

# What Instills Trust? A Qualitative Study of Phishing

Markus Jakobsson<sup>1</sup>, Alex Tsow<sup>2</sup>, Ankur Shah<sup>2</sup>, Eli Blevis<sup>2</sup>, and Youn-Kyung Lim<sup>2</sup>

<sup>1</sup> Indiana University, Bloomington and RavenWhite Inc.  
email: markus@indiana.edu, markus@ravenwhite.com

<sup>2</sup> Indiana University, Bloomington  
email: {shahak,atsow,eblevis,younlim}@indiana.edu

**Abstract.** This paper reports the highlights of a user study which gauges reactions to a variety of common “trust indicators” – such as logos, third party endorsements, and padlock icons – over a selection of authentic and phishing stimuli. In the course of the *think-aloud* protocol, participants revealed different sensitivities to email messages and web pages. Our principal result is the analysis of what makes phishing emails and web pages appear authentic. This is not only of interest from a pure scientific point of view, but can also guide the design of legitimate material to avoid unnecessary risks. A second result of ours are observations of what makes legitimate content appear dubious to consumers. This is a result with obvious applications to online advertising.

**Keywords:** authenticity, design, email, experiment, phishing, psychology, stimuli, think-aloud, user interface design, web pages.

## 1 Introduction

Over the last few years, the problem of phishing has grown at an alarming rate. As service providers are becoming increasingly aware of the threat, more and more effort is spent on countermeasures—whether technical, educational or legal. In order to keep pace with such countermeasures, and with increasing competition between groups, phishers are also becoming more sophisticated. There are both social and technical examples of this trend—better spelling, use of subdomains and cousin domains to deceive users, and improved psychological design of the request all fall within the former class. The use of DNS modifications to avoid takedown, and of keyboard loggers to capture information belongs to the latter. With research in social aspects of phishing—the deceit component—lagging behind efforts dealing with technical aspects, the next wave in phishing may very well rely on an attack that has a better deceptive component. In order to understand what direction such an attack may take, we study the role of authenticity in phishing.

We performed an experiment to determine what aspects of email and web pages effectively convey authenticity to visitors. Users are presented with a collection of live, but locally hosted, web pages and email messages. Subjects followed a think-aloud protocol, verbalizing sources of doubt, concern, confidence, and confusion. Participants interacted with web pages using actual client software, however all hyperlinks were link redirected to a page asking them to describe how they expected the result to influence their evaluation.

While our experiment does *not* answer the question of the real impact of these indicators of trust, our experiment *does* indicate what typical computer users are able to detect when they are carefully watching for signs of phishing. This *security first* approach approximates a lower bound on vulnerability to phishing attacks.

The deceit component of phishing does not take advantage of technical vulnerabilities, nor can it always be properly addressed by technical countermeasures. As an example of this, it has been shown by Jakobsson and Ratkiewicz [6] that the use of IP addresses as the domains of phishing web pages is considered legitimate by a smaller portion of users than the use of relevant subdomains in conjunction with non-descriptive or related domains.

Dhamija, Tygar and Hearst [2] studied how computer users fall victim to phishing attacks based on a lack of understanding of how computer systems work; due to lack of attention; and because of visual deception practiced by the phishers. They also adopt the security first assumption, asking participants to evaluate a collection of web pages.

Downs, Holbrook, and Cranor [3] have performed an assessment of risk familiarity and phishing vulnerability in a lab setting. Their participants are given fictional identities to role play a suite of web and email interactions – a method focusing on task completion rather than security. They find that subjects are good at protecting themselves against known scams, but have difficulty adapting to new tactics.

In a study by Whalen and Inkpen [7], the authors used eye-tracking and surveys to determine that people see the lock icon but rarely interact with it. This indicates a vulnerability to forged lock icons, and potentially other trust indicators as well. Friedman et al. [4] studied user perceptions of secure connections and determined that the presence of a lock or key icon was the primary perceived indicator of “correct” evaluations of security, distantly followed by the presence of an “https” designation in the URL, and the type of information requested.

## 2 Experimental Design

**Subject selection.** Subjects were chosen both among college students and university staff and faculty. A total of 17 subjects participated in the study. Ages ranged from 18 to approximately 60. Computer science students and staff/faculty with computer science backgrounds were excluded, as was anybody who had taken one or more computer security classes.

