Vdetector: Attacking the Attacker towards Combating Phishing and Identity Thefts on the Internet

Olufade F. W. Onifade¹  Kolawole John, Adebayo²
University of Ibadan
Nigeria

Abstract

Information security remains a critical issue in internet adoption and diffusion studies. However, the security of information has been compromised on the internet due to attacks by cyber criminals and hackers.

This paper discusses the menace of identity theft, phishing and pharming. We give a brief history of phishing, some statistics about phishing attempts and its social impact. We also present a quick survey of existing solutions’ approaches and their strength and weaknesses. We observed that even with the solutions presented, attacks have been carried out successfully due to the vulnerabilities of the web application in use. We present a novel approach of identifying the vulnerability risks from websites visited by users. The proposed model, an optimized automated web application scanner tagged Vdetector will greatly complement existing solutions.

Keywords: Pharming, security, vishing, optimized web application scanner.

Introduction

The emergence of internet has turned out to be one of the greatest benefits of technology to mankind, with its alarming growth within a very short time, it continues to transform the way we live, relate and see things around us. With the internet the world is gradually snowballing into a global village, information flying over the air in seconds, online real-time business transactions, social networking and many more, and yet, this is just the beginning; the cyberspace has neither limits nor boundaries. Each day brings to us the numerous opportunities that abound with the use of this technology and of course, it’s many challenges which must be tackled head on. One of the many challenges posed by the use of the internet is that of information security. Information security means protecting information from unauthorized access, use, disclosure, disruption, modification and denial of service (Wikipedia, 2010). As some people are exploiting the opportunities offer by the internet for the right cause, many people see it as a means of propagating their obnoxious act. They take advantage of the many unsuspecting innocent internet users, insecure web applications vulnerable to attacks and careless corporate organizations as their targets. Today, cybercrime has become the fastest growing form of crime (Phair, 2007) and it mostly involves unauthorized access to personal information of someone else.
A major challenge to information security on the internet is identity theft, phishing and pharming. As part of the growing internet community, none can be said to be immune to the threats these posed, most especially with the increase adoption of e-commerce, online banking and business transactions by corporate organizations and governments. The earlier and better we know the reality of the threat, the more equipped we will be to tackle it and the safer the cyberspace will be.

Presently, no single anti-phishing solution can adequately guaranty the users’ information security, more so, a sizable number of these attacks have been possible due to vulnerabilities of the web application in use even with the presence of an anti-phishing solution. Attacks on vulnerabilities in web applications began appearing in the mid-1990s. Attacks are usually based on fault injection, which exploits vulnerabilities in a web application's syntax and semantics (Verbjoen and Klingsheim, 2009). Using a standard browser and basic knowledge of HTTP and HTML, an attacker attempts a particular exploit by automatically varying a Uniform Resource Locator (URL) link, which in turn could trigger an exploit such as SQL injection or cross-site scripting. Web application security vulnerabilities usually stem from programming errors from a web application programming language (like JavaScript, JSP, PHP, etc.), code library, design pattern and architecture. The use of many techniques to ward off these attacks has been encouraged.

In this paper, we briefly discuss identity theft, phishing and pharming, and the pros and cons of the existing solutions available. For details about phishing and some of its statistics, the reader can consult (David, 2005; Knickerbocker, 2008; Antonio and Xavier, 2009; Price, 2009; Aslam, Lei and Cliff, 2009) for extensive reading.

The concluding part of the paper is divided into four sections. The section that follows is a review of identity theft and phishing, the next section contains our research methodology. Section three contains our observations and the last section is the summary and conclusion.

**Identity Theft and Information Security**

Contemporary insecurity now extends well beyond individual bodies, national borders and material infrastructures. It includes, as well, the electronic data required for people to function in societies. Identity theft, which is perceived as being one of the greatest threats to people’s critical data, now represents the largest category of fraud-related complaints not only in the developed countries of America and Europe but also in Africa.

Identity theft can be regarded as theft of personal information such as somebody’s credit card details etc. According to the Federal Trade Commission (FTC), roughly 9 million US adults are victims of identity theft each year (FTC, 2008). These include cases of credit card theft, illegal wire transfers, Internet scams, phone and utilities fraud and theft of business data.

