By Tom N. Jagatic, Nathaniel A. Johnson, Markus Jakobsson, and Filippo Menczer

SOCIAL PHISHING

Sometimes a "friendly" email message tempts recipients to reveal more online than they otherwise would, playing right into the sender's hand.

Phishing is a form of deception in which an attacker attempts to fraudulently acquire sensitive information from a victim by impersonating a trustworthy entity. Phishing attacks typically employ generic "lures." For instance, a phisher misrepresenting himself as a large banking corporation or popular online auction site will have a reasonable yield, despite knowing little to nothing about the recipient. In a study by Gartner Group [9], about 19% of all those surveyed reported having clicked on a link in a phishing email message, and 3% admitted to giving up financial or personal information. The research project described here was designed to provide us with a baseline success rate for individual phishing attacks, and was, when it was performed in 2005, the first study to achieve this goal.

Illustration by Robert Neubecker

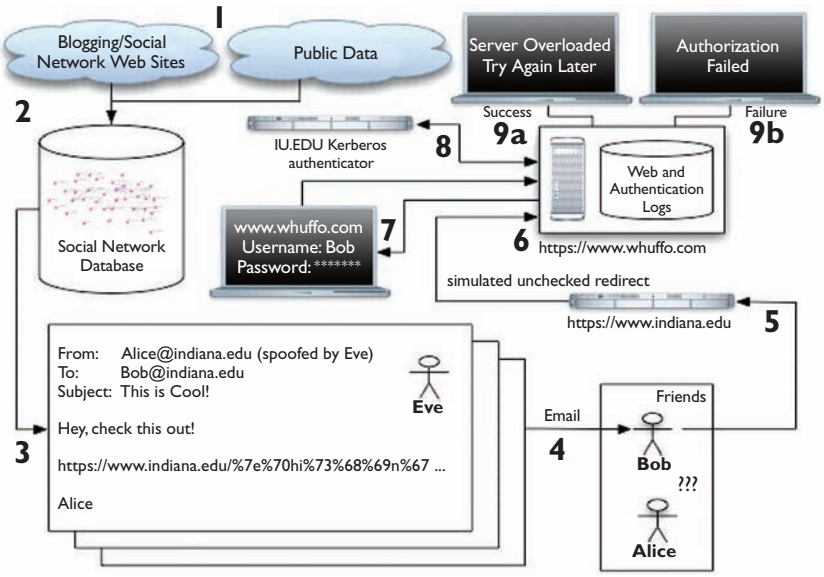

Figure 1. Illustration of phishing experiment: 1. Blogging, social network, and other public data is harvested; 2. Data is correlated and stored in a relational database; 3. Heuristics are used to craft spoofed email message by Eve "as Alice" to Bob (a friend); 4. Message is sent to Bob; 5. Bob follows the link contained within the email message and is sent to an unchecked redirect; 6. Bob is sent to attacker whuffo.com site; 7. Bob is prompted for his University credentials; 8. Bob's credentials are verified with the University authenticator; 9a. Bob is successfully phished; 9b. Bob is not phished in this session; he could try again.

It is worth noting that phishers are getting smarter. Following trends in other online crimes, it is inevitable that future generations of phishing attacks will incorporate greater elements of context to become more effective and thus more dangerous for society. For instance, suppose a phisher were able to induce an interruption of service to a frequently used resource, for example, to cause a victim's password to be locked by generating excessive authentication failures. The phisher could notify the victim of a "security threat." Such a message may be welcomed or expected by the victim, who would then be easily induced into disclosing personal information.

In other forms of so-called *spear phishing* or *context aware phishing* [10], an attacker would gain the trust of victims by obtaining information about their bidding history or shopping preferences (freely available from eBay), their banking institutions (discoverable through their Web browser history, see browser-recon.info and [12]), or their mothers' maiden names (which can be inferred from data required by law to be public [8]). Avi Rubin, a computer science professor at Johns Hopkins University, designed a class project for his graduate course, "Security and Privacy in Computing," to demonstrate how a database can be built to facilitate identity theft. The project focused on residents of Baltimore using data obtained from public databases, Web sites, public records, and physical world information that can be captured on the computer.

