

Phishing @ IU

BY JONATHAN BAUMANN

Based upon research by Markus Jakobsson and the Phishing @ IU team

Team:

The team is comprised of the following individuals (taken from the Phishing @ IU website):

- Markus Jakobsson
- Steven Myers
- L. Jean Camp
- Minaxi Gupta
- Filippo Menczer
- Sid Stamm
- Jacob Ratkiewicz
- Ruj Akavipat
- Tom N. Jagatic
- Nathaniel A. Johnson
- Virgil Griffith

Goals:

The goal of this research was to better understand phishing¹ and the attacks phishers² use in order to prevent identity theft³. It was completed through a variety of mediums, to diversify the results, and gain a broader image of the threat both immediate and future. The immediate result was a study of how effective a phishing expedition could be, and what the ramifications of that would be. Finally, by identifying new way to phish before attackers learn them, prevention would be much easier.

Previously Known Results:

This research was undertaken with the knowledge that phishing was becoming increasingly popular and successful. Sites such as eBay.com had been the targets of many phishing campaigns, and identity theft was on the rise. Throw previous studies, and the knowledge of the larger community, authentication techniques such as Mother's Maiden Name⁴ were known to be insecure, and the subsequent studies utilized those results to test just how effective an attack could be.

1. Phishing - Utilizing

2. Phishers - People that go on Phishing expeditions, explained above, in order to exploit the data.

3. Identity Theft - Taking identifying information about someone and using it to exploit his monetary or physical resources.

4. Mother's Maiden Name (MMN)- "Secure" scheme to identify a person using his mother's maiden name. Also the name of a study done by Markus Jakobsson on the same subject.

Markus and the Phishing @ IU team started by listing what people already know about Phishing and how it works. Three techniques are common today:

- Spamming⁵
- Spoofing⁶
- Password Reuse⁷

The easiest is spamming, where the phisher, or common spammer in this case, just sends out a lot of email and hopes that someone clicks on the link. While this is “super-liminal”, as the Simpsons would put it, it can pay off due to the small cost of sending out bulk spam, and the high payoff for valid emails if people respond. This technique is the easiest to combat in the present day because of the proliferation of anti-spam and spam-blocking software.

In order to get more spam by, spammers then turned to spoofing, where they assume the IP or domain name of a valid entity, such as a bank, or software company. This is harder to verify, as Markus demonstrated, as a simple telnet session can hide the real information from a non-discerning individual⁸.

By telnetting to port 25 on a mail server, you can submit mail by hand. Simply filling in different information in the from field will effectively spoof your identity from most users. Write that up as a script, and you can do it thousands of times a minute. Assuming the spam blocking software only looks at the From: field, and not the real IP, this works perfectly.

I happen to know from personal research along with my father that a big ISP as recently as two years ago was completely vulnerable to this attack, and had no plans to prevent it. They checked only that the domain name in the From: field ended with “@ISP-name.com” and never did reverse lookup to make sure the IP was from a subscriber, a simple task.

The final trick, is password reuse. This can occur after a successful phishing expedition that gathers a username/password combo, or after cracking a database that yields hundreds if not thousands of said combinations. The phisher would then take this list and run it against a number of known sites, such as eBay.com and major banks, to, hopefully, successfully login, and gather information about the person for identification theft, or to make purchases with their money. This is also a very effective tactic, as most non-discerning individuals will use the same username/password combinations for many sites, to make it easy to remember.

In phishing, there are even more types of attacks. The one most commonly seen today is the dumb, “Click me” approach, which relies on the hope that the target will click a link they do not know, or is poorly disguised. Markus suggests this gets approximately 3% of the Americans who get it. A disturbing thought, since a spambot could send out a hundred thousand emails from one computer without using anywhere near a noticeable amount of any resource. If 3% responded, that’s 3000 email addresses harvested, and if only 3% of that 3000 actually enter valuable data, you still have 90 username and password combinations. Now factor in that it is not unreasonable for a botnet (spam or trojan) to have 100,000 computers, you could gather 9 million username/password combinations, and 3 BILLION valid email addresses. That would be worth quite a lot on the underground market.

It gets worse, when the team’s studies showed that a context aware attack⁹ could yield a 35%-50% return. On top of the size of the number gathered, the data gathered would be more valuable, as it would personally identify a person to their email. This attack scheme is detailed below, and has been part of the work the team at IU has done recently.

5. Spamming - Sending out massive amounts of unsolicited email in the hopes of garnering data to exploit, be it for identity theft or just valid emails addresses to sell to other spammers and phishers.

6. Spoofing - The act of hiding one’s identifying information and putting in data that appears to be someone else.

7. Password Reuse - Website login attack tool that utilizes the knowledge that most people reuse username/password combinations.

8. Non-Discerning Individual (NDI)- The average user with little working knowledge of the Internet, to whom I consider this information to be most beneficial.

9. Context Aware Attack - An attack that takes information from some medium or exchange in order to identify someone personally, or who they do business with with whom he does business.

Work Completed So Far:

While at IU, the team has done a few different experiments. The MMN study was done to prove that a MMN approach to security can be easily beaten. This experiment took birth certificates and marriage licenses, available off the web from Texas, and used location data to fairly accurately guess a maiden name. This information could be used along with information below to mount an effective phishing campaign, because the NDI would not realize the information is so readily available.

