Editors: Matt Bishop, bishop@cs.ucdavis.edu
Deborah A. Frincke, deborah.frincke@pnl.gov

Why and How to Perform Fraud Experiments

Markus Jakobsson
Palo Alto Research Center

Peter Finn and Nathaniel Johnson
Indiana University

Fraud isn't new, but in the eyes of many experts, phishing and crimeware threaten to topple society's overall stability because they erode trust in its underlying computational infrastructure. Most people agree that phishing and crimeware must be fought, but to do so effectively, we must fully understand both types of threat; that starts by quantifying how and when people fall for deceit.

Fraud experiments using human subjects can benefit us in three ways. One, we can anticipate trends in online fraud by identifying yet-untapped behavioral vulnerabilities, which helps security-tool designers focus their efforts on the most acute problems. Knowing that typical Internet users don't understand cousin-name attacks, for example, suggests the use of heuristic spam filters that can scan for common cousin-name variants.[1] (Cousin-name attacks use URLs that resemble valid domains, such as paypa1.com, which puts a numeral 1 in place of the letter L, or mycitibank.com, which looks like a user's personal CitiBank Web page.) Similarly, knowing that users are willing to execute software sent by a friend—even if the software is self-signed by an unknown authority—suggests that operating systems should always reject self-signed certificates, except when expert users have configured their systems to explicitly allow them (www.indiana.edu/~phishing/verybigad/).

Understanding what types of attacks will succeed also helps designers build better user interfaces. Knowing that typical Internet users don't notice the absence of valid information but do notice the presence of incorrect information suggests that user interfaces should make it very difficult to ignore the absence of information needed to authenticate a service provider.[1] Bank of America's SiteKey, for example, presents users with personally chosen images when they visit a secure part of its site. Making the absence of such images truly difficult to miss (such as by embedding the login window in the image[2]) leaves users less likely to fall victim to phishing attacks because the phisher doesn't know what image to use. Finally, understanding why attacks succeed promotes better user habits. Typical online banking customers won't reject an email in which the apparent sender—their bank—authenticates itself by stating the first four digits of the recipient's account number instead of the last four.[3] However, the first four digits of a credit-card or bank account number aren't random; they're directly related to the issuer, and thus are easy for a phisher to guess. Educating users about simple tricks such as these encourages them to develop good habits rather than bad ones—SecurityCartoon.com, for example, is an educational effort that uses knowledge derived from phishing experiments to identify common misconceptions and security vulnerabilities, and explain to typical users what to do and what not to do.[4] Now that we've explained why fraud experiments are a good idea, let's look closer at how to perform them.

How to experiment

Researchers typically use three approaches to quantify fraud: surveys, in-lab experiments, and naturalistic experiments. Surveys tend to underestimate damages—many victims are unaware that an attack occurred or are unwilling to disclose that they fell for it. Conversely, in-lab experiments can overestimate attack awareness because of expectancy bias—the mere knowledge of the study's existence biases its likely outcome.[5] Subjects can't distinguish naturalistic experiments from reality, so this approach offers the most accurate picture because it mimics real attacks and thus measures actual success rates. However, such experiments pose a thorny ethical issue: if a study is identical to reality, then the study itself constitutes a real-world fraud attempt.

Before they even start a naturalistic fraud experiment, researchers must first face review and approval processes before institutional review boards (IRBs). This requirement can be daunting because fraud researchers must typically request waivers for certain aspects of the informed consent process as well as the use of deception (so that subjects don't

know they're participating in a study). The ethical issues related to waiving aspects of informed consent are controversial, with little consensus existing among IRB members and ethicists. Such issues are particularly controversial in the domain of online research, which is relatively new to IRBs and ethicists in general. Fraud research involves additional challenges as well. Although deception and the complete waiver of informed consent are a necessity in some types of studies on human subjects, researchers usually avoid them to the extent possible, and IRBs allow them only when a study's expected benefits outweigh the anticipated risks or if it meets certain conditions outlined in federal regulations governing human subject research.[6]