Given that phishing attacks take advantage of both technical and social vulnerabilities, there are a large number of different attacks; an excellent overview of the most commonly occurring attacks and countermeasures can be found in [4]. A more in-depth treatment—also covering attacks that do not yet exist in the wild—is offered in [11]. Here, we discuss how phishing attacks can be honed by means of publicly available personal information from social networks. The idea of using people's social contacts to increase the power of an attack is analogous to the way in which the "ILOVEYOU" virus used email address books to propagate itself. The question we ask here is: How easily and effectively can a phisher exploit social network data found on the Internet to increase the yield of a phishing attack? The answer, as it turns out, is *very easily and very effectively*. Our study suggests that Internet users may be over four times as likely to become victims if they are solicited by someone appearing to be a known acquaintance.

To mine information about relationships and common interests in a group or community, a phisher need only look at any one of a growing number of social network sites, such as MySpace (myspace.com), Facebook (facebook.com), Orkut (orkut.com), and LinkedIn (linkedin.com). All these sites identify "circles of friends" that allow a phisher to harvest large amounts of reliable social network information. The fact that the terms of service of these sites may disallow users from abusing their information for spam, phishing and other illegal or unethical activities is, of course, irrelevant to those who would create fake and untraceable accounts for such malicious purposes. An even more accessible source, used by online blogging communities such as LiveJournal (livejournal.com), is the Friend of a Friend project (www.foaf-project.org), which provides a machine-readable semantic Web format specification describing the links between people. Even if such sources of information were not so readily available, one could infer social connections from mining Web content and links [1].

In the study described here we simply harvested freely available acquaintance data by crawling social network Web sites. This way we quickly and easily built a database with tens of thousands of relationships. This could be done using off-the-shelf crawling

	Successful	Targeted	Percentage	95% C.I.
Control	15	94	16%	(9–23)%
Social	349	487	72%	(68–76)%

Table 1. Results of the social network phishing attack and control experiment. An attack was "successful" when the target clicked on the link in the email and authenticated with his or her valid IU username and password to the simulated non-IU phishing site. From a t-test, the difference is very significant ($p<10^{-25}$).

and parsing tools such as the Perl LWP library, accessible to anyone with basic Web scripting. For the purposes of our study, we focused on a subset of targets affiliated with Indiana University by cross-correlating the data with IU's address book database. This was done to guarantee that all subjects were IU students, which was part of the approval to perform experiments on human subjects; of course a real phisher would not need to perform such a weeding of victims.

We launched an actual (but harmless) phishing attack targeting IU students aged 18 to 24 years old. Targets were selected based upon the amount and quality of publicly available information disclosed about themselves; they were sampled to represent typical phishing victims rather than typical students. Much care in the design of the experiment and considerable communication and coordination with the university's IT policy and security offices were required to ensure the experiment's success. The intent in performing such an experiment was to quantify, in an ethical manner, how reliable social context would increase the success of a phishing attack. Standards involving federal regulations in human subject research and applicable state and federal laws had to be carefully considered. We worked closely with the University Institutional Review Board in designing the protocols of this unprecedented type of human subject study;[1] these efforts are described in [6].

As illustrated in Figure 1, the experiment spoofed an email message between two friends, whom we will refer to as Alice and Bob. The recipient, Bob, was redirected to a phishing site with a domain name clearly distinct from IU; this site prompted him to enter his secure university credentials. In a control group, subjects received the same message from an unknown fictitious person with a university email address. The design of the experiment allowed us to determine the success of an attack without truly collecting sensitive information. This was accomplished using IU authentication services to verify the passwords of those targeted without storing the passwords. Table 1 summarizes the results of the experiment. The relatively high success in the control group (16%) may perhaps be due to subtle context associated with the fictitious sender's email address and the university domain name identified in the phishing hyperlink. While a direct comparison cannot be made to Gartner's estimate of 3% of targets falling victim to phishing attacks, the 4.5-fold difference between the social network group and the control group is noteworthy. The social network group's success rate (72%) was much higher than we had anticipated. However, the figure is consistent with a study conducted among 400 cadets of the West Point Military Academy, where 80% were deceived into following an embedded link regarding their grade report from a fictitious colonel [5].

