International Journal of Information Technology and Management

Vol. V, Issue No. I, August-2013, ISSN 2249-4510

AN ANALYSIS UPON VARIOUS MEASUREMENTS FOR DETECTION AND PREVENTION OF PHISHING ATTACK

AN INTERNATIONALLY INDEXED PEER REVIEWED & REFEREED JOURNAL

www.ignited.in

International Journal of Information Technology and Management
Vol. V, Issue No. I, August-2013, ISSN 2249-4510

An Analysis upon Various Measurements for Detection and Prevention of Phishing Attack

Ashish Gupta

Research Scholar, Sai Nath University, Ranchi, Jharkhand

Abstract – Phishing is form of identity theft that combines social engineering techniques and sophisticated attack vectors to harvest financial information from unsuspecting consumers. Often a phisher tries to lure her victim into clicking a URL pointing to a rogue page. In this paper, we focus on studying the structure of URLs employed in various phishing attacks.

We find that it is often possible to tell whether or not a URL belongs to a phishing attack without requiring any knowledge of the corresponding page data. We describe several features that can be used to distinguish a phishing URL from a benign one. These features are used to model a logistic regression filter that is efficient and has a high accuracy. We use this filter to perform thorough measurements on several million URLs and quantify the prevalence of phishing on the Internet today.

Phishing is a type of cyber-attack where the attacker creates a replica of an existing web page with an aim to acquire information such as usernames, passwords and credit card details of the users by fooling them into submitting personal data to what they think is their service provider's website. In this work, we present, a novel, algorithm to detect phishing web sites, based on the characteristics of the hyperlinks used in the phishing attacks and the content in the website.

----------♦----------

INTRODUCTION

Phishing is a form of identity theft that occurs when a malicious Web site impersonates a legitimate one in order to acquire sensitive information such as passwords, account details, or credit card numbers. Though there are several ant phishing software and techniques for detecting potential phishing attempts in emails and detecting phishing contents on websites, phishers come up with new and hybrid techniques to circumvent the available software and techniques.

Phishing is a deception technique that utilizes a combination of social engineering and technology to gather sensitive and personal information, such as passwords and credit card details by masquerading as a trustworthy person or business in an electronic communication. Phishing makes use of spoofed emails that are made to look authentic and purported to be coming from legitimate sources like financial institutions, ecommerce sites etc., to lure users to visit fraudulent websites through links provided in the phishing email. The fraudulent websites are designed to mimic the look of a real company webpage.

The phishing attacker's trick users by employing different social engineering tactics such as threatening to suspend user accounts if they do not complete the account update process, provide other information to validate their accounts or some other reasons to get the users to visit their spoofed web pages.

Why is it important to tackle the problem of phishing? According to the Anti-Phishing Working Group, there were 18,480 unique phishing attacks and 9666 unique phishing sites reported in March 2006. Phishing attacks affect millions of internet users and are a huge cost burden for businesses and victims of phishing. Gartner research conducted in April 2004 found that information given to spoofed websites resulted in direct losses for U.S. banks and credit card issuers to the amount of $1.2 billion (Litan 2004). Phishing has become a significant threat to users and businesses alike.

Over the past few years, much attention has been paid to the issue of security and privacy. Existing literature dealing with the problem of phishing is scarce. Fette et al proposed a new method for detecting phishing emails by incorporating features specific to phishing (Fette et al. 2006).

We applied different methods for detecting phishing emails using known as well as new features. We employ a few novel input features that can assist in discovering phishing attacks with very limited a-prior knowledge about the adversary or the method used to launch a phishing attack. Our approach is to

classify phishing emails by incorporating key structural features in phishing emails and employing different machine learning algorithms to our dataset for the classification process. The use of machine learning from a given training set is to learn labels of instances (phishing or legitimate emails). Our paper provides insights into the effectiveness of using different machine learning algorithms for the purpose of classification of phishing emails.

Phishing is as an act of sending an e-mail to a user falsely claiming to be a legitimate business establishment in an attempt to scam or trick the user into surrendering private information that will be used for identity theft. It is a type of network attack where the attacker creates a replica of an existing legitimate commercial website to deceive users to submit personal, financial, or confidential data to what to they think is their genuine business provider's site. It is a security attack that involves obtaining private and classified data by presenting oneself as a reliable and genuine entity. The damage caused by phishing ranges from loss of access to email to significant and considerable financial loss.

A proactive approach to minimizing phishing has been conducted where the system removes a phishing page from the host server rather than just filtering email and flagging suspected messages as spam. The study in however, assumes that emails have already been classified as a phishing email or legitimate email. The study has ignored the phishing email classification and was more concerned with how to deal with the Phisher once a phishing email has been detected.

