12-1-2006

Don't Be a Phish: Steps in User Education

Stefan Robila
Montclair State University, robilas@mail.montclair.edu

James W. Ragucci
Montclair State University

Don't be a Phish: Steps in User Education

Stefan A. Robila
Montclair State University
RI 301, Computer Science
Montclair, NJ 07043
001-(973)-655-4230
robilas@mail.montclair.edu

James W. Ragucci
Montclair State University
RI 301, Computer Science
Montclair, NJ 07043
001-(973)-655-4166
raguccij1@mail.montclair.edu

ABSTRACT
Phishing, e-mails sent out by hackers to lure unsuspecting victims into giving up confidential information, has been the cause of countless security breaches and has experienced in the last year an increase in frequency and diversity. While regular phishing attacks are easily thwarted, designing the attack to include user context information could potentially increase the user's vulnerability. To prevent this, phishing education needs to be considered. In this paper we provide an overview of phishing education, focusing on context aware attacks and introduce a new strategy for educating users by combining phishing IQ tests and class discussions. The technique encompasses displaying both legitimate and fraudulent e-mails to users and having them identify the phishing attempts from the authentic e-mails. Proper implementation of this system helps teach users what to look for in e-mails, and how to protect their confidential information from being caught in the nets of phishers. The strategy was applied in Introduction to Computing courses as part of the computer security component. Class assessment indicates an increased level of awareness and better recognition of attacks.

Categories and Subject Descriptors
K.3.2 [**Computer and Education**]: Computer & Information Science Education – *Computer Science Education, Curriculum*. C.2.0 [**Computer-Communication Networks**]: General - *Security & Protection*, K.6.5 [**Management of Computing & Information Systems**]: Security & Protection Education

General Terms
Reliability, Experimentation, Security, Human Factors

Keywords
Phishing, information security, computer education

1. INTRODUCTION
Evolution does not only apply to plants and animals, it also applies to human technology. Armor, for instance, was designed to protect the wearer from being wounded. Technology eventually evolved to find ways of piercing or breaking that armor. In response, various types of armor have been developed to protect the wearers from the new methods of destroying the armor, creating an endless cycle of improvement within technology. When one side improved its defenses, the other side would improve its offences. Computer security and security threats have evolved in the same manner. Throughout this ongoing race, one thing remained constant as the weakest link, the human factor [1].

Spam, defined as the unsolicited sending of commercial e-mail advertisements [2], which combined with online fraud has caused losses of over 200 million dollars in 2003 alone [3]. A defense against the overabundance of spam is the implementation of spam filters currently used within most of the organizations or free mail systems. Social engineering, defined by Kevin Mitnick as "using manipulation, influence and deception to get a person, a trusted insider within an organization to comply with a request, and the request is usually to release information or perform some sort of action item that benefits that attacker"[1] transforms spam to the next level in the evolutionary chain. A combination of fraudulent spam e-mail with social engineering creates a relatively new tactic called Phishing. Phishing is using social engineering to send spam e-mail(s) to unsuspecting victims, known as phishes [4]. The e-mails are disguised ("spoofed") as coming from legitimate corporations and aim at directing users to copies of legitimate websites. The "phishers" goal is to "fish" for confidential information that the phishes have access to, such as bank account numbers, usernames, passwords and social security numbers [2]. Phishing attacks come in a large variety of flavors. Some attacks may masquerade as security upgrades or information verification from the Bank of America. In more recent times, some phishing attacks have appeared claiming to be a charity organization collecting money to benefit the victims of Hurricane Katrina or the 2004 Indian Ocean Tsunami [5] as seen in Fig 1 [6]. Damage from these attacks is immense. Between May 2004 through May 2005, approximately 1.2 million U.S. computer users suffered losses of nearly one billion dollars [7]. The United States is not the only target. In the United Kingdom, it was reported in March 2005 that damage jumped 20% to cause 504 million pounds in damage [8]. In addition, phishing attacks were also reported in non-English speaking countries as early as 2004 [9].

Although the frequency of phishing attacks has recently stabilized [10], context aware phishing is of particular concern for the future. A context aware attack consists of the phisher gaining knowledge of what sites and services the phish uses and customizing an attack that appears to be from the target's service [11]. While currently, a phishing attack success rate is under 1% [1] a context aware attack would result in much higher rates [11].