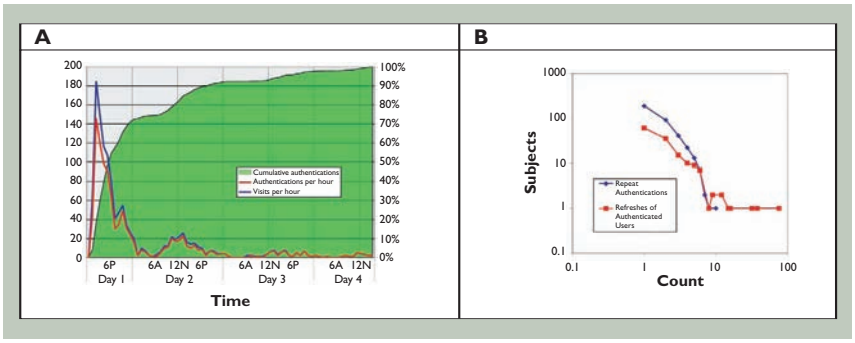

Figure 2. (a) Unique visits and authentications per hour. All phishing messages were delivered within three minutes. The experiment commenced on a Thursday afternoon and concluded Sunday afternoon. (b) Distributions of repeat authentications and refreshes of authenticated users. These results were interpreted from sequential accesses in the Web server logs. When each subject first attempted to authenticate, the request was coded with a unique identifier. Upon authenticating, their network ID was additionally associated with this unique identifier, marking the beginning of a session. Each successive request made after authenticating in the same session was considered a refresh.

Some insight is offered by analyzing the temporal patterns of the simulated phisher site's access logs. Figure 2a shows that the highest rate of response was in the first 12 hours, with 70% of the successful authentications occurring in that time frame. This supports the importance of rapid *takedown*, the process of causing offending phishing sites to become non-operative, whether by legal means (through the ISP of the phishing site) or by means of denial-of-service attacks—

[1]Two research protocols were written. The first protocol for mining the data was determined exempt from human subjects committee oversight. The second protocol for the phishing experiment underwent full committee review. A waiver of consent was required to conduct the phishing attack. It was not possible to brief the subjects beforehand that an experiment was being conducted without adversely affecting the outcome. The human subjects committee approved a waiver of consent based on the Code of Federal Regulations (CFR) 46.116(d). A debriefing email message explained the participants' role in the experiment after the fact and directed them to our research Web site for further information.

both prominently used techniques. Figure 2b reports the distributions of the number of times that victims authenticated or refreshed their credentials.

The reason for repeated visits to the simulated phisher site is that, as shown in Figure 1, victims who successfully authenticated were shown a fake message indicating the server was overloaded and asking them to try again later. A real phisher would not need to do this, of course, but we wanted to count how many victims would catch on or continue to be deceived; those who repeatedly authenticate give us a lower bound on the number of victims who continue to be deceived. The log-log plots in Figure 2b highlight distributions with long tails—some users visited the site (and disclosed their passwords) a large number of times. This, in spite of many ways to detect the phishing attack, for example, mouse-over, host name lookup, whois registrant database lookup, a bogus authentication message, and most tellingly the non-university URL in the browser's address bar. We must conclude that the social context of the attack leads people to overlook important clues, lowering their guard and making them significantly more vulnerable.

Additional interesting observations stem from analyzing the gender of the subjects who fell victim to the social phishing attack, as illustrated in Table 2. We see that females were more likely to become victims overall (77% versus 65% for males). Furthermore, the attack was more successful if the spoofed message appeared to be sent by a person of the opposite gender. This was true for both males and females, but the effect was more marked for males (68% if the message was from a female versus 53% if from another male). This suggests yet another vulnerability factor that phishers can easily exploit.

Finally, let us look at some demographics of the victims. Figure 3a shows a correlation between attack success rate and age, with younger targets being slightly more vulnerable. Figure 3b reports on success rates for students in different majors. All majors show a significant gap between the success rate in the social network group versus the control group. The gap is highest for science students, none of whom fell victim to the attack in the control case (out of 17) while 77 out of 96 did when the email message appeared to come from a friend. Somewhat reassuringly, students in technology majors (computer science, informatics, and cognitive science) seemed to be the least vulnerable group.

