Phishing – What to Expect Next

Markus Jakobsson
School of Informatics
Indiana University
Bloomington, IN 47406, USA
markus@indiana.edu

Abstract

We argue that phishing is not an *event*, but a *tool*. Rather it is a combination of technical and psychological methods used to steal restricted information. Phishing is not limited to a small set of poorly spelled email messages demanding login to an irrelevant online bank. It is a manifestation of human desire to deceive. Phishing tactics will adapt to new situations – whether technical or social. As it evolves and matures, the phishing industry will employ bright and specialized minds driven to improve attack yields. New principles will be developed and known principles will be refined. The battle against phishing does not end when the first effective anti-phishing product becomes ubiquitous – the scenery will simply change. This article explores how phishing threats may change in future scenarios.

1 Introduction

Due to the large number of technically uninformed and unprotected users, today's phishers obtain satisfactory yields and profits at minimal risk and effort. While people agree that the direct and indirect costs of phishing are significant, we lack a precise account of its damages. In fact, official numbers such as [12], may *both* underestimate and overestimate the costs to society; the former due to embarrassment among victims or a failure to realize that they were taken advantage of; the latter due to popular misunderstandings of what exactly counts as identity theft, phishing, or simply common fraud. Moreover, surveys such as [13] may overestimate just how suspicious users are: users who *know* that they are being tested for their ability to recognize phishing attempts are more likely to classify everything as a phishing attempt – which seems to be confirmed by the low success rate of classifying legitimate emails in the study in question. Nevertheless, the ease of defrauding people has permitted phishers to make a good profit simply using copycat attacks; the most common of which is a request for credentials in order for the victim's bank account to remain in good standing. As technical tools – most prominently spam filters and browser

toolbars – are likely to decrease the yield of such attacks onwards, we believe that phishers will respond with an increased degree of sophistication.

In this article, we will describe four possible and independent directions in which phishing may evolve; for an excellent survey of the *current* threats, we refer to [1]. The first of these directions relate to a likely increase in the use of contextual information, corresponding to a type of attack that is sometimes referred to as "spear phishing" [4, 7]. Attacks of this type will fuel privacy intrusions in the form of data mining, whether from public or private databases, the browsers of potential victims, or behavioral information accessible to the phishers. A second direction describes ways for attackers to avoid that their messages are classified as phishing messages – whether by humans or machines. A third direction correspond to how the attack is delivered to the victim. While spoofed email is the delivery mechanism of choice today, this may change drastically if spam filters become increasingly successful. Phishers may utilize other media, such as voicemail, in order to deliver the come-ons; alternatively, they may rely on social propagation of malware to corrupt machines and steal information. Finally, a fourth direction is an expansion of the types of information phishers aim to obtain. For example, phishing attacks that aim to capture credentials used for online gaming are becoming increasingly common. We believe that highly targeted phishing attacks may also be used for personal and industrial espionage, and potentially also for collection of military intelligence. To evaluate such threats and anticipate trends, a good understanding of who may want to obtain what information may be valuable.

2 Using Context for Spear Phishing

Attack lifetime decreases as the set of targets grow, while context improves yield without decreasing attack lifetime. More particularly, we believe that the risk of *takedown* (the removal of sites impersonating legitimate sites) becomes problematic to phishers in situations where the phisher has to offset a relatively low response rate by targeting a huge number of potential victims. Namely, as the set of targeted users grows, the likelihood that one of them (whether a real user or a honeypot [5]) will alert authorities will grow as well, which in turn decreases the expected time until takedown. This caps the number of credentials that can be collected from gullible users. Phishers who can target a smaller set of victims in a way that is more likely to succeed (on a per-victim basis) can limit the threat of takedown and improve their profits.

In [8, 10], for example, it is shown how an attacker can determine what bank a given user has a relationship with, simply by probing the user's browser history. This, in turn, allows the phisher to send victims emails appearing to come from their *actual* banks as opposed to from arbitrary banks. While browser caches and histories constitute a gold-mine of private data, they are far from the only source of personal information. For example, it was shown in [3] that one can derive mothers maiden names from databases containing marriage and birth records – such databases are by law required to be accessible

to the public. Imagine now the combination of these two attacks: The phisher first determines whom a given victim banks with, and then sends the user a message appearing to come from this bank. The bait message contains the client's mothers maiden name – supposedly for purposes of authentication and phishing prevention. The message might then request that the user changes his or her PIN or password, after having correctly authenticated with his or her current credentials. This well prepared attack collects access credentials with a much better success rate than current attacks. However, as a critical reader might point out, this combination attack requires the phisher to determine the victim's full name, and subsequently to lure the victim to a site where his or her banking affiliation could be extracted from the browser. This may be easier than it may first seem. An attacker may determine names, email addresses and personal relationships from social networks and then spoof emails from a friend of a victim, suggesting to the recipient that he visits a given site (where his browser history and cache will be mined.) As was demonstrated in [6], such an attack has a success probability of over 80%, where an attack is said to succeed if the victim followed a given URL. Moreover, the same experiment was shown to succeed in obtaining the victim's password for over 70% of the subjects. These are just a few examples of attacks in which contextual information could be used to increase the yield – whether the contextual information was obtained from browsers, public databases, or websites.