ITiCSE'06, June 26-28, 2006, Bologna, Italy.

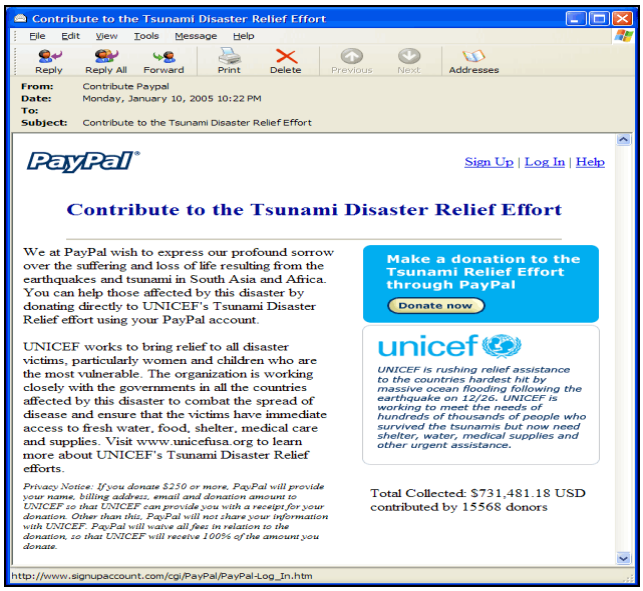

Fig 1. Phishing Example

Faced with these prospects, one must see what tools are available to fight phishing. On one hand, application development will continue to improve spam and phishing detection. On the other hand, human factor risks can be reduced by spam and phishing education. In this paper we analyze various efforts in educating students related to phishing and we discuss our experience with teaching based on phishing IQ tests and context information. The paper is organized as follows. In Section 2 we briefly look at previous phishing education efforts. In Section 3 we discuss context aware phishing and the design of context aware attacks and tests. In Section 4 we discuss the pedagogical setting for phishing education. Section 5 presents a summary of the IQ tests as well as results of the educational assessment survey. We end the paper with Conclusions and Future Work (Section 6) and References.

2. PHISHING EDUCATION

Anti-phishing action is supported by a wide group of interesting parties including most of the financial institutions. Private and government institutions have developed phishing awareness websites including [12], [13], [14].

According to [3], there are five individual issues that have to be addressed in order to combat phishing: education, preparation, avoidance, intervention and treatment. Within these groupings, education is given the least attention. The paper only states that users need to be educated in how to recognize suspicious requests in their e-mail. There is minimal instruction of how users should learn how to identify phishing attacks. In fact, both [3] and [15] briefly list what should be taught to users, but they do not list a vehicle for this education. The websites listed above, while relatively up to date, mainly provide a description of phishing attacks and some good sense advice.

An alternative approach is described in [16]. Researchers at Indiana University conducted a study of 1,700 students in which they collected websites frequently visited by students and either sent them phishing messages or spoofed their e-mail addresses. The respondents were provided with a discussion forum and were informed of their participation in the study. Nevertheless, while the authors indicate that public awareness was of concern, the only clear outcome was an increased effort on the university IT to protect against phishing. The large success rate of the attacks (over 50% in some experiments) was mainly due to the use of social context information.

One tool that has been developed by [17] to educate users is a phishing IQ test. The test provides users with a combination of actual phishing attacks and legitimate e-mails and users are asked to distinguish between them. Once finished the user is given a score of how well he or she was able to identify the e-mail. As a result of this IQ test, users' awareness of phishing increased. In a period of a year, the test has seen an average score increase of 14 points, with this year the average score being a 75% [18]. Although this increase is a good sign, the average score shows that 25% of the e-mails are wrongly identified. Results show that 82% of the test takers identify phishing e-mails correctly, but legitimate e-mails are only identified correctly worldwide 52% of the time [18]. This alarmingly low number compared to the correct number of detected phishing attempts could probably be attributed to users classifying all of the e-mails on the test as phish [19].