Traditional security has focused mainly on authentication and encryption, with inroads to topics like privacy, robustness, and security against mobile adversaries. In all such cases, the security modeling has ignored the human factor and the impact on security that such attacks may have when combined with social engineering. The recent tide of so-called phishing attacks gives ample evidence that it is necessary to include the human factor in security modeling. These are attacks in which, typically, the victim is deceived to give out secret information such as passwords or other information enabling access to a given resource. Even though most attacks are surprisingly straight-forward – such as point-blank asking a victim for his bank account number and PIN – they are also rather successful. A recent study by Gartner (April, 2004) shows that around 3% of all those surveyed reported giving up financial or personal information in a phishing scam.

While it is likely for the very straight-forward attacks to become less successful as public awareness increases, phishing attacks are also likely to become more sophisticated in response. Also, with the development of do-it-yourself kits for phishing, most anybody who wants to can become a phisher. Here, we should note that while the term phishing typically is used for automated attacks that are performed en masse, we extend the use of the term to also mean automated attacks that target smaller sets of victims, but which succeeds with much higher probabilities. We believe this is a likely course of events, as it involves taking advantage of partial information of potential victims.

Apart from increasing the success ratio, this approach will lower the visibility of the attacks. In particular, by performing targeted attacks, phishers may to a much larger extent avoid phishing honeypots; these are identities and accounts created solely for the purpose of attracting attackers, and are used by service providers to detect from where attacks are performed.

Improved public awareness of threats is a necessary component in building a system that is secure against phishing attacks. Not surprisingly, though, public awareness is not sufficient, but must be accompanied by the development and employment of technical security mechanisms.

PHISHING PREVENTION

Websites used for phishing are detected by analysing end user confidential data submission statistics. A central process such as Google bots receives data indicating confidential information submitted to websites from a plurality of user computers. The received data is aggregated and analysed, for example through statistical profiling. Through the analysis of the aggregated data, anomalous behaviour concerning submission of confidential information to websites is detected, such as an unexpected, rapid increase in the amount of confidential information submitted to a given website.

Such anomalous behaviour indicates that the website is being used for phishing. Responsive to detecting the anomalous behaviour, further action is taken to protect users from submitting confidential information to that website. For example, an alert can be sent to an appropriate party or automated system, a protective measure against the site can be published, the site can be added to a blacklist, or a procedure to have the site shut down can be initiated.

From a business perspective**,** there are a number of things you can do to help avoid becoming a victim of phishing and to minimise damage o not only an individual, but to the organisation should it become a target. Some of the following could be considered, but may not be suitable for all types of organisations.

Figure 1 : Example of a warning message from an unsafe site.

- **Use Dedicated Systems for Payments** including requests and approval processes. Consider disabling email access on any system involved with payment processing. If a hacker cannot compromise the systems in payment processing, they will have a harder time obtaining payment usernames and passwords, and an even harder time actually requesting/approving a transfer.

- **Use a Strong Authentication Mechanism** on all payment processing systems. This would include replacing or augmenting username/password combinations with a hardware token and PIN, or with biometrics such as a fingerprint reader. An attacker will be unable to copy and reuse strong authentication such as a token or biometrics.

- **Block Internet Access** for systems involved in payment processing. If the system genuinely has no Internet access, malware would be unable to talk back to its controlling systems and attacker.

- **Disable the use of USB Flash Drives** in payment processing systems. In some circles USB flash drives are often referred to as "malware delivery devices". Disabling USB flash drives removes one more potential avenue for infection. Use tools available in your email client. Outlook, for instance, has the ability to help filter potentially harmful links.

- **Be diligent in your use of anti-virus** and anti-malware software, including regular updates and scans. Most of the malware used as part of a phishing attack is not detected by standard anti-virus software, but some of it is. Some malware indicators may not be changed before an anti-virus update is available, and sometimes older versions of malware are distributed. Additionally, antivirus software can help identify secondary infections that may be related to an attack.

- **Use reputation-based website**, IP address, and URL filtering to help ensure that any systems accessed from within the company are not considered "bad" sites. You can extend this further by allowing only "white-list" access – access to addresses that have specifically been recognized as "good" sites, (note that this has the potential to inhibit some Internet capability).

- **Enforce time-of-day login** and payment processing. Many fraudulent transactions occur after normal working hours. For instance, a series of large transfers that completed at 7:00pm Friday evening might be functionally ignored until staff return and see abnormal activities Monday morning.

- **Limit access to payment** processing systems from mobile devices, laptops, and systems based in home offices. These distributed systems are typically more vulnerable to threats.