**Stimulus design.** Subjects were shown both email stimuli and web page stimuli, some of which modeled phishing attempts while others represented authentic media. Most of the stimuli were based on actual emails or web pages, whether authentic or not; others were designed by the researchers. Minor modifications were made to email stimuli: All emails were modified (or designed) to have the same apparent recipient.

A number of features were tested: legitimate endorsement logos (Verisign, BBB, and TrustE), made-up endorsements, cousin domain names, naked IP addresses, padlocks in various locations (favicon, content body, browser frame), spelling and grammatical irregularities, `https` and `http` hyperlinks, and personalization (salutations and account data).

**Procedural overview.** Participants sat in front the computer and display. The subjects were asked to rate 26 stimuli each stimulus in terms of their perceived phishiness/authenticity. The rating was done using a 5-point Likert scale [8]: *certainly phishing*, *probably phishing*, *no opinion*, *probably not phishing*, and *certainly not phishing*. Subjects were allowed to scroll stimuli windows up and down and perform mouse-over of hyperlinks, but *not* allowed to click on any hyperlinks. We required classification of a given stimulus before presenting the next one. As subjects observed and judged the stimuli, they were asked to verbalize their thoughts – whether relevant to the decision or not. Voice and screen actions were recorded using *Camtasia Studio* [1]. After completing the stimuli evaluation, participants discussed three questions in an exit interview: *Do you think you have been fooled by any of these stimuli in your daily computer usage?*, *What stimulus features inspired confidence in authenticity?*, and *What stimulus features generated suspicion in authenticity?*

### 3 Summary of Findings

The think-aloud protocol recorded several insightful comments:

- “If the URL looks hinky, I’m not going to trust it.”
- “Well I hate to see a statement says ‘this message is all authentic.’”
- “When I see Verisign, I would probably go with it and say that it’s certainly not phishing.”
- “There’s no copyright thing which is usually there on other banks.”
- “Probably any of these emails I got I would have just deleted ... I wouldn’t read anything that looks not important to me.”

The full version of this paper showing all the stimuli is available at:  
<http://www.indiana.edu/~phishing/papers/JakEtal.pdf>

When reading the conclusions, it is important to realize that these were made in a security first context; therefore, they describe the *abilities* of the subjects rather than the *habits* of the subjects.

The conclusions from our analysis are:

1. **Document layout matters.** Security awareness was generally well received when confined to a specific portion of a web page. The Chase *Security Center Highlights* and USBank *Online Security* area both use high resolution padlocks to draw visitor attention; these two frequently evoked positive reactions from subjects. Some legitimate providers (such as Keybank) were given a low rating due to “unprofessional design.” In the case of Keybank, subjects cited the absence of the institutional name on the login-page, along with the fact that the fields for user name and password were of different length. The presence of copyright information and legal disclaimers, typically at the bottom of the stimulus in small print, enhanced trust for many subjects.
2. **Too much emphasis on security can backfire.** Some stimuli, in particular the (legitimate) IUCU (Indiana University Credit Union) website with its blinking phishing banner, were criticized for their overwrought concerns about online security. Subjects did not like that the IUCU website said “phishing attack in progress” in three different locations. Some commented that “phishing” is too obscure a term for a financial institution to use in their communications – the phrase “identity theft” was offered as a plausible substitute. Explicit assertions of safety, such as “This message is authentic,” and “Phishing Protected,” sounded implausible to many subjects.
3. **People look at URLs.** It was found that subjects looked carefully at URLs of web pages, and on the URLs obtained by mouse-over in emails. Subjects were good at detecting IP addresses as being illegitimate, but were not highly suspicious of URLs that were well-formed, such as `www.chase-alerts.com`. On the other hand, subjects were good at detecting syntactically peculiar addresses, such as `www-chase.com`.
4. **Third party endorsements depend on brand recognition.** The stimuli deployed a range of third party endorsements, from well established brands like Verisign to made-up endorsements like *Safe Site*. We found that endorsements from Verisign were taken with the most gravity. Almost every subject mentioned Verisign by name as a positive factor in their trust evaluation. BBBOnLine and TRUST-e endorsements had no significant effect. On the other hand, the three made-up endorsements evoked consistent criticism. Some subjects noticed third party endorsements on stimuli they clearly believed to be phishing, and deduced that the graphics could be rendered on any page. One subject observed “Probably now that I see all these [stimuli], I should not believe in Verisign,” but later dismissed a web page because “it’s not Verisign protected, but it says something which I’ve never seen, ‘TRUST-e’. I don’t know, so probably I wouldn’t go in this account.”
5. **People judge relevance before authenticity.** Subjects often decided whether a stimulus was legitimate or not based on the content, as opposed to the signs of authenticity. In particular, any stimuli that offered a monetary reward was considered phishy, independently of whether it was authentic or not. Likewise, emails that requested passwords up-front were considered phishy, whereas emails that only appeared to contain information were considered safe. This is a problem, as users could be drawn to a site by an

email that appears to be for information only, and once at the site asked for credentials.