Consent and deception

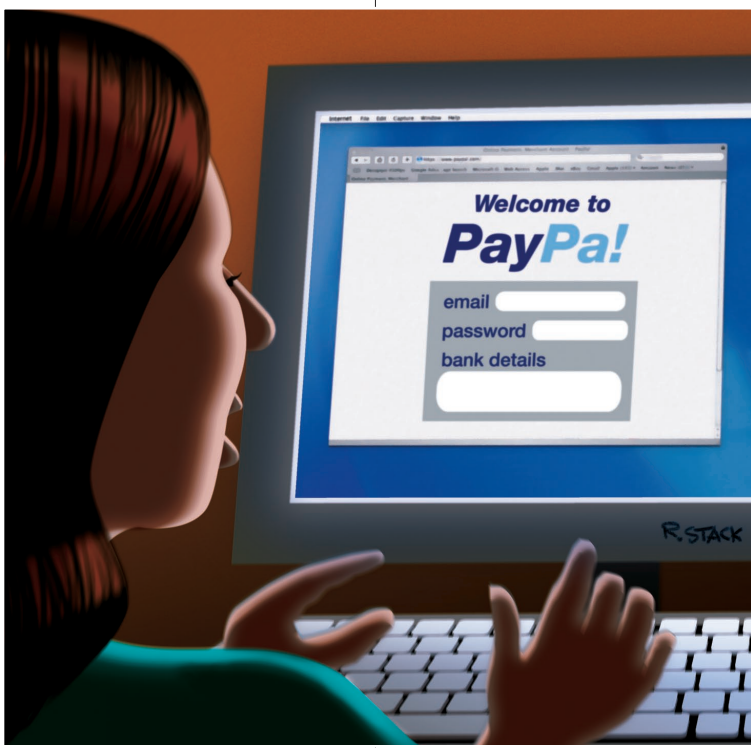
As an example of a naturalistic phishing experiment, let's look at the social phishing study that Tom Jagatic and his colleagues conducted at Indiana University to determine how social context increases the likelihood of success in a real phishing attack.[7] The experiment consisted of three phases and their corresponding protocols. The first step in the study described how to collect contextual data: because no human contact was necessary for this task, Indiana University's IRB determined that this step was exempt from a full IRB review process. The second phase described how to conduct the experiment on human subjects, so it required a full committee review. The IRB granted a waiver of consent because it determined that this part of the experiment would cause minimal to no real harm to the participants and that prior knowledge of the study would adversely affect its outcome. The third phase described the debriefing.

The study consisted of sending an email message to two groups of participants; under the subject line "This is cool!", the message's contents included a short amount of body text ("Hey, check this out"), a link to the experimental phishing Web site, and the spoofed sender's name. The link prompted the participant for university credentials, which the university's authentication services verified securely. Although these services verified the credentials, they didn't actually store usernames or passwords on the researchers' computers. The verification step was extremely important for gathering statistics about people who give away real usernames and passwords in bona fide phishing attacks, but if this were a real attack, the phisher would have stored and used the information against the participants.

A control group of 94 people received the email message from an unknown (fictitious) person; the second group of 487 people received the message from a known (spoofed) friend. The researchers determined the friend relationship in the experiment's first step, in which they collected data from a popular social networking site used by college students. Messages posted on one person's profile from someone else on the site indicated strong links between the two people, and when combined with email addresses harvested from other institutional Web sites, the researchers could build spoofed email messages with a strong social context. The study yielded surprising results: not only was the 72 percent success rate for authentications much higher than expected (the control group yielded 16 percent), but the ethical debate that followed on the topic of using human subjects without their consent offered valuable insights into the reactions of potential phishing victims. Although the initial reaction to the study was very negative, the overall consensus (among the more than 400 messages posted to the study's blog) was that this research was both necessary and beneficial for learning how to prevent contextual phishing attacks in the future.

Debrief or not?

In a follow-up to this experiment, one of us (Markus Jakobsson)



worked with Jacob Ratkiewicz[1] at Indiana University to study how unsuspecting users interpret cousin domains and IP address links and to what extent correct salutations affected their security decisions. We found that many users react suspiciously to URLs with an IP address, but that cousin domains don't arouse any concern at all. Similarly, we found that the presence of an incorrect salutation alerted users to a threat, but they didn't detect the absence of correct information.