The wave of identity theft has virtually reached every nook and cranny of the cyberspace, the developing countries of Africa not being left behind. For instance in Nigeria, when banks and other organizations began to be computerized over a decade ago, little did they know that they were setting the pace in computer crimes age. Essentially, computers and the Internet in banks facilitate records of customers’ transactions and transfer of monetary values, through it, online
real-time transactions have been made possible. With the computer and Internet facilities put in place, customers communicate directly with their banks to pay bills, transfer funds, inquire about account balances, and perform all sorts of services offered by such banks.

This development undoubtedly improved their business processes and efficiency but also opened an avenue for the criminally-minded to illegally gain access to someone else banking details including security keys like passwords and thus posing as a legal customer to make illegal transactions.

The political and economic context for identity theft is one of post-industrialization, particularly within regions that have lost their stable industries and have gained a host of social instabilities, such as the manufacture and use of hard drugs. The second type of environment where identity theft thrives is in regions of heightened economic polarization and socio-spatial segregation, such as large cities, these environments are readily obtainable in Africa thus making Africa to stand out as a potential breeding ground for phishers. The technological systems that facilitate identity theft are those of large-scale databases that are poorly regulated and vulnerable to attack, also web applications that are not robust and fault-tolerant are also an agent.

It is thus obvious that the two settings that brew identity theft are typical of African countries. Today, several cases of internet scam by fraudsters have been reported in Africa, most especially Nigeria, thus necessitating the establishment of government agency like economic and financial crime commission (EFCC) which has fairly tracked some of these perpetrators. While some of these attacks have been directed towards individuals through mass phishing emails, notable parts of it have been directed towards corporate organizations including the Central Bank of Nigeria (CBN). The fact is that the advent of the internet has only provided another means and opportunity to criminals. In the section that follows, we discuss phishing and pharming which are two great forms of identity theft and solutions already presented to prevent them.

**Phishing and Pharming**

Phishing is no more a new concept, the first phishing attempt involving AOL attempt was reported some years ago, with the rapid growth of the internet, more cases are being reported even in Africa. Few years ago many financial institutions started offering internet banking and online payment systems making them to be prime victims of phishing attacks. It is estimated that the annual direct financial loss in the Africa because of phishing attacks is millions of dollar. Reliable indications of indirect losses are not available but these may be much higher (Wikipedia, 2010). InJune2006 the first phishing attack that involved cross site scripting was carried out (Mutton, 2006). The attackers sent a specially crafted (Uniform resource locator) URL to several PayPal users, exploiting PayPal's lack of input validation, the users that visited this URL encountered a page on the legitimate paypal.com domain saying that their account had been suspended due to abuse. When a victim clicked the button on this page, he was redirected to a login page on a host controlled by the attacker. Note that the difference with traditional phishing attacks is the failure message on the domain of PayPal. Hence, it was part of a legitimate SSL (Secure socket layer) connection with PayPal. This form of phishing is known as pharming.
In February 2010, some attackers craftily cloned the (Central Bank of Nigeria) CBN site, they periodically sent email to bank customers requesting them to update their records with the CBN for a new exercise being carried out to create a database of all the banks customers in Nigeria, the victims were to submit their various account numbers and ATM pins before a deadline date. Victims who clicked on the link in the email were taken to a clone of the CBN’s site, thereby posing as the legal CBN site. During the last five years phishing has been growing rapidly, with an estimate citation of approximately 8 million daily phishing attempts all over the world (Donnel, 2009).

According to (Jakobsson, 2006) phishing is a form of social engineering in which an attacker, attempts to fraudulently retrieve legitimate users' confidential or sensitive credentials by mimicking electronic communications from a trustworthy or public organization in an automatic fashion. It as a form of internet scam in which the attackers try to trick consumers into divulging sensitive personal information usually through fraudulent E-mail and websites that impersonate both legitimate E-mail and websites (Tally, Thomas and Vanvleck, 2004). Also, (Clayton, 2004) defined it as an attack in which victims are lured by official looking email to a fraudulent website that appears to be that of a legitimate service provider. In this paper we will define phishing as the process in which an attacker/enemy attempts to steal and exploit confidential information by leading a human being into believing to be involved in an electronic transaction with a legitimate party while actually the attacker exerts influence on, and controls the transaction.