6. **Faux-personalization creates trust.** Personalization, even with well-known data, increases the trustworthiness of stimuli, whether email or web pages. One subject said that presentation of ZIP code and mother’s maiden name would enhance trust in an email message. Yet, this data could be gathered by an attacker using IP to zip code mapping software and publicly available databases [5]. Some expressed comfort with an email stating it was intended for a user whose account starts with 4546; account number prefixes are easily guessed because they come from a small pool of possibilities when the institution is known. While many subjects insisted that the last four digits are a more trustworthy marker, they did not explicitly penalize the message for using the first four digits.
7. **Emails are very phishy, web pages a bit, phone calls are not.** Overall, email stimuli were considered more phishy than web stimuli to participants in the study. Many subjects said that following links from email was a risky activity, and consciously avoid the practice. Since very few admit to following links given in phishy emails, their exposure to phishy web pages is inherently more limited. Many participants said that they would try to independently verify email contents by calling the institution directly. Few participants specified how they would obtain the correct phone number and therefore could expose themselves to fraudulent customer service numbers; most systems prompt users to dial in their account number and zip code prior to speaking with a representative. Several participants also said that email is an inappropriate alert medium for urgent matters, such as password changes and account lock-outs, and expected a phone call from the institution.
8. **Padlock icons have limited direct effects.** Large padlock graphics were effective at drawing attention to specific portions of the stimulus. By themselves, they did not cause any subject to express an improvement in trust. Small padlock icons in the content body were never commented on by subjects. Their ineffectiveness was supported by the nearly identical rating distributions of two Chase web pages that differ only by the presence of the SSL-post padlock icon in the login area. The SSL padlock at the bottom of browser frame enhanced trust in many subjects, however two subjects lost trust when mouse-over revealed a made-up certification authority, *Trust Inc.*
9. **Independent channels create trust.** If a stimulus suggested that the subject could call to verify the authenticity of the email or web page, then the very existence of this possibility strengthened the trust the subjects had in this stimuli. Subjects stated that they would not call the number to verify the authenticity, but *someone else would*.

## Acknowledgments

We wish to thank Sukamol Srikwan for her guidance on statistical methods and for her helpful discussions. Thanks to Sid Stamm and Liu Yang for suggestions improving the readability of the paper.

## References

1. “Camtasia Studio Screen Recorder for Demos, Presentations and Training,” Camtasia Studio 4, TechSmith Corporation. <http://www.techsmith.com/camtasia.asp>, last visited October 31, 2006.
2. Rachna Dhamija, J. D. Tygar, and Marti Hearst. Why phishing works. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 581–590, New York, NY, USA, 2006. ACM Press.
3. Julie S. Downs, Mandy B. Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security*, pages 79–90, New York, NY, USA, 2006. ACM Press.
4. B. Friedman, D. Hurley, D.C. Howe, E. Felten, H. Nissenbaum, “Users’ conceptions of web security: a comparative study,” In: *CHI '02 extended abstracts on Human factors in computing systems*. Minneapolis, Minnesota, USA: ACM Press; 2002:746-7.
5. V. Griffith and M. Jakobsson, “Messin’ with Texas, Deriving Mother’s Maiden Names Using Public Records,” *ACNS'05*, 2005.
6. M. Jakobsson and J. Ratkiewicz, “Designing Ethical Phishing Experiments: A Study of (ROT13) rOnl Query Features,” *WWW '06*.
7. T. Whalen and K.M. Inkpen, “Gathering evidence: use of visual security cues in web browsers,” In *Proceedings of the 2005 Conference on Graphics interface* (Victoria, British Columbia, May 09 - 11, 2005). ACM International Conference Proceeding Series, vol. 112. Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, pp. 137–144.
8. Likert, Rensis (1932) A technique for the measurement of attitudes. *Archives of Psychology*, 140 (June).