The team has also examined the emails of banks and popular websites, in order to accurately pass themselves off as the group they are targeting. They used a popular website, <http://thefacebook.com>, to create a phishing expedition which yielded a high percentage of people giving up information.

“About 70% of recipients fell victim to the attacks using contextual information from social networks; this is an increase by a factor of 23 compared to known phishing attacks, and by a factor of four compared to the case where the sender is unknown but appears to be in the same domain as the victim”

From the blog for the Phishing @ IU team.

This experiment shows the thin ethical line the team has had to skirt while doing these projects. While they would like to test these attacks in a live environment, they can't actually gather the data as a real attacker would. That would be a major ethical breach of IU resources, and as such they have limited themselves to merely testing their data against servers to verify if the logins are correct, then deleting it. All that is saved is the raw data about how many fell for it and responded accurately. As Markus stressed to us in the presentation, this is a hard line to skirt, and just the one phishing expedition which did not necessarily violate any morals created many hard feelings and even led to a Slashdot¹⁰ article on the study.

Next, the team put together a website to show how popular scripting languages, in this case javascript, can be exploited to view your cache, and gather personal data about the viewer. This was termed “Context Aware Phishing” by Markus Jakobsson, in the pdf of background, given below. It reads the cache stored locally on the target machine and will tailor the attack to something it finds. For instance, the website could recognize a known bank in the target's cache, and present what appears to be the users own bank, when in fact, it is a fake page designed to trap a username and password, and pass the user onto the bank's real site. Or, as he presented in his presentation to us, tailor an email that would utilize all of the above to capture an even higher percentage. He speculated that perhaps as high as 90% would fall for the following scheme.

1. Use spam from many accounts to send people links to a site that will use the javascript trick to obtain information on who they bank with.
2. Take the people that responded, and send mail that's supposedly from a spouse, or family member, using a fake URL to get login information, and ask for more information. Here the phisher could ask for MMN, Social Security Number, or any information the phisher wanted.
3. It's important at this step to reassure the user that this site is secure, so a generic “we protect your identity” message is recommended.
4. After the above, the attacker will have a login, password, SSN, and MMN for someone. These will be used by many sites as the only authentication.

To prevent this problem, Markus presented a few possibilities. The first idea is that a bank's logo or other identifying picture could be checked against a database. It would have to originate from a known bank server in order to be valid. However, this would not stop someone from sending an email that links to the bank's logo. To prevent that, the server would have to make sure any computer serving up the image was authenticated, and allowed to. Many image hosting companies online do this already, presenting a static “this is copied” image instead of the linked-to image, and would not be hard to implement.

10. Slashdot - A tech news website <http://www.slashdot.org> which allows discussion of news.

Also, the email itself could be verified using a sort of PKI¹¹. The From: field would be checked against the real originating address, and in the event of a discrepancy, alert the appropriate people. The appropriate people being, in this case, the user, and some sort of electronic fraud support at the bank.

Anticipated Results:

So where does the team go from here? Markus then presented a few possibilities of attacks that might spring up. Attackers could set up many Paypal accounts to funnel money from legitimate sources (say, software developers that ask for simple 5 dollar donations) through a network of many accounts of really small amounts, to an end source that could be anything, terrorism being the biggest “threat” right now.

The end goals of this team are the following (Taken from Markus Jakobsson’s presentation):

1. Understand how things go wrong today.
 - We must understand how attacks are being carried out today in an effort to curb them.
2. Anticipate what will go wrong tomorrow!
 - Someone has to take the next step and do what the team did with facebook. If we don’t think of new exploits, the attackers will.
3. Fix technical problems.
 - Easy exploits such as browser vulnerabilities need to be patched, so that the easy attacks are unavailable to attackers.
4. Develop better authentication techniques and techniques to communicate warnings and reassurance to the users.
 - The present systems are too vulnerable, see the above bullet, and allow easy attacks.
 - New systems should allow for much more secure authentication, definitely not SSN or MMN so that identity theft is not a problem.

Further Readings:

- The website and blog for the Phishing @ IU experiments.
<http://www.indiana.edu/~phishing/>
<http://www.indiana.edu/~phishing/blog/>
- A website put up to demonstrate how simple javascript can be used to phish.
www.browser-recon.info
- The pdf file of the background of phishing trips.
www.markus-jakobsson.com/papers/phishing_jakobsson.pdf
- Studies on finding Mother’s Maiden name.
www.markus-jakobsson.com/papers/mmn.pdf

11. PKI - Public Key Infrastructure. A way for third-party software to verify the identity of a user.

Griffith and Jakobsson, "Messin' with Texas: Deriving Mother's Maiden Names Using Public Records."

- The speaker's website.

<http://www.informatics.indiana.edu/markus/>

- Slashdot news thread about this study.

<http://it.slashdot.org/it/05/04/26/1959256.shtml?tid=172&tid=146&tid=95>

Thanks:

I'd like to thank Markus Jakobsson for his wonderful talk, as I learned a lot, and reinforced what I already knew, as well as the entire Phishing @ IU team, for their work to help keep us safe. I'd also like to thank George Springer for the chance to take this course, and it was great fun to hear what is going on at my own university.