Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 2, February 2015, pg.166 – 170

REVIEW ARTICLE

A REVIEW ON PHISHING DETECTION APPROACHES

Ms. Pallavi D. Dudhe[1], Prof. P.L. Ramteke[2]

[1]M.E. Second Year CSE, HVPM C.O.E.T. Amravati, India

[1] pallavidudhe2@gmail.com

[2]Prof IT Department HVPM C.O.E.T. Amravati, India

[2] pl_ramteke@rediffmail.com

Abstract - Phishing is web-based attack that uses social engineering techniques to exploit internet users and acquire sensitive data. Most phishing attacks work by creating fake version of the real site's web interface to gain user's trust. Despite the fact that these phishing sites look identical or nearly identical to real sites they imitate, user studies have shown that they ignore browser-based indicators and often use the appearance of site to judge the authenticity of sites, just as they use the appearance of physical sites to judge their authenticity. To protect users against phishing, various anti-phishing techniques have been proposed that follows different strategies. We applied different methods for detecting phishing using known as well as new features. In this paper we used the different approaches for detecting phishing websites such as heuristic approach, Blacklist-Whitelist based approach, Fuzzy rule based approaches, Machine learning approaches, CANTINA based approaches, Image based approaches, etc.

Keywords- Phishing, heuristic based, anti-phishing, fuzzy rule based, CANTINA

I. INTRODUCTION

Phishing is a new word produced from 'fishing', it refers to the act that attacker allure users to visit a faked Web site by sending them faked e-mails (or instant messages), and stealthily get victim's personal information such as user name, password, and national security ID,etc. Phishing is a form of identity theft that occurs when malicious Website impersonates a legitimate one in order to acquire sensitive information such as passwords, account details, or credit card numbers [1]. Though there are several antiphishing 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.

The internet is not only important for individual users but also for organizations doing business online [2]. Many of the organizations offer online trading and online sales of services and goods. Nevertheless, internet-users may be vulnerable to different types of online threats that may cause financial damages, identity theft, and loss of private information [3],[4]. Therefore, the internet suitability as channel for commercial exchanges comes into question. Phishing is considered a form of online threat that is defined as the art of impersonating website of an honest firm aiming to acquire user's private information such as usernames, passwords and social security numbers. Phishing websites are created by dishonest individuals to imitate genuine websites [5]. These websites have high level of visual similarities to the legitimate ones in an attempt to defraud honest internet- users.
In general, phishing attacks are performed with the following four steps
1) A fake web site which looks exactly like the legitimate Web site is set up by phisher.
2) Phisher then send link to fake web site in large amount of spoofed e-mails to target users in the name of legitimate companies and organizations, trying to convince the potential victims to visit their websites
3) Victims visit the fake web site by clicking on link and input its useful information there.
4) Phishers then steal personal information and perform their fraud such as transferring money from the victims' account.

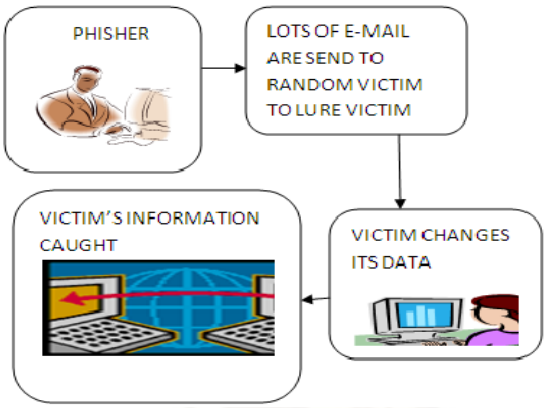

Fig 1: Process of phishing

Below, we explain the two most popular approaches in designing technical anti-phishing solutions
1) Blacklist approach: where the requested URL is compared with predefined phishing URLs. The drawback of this approach is that the blacklist usually cannot cover all phishing websites since newly created fraudulent website takes considerable time before it can be added to the list.
2) Heuristic Search approach: the second approach is based on search/heuristic methods, where several website features are collected and used to identify type of the website. In contrast to blacklist approach, the heuristic-based approach can recognize newly created fake websites in real-time.

