the subjects time to read over an example of phishing education intended for the layman. An example is shown in Figure 2. After reading the education material provided, subjects repeated the first step on the remaining five examples they had not yet seen.

A second phishing IQ test, and analysis. After the educational step, a second phishing IQ test was administered. This contained different stimuli than what the subject saw in the first phishing IQ test; however, each given stimulus that was shown to one group in the second phishing test was also shown to another group in the first phishing test. Thus, the selection of stimuli had no impact on our results.



Figure 2: The figure shows the phishing education we used as part of our experiment. It describes bait messages, warns against replying or clicking links, and warns users that telephone area codes can be faked. This was taken from ftc.gov [8]. This text was shown to subjects between the two tests they were given. The effect of the education can be seen in Figure 3; therein it can be seen that subjects became more suspicious of stimuli, but not more competent at distinguishing between “good” and “bad” ones.

Figure 3 shows, among other things, how many of the examples they were shown that subjects labeled as phishing, and the number required to get a perfect score. It can be seen that the experiments support the hypotheses described above. Consistent with our first hypothesis, the number of times subjects labeled an example as phishing appears to have no correlation with the number of actual phishing examples they were shown. The individual responses have a practically zero linear correlation coefficient with the number of actual phishing examples, suggesting that the number of times a subject labels an example as phishing does not depend on the number that actually *are* phishing.

Figure 3 also shows how many of the stimuli subjects labeled as phishing in the post-education test, compared to the actual number shown. This time, we see the overestimation we predicted in our hypothesis; after reading about how to identify phishing, subjects started seeing more instances of phishing than were necessarily there. Here we see the high rate we expected from the subject-expectancy effect; the subjects' likelihood of labeling a stimulus as phishing in the context of the test increased, from generally underestimating the number to generally overestimating.

In the short term at least, the education does appear to affect the subjects' judgment, but more in terms of making them more suspicious than in improving their ability to distinguish phishing from legitimate emails. The number of times

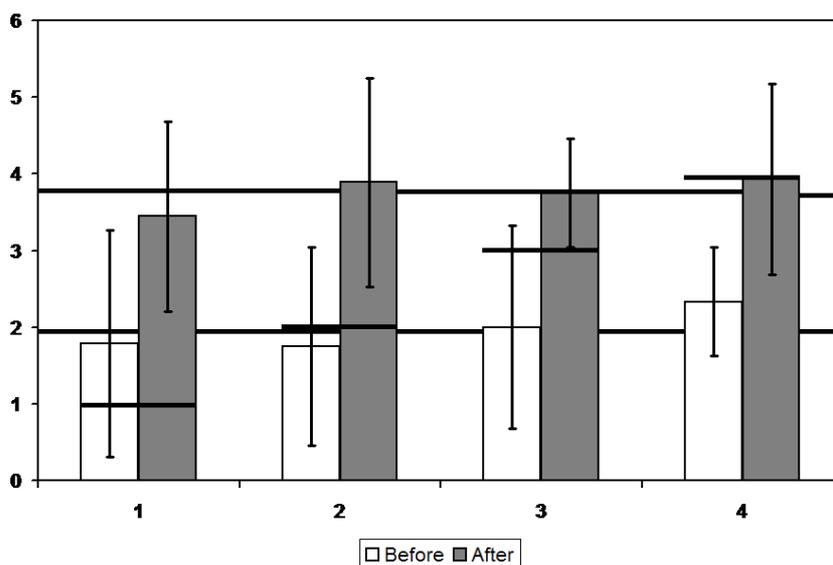


Figure 3: Subject responses before and after education. On the x-axis, the number of stimuli that were *actually* phishy are shown; the diagram reads the number of stimuli that subjects indicated as phishy. The left bars show the experimental results of the first phishing IQ test, whereas the right bars show the results of the second test. The standard deviation is indicated for each measurement. For each pair of bars, we include a horizontal line showing what the results should have been if subjects were truly able to distinguish authentic stimuli from phishing attempts. It is evident that this is not the case. Note also the effect of the education: Subjects were not better at taking the second test than the first; however, they were more suspicious of all stimuli shown during the second test and hence falsely labeled many legitimate stimuli as phish.

that subjects labeled a stimulus as phishing increased from the first to the second test for most subjects, even though two of the four groups were actually shown fewer instances in the second round than in the first; however, the number of instances labeled as phishing still did not correlate with the number which were phishing.