3. CONTEXT AWARE PHISHING AND STUDENTS

While benign in nature, the MailFrontier's IQ tests have limited usability in a classroom. Given its generality, it is likely that most of the students will have never dealt with the companies listed in the tests. As such, the corresponding messages would have end up being classified as SPAM from the beginning, would have been placed in a bulk folders, or deleted without being opened [15]. Alternatively, the Social Phishing experiments described in [16] would have a stronger impact since they addressed entities familiar to the users. However, a direct phishing attack targeting the students, while interesting from the point of research, would have limited educational impact.

We suggest a hybrid approach where we employ IQ tests enriched by social contexts. Borrowing from the idea of context aware phishing attacks, described in [11], if an IQ test is developed using known services that a group of users have a high probability of using, then the element of inexperience with the company/service is eliminated. Instead, users would be familiar with the source that they are being tested on and should be better able to identify real e-mails from fraudulent phishing attempts. Once this is accomplished, then a different set of reasons can be assumed for a user to falsely identify an e-mail. Reasons for false identification might include any of the following:

- User is unaware of what phishing is.
- User is aware of phishing, unaware how to search for it.
- User does not suspect a phishing attempt from that company/service.

There are several approaches that would allow collection of social context information. These include collection of browser data (such as history collection, auto-fill properties, etc. - for some excellent examples, visit [20]), exploit of the browser tabs [21], access to user's labs, and crawling social network websites [16].

We pursued to design a customized phishing IQ test focused on Montclair State University students. Picking companies and online services that a large portion of the student body uses per the specifications addressed above in section three required different strategies of data collection. Most of the data was collected from the computer laboratory where the test was administered and the campus computer centers. All of the computers examined in these locations underwent browser analysis from tools found at [20] followed by analysis of the browsers cache and local history to determine if people used the machines to visit any additional websites of interest. Other websites that were used for this project were included after looking at the resources available to students on campus. For instance, the ATM machines on the MSU campus are run by BankX. Therefore, it is logical to assume that many, if not most of the students who attend the university have an account at BankX to avoid service charges when using the ATM machines. We also note that while the use of social network websites is very attractive, generating user personalized IQ tests was beyond the scope of our work.

Once the data of what MSU students use was gathered, phishing and legitimate e-mails were generated and mixed with others from a variety of sources including [10] and [14] and then grouped in an IQ survey that contained 12 questions, with six answers that are legitimate and six that are phishing attacks. At the end, the survey was formed with messages from: University registrar (P), Ebay.com (L), BankX (L), University Parking (L), MySpace.com (P), BankY (L), Fastweb.com (L), Amazon.com (P), Paypal.com (P), Yahoo.com (P), Blackboard.com (P), Facebook.com (L) where P and L indicate a phishing or a legitimate e-mail respectively. Fig 2 shows a sample question related to Facebook.com. To implement the survey and record the results we used an in-house mechanism based on javascript and an Apache Tomcat server running on a Sun Solaris box. A survey taker would be presented with images of an email at a time and asked to indicate if it is a phishing attempt or not. Upon selecting and confirming an answer, the next e-mail image was displayed. At the end, the application displayed the number of correct answers.

Fig 2. Phishing e-mail created from a confirmation e-mail sent by Facebook.com to a dummy account

1. Learn what phishing is [4]
2. Understand the implications of phishing [12]
3. Discuss the ways phishing can be detected [12], [13]
4. Discuss the ways phishing information can be collected.[20]
5. Discuss ways to evaluate phishing education [12], [20]
6. Fill Phishing IQ
7. Discuss general results and evaluate the session.

Fig. 3 Phishing Education Steps

4. EDUCATIONAL APPROACH

4.1 Steps

Fig. 3 describes the steps taken in discussing and testing phishing knowledge. We started by presenting what phishing is and discussing its implications. Following, we discussed the nature of the phishing emails and we discussed the methods that can be employed by attackers to generate better targeted messages as well as identify many browser and web vulnerabilities. Next, we discussed ways to increase public awareness on phishing and debated on aggressive (such as the phishing emails used in [12]) versus passive (such as the IQ tests [20]) education.

Steps six and seven were used to assess the quality of the instruction. First, an IQ survey was administered followed by a class assessment survey. The session was timed at approximately 30 minutes. A control number was presented to the survey taker. The number could be used to show that the student has taken the survey; however, it could not be related to the answers, in order to ensure anonymity.