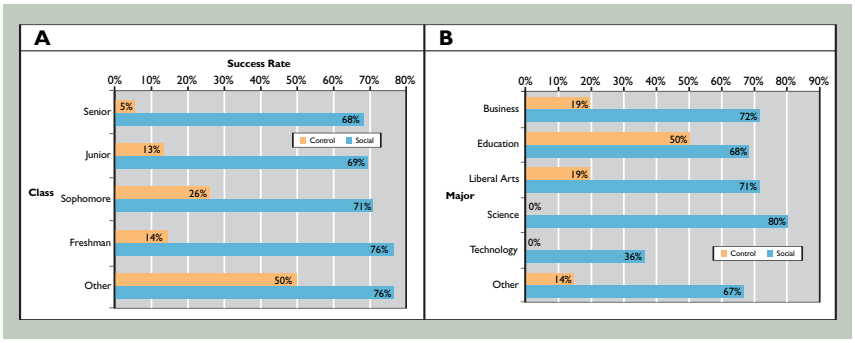

Figure 3. (a) Success rate of phishing attack by target class. "Other" represents students who did not provide their class or students who are classified as graduate, professional, and so on. While according to t-tests the differences in success rates are significant for all classes ($p \leq 0.01$), x^2 tests reveal that the success rate of neither the control nor the social attack depends significantly on the class. (b) Success rate of phishing attack by target major. "Other" represents students who did not provide information about their major. While according to t-tests the differences in success rates are significant for all majors ($p \leq 0.02$), x^2 tests reveal the success rate depends significantly on the major only in the social attack ($p=0.05$).

We also performed a third experiment that included an element of greater context—a message forwarded to a friend from a group of friends. It was hypothesized that the stronger context would yield a greater success rate. Highly connected subjects were chosen as the spoofed sender of the forwarded message. Unfortunately, due to a coding error, the results are not representative of the intended design of the experiment. Suppose Bob is friendly with Alice, Carol, Dave, and Ed. A message from Bob to Alice, Carol, Dave, and Ed was supposed to be forwarded to Frank but was instead (due to the coding error) sent to Bob with Frank's name. The subject (Bob) would likely be more suspicious of acting upon a message that was not addressed to him, and of which he was the purported originator. The success rate of the flawed experiment was 53% (139 success of 260 targeted). In some initial reactions the flawed messages were interpreted as an "email virus" according to comments posted on the project blog.

A total of 1,731 participants were included in the study—921 subjects received phishing attacks and 810 had their email addresses spoofed. To provide all of these participants and the campus community with a public discussion forum for anonymous comment and feedback about the experiment, a debriefing message invited subjects and participants to visit the project Web site and blog. Some media coverage from the student newspaper and a popular technology Web site (slashdot.org) attracted other visitors. After three days, the blog counted 440 posts, the majority of which were supportive of the experiment and the lessons learned from it. The number of complaints

made to the campus support center was also small (30 complaints, or 1.7% of the participants). Only seven participants (0.4%) requested to be excluded from the study (which they were).

Despite the relatively small number of complaints, the critics among the experiment participants were very vocal as demonstrated by the messages posted on the blog in the first several hours following the debriefing message. Significant insight can be gathered from these reactions, not only toward the ethical aspects of conducting such a study but also toward a better understanding of phishing victims, their vulnerabilities, and their feelings following an attack. Here, we report on some of the observed reactions along with lessons that can be learned from them:

	To Male	To Female	To Any
From Male	53%	78%	68%
From Female	68%	76%	73%
From Any	65%	77%	72%

Table 2. Gender effects. The harvested profiles of potential subjects identified a male/female ratio close to that of the general student population (18,294 males and 19,527 females). The number of females eligible for participation in the study (based on the requirement that subjects identify their age or birth date) was higher than males. x^2 tests of independence reveal that while the gender of the sender alone did not have a significant effect on the success rate ($p=0.3$), the gender of the receiver was significant ($p<0.005$) and the combination of sender-receiver genders was also significant ($p<0.004$).