One way phishers are likely to increase their yield is by collecting information in one domain – such as a domain with geographic relevance – and use it in another domain – such as one with a logical structure, e.g., based on email addresses, IP addresses, etc. As an example of an attack of this type, consider a situation in which an attacker mines records about political contributions (these are by law public in the United States, and can be accessed online). Then, the attacker creates a huge database of triplets `<name, address, email address>`; this can be done by mining publicly available web pages, and looking up names in an online phone book. If this is done in a manner that takes the geographic location of the entity hosting the web page in question, that allows for a focused search of phone numbers and – importantly – addresses. (The approximate geographic location can be inferred in many cases, e.g., for corporations and universities, and from information posted in social networks.) Finally, the attacker would send a spoofed email to a potential victim for which he has found the record of a prior campaign contribution, asking the victim to contribute again by following a given link. At the site where the victim will be taken, he or she will be asked for information, such as social security number to allow for a receipt to be generated, a credit card number, etc. Furthermore, the attacker may offer the victim to perform a bank transfer directly from his or her account, or to use his or her PayPal account to make a quick contribution. In the latter case, the victim would be taken to a site that looks like PayPal, but which is controlled by the attacker, and which would harvest the login credentials of the victim.

It should be noted that context aware attacks not only rely on *inferring* the context of an intended victim, but may also try to *impose* a context on the same.

For example, by sending the intended victim a first sequence of messages (e.g., confirmations of claimed online transactions, such as shipping notifications or PayPal payment confirmations), the victim can be primed to later believe in a spoofed message from its bank, in which the bank claims that there are suspect transactions the customer needs to pay attention to (after having logged in at the provided URL).

3 Avoiding Detection

With the development of various techniques to detect attempts to spoof messages, phishers are likely to increasingly rely on valid domains that they register and control, and for which they do not have to spoof messages – thereby avoid machine detection of the attack. At the same time, the choice of domains must be such that it also avoids that the human recipient of the message detects that it is fraudulent. We believe this will the use of two classes of domains: what one may call *doppelganger domains* and what we will refer to as *colleague domains*. The former class corresponds to cousin-domains (such as `www.chase-rewards.com`) and domains intended for use with subdomains (e.g., `www. ebay.secure-connection.com`); attacks using this type of domains are becoming increasingly common already. The second class – not yet witnessed in attacks in the wild as of yet – corresponds to services claiming a relationship or association with the targeted domain. Both of these classes rely on deception of users, which in turn relies on psychological observations of user behavior. Recent work on this topic, e.g., [11, 9, 15, 16, 17] indicate that deceptive practices can have notable impact.

Example: A doppelganger domain attack. Phishers may send messages to consumers, purportedly from a given bank, in which they describe that the new security policies requires six-digit PINs instead of four-digit PINs. The phisher may then request that in order to comply with this change, the consumer chooses a new (and compliant) PIN after first having authenticated themselves using the old PIN. Without the common threat of a time-limit to respond (as is often seen in current phishing messages), consumers may be naturally enticed to act quickly after all, it is for their own good. In addition, though, the messages may speak of an imminent roll-over requirement for longer PINs for all stores and ATMs vague enough that the recipient will feel nervous about not making immediate updates. To avoid using URLs that tip consumers off that they are being attacked, phishers are likely to want to use domains that match the request. Domains such as `PIN-update.com` may offer such advantages, especially if used with a subdomain matching the name of the bank that the email purportedly comes from.

Example: A colleague domain attack. Consider a class action suit in which an attorney-sounding entity promises the user some desirable but not unbelievable benefit given a possible wrongdoing of some entity – whether a

financial service provider or not. In the case of purported litigation against a bank, the authentication would serve as evidence that the recipient is a rightful beneficiary; in the case wherein a third party would be the defendant, the authentication would serve the purpose of allowing for direct deposit of the promised amount. Phishers could use existing class action suits as a foundation for the scam, thereby potentially benefiting from a basic familiarity with the existence of the suit. For example, the recent litigation against Netflix could be used as the ground for a phishing email; while this real suit promised no financial benefits (but only service benefits), the email could claim that this was ruled as unfair, and the new and improved suit permits for financial reimbursement. Note that the approach in which no bank is specified helps a phisher, as he only would have to identify the correct bank of the victim, but only the right third party provider selecting Netflix, eBay or AT&T would make for a believable scam for a large fraction of recipients. Phishers may use any available domain that evoke the image of a law firm. Preemptive registration of suitable domains poses a difficulty in this scenario.