According to (Liu, Deng, Huang and Fu, 2001) a phishing attack is generally characterized by the ‘lure’ which involves an enticement delivered through email containing a message encouraging the recipient to follow an included hypertext link. The hyperlink often masks a spoofed uniform resource locator (URL) of a legitimate website. The ‘hook’ which is a malicious website designed to look and feel like a legitimate website. The authentic-looking website asks the victim to disclose privacy-related information, such as user identification and password. Most times, the hook is an obfuscated URL that is very close to one the victim finds legitimate and is really a site under the attacker’s control. And finally, the ‘catch’ which is when the originator of the phishing message uses the information collected from the hook to masquerade as the victim and conduct illegal financial transactions. Unfortunately, in Africa, people still succumb to these attacks. Below, we take a look at some common forms of phishing.

**Spear phishing** is a directed type of attack that targets specific groups of people. With this attack, the phisher sends an email to group of people who are often in the same organization. Frequently, the phishing email is spoofed to appear to be from an actual member of the group.

A more subtle form of phishing is called pharming; it involves implanting fraudulent Domain Name System (DNS) records to provide users with IP addresses to phishing websites when they request the IP addresses of legitimate sites. **Pharming** is usually accomplished by poisoning the cache on a compromised DNS server, or by re-configuring a victim's DNS server list to point to a DNS server set up by the pharmer. Re-configuring the victim's DNS list can be done using viruses, Trojan horses or by compromising the network's DRCP server (like in drive-by pharming, which exploits un-configured wireless routers (Jakobsson et al, 2001).
Recently, there has also been a rise in **vishing**, where the personal information is collected over the phone and not through a fraudulent website. Similar to standard phishing, vishing uses mass e-mails to compel victims to call a phone number and give their account information to an operator which is in fact an identity thief. The rise in vishing may be related to the expansion of IP telephony, providing voice lines which can be set up and run with minimal exposure for the visher. **Man in the Middle (MITM)** is a form of active eavesdropping in which the attacker makes independent connections with the victims and relays messages between them, making them believe that they are talking directly to each other over a private connection when in fact the entire conversation is controlled by the attacker. Today, it is a reality in Africa; we are witnessing increase in all forms of cybercrimes every time mostly fuelled by the high rate of un-employment among youths and widespread poverty.

**Research Methodology**

The existing solutions to pharming can be categorized to three, i.e. the **password management based**, **heuristic based** and **website credential systems**. Most anti-phishing systems usually incorporate strategies from several of these groups to counterbalance each group's relative weaknesses. The most popular tools use a combination of phishing detection and website credentials to alert the user to potential phishing threats during web browsing. Increasingly these tools have become an integral part of the web browser.

**Password Management** involves automating the password entry mechanism through the use of password managers. Users expose their passwords by submitting them to websites that appear legitimate. Password managers prevent accidental exposures by utilizing technical identification procedures that are not influenced by human perception. Password managers take the user out of the password submission process and so limit the possibility of submission to an unauthorized server. They provide control on the release of passwords by managing password records based on the Fully Qualified Domain Name (FQDN) of the servers the password was bound to. While phishing URLs can be constructed to appear to be related to the mimicked site, the mechanical matchup between the Fully Qualified Domain Name (FQDN) associated with a password in a password management system is immune to these deceptions. Obfuscation is also used under this scheme; it involves use of cryptographic hashes. It combines hash of a user-selected password and a piece of data unique to the legitimate site, often the Domain Name System (DNS) domain name, as the password to the site. When the user authenticates, the obfuscators recreate the hash based on the user's password and the data from the site being submitted to. If the site is a phishing site the password submitted will not be the password to the legitimate site because the seed from the website used in creating the password is different from the legitimate site, producing a different result. The demerit is that due to assumptions about the validity of Domain Name System (DNS) entries by the managers, users are still susceptible to attack. Reliance on the managers by users for passwords also poses some problems.