- **Anger:** Some subjects called the experiment unethical, inappropriate, illegal, unprofessional, fraudulent, self-serving, and/or useless. They called for the researchers conducting the study to be fired, prosecuted, expelled, or otherwise reprimanded. These reactions highlight that phishing not only has the potential monetary costs associated with identity theft, but also a significant psychological cost to victims. Even though no sensitive information about the victims was retained (or even ever stored) in this study, some victims were clearly upset that the phishers had tricked them and violated their privacy.
- **Denial:** No posted comments included an admission that the writer had fallen victim to the attack. Many posts stated that the poster did not and would never fall for such an attack, and they were speaking on behalf of friends who had been phished. This natural denial reaction (as well as the anger and blaming of researchers mentioned earlier) suggests that we may find it difficult to admit to our own vulnerability. As a consequence many successful phishing attacks may go unreported, making phishing success rates from surveys severely underestimated.
- **Misunderstanding of email:** Many subjects were convinced the experimenters (with the complicity of IU officials) had hacked into their email accounts. They believed this was the only possible explanation for the spoofed email messages. This reaction highlights two concerns: first, few people understand how easy it is to spoof messages; second, many users overestimate the security and privacy of email.
- **Underestimation of dangers of publicly posted personal information:** Many subjects did not understand how the researchers had obtained information about their friends, and assumed the researchers had accessed their address books. Others, understanding that the information was mined from social network sites, objected that their privacy had been violated by the researchers who accessed the information that they had posted online. These reactions highlight that some users do not appreciate the potential ramifications of the information they willingly disclose on the Web. Some believe the information on social network sites is not public, either because it should be protected by terms of service or because it should only be accessible to their friends. It is not clear to them that anyone (without ethical concerns) can easily gather their personal information and that in most cases there are no consequences for the offender.

Another valid concern was expressed by some subjects whose names and email had been spoofed as senders. These subjects were notified by the corresponding receivers and initially believed that their computers were infected by an email virus. Some of these participants may have needlessly changed their campus account passwords and installed anti-virus software. While these actions may be positive protective steps in general, they may also have caused undue stress. Furthermore, since the study was part of a class project, the attack was carried out near the end of the semester. This may have intensified the stress felt by some students. Finally, anonymous blogs lend themselves to abuse; we spent considerable efforts censoring inappropriate messages and eventually were forced to shut down the blog three days after the end of the experiment. Some offensive posts were directed by people external to the university (the Slashdot crowd) toward the subjects who complained about the study. It is not clear how one might address this problem in a future experiment. Any kind of filter that disallows posts from outside the university implies a loss of anonymity, either by requiring a login or by monitoring IP addresses. A feedback mechanism balancing

between the conflicting needs for anonymity guarantees and abuse prevention remains elusive.

These issues must be addressed as we strive to find ethical ways to conduct experiments that can shed light into social engineering attacks, such as phishing, and help us design effective countermeasures. The study reported here established for the first time a baseline for the success rate of phishing attacks, both traditional and with social context. The astonishing percentages of victims who disclosed their university credentials to a non-university site underscore the need to strengthen our efforts in developing phishing prevention techniques.

In terms of technical countermeasures, there are several noteworthy efforts that would have reduced the success rate of our particular experiment, and of many real phishing attacks. Namely, if digitally signed email became commonplace, this would reduce the likelihood of users falling victim to attacks of this type, as many users would have realized that the messages were not sent by the apparent senders. However, as indicated in a user study by Garfinkel and Miller [7], many users may still be vulnerable. A second line of defense might be a browser toolbar [2], which alerts users of likely Web spoofing attempts, Indeed, this might have allowed many subjects to detect a phishing attempt corresponding to the experiment we performed. A technique to provide users with a secure path for entering passwords [3] could be used to alert users that they are attempting to authenticate to an unknown site; whereas this would not have affected our experiment (since we let users authenticate to the real IU authentication server), it would have alerted victims in a real phishing attack. Simple spam filters, on the other hand, are not likely to have an impact on attacks like ours, unless they also detect whether email messages have been spoofed.