A second and orthogonal approach that phishers may start to use to bypass filters is that of obfuscation of content. This, too, has already started to happen, although not yet for spam messages aimed at phishing, but only generic spam. Two techniques are commonplace: the use of images that look like text, potentially interspersed with actual text; and the use of Cascading Style Sheets to arrange the displayed information in a way that is hard to anticipate unless it is rendered.

4 Delivery

Current phishing attacks are almost exclusively mounted by email. There is no reason why attackers will not branch out to take advantage of other delivery mechanisms, including instant messaging, telephony and rogue captive portals. For example, if an attacker can place massive number of phone calls (with recorded messages) at a very low cost, and with a low risk of being traced, then this will start to become a viable delivery mechanism. This may involve the compromise of a router or a VoIP supernode. Alternatively, attackers may simply place standard phone calls using the same techniques used by telemarketers. The attackers may then request the targeted victims to call their bank to confirm some unusual transaction – whether real or fictional. The attacker may leave both the number of the local branch and a number he controls – at a time when the local branch is not going to be open. The automated message played to the user as he or she places the phone call may prompt the user to enter information from a check on his or her phone – thus allowing the attacker later to perform Automated Clearing House (ACH) transfers from the associated account. Alternatively, the user may be prompted to enter his or her ATM PIN on the keypad, or the currently displayed number on a login token. Such information could be immediately and automatically used by the attacker to gain access to the user's account.

As suggested by the above example, we do not believe phishing will be limited to the Internet. Consequently, concerned users and organizations will not be able to fully isolate themselves from attacks by disconnecting themselves from the Internet: *There is nowhere to hide!*

Another example of an avenue of delivery of the phishing attack is a rogue captive portal. Imagine, for example, a rogue node operating in an airport, and allowing anybody to establish a network connection – potentially for free. If the user would not rely on a Virtual Private Network (VPN) to protect the content delivery from or to himself or herself, then the rogue node may eavesdrop on traffic and perform content injection attacks. If the user relies on a spam filter that resides with his Mail Transfer Agent (MTA) or elsewhere on the network, then this content injection attack will succeed. Similarly, by performing the actions of a Domain Name Server (DNS), the rogue node would be able to mount a pharming attack on the user. This, combined with a man-in-the-middle attack, would allow the victim to believe he is protected by secured sessions, when in fact the attacker is the other endpoint to and from which the traffic is secured. Here, the captive portal could demand that the user would accept this certificate at the beginning of the session.

Yet another delivery mechanism is examplified by recent work [2] which shows a success rate of above 50% of an attack relying solely on social propagation of malware, in spite of requiring victims to accept self-signed certificates before corruption succeeds. In the study in question, entertaining short movies (that were executable and which *could have* contained malicious payload) were spread – not by the attacker, but by users who received the movie and passed it on to their friends. It is worth noting that such an attack does not have to rely on technical vulnerabilities or be platform-specific, in contrast to most malware.

Delivery mechanisms do not have to operate in a traditional electronic way. In particular, recent work [14] suggests a grave threat that has not yet been considered in the context of phishing. In his work, the modifications to embedded software in consumer routers is considered; these are to an increasing extent reprogammable Linux machines, and have large quantities of available flash memory space. This allows for an attacker to modify routers and resell them at a desirable price, thereby allowing for a new type of pharming (or DNS spoofing attack) by shortwired rerouting of select requests.

Thus, while almost all phishing attacks today would be prevented if all users were to use a perfect spam filter, such a device would not mark the end of phishing. In particular, if email spam filters at any point were to become so successful that the yield of phishing attacks is severely reduced, then the pressure to develop and deploy alternative delivery mechanisms will increase.

5 Who Will Phish and Why?

Let us think of phishing as a set of tools useful to obtain information from people who may not have in their best interests to share this information. The information does not have to be banking credentials (as it currently is with

almost no exception) but could be *anything*. For example, the information may be the plan for a political campaign; a requisition for troops or materials; the text for a patent application to be filed; the answers to a homework assignment; or confidential hiring discussions between employees. Accordingly, we can easily see that the potential attacker (and his resources and capabilities) will vary.

Among "professional" phishers, one can already see the trend of specialization and subcontracting of tasks. This may be done within one set group of people, or in a more ad-hoc manner. The latter is made possible by bulletin boards that are springing up with the goal of matching skills to needs – in this sense the job market within the niche of phishing is not very different from the more general job market. Ignoring the legality of the activity of phishing, and focusing on the tasks involved, it is clear that this is an activity very well suited for telecommuters. Consequently, and just as can be seen from current trends, attacks are typically mounted from low-wage countries, and in particular those with a highly computer-literate but underemployed workforce. Law enforcement in such countries may not prioritize efforts to discourage phishing.