II. LITERATURE REVIEW

Phishing website is a recent problem, nevertheless due to its huge impact on the financial and on-line retailing sectors and since preventing such attacks is an important step towards defending against website phishing attacks, there are several promising approaches to this problem and a comprehensive collection of related works. In this section, we briefly survey existing anti-phishing solutions and list of the related works [7].

There are various techniques which defend against phishing. Some techniques give e-mail level protection and some provide security toolbars embedded with anti-phishing tools.

III. DIFFERENT APPROACHES FOR PHISHING DETECTION

1. Heuristic approach

In this approach, researchers try to understand the anatomy of phished web sites and detect attacks based on several features. Features used in this approach include url, domain name, age of domain, spelling error, source of the images, links, etc [6],[7]. For example, Spoof Guard first checks the current domain name, then the full URL is analyzed to detect obfuscation as well as non-standard port numbers, then spoof guard analyzes the contents, making note of any password fields, embedded links and images.

2. Blacklist approach

Blacklists hold URLs (or parts thereof) that refer to sites that are considered malicious. Whenever a browser loads page, it queries blacklist to determine whether currently visited URL is on this list. If so, appropriate countermeasures can be taken. Otherwise, the page is considered legitimate [8]. The blacklist can be stored locally at the client or hosted at central server. Obviously, an important factor for the effectiveness of blacklist is its coverage. The coverage indicates how many phishing pages on the Internet are included in the list [13]. Another factor is the quality of the list. The quality indicates how many non-phishing sites are incorrectly included into the list. For each incorrect entry, the user experiences a false warning when she visits a legitimate site, undermining her trust in the usefulness and correctness of the solution. Finally, the last factor that determines the effectiveness of blacklist-based solution is the time it takes until a phishing site is included [9]. This is because many phishing pages are short-lived and most of the damage is done in the time span between going online and vanishing. Even when a blacklist contains many entries, it is not effective when it takes too long until new information is included or reaches the clients. Whitelisting approaches seek to detect known good sites but the user must remember to check the interface every time they visit the site.

3. Fuzzy rule based approach

One approach is based on experimentally contrasting few rule based classification algorithms after collecting dissimilar features from a range of websites as revealed [10] .Those features varied amongst three uncertain values ‘‘Legitimate & Genuine’’ and ‘‘Doubtful’’.

4. Machine learning approach

The majority of methods developed to deal with the phishing problem are based on support vector machine (SVM). SVM is known machine learning technique that has been used effectively to solve classification problems[10]. Its popularity comes from the accurate results it produced particularly from unstructured problems like text categorization. An SVM in general can be seen as a hyper-plane that splits the objects (points) belonging to a class (positive objects) from those that do not belong to that class (negative objects). This split is implemented by the SVM algorithm during the learning step where the hyperplan is obtained to divide positive and negative objects with maximal margins. The margin denotes the space from hyper-plane to the closest positive and negative object.

5. CANTINA based approach

CANTINA makes use of TF-IDF for detecting phishing sites. TFIDF is well-known information retrieval algorithm that can be used for comparing and classifying documents, as well as retrieving documents from a large corpus. In this section, we first review how TF-IDF works [11],[12]. We then introduce an application of TF-IDF called Robust Hyperlinks. Finally, we describe how we adapted Robust Hyperlinks for detecting phishing web sites.
Roughly, CANTINA works as follows:

· Given a web page, calculate the TF-IDF scores of each term on that web page.
· Generate a lexical signature by taking the five terms with highest TF-IDF weights.
· Feed this lexical signature to a search engine, which in our case is Google.
· If the domain name of the current web page matches the domain name of the N top search results, we consider it to be a legitimate web site. Otherwise, we consider it a phishing site.

6. Image based approach

This approach detected the type of websites by comparing phishy sites with the non-phishy sites based on visual similarity. This technique breaks down the webpage into block regions depending on ‘‘visual cues.’’ The visual similarity between a phishy webpage and non- phishy one is evaluated using three metrics: block level similarity, layout similarity, and style similarity [14]. A webpage is considered phishy if any metric has a value higher than a predefined threshold.