4.2 Course Description

We have applied the phishing education module to two sections of our Introduction to Computing course. Approved to satisfy the General Education requirements for computer fluency in a liberal arts college, the course includes one lecture and one lab session every week for a semester. The lectures are dedicated to various computing issues such as computer organization, structured programming, networks, privacy and ethics, while the labs are focused on acquiring skills on various productivity packages as well as learning basic concepts in programming, web development and netiquette. The course is attended by large numbers of non-science majors and it is organized in sections of up to 24 students (due to lab space limitations).

As per the arguments in [22], the Introduction to Computing course was a natural choice for piloting the phishing education module since the vast majority of students will have computing experience mainly from the user's point of view. In addition, one component of the course is teaching students about basic computer and network security issues, a broad topic that includes phishing. Finally, the lab component of the course allowed us to have the students take the surveys in a set timeframe and to discuss the general results immediately.

5. RESULTS

Of the 48 IQ surveys completed, we recorded an average correctness rate of approximately 6.87 corresponding to an

average IQ of 57.29%. This means that approximately one out of two e-mails, the students would erroneously identify a legitimate message as a phishing attack, or a phishing attack, as a legitimate e-mail.

Next, we analyzed the success rate for the legitimate and phishing e-mails respectively. In case of the legitimate e-mails, the students we able to correctly identify on average 3.64 out of 6 (60%) while the phishing e-mails were recognized on average 3.22 out of 6 (53%). Interestingly, the messages easiest recognized as legitimate was one from Fastweb.com (81% correct match rate) while the lowest legitimate recognition was received by a message from BankX (50% correct match rate). For phishing, the lowest correct identification was achieved by a message purporting to be from the university's registrar office (35% correct match rate) while the highest was built from scratch claiming to be from an instructor and referring students to Blackboard.com (75%). We note that the question with the lowest overall correct match rate, the phishing attempt based on the university's registrar's office, was requesting the student to login and verify personal information. At the time of the survey administration, the university was revealing that a large number of student data was inadvertently disclosed publicly and that identity theft risk for the students was increased. We believe that the students were aware of the incident and that this played a major role in their choice for an answer. If such a phishing attack would have really occurred, a 65% success rate would have been disastrous. Also note that a month after the survey was administered, the university sent a legitimate version of the e-mail to the student body.

A second survey was administered to students to evaluate the educational value of the session. The surveys were anonymous and no points were awarded based on its completion. 78% of the respondents indicated that they were not aware of phishing prior to the class. At the same time, 93% acknowledged receiving possible phishing e-mails in the past and 28% revealed that they probably answered to phishing attacks in the past.

Fig. 4 and 5 are examples of the questions intended to test understanding of social context factors for phishing. We note that, as previously reported, most of the students still consider messages from a friend as legitimate, although a large minority (40%) now has second thoughts. In addition, most of the participating students have become aware of the vulnerabilities existing in browser data (see Fig 5). The students have positively perceived both the IQ surveys as well as the overall phishing session, with 94% agreeing that the IQ survey was helpful (Fig 6) and all agreeing that the session was helpful (Fig. 7).

Finally, class discussions have revealed that most of the students place significant trust in the educational institutions as well as the social network websites. As such, while many have easily declared that they would ignore banking or auction/payment site emails, they were surprised to discover that attacks could occur by spoofing the university or social network website communication. We believe that this is explainable by the current focus of the anti-phishing and anti ID-theft campaigns that put emphasis on protection of financial and health information and ignore popular web destinations. Unfortunately, such destinations also handle private information (such as names, addresses, date of birth) which can be used for identity theft. Successful phishing attacks based on them would be equally disastrous.

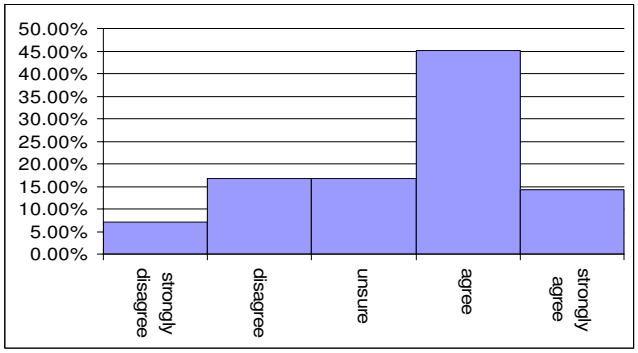

Fig 4. Receiving a message from a friend makes me less likely to assume it is a phishing attempt.