Similarly, we found that the presence of an incorrect salutation alerted users to a threat, but they didn't detect the absence of correct information.

In our study, we got a large number of eBay users to unwittingly participate. We first determined the relationship between their email addresses, usernames, and salutations.[1] The subjects then received spoofed emails that seemed to be legitimate questions from other eBay users, to which they could respond only by visiting a site where they had to enter their eBay usernames and passwords. (At the time of the study, this type of phishing attack wasn't commonly known but did exist, according to follow-up discussions with eBay employees, who also determined that the explosion of abuse of this kind wasn't a result of our experiment.) Although the server verifying the credentials was an eBay server, we used obfuscation techniques to make it seem like a selected domain so that we could see the subjects' reaction (if any). We did this without eBay's knowledge or collaboration, but we proved afterward that we couldn't obtain any of the credentials. Many subjects detected the apparent attempt to fraud and didn't enter their credentials. To them, nothing unusual had happened—with the greatest likelihood, this was "just another" phishing attempt. A second group of subjects did enter their credentials—to them, too, nothing unusual had happened: they believed that they had just responded to a query from a peer eBay user. In this case, debriefing subjects of either type would significantly increase the harm done to them, which is the opposite result of a typical debriefing. After careful IRB review, we avoided the debriefing phase,[6] in spite of our use of deceit in the study.

We've argued here for naturalistic fraud experiments and illustrated how to design them. We've also described some of the ethical issues associated with such experiments. Although the intersection of technology and ethics poses difficult questions that have yet to be answered, we believe that the need to better understand how people interact with computers makes such research extremely worthwhile. □

References

1. M. Jakobsson and J. Ratkiewicz, "Designing Ethical Phishing Experiments: A Study of (ROT13) rOnl Auction Query Features," *Proc. 15th Int'l Conf. World Wide Web*, ACM Press, 2006, pp. 513–522.
2. R. Dhamija and J.D. Tygar, "The Battle against Phishing: Dynamic Security Skins," *Proc. 2005 Symp. Usable Privacy and Security*, ACM Press, 2006, pp. 77–83.
3. M. Jakobsson et al., "What Instills Trust? A Qualitative Study of Phishing," *Proc. 1st Int'l Workshop on Usable Security*, Springer-Verlag, 2007; www.informatics.indiana.edu/markus/papers/trust_USEC.pdf.
4. S. Srikwan and M. Jakobsson, "Using Cartoons to Teach Internet Security," *Cryptologia*, vol. 32, no. 2, 2008; www.securitycartoon.com/cryptologia08.
5. V. Anandpara et al., "Phishing IQ Tests Measure Fear, not Ability," *Proc. 1st Int'l Workshop on Usable Security*, Springer-Verlag, 2007; www.informatics.indiana.edu/markus/papers/phish6.pdf.
6. P. Finn and M. Jakobsson. "Designing and Conducting Phishing Experiments," *IEEE Technology and Society*, special issue on usability and security, vol. 26, no. 1, 2007, pp. 46–58.
7. T. Jagatic et al., "Social Phishing," *Comm. ACM*, vol. 5, no. 10, 2007, pp. 94–100; http://portal.acm.org/citation.cfm?doid=1290958.1290968.

Markus Jakobsson *is a principal scientist at the Palo Alto Research Center. His technical interests include Internet fraud, user education, and applied cryptography. Jakobsson has a PhD in computer science from the University of California, San Diego. Contact him at markus.jakobsson@parc.com.*

Nathaniel Johnson *is a principal systems analyst in the Division of Enterprise Software at Indiana University. His technical interests include open source software design and development, computer security, and phishing prevention. Johnson has an MS in computer science from Indiana University. Contact him at natjohns@indiana.edu.*

Peter Finn *is a professor in the Department of Psychological and Brain Sciences at Indiana University. His technical interests include the mathematical modeling of decision-making, developing and maintaining cybersecurity measures that model known processes that affect human decision-making, and developing secure methods of research data collection online. Finn has a PhD in psychology from McGill University. Contact him at finnp@indiana.edu.*