IV. CONCLUSION

Phishing has becoming a serious network security problem, causing finical lose of billions of dollars to both consumer send e-commerce companies. Detecting the phishing websites is one of the crucial problems facing the internet community because of its high impact on the daily online transactions performed. In this paper we observed the different approaches for phishing detection, the Heuristic approach is better than the other approaches.

As a future work on phishing we can do more work on server side security. In the server side security policy we use dual level of authentication for user by which only authentic user can get the access of his account, and to educate the user about this policy will results in avoiding user to give his sensitive information to phished website.

REFERENCES

[1] Hicham Tout, William Hafner “Phishpin: An identity-based anti-phishing approach” in proceedings of international conference on computational science and engineering, Vancouver, BC, pages 347-352, 2009.

[2] J. Zdziarski, W Yang, and P. Judge, spam conference, Phishing activity trend report 1st half 2011

[3] Mather Aburrous, M.A. Hossain, Keshav Dahal, Fadi Thabtah “Prediction phishing websites using classification mining techniques with experimental case studies” in proceedings of Seventh International Conference on Information Technology, Las Vegas, NV, pages 176-181, 2010.

[4] Aaron, G., & Manning, R.APWG phishing reports 2012.

[5] Abdelhamid, N., Ayesh, A., & Thabtah, F. Associative classification mining for website phishing classification. In Proceedings of the ICAI(pp. 687–695), USA, 2013.

[6] Abdelhamid, N., Ayesh, A., Thabtah, F., Ahmadi, S., & Hadi, W. MAC: A multiclass associative classification algorithm. Journal of Information and Knowledge Management, 11(2), 1250011-1–1250011-10, 2012.

[7] Garera, S., Provos, N., Chew, M., and Rubin, A. D. A framework for detection and measurement of phishing attacks. In Proceedings of the 2007 ACM Workshop on Recurring Malcode (WORM’07). 1–8, 2007

[8] Xiang, G., Pendleton, B. A., Hong, J. I., and Rose, C. P.A hierarchical adaptive probabilistic approach for zero hour phish detection. In Proceedings of the 15th European Symposium on Research in Computer Security (ESORICS’10). 268–285, 2010.

[9] Chou, N., R. Ledesma, Y. Teraguchi, D. Boneh, and J.C. Mitchell. Client-Side Defense against Web-Based Identity Theft. In Proceedings of The 11th Annual Network and Distributed System Security Symposium (NDSS'04).

[10] Phinding Phish: Evaluating Anti-Phishing Tools, Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong, In Proceedings of the 14th Annual Network &
Distributed System Security Symposium (NDSS 2007).

[11] CANTINA: A Content based approach to detecting phishing web sites, Yue Zhang, Jason Hong, Lorrie Cranor, In Proceedings of the 16th International conference on World Wide Web, Ban_, Alberta, Canada, May 8-12, 2007.

[12] A Proposal of the AdaBoost-Based Detection of Phishing Sites, Daisuke Miyamoto, Hiroaki Hazeyama, and Youki Kadobayashi (Internet Engineering Laboratory, Graduate School of Information Science, Nara Institute of Science and Technology, Japan),In JWIS, August 2007.

[13] Golle, Phillipe, Brent Waters, et Jessica Staddon. «Secure Conjunctive Keyword Search over Encrypted Data.» Applied Cryptography and Network Security (ACNS '04).31-45, 2004.

[14] Tyler Moore, Richard Clayton, and Henry Stern. Temporal correlations between spam and phishing websites. In Proceedings of the 2nd USENIX conference on Large-scale exploits and emergent threats: botnets, spyware, worms, and more (LEET'09). USENIX Association, Berkeley, CA, USA, 5-5, 2009.

[15] Maher Aburrous, M. A. Hossain, Keshav Dahal, Fadi Thabtah, "Predicting Phishing Websites Using Classification Mining Techniques with Experimental Case Studies, "Information Technology: New
Generations, Third International Conference on, pp. 176-181, 2010.