Our study also points to the need for extensive educational campaigns about phishing and other security threats. Efforts such as SecurityCartoon.com aim to make typical Internet users less vulnerable by a heightened awareness of the dangers of phishing, the importance of reporting attacks to which they fall victims, the ease of spoofing, and the possible (mis)uses of personal information posted on the Web. At IU, the IT policy and security offices have rapidly put these lessons to use through a campuswide campaign that, among other things, warns students that phishing attacks may appear to come from anyone—*even friends with IU addresses.* It remains to be seen whether such educational campaigns work in the long run. One way to evaluate their effectiveness would be to repeat the social phishing experiment in the future.

Phishing has become such a prevalent problem due to its huge profit margins, ease in performing an attack, and difficulty bringing those responsible to justice. As countermeasures to combat phishing improve, it is likely that phishers will incorporate greater contextual information in their attacks to appear more convincing.[2] We now know that social networks can provide phishers with a wealth of information about unsuspecting victims. These so-called social phishing attacks underscore the dangers of publicly disclosing too much personal information and emphasize the need for adequate countermeasures. C

References

1. Adamic, L.A. and Adar, E. Friends and neighbors on the Web. *Social Networks 25*, 3 (July 2003), 211–230.
2. Chou, N. Ledesma, R., Teraguchi, Y., Boneh, D., and Mitchell, J.C. Client-side defense against Web-based identity theft. In *Proceedings of the 11th Annual Network and Distributed System Security Symposium*. (Feb. 2004).
3. Dhamija, R. and Tygar, J.D. The battle against phishing: Dynamic security skins. In *Proceedings of the ACM Symposium on Usable Security and Privacy.* (2005), 77–88.
4. Emigh, A. Online identity theft: Phishing technology, chokepoints and countermeasures. ITTC Report on Online Identity Theft Technology and Countermeaures (Oct. 2005); www.anti-phishing.org/Phishing-dhs-report.pdf.
5. Ferguson, A.J. Fostering e-mail security awareness: The West Point carronade. *Educause Quarterly 28*, 1 (2005).
6. Finn, P. and Jakobsson, M. Designing and conducting phishing experiments. *IEEE Technology and Society Magazine, Special Issue on Usability and Security* (2005).
7. Garfinkel, S. and Miller, R. Johnny 2: A user test of key continuity management with s/mime and outlook express. Presented at the Symposium on Usable Privacy and Security (July 6–8, 2005, Pittsburgh, PA).
8. Griffith, V. and Jakobsson, M. Messin' with Texas: Deriving mother's maiden names using public records. *RSA CryptoBytes 8*, 1 (2006).
9. Gartner Group. Gartner study finds significant increase in e-mail phishing attacks (Apr. 2004); www.gartner.com/5_about/press_releases/asset_71087_11.jsp, April 2004.
10. Jakobsson, M. Modeling and preventing phishing attacks. In Phishing Panel at Financial Cryptography. Feb. 2005.
11. Jakobsson, M. and Myers, S. *Phishing and Counter-Measures.* John Wiley and Sons, 2006.
12. Jakobsson, M. and Stamm, S. Invasive browser sniffing and countermeasures. In *WWW '06: Proceedings of the 15th International Conference on World Wide Web.* ACM Press, NY, 2006, 523–532.

Tom N. Jagatic (tnj@mit.edu) is a senior IT security consultant at the Massachusetts Institute of Technology. He was formerly Manager of Incident Response for the Indiana University IT Policy Office.
Nathaniel A. Johnson (natjohns@indiana.edu) is a software architect for the Division of Enterprise Software at Indiana University, Bloomington.
Markus Jakobsson (markus@markus-jakobsson.com) is an associate professor of informatics at Indiana University, Bloomington. He is also associate director of the Center for Applied Cybersecurity Research, and co-founder of RavenWhite.
Filippo Menczer (fil@indiana.edu) is an associate professor of informatics, computer science, and cognitive science at Indiana University, Bloomington. The experiment described here was carried out as a class project for the graduate course that he teaches on Web mining.

[2]In fact, social spoofing has been reportedly used to deliver spam to MySpace users in late 2006.