- **Do not allow access to** any internal organization system, especially payment processing systems, from a personally owned home computer. There is simply no way the organization can enforce proper control over such a system.

- **Conduct employee security awareness** sessions to instruct employees on how to identify phishing emails and avoid falling victim to them. Any reduction in exposure slows compromise and increases your organization's capability to identify an escalating threat.

APPROACHES TO PREVENT PHISHING ATTACKS

There are several (technical or non-technical) ways to prevent phishing attacks: 1) educate users to understand how phishing attacks work and be alert when phishing-alike e-mails are received; 2) use legal methods to punish phishing attackers; 3) use technical methods to stop phishing attackers. In this paper, we only focus on the third one.

Technically, if we can cut off one or several of the steps that needed by a phishing attack, we then successfully prevent that attack. In what follows, we briefly review these approaches.

1) **Detect and block the phishing Web sites in time:** If we can detect the phishing Web sites in time, we then can block the sites and prevent phishing attacks. It's relatively easy to (manually) determine whether a site is a phishing site or not, but it's difficult to find those phishing sites out in time. Here we list two methods for phishing site detection.

a) The Web master of a legal Web site periodically scans the root DNS for suspicious sites (e.g. www.1cbc.com.cn vs. www.icbc.com.cn).

b) Since the phisher must duplicate the content of the target site, he must use tools to (automatically) download the Web pages from the target site. It is therefore possible to detect this kind of download at the Web server and trace back to the phisher. Both approaches have shortcomings. For DNS scanning, it increases the overhead of the DNS systems and may cause problem for normal DNS queries, and furthermore, many phishing attacks simply do not require a DNS name. For phishing download detection, clever phishers may easily write tools which can mimic the behavior of human beings to defeat the detection.

2) **Enhance the security of the web sites**: The business Web sites such as the Web sites of banks can take new methods to guarantee the security of users' personal information. One method to enhance the security is to use hardware devices. For example, the Barclays bank provides a hand-held card reader to the users. Before shopping in the net, users need to insert their credit card into the card reader, and input their (personal identification number) PIN code, then the card reader will produce a onetime security password, users can perform transactions only after the right password is input. Another method is to use the biometrics characteristic (e.g. voice, fingerprint, iris, etc.) for user authentication. For example, Paypal had tried to replace the single password verification by voice recognition to enhance the security of the Web site. With these methods, the phishers cannot accomplish their tasks even after they have gotten part of the victims' information. However, all these techniques need additional hardware to realize the authentication between the users and the Web sites, hence will increase the cost and bring certain inconvenience. Therefore, it still needs time for these techniques to be widely adopted.

3) **Block the phishing e-mails by various spam filters:** Phishers generally use e-mails as 'bait' to allure potential victims. SMTP (Simple Mail Transfer Protocol) is the protocol to deliver e-mails in the Internet. It is a very simple protocol which lacks necessary authentication mechanisms. Information related to sender, such as the name and email address of the sender, route of the message, etc., can be counterfeited in SMTP. Thus, the attackers can send out large amounts of spoofed e-mails which are seemed from legitimate organizations.

The phishers hide their identities when sending the spoofed e-mails, therefore, if anti-spam systems can determine whether an e-mail is sent by the announced sender (Am I Whom I Say I Am?), the phishing attacks will be decreased dramatically. From this point, the techniques that preventing senders from counterfeiting their Send ID can defeat phishing attacks efficiently.

4) **Install online anti-phishing software in user's computers:** Despite all the above efforts, it is still possible for the users to visit the spoofed Web sites. As a last defense, users can install anti-phishing tools in their computers. The anti- phishing tools in use today can be divided into two categories: blacklist/whitelist based and rule-based.

DETECTING AND CLASSIFYING PHISHING EMAIL

The proposed methodology will apply fuzzy logic and data mining algorithms to classify phishing emails based on two classification approaches such as content-based approach and non-content based approach. Specific categories or criteria are selected for each approach. The components or selected features are then identified for each category. The list of the classification approaches with the identified criteria and specific features is listed in the table below. The list will be used as basis for in the simulation and determination of phishing emails. The main characteristics of phishing emails are listed in Table 1.

Classification Approach	Category/Criteria	Component	Stage/Layer
Non-content Based Approach	URL	IP URL	Stage 1
		Redirect URL	
		Non-matching URL	
		Crawler URL	
		Long URl address	Weight = 0.5
		URL prefix/suffix	
Content-based Approach	Email Message	Spelling Errors	Stage 2
		Keywords	
		Embedded links	Weight = 0.5
Overall Weight			1.0

Table 1. Characteristics and stages of the components of phishing emails.