One can also read out the observed mean responses for both before and after the education from Figure 3. Neither set of responses correlates with the number actually shown to the subject, but the two sets of values are different. As mentioned before, the subjects became more likely to label an example as phishing after education than they were before.

References

- [1] Anti-Phishing Working Group: Phishing Activity Trends Report November 2005 (2005)
- [2] Chou, N., Ledesma, R., Teraguchi, Y., Boneh, D., Mitchell, J. C.: Client Side Defense Against Web-based Identity Theft. <http://crypto.stanford.edu/SpoofGuard/#publications>
- [3] The Crimeware Landscape: Malware, Phishing, Identity Theft and Beyond. http://www.antiphishing.org/reports/APWG_CrimewareReport.pdf
- [4] Dhamija, R., Tygar, J. D.: The Battle Against Phishing: Dynamic Security Skins. Proc. SOUPS (2005)
- [5] Dhamija, R., Tygar, J. D., Hearst, Marti: Why Phishing Works. to appear in the Proceedings of the Conference on Human Factors in Computing Systems (CHI2006)(2006)
- [6] Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., Tauber, E. R.: How Do Users Evaluate the Credibility of Web Sites?: A Study with Over 2,500 Participants. Proc. DUX(2003)
- [7] Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M.: What Makes Web Sites Credible?: A Report on a Large Quantitative Study. Proc. CHI (2001) 61-68
- [8] FTC.gov Alert. <http://www.ftc.gov/bcp/edu/pubs/consumer/alerts/alt127.htm>
- [9] Herzberg, A., Gbara, A.: Protecting (even) Naive Web User, or: Preventing Spoofing and Establishing Credentials of Web Sites. DIMACS Technical Report 2004-23 (May 2004)
- [10] Jagatic, T., Johnson, N., Jakobsson, M.: Phishing Attacks Using Social Networks. Indiana University Human Subject Study 05-9892 & 05-9893 (2005)
- [11] Jakobsson, M.: Modeling and Preventing Phishing Attacks. Phishing Panel in financial Cryptography '05 (FC'05) (2005)
- [12] *Phishing and Countermeasures: Understanding the Increasing Problem of Electronic Identity Theft*. M. Jakobsson and S. A. Myers (editors). ISBN 0-471-78245-9, Hardcover, 739 pages, December 2006.
- [13] Loftesness, S.: Responding to "Phishing" Attacks. Glenbrook Partners (2004) <http://www.glenbrook.com/opinions/phishing.htm>
- [14] MailFrontier Phishing IQ Test II. http://survey.mailfrontier.com/forms/msft_iq_test.html

- [15] MailFrontier Phishing IQ Test – Deutsche Edition.
<http://german.mailfrontier.com/survey/phishing-de.jsp>
- [16] MailFrontier Phishing IQ Test – UK Edition.
http://survey.mailfrontier.com/survey/phishing_uk.html
- [17] MailFrontier/Sonicwall Phishing. <http://www.sonicwall.com/phishing/>
- [18] PassMark Security: Protecting Your Customers from Phishing Attacks - An Introduction to PassMarks. <http://www.passmarksecurity.com/>
- [19] The Phishing Guide - Understanding & Preventing Phishing Attacks.
<http://www.ngssoftware.com/papers/NISR-WP-Phishing.pdf>
- [20] RSA Security: Protecting Against Phishing by Implementing Strong Two-Factor Authentication. (2004)
https://www.rsasecurity.com/products/securid/whitepapers/PHISH_WP_0904.pdf
- [21] Wu, M., Miller, R., Garfinkel, S.: Do Security Toolbars Actually Prevent Phishing Attacks?. Proc. CHI (2006)
- [22] M. Jakobsson: The Human Factor in Phishing
Privacy & Security of Consumer Information,2007.