**Heuristic-based Detection** identifies phishing by looking for phishing characteristics in a website's HTML or in the text of e-mail messages. Most heuristics are simplistic and check for common phishing tactics such as misleading URLs or hyperlinks to IP addresses. Spoofguard (Chou, Ledesma, Teraguchi, Boneh, Mitche, 2009) and numerous browser plug-ins (Calling ID
ltd, 2007; Mozilla foundation, 2007; Microsoft Corp, 2007) use machine-readable properties of a website and compare them with the websites in the browsers history to determine whether a site is legitimate (previously visited sites assumed to be more legitimate). Similar heuristic-based schemes are used to identify phishing e-mails that lead users to fraudulent sites through deceptive hyperlinks (Fette and Sadeh, 2007; Hohenberger, Adida, Rivest, 2005). Contextual analysis (Zhang, Hong, Cranor, 2007) and comparisons of the structure of rendered web pages to pre-established signatures (Deng et al, 2006; Rosiello, Kirda, Kruegel, Ferrandi, 2007) are more advanced, detection techniques proposed. As good as it is, it is incapable of handling pharming due to its assumption on DNS integrity, and also, it needs constant revision of the heuristics due to changes in phishing techniques adopted by attackers (Antonio et al, 2009).

**Website Identification** establishes positive website identification before allowing users to enter their authentication credentials. It uses technological means (such as certificates) which positively identifies the site. **Cueing** prevents phishing by using cues to signal their legitimacy to the user. Like others, it is also not efficient. **Verified credentials** automatically discriminate between legitimate sites and phishing sites, usually through the use of a trusted third party. It aims to enable legitimate websites to prove their own identities in a way phishing sites could not mimic, thereby detecting phishing sites by their absence of credentials. Public Key Infrastructure (PKI) is the most widespread example of a credential system to identify legitimate websites. When a user establishes a secure connection with a website, typically through the Secure Socket Layer (SSL) protocol, the user can verify the public key certificate of the website in order to make sure the website is legitimate. However, as effective as it may seem, recent surveys established that most users do not understand how a website is verified using certificates or understand the connection between encrypted communications and website security. **Restriction Lists** involves use of whitelists and blacklists with blacklists populated through reports of phishing activity. The lag before a phishing site is detected and added to a blacklist limits the usefulness of blacklists against rapidly emerging threats. Studies have shown that blacklists from Google and Microsoft only list 55-65% of the active phishing sites reported in another well-known blacklist (Kruegel et al, 2007). Most of the existing solutions rely partially on users’ vigilance and they mostly focus on detection rather than prevention.

From our research, we found out that most of the existing solutions are not highly efficient; in fact, no single technique can give a success rate of 90%. Furthermore, phishing was still possible even with the use of these solutions. We found that most of the successful attacks when some of these solutions were in use were possible because of some vulnerabilities in the web application in use, especially in cases where the websites are fairly large and whose pages contains dynamic contents. Most of these attacks exploit vulnerability in syntax and semantics of the language used. We classify the vulnerabilities to be

- **Authentication** i.e. stealing user identities
- **Authorization** i.e. illegal access to application.
- **Client side attacks** i.e. illegal execution of foreign codes.
- **Command execution** i.e. hijacking web application’s control e.g. SQL injections
- **Logical attacks** which involves interference with application usage. For a more and full list of possible vulnerabilities, see (www.owasp.org).
The Vdetector model

Detection of some of these vulnerabilities may involve source code inspection and/or on-site penetration testing. We assumed that a web application is vulnerable if a script or an html code could be inserted into a dynamic page or there is possibility of manipulating the structure of the SQL query of the dynamic page.

We then propose our model as a two way automated scanning tool called Vdetector. Vdetector does not have access to any web application's source code; it detects vulnerabilities by simulating and performing likely attacks on the target application. Time required for scanning however varies i.e. performing a broad simulated attack on an application takes significantly longer than performing a network vulnerability scan against a single IP. A major requirement for the Vdetector is comprehensive coverage of the target application's functionality. Incomplete coverage will cause it to overlook existing vulnerabilities. Running inside a browser, it crawls the web for input vectors used by the server side script, it then test for vulnerabilities of the site by using some set of input data against the dynamic pages, if the pages are susceptible then the site may be vulnerable and the user is notified. The Vdetector works thus, when a user enters the URL of a website to be visited, the Vdetector automatically makes a virtual connection with the host server where the page resides, the host server is obviously known from the URL supplied by the user which the DNS transform to an IP address. The Vdetector then inserts some illegal characters/code into the script (JavaScript, PHP) of the website; the aim is to know if the generated script will also contain the maliciously inserted codes after manipulation. Since a vulnerable proof dynamic site should prevent manipulation of its script by external users other than its developers, it assumes that the website is vulnerable to attacks if the generated script after illegal code insertion also contains the illegal code inserted e.g. to test for cross site scripting, an input such as &lt;script&gt; alert ('XSS') &lt;/script&gt; is intentionally inserted into the server side script by the Vdetector. If this string is included in a dynamically generated page, the Vdetector will detect the text XSS and warns the user of site’s likely being vulnerable. Also, to test for SQL injection vulnerability, for example, a single quote (') is used as input, the single quote is a special character in SQL and if it is included in SQL query of the dynamic script at the server side then Vdetector assumes a vulnerability and will pop-up a message to the user through the browser to that effect. The Vdetector is compact so as to be integrated into the browser. It should be noted that our assumption is that the script of a vulnerability free website should not be prone to external manipulation else we classify it as being vulnerable since attackers can easily implant malicious code to exert influence on the site’s visitors. The Vdetector can also validate user input by parsing all user inputs, handling the dangerous characters/commands or rejecting the user input before being sent to the host server where the web pages and script resides if used as web applications profiler. If the user wants to maliciously insert illegal script into a server side script, the Vdetector automatically detects and reject the user input.