FEATURES USED

There exist a number of different structural features that allow for the detection of phishing emails. In our approach, we make use of sixteen relevant features. The features used in our approach are described below.

- HTML Email: HTML-formatted emails are mainly used for phishing attacks, because plaintext emails do not provide for the scale of tricks afforded with HTML-formatted emails. Hyperlinks are active and clickable only in htmlformatted emails. Thus, a HTML-formatted email is flagged and is used as a binary feature.

- IP-based URL: One way to obscure a server's identity is achieved through the use of an IP address. Use of an IP address makes it difficult for users to know exactly where they are being directed to when they click the link. A legitimate website usually has a domain name for its identification. Phishers usually use some zombie systems to host phishing sites. When a link in an email contains a link whose host is an IP address (for example, http://81.215.214.238/pp/) we flag the email and is used as a binary feature.

- Age of Domain Name: The domain names (if any) used by fraudsters are usually used for a limited time frame to avoid being caught. We can thus use this feature to flag emails as phishing based on the fact that the domain is newly registered and set a criteria of being new if it is less than 30 days old. This can be achieved by performing a WHOIS query on the domain name in the link. A WHOIS query provides other information such as the name or person to which the domain is registered to, address, domain's creation and expiration dates etc. This feature is a binary.

- Number of Domains: We make use of the domain names in the links that we extract and do a count of the number of domains. Two or more domain names are used in an URL address to forward address from one domain to the other.

- Number of Sub-domains: Fraudsters make use of sub domains to make the links look legitimate. Having sub domains means having an inordinately large number of dots in the URL. We can make use of this feature to flag emails as phishing emails.

- Presence of JavaScript: JavaScript is usually employed in phishing emails, because it allows for deception on the client side using scripts to hide information or activate changes in the browser. Whenever an email contains the string "JavaScript", we flag it as a phishing email and use it as a binary feature.

- Presence of Form Tag: HTML forms are one of the techniques used to gather information from users.

- Number of Links: Most often phishing emails will exploit the use of links for redirection. The number of links in email is used as a feature. A link in an email is one that makes use of the "href" attribute of the anchor tag. This feature will be continuous.

- URL Based Image Source: To make the phishing emails look authentic, images and banner of real companies are used in the emails. Such images are usually linked from the real companies' web pages. Thus, if any of the emails make use of such URL based images we flag it as a phishing email. This feature is binary.

CONCLUSION

In this paper we have identified several new and generic features for identifying phishing URLs. We use our features in a logistic regression classifier that achieves a very high accuracy.

Phishing is a form of online identity theft that aims to steal sensitive information from users such as online banking passwords and credit card information. The last years have brought a dramatic increase in the number and sophistication of such attacks. Although phishing scams have received extensive press coverage, phishing attacks are still successful because of many inexperienced and unsophisticated Internet users. Attackers are employing a large number of technical spoofing tricks such as URL obfuscation and hidden elements to make a phishing web site look authentic to the victims.

There are many preventative measures and precautions that organizations can also implement to ensure their websites and networks remain secure. This has now become more vital than ever in ensuring that customer trust is not damaged.

REFERENCES

- Afroz, S.; Greenstadt, R., "PhishZoo: Detecting Phishing Websites by Looking at Them", IEEE Fifth International Conference on Semantic Computing (ICSC), 2011, pp. 368-375

- Fette, I., Sadeh, N., Tomasic, A.: Learning to Detect Phishing Emails. Technical Report CMUISRI- 06-112. Institute for Software Research International, Carnegie Mellon University (2006)
- Juan Chen and Chuanxiong Guo, “Online Detection and Prevention of phishing attacks”, IEEE conference,2009.
- Litan, A.: Phishing Attack Victims Likely Targets for Identity Theft. Gartner Research (2004)
- Rachna Dhamija and J. D. Tygar. Phish and HIPs: Human Interactive Proofs to Detect Phishing Attacks. In Human Interactive Proofs, pages 127–141, 2005.
- Shah, R.; Trevathan, J.; Read, W.; Ghodosi, H.,”A Proactive Approach to Preventing Phishing Attacks Using the Pshark Model”, IEEE Sixth International Conference on Information Technology: New Generations, March 2009, pp. 915- 921
- Sujata Doshi, Niels Provos, Monica Chew, and Aviel D. Rubin. A Framework for Detection and Measurement of Phishing Attacks. Technical report, Johns Hopkins University, SPAR,
- Y. Zhang, J. Hong and L. Cranor, “A Content based approach to detecting pharming websites” in proceedings of international world wide web conference(WWW),2007.