If phishing efforts were to be taken up on a governmental level – whether as a technique to infiltrate or spy on other governments, or in order to perform corporate espionage – then it may become more important for phishers to hide their tracks, and make it possible to cause the trail to lead to other organizations in other countries. Again, given the very "transportable" nature of this threat, and the possibility to go through proxies, this is not impossible to imagine.

Finally, small-time phishers are going to be empowered by phishing toolboxes, which show evidence of being developed. There are currently automated tools for spamming, malware generation and control, and in particular, for rootkitting. It may not be long until these are bundled and made available to a wider base of would-be phishers.

Clearly, everybody whose infrastructure or activities rely on electronic transmission and storage of valuable information are at risk, where the risk increases with the degree and frequency of connectivity to public resources – the Internet, in particular. This turns phishing into an equalizer of power, a tool for asymmetric warfare – whether between governments, organizations or individuals. Furthermore, the stunningly low entrance costs to perform phishing (in particularly in comparison to the costs of avoiding the same!) could make phishing a tool in the hands of one-man terror organizations pursuing acts of aggression against corporations and governments. This aggression may involve getting access to valuable and proprietary information; gaining access rights to internal resources (including human capital); and being able to perform fundraising (or large-scale theft) to fund unrelated operations aimed at furthering the damage to the victim(s). In many cases, it may be enough for an attacker to compromise *one* node in a vast network in order to gain access to resources of the network. When the weakest link of the chain is not technology, but human agreeability and gullibility, this may be very severe, and emphasizes the need for internal firewalling and compartmentalization of information and resources.

References

[1] A. Emigh, "Online Identity Theft: Phishing Technology, Chokepoints and Countermeasures," Report of the US Department of Homeland Security - SRI International Identity Theft Technology Council, `www.anti-phishing.org/Phishing-dhs-report.pdf`, November, 2005.

[2] M. Gandhi, S. Stamm, M. Jakobsson, "Social Propagation of Malware," Manuscript in preparation. Demo available at `www.verybigad.com`.

[3] V. Griffith, and M. Jakobsson, "Messin' with Texas: Deriving Mother's Maiden Names Using Public Records," RSA CryptoBytes, Vol. 8, No. 1, 2006.

[4] B. Grow, "Spear-Phishers are sneaking in," BusinessWeek (07/11/05) No. 3942, p. 13

[5] The Honeynet Project and Research Alliance, "Know your enemy : Phishing. Behinds the scenees of phishing attacks," `www.honeynet.org/papers/phishing`, 2005.

[6] T. Jagatic, N. Johnson, M. Jakobsson and F. Menczer, "Social Phishing," To appear in the Communications of the ACM, 2006.

[7] M. Jakobsson, "Modeling and Preventing Phishing Attacks," Panel on Phishing at Financial Cryptography, 2005. `www.markus-jakobsson.com`

[8] M. Jakobsson, T. Jagatic, S. Stamm, "Phishing for Clues," `www.browser-recon.info`, Referenced May 2006.

[9] M. Jakobsson, J. Ratkiewicz, "Designing Ethical Phishing Experiments: A Study of (ROT13) rOnl Query Features," Proceedings of The 15th annual World Wide Web Conference (WWW2006), 2006

[10] M. Jakobsson, S. Stamm, "Invasive Browser Sniffing and Countermeasures," Proceedings of The 15th annual World Wide Web Conference (WWW2006), 2006

[11] S. Garfinkel, R. Miller, "Johnny 2: A User Test of Key Continuity Management with S/MIME and Outlook Express," presented at the Symposium on Usable Privacy and Security (SOUPS 2005), July 6-8, 2005, Pittsburgh, PA.

[12] A. Litan, "Phishing Attack Victims Likely Targets for Identity Theft," FT-22-8873, Gartner Research, 2004.

[13] Mailfrontier. Mailfrontier phishing IQ test expands to the U.K.: U.S. test reveals email users less savvy at identifying legitimate emails. `http://mailfrontier.com/press/press.phishingtest.expands.jsp`, March 2005

[14] Alex Tsow, "Phishing with Consumer Electronics – Malicious Home Routers," To appear in Models of Trust for the Web, a workshop at the 15th International World Wide Web Conference WWW2006, 2006.

[15] T. Whalen, K. M. Inkpen, "Gathering evidence: use of visual security cues in web browsers," ACM International Conference Proceeding Series; Vol. 112, Proceedings of the 2005 conference on Graphics interface, pp. 137 – 144, 2005

[16] A. Whitten and J. D. Tygar, "Why Johnny can't encrypt: A usability evaluation of PGP 5.0," 8th USENIX Security Symposium, 1999

[17] M. Wu, R. Miller, S.L. Garfinkel, "Do Security Toolbars Actually Prevent Phishing Attacks?," Conference On Human Factors In Computing Systems.