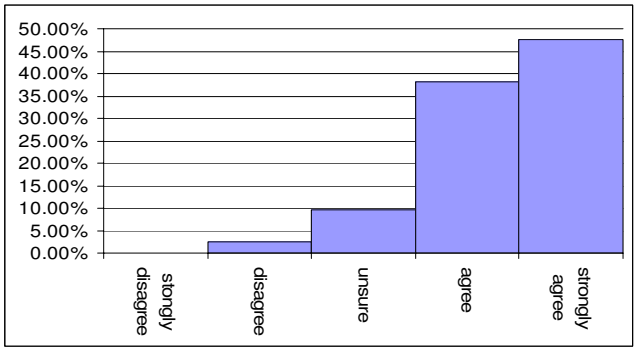

Fig. 5 Information stored by the Internet browser can be used to refine a phishing attack.

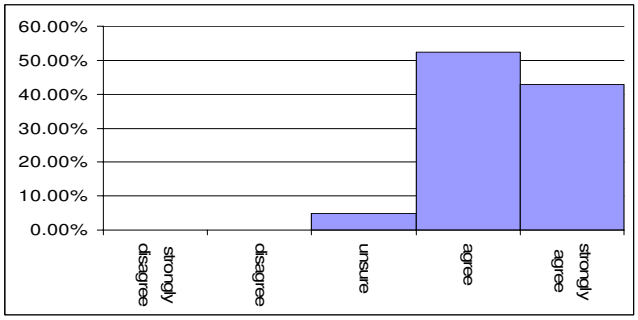

Fig. 6 The phishing IQ survey was helpful in understanding the topic.

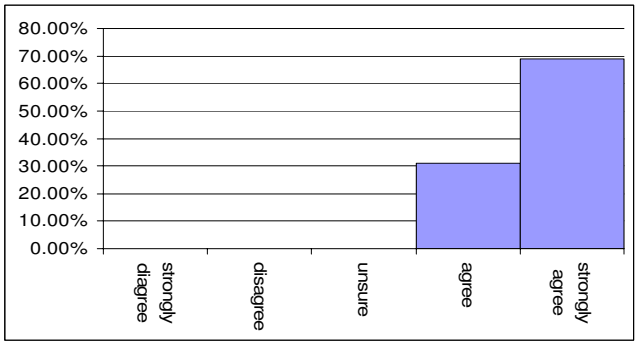

Fig. 7 Attending this session made me more informed about phishing and less likely to fall prey to such an attack.

6. CONCLUSIONS

Phishing has become a significant problem for internet users. While most of its effects are noticeable in the United States, it is expected that phishing will continue to expand all over the world. Recent reports such the ones produced by the Antipishing Working Group [23] and the Korea Internet Security Center [24] reflect these trends and predict a continuous increase in attacks and diversity. While technology advances continue to fight the problem, user education continues to constitute a significant component.

In this paper we have discussed an approach to user education that involves quantitative testing and social context aware examples. The strategy allowed us to include phishing topics in an Introduction to Computing course aimed at students pursuing a non-computer science education. The phishing IQ test and the session evaluation survey reveal that the current student body is mostly oblivious to phishing threats. Upon being exposed to the topics and shown how to analyze a message for phishing characteristics, students are able to correctly identify most of the threats. The students have positively appreciated the session and its format and have acknowledged its usefulness.

More work remains to be done. Given a predicted increase in tools available to fight phishing, it is expected that future attacks will continue to be more and more refined in user and event specificity. Previous work together with our experiments show that such attacks have an extremely high success rate since they most likely appeal to the user's emotions. Accordingly, phishing education will continue to be improved by use of user specific tools. A social aware IQ test, that would be personalized on each user, could be such a tool.