To further ensure security on the internet, user’s information on any website should be encrypted. Encryption is the process of transforming information to make it unintelligible to all but the intended recipient, secure socket layer (SSL) which supports encryption is already built into every major operating systems and browsers and thus making encryption appealing. We also advocate that government should put up campaign against cybercrimes while enacting stringent laws which will enable prosecution of perceived cyber criminals. Also, in this day of explosion of
mobile telecommunication powered modem usage, governments should make a consolidated database of all modems users irrespective of their service provider such that criminal activities if and when discovered will be tagged to a particular individual as the perpetrator and such person can then be prosecuted. The figures below show the model for detecting vulnerabilities.

![Figure 1. The framework for web-applications’ vulnerabilities detection](image)

**Results Obtained**

The Vdetector relieved the application’s developer the daunting task of multiple verifications and testing involved in manual checking of dynamic pages for vulnerabilities, while also saving time and cost the manual process would have incurred. It performs authenticated crawling and profiles the application. Also, since it identifies vulnerabilities of syntax and semantics in custom web applications, it ensures accuracy by reducing the level of false positives and false negatives in the application.

Users can easily detect cases of threat because the model pops up a warning message in respect of that. The model can further be combined with some existing anti-phishing solutions.
Combating Phishing and Identity Thefts

The Vdetector cannot detect any phishing attempt apart from those based on the vulnerability of the web application. It only warns the user that the website that is being visited might have been compromised, in this; the false positive is low because it directly detects the vulnerability by actually simulating an attack on the application. Also, presently, the Vdetector works within a browser and not standalone, it relies on the browser for end-to-end connection to the host server and its display of warning messages to the users. However, it should be noted that the model will perform optimally when combined with any anti-phishing solution.

Conclusions

Identity theft and phishing poses a great threat to trustworthy transactions on the internet, we are beginning to see it permeate every nook and cranny of the cyberspace and thus the earlier we stand up to it, the better it will be. In this paper we discussed several forms of identity theft and phishing; we also make a quick survey of available solutions to phishing. We then propose a model which uses an automated scanner, Vdetector to simulate some possible attacks on the application, if any vulnerability is found, the user is notified, and also the developer can thoroughly redesign the web application.
We also give some preventive measures and advocate that governments and corporate organizations should rise to the challenge of sensitizing the users about the threats and realities of phishing. However, no one tool should be relied upon exclusively. The proper application of defence-in-depth measures involving people and technology will be needed to counter these attacks as they appear.

Acknowledgements

We thank the senior faculty members of the computer science department, University of Ibadan for their immense contributions; timely advice and guidance in making this work a success.

References


Calling ID Toolbar. www.callingId.com/Desktopsolutions/callingidtoolbar.aspx


Combating Phishing and Identity Thefts


© JITI 2011


1 Dr. Olufade F. W. Onifade is a lecturer at the Department of Computer Science, University of Ibadan, Nigeria. He can be reached at fadowilly@yahoo.com or olufadefwilliams.onifade@loria.fr; Phone: + (234) 807-401-0558; + (234) 803-284-1817.

2 Mr. Kolawole John Adebayo is a research student at the Department of Computer Science, University of Ibadan, Nigeria. He can be reached at Collawolley3@yahoo.com; Phone: + (234) 806-607-8619