Our work can be expanded for use in other courses as well for general public training. Coupled with context information, one can always design specific tests, targeted at specific user groups. Overall, our approach to phishing education was shown to be an attractive tool.

7. ACKNOWLEDGMENTS

We would like to thank Dr. Carl Bredlau for his vital part in the construction of the creation of the survey using jsp files, as well as hosting the files on his server.

A sample of the phishing test together with other materials is available at http://www.csam.montclair.edu/~robila/RSL/Phish/

8. REFERENCES

[1] CNN. com, "A convicted hacker debunks some myths." http://www.cnn.com/2005/TECH/internet/10/07/kevin.mitnick.cnna/index.html 2005, accessed 01/06/06

[2] Duntemann J., *Degunking Your Email, Spam, And Viruses. Scottsdale*, Arizona: Paraglyph Press, 2004

[3] Merwe A, Loock M., and Dabrowski M.. "Characteristics and responsibilities involved in a Phishing attack." Proc. ACM WISCT 05, **92**, 249-254, 2005

[4] http://en.wikipedia.org/wiki/Phishing, accessed 30 Nov 2005

[5] Roberts, Paul F. "Cyber-looters Capitalize on Katrina." eWeek. 12 Sept. 2005: 11-12

[6] MailFrontier Phishing IQ, "Paypal Tsunami" example, http://www.mailfrontier.com/quiztest2/S2img/Q22_tsunami.gif, accessed 3 Nov. 2005.

[7] Kerstein P.L., "How Can We Stop Phishing and Pharming Scams?" http://www.csoonline.com/talkback/071905.html, accessed 27 Nov 2005

[8] Richardson T., "Brits Fall Prey to Phishing." *The Register*. http://www.theregister.co.uk/2005/05/03/aol_phishing/, accessed 27 Nov 2005

[9] Sunday Morning Herald, "Phishing Spreads in Europe", http://www.smh.com.au/articles/2004/05/10/1084041315645.html, accessed 5 Jan 2006

[10] Anti-Phishing Working Group, *October 2005 Report,* http://antiphishing.org/apwg_phishing_activity_report_oct_05.pdf, accessed 27 Nov 2005

[11] Jakobsson M., Modeling and Preventing Phishing Attacks. *Phishing Panel in Financial Cryptography* '05.

[12] Anti-Phishing Working Group, http://www.antiphishing.org/, accessed 27 Nov 2005

[13] Better Business Bureau, http://www.bbbonline.org/idtheft/phishing_cond.asp, accessed 4 Jan 2006

[14] Microsoft, Consumer Awareness Page on Phishing http://www.microsoft.com/athome/security/email/phishing.mspx, accessed 6 Jan 2006

[15] Emigh A., Online Identity Theft: Phishing Technology, Chokepoints, and Countermeasures. Radix Labs. 3 Oct, 2005.

[16] Jagatic T., Johnson N., Jakobsson M., and Menczer F., "Social Phishing", *Communications of ACM*, to appear, http://www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf, accessed 3 Jan 2006

[17] Mail Frontier. *Phishing IQ*, http://www.mailfrontier.com, accessed 3 Nov 2005

[18] Horgan D.,."The Phishing Phleet" Courant.com. http://blogs.courant.com/travel_columnists_horgan/2005/11/the_phishing_ph.html, accessed 2 Dec 2005

[19] Brandt A., "Phishing Anxiety May Make You Miss Messages" *PCWORLD*. October 2005: 34

[20] IU Phishing Research, http://www.indiana.edu/~phishing/, accessed 6 Jan 2006

[21] CNETNews.com, "Browser Phishing Flaw Could Hook Users", http://news.zdnet.com/2100-1009_22-5484315.html, accessed 15 Dec 2005

[22] Werner, Laurie. "Redefining Computer Literacy in the Age of Ubiquitous Computing." Proc. ACM SIGITE 05, 95-99, 2005

[23] Anti-Phishing Working Group, "Phishing Activity Trends Report", http://www.antiphishing.org/reports/apwg_report_DEC2005_FINAL.pdf, accessed 20 March 2006

[24] Korea Internet Security Center, "Korea Phishing Activity Trends Report", http://www.antiphishing.org/reports/200601_KoreaPhishingReport_Jan2006.pdf, accessed 20 March 2006