RESEARCH ARTICLE OPEN ACCESS

Identification of Fraudulent Phishing Emails Based On CSS Standard Technique to Explore Similarities in Web Pages

Aishwarya Chavan [1], Raadhieca Iyer [2], Aparna Ramtirthakar [3]
Mrs. Shanthi K. Guru [4], Ms. Pallavi Khude [5]
Assistant Professor [4] & [5]
D.Y. Patil College of Engineering
Akurdi, Pune
Maharashtra - India

ABSTRACT
An individual or an organization set out to procure sensitive information, pertaining to another individual or organization, with questionable intents is "phishing" in a broad sense. Technologies to combat phishing that are currently being used are: better mutual authentication, spam filtering, detecting infringed domain names and alerting consumers when they are being directed to fake websites. Better mutual authentication requires awareness on the part of the user, which needs exhaustive efforts both from the users' and the organizations' side. As for other approaches, phishing websites have short lives and this makes it even more difficult to track them down. The proposed paper aims to use visual properties of a webpage namely page layout and page content as the way to find web page similarities. Since CSS is a commonly used technology used to define appearances of web pages, this paper uses it as a means to compare genuine websites against phishing websites which in turn prompts us which websites are fraudulent. This paper takes us through the details of the above mentioned approach.

Keywords:- Phishing, Spam filtering, Visual features, CSS, Web page similarities

I. INTRODUCTION

Internet has come a long way - from being just an information resource to becoming an everyday need. Gone are the days when users took to internet as an alternative to traditional letters and telegraphs. Its purpose now has expanded in multiple dimensions as opposed to it being merely a knowledge repository. From booking airplane tickets to ordering grocery items to be delivered at footsteps, the Internet has truly become multi-purpose.

According to security expert Chuck Wade of Interisle Group, "Technology is the rising tide that lifts all ships—including pirate ships." Deception paired with automation over the Internet, done to steal authentication credentials like passwords and account numbers for malicious purposes is on the rise. This concept is termed as "phishing". Credulous and unaware individuals are the easiest preys. Individuals and institutions incur heavy losses due to phishing. Another major side-effect of phishing is that online-transactions are slowly losing customers' trust.

Perpetrators intercept the on-going exchange of information for their vested interests. If this information lands in unworthy hands, it will surely be misused. This can lead to irreparable losses. As technology prospers, people grow with it. Even bad tendencies thrive. Phishing is one of the many ill-doings of people who seek to gain out of others' losses. Deception paired with automation over the Internet, done to steal authentication credentials like passwords and account numbers for malicious purposes is called "phishing". Phishing propagates largely through emails. Rogue URLs asking for authorization details is nothing new to us.

Once the credentials are submitted, they hack into bank accounts, email accounts, social networking site accounts etc. to carry out their plans. Credulous and unaware individuals are their easiest preys. Individuals and institutions incur heavy losses due to phishing. Billions of dollars' worth of loss has to be faced every year around the world. Another major side effect of phishing is that online-transactions are slowly losing customers' trust.

Fraudsters, apart from email, have taken to the Web, chat rooms, instant messaging, interactive games and keyboard logging programs (captures passwords

typed into web pages of legitimate sites) to execute their plans. User awareness regarding malicious activities is the highest-priority measure against phishing. But creating awareness requires conscious and continuous efforts and cannot be achieved in a fortnight. There are other counter-measures available to combat phishing.

Blacklisting phishing websites is one of the majorly used remedies on phishing. Blacklist, in Internet terminology, is a list of websites, or more popularly emails, that are traced back to spamming origins. Blacklists help us filter out notorious attempts at phishing. But the short lifespan of phishing websites makes the process of blacklisting inefficient. Also, URL variations used in phishing activities, again, make blacklisting a less useful option as blacklisting demands exact match for a URL. Spam filtering suspicious mails, better authentication processes by organizations etc. are other such steps being taken in the same direction.

Anti-Phishing Working Group (APWG) – the international consortium that brings takes under its wings businesses that have been victims of phishing attacks, security products and services companies, law enforcement agencies, government agencies etc. releases reports of the evolution, proliferation, and propagation of crimeware by drawing from the research of the member companies.

The excerpts of fourth quarterly APWG report [11] of 2014 suggested the following statistics:

- Average number of malware variants detected - 255,000 new threats each day
- 197,252 unique phishing reports submitted to APWG during Q4 indicating an increase of 18% from the 163,333 received in Q3 of 2014
- 46,824 phish observed in Q4
- 437 brands targeted by phishers in Q4
- The United States continues to be the top country hosting phishing sites

The figure below goes on the exhibits the victimized industries based on its vulnerability to phishing:

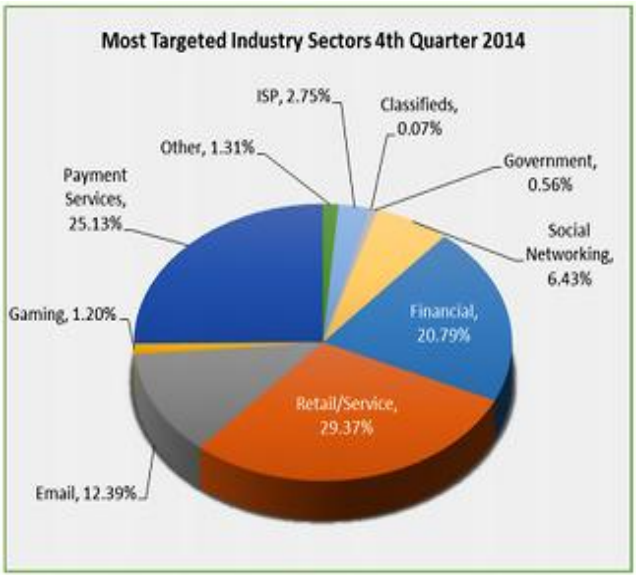

Fig. I

APWG Report 4th Quarter 2014 [11]

II. LITERATURE SURVEY

Home computer users often overlook potential threats while surfing the web which make them the weakest links in computer security. This paper [1] reviews factors that influence the decisions of security for home computer users. The review is presented in four sections: understanding of threats, perceptions of risky behavior, efforts to avoid security breaches and attitude towards security interventions.

Some mischief-makers in cyber world replicate originally safe and authorized websites and others overlay a deceptive appearance over unsafe websites. Nevertheless, similarities often persist. This paper [2] presents a combined clustering method that links together replicated scam websites, even when the criminal has taken steps to hide connections.

Ye Cao et. al [3] present a novel anti-phishing approach - Automated Individual White-List (AIWL). Phishing attacks involves tricking of users to submit their relevant personal and private information such as bank details into websites meant for phishing which appears similar to the genuine emails. Phishing attackers use many techniques to achieve their economic

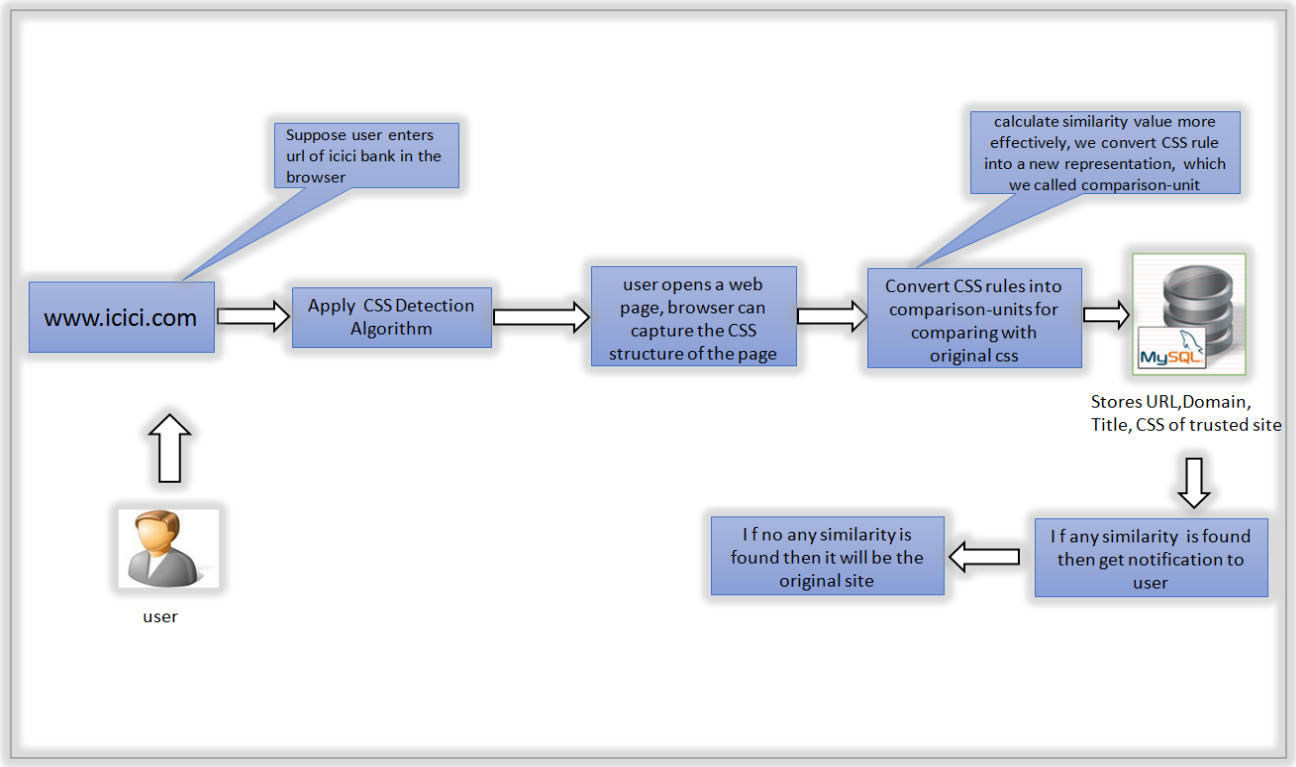

Fig. II
System Architecture

goals by posing great loss to the users who thereby hesitate to perform e-commerce or online transactions. These techniques may use both social engineering and technical subterfuge .The traditional work proposed an approach named AIWL which notifies users against phishing and pharming attacks with the help of white list which contains user's all familiar Login User Interfaces (LUIs) of websites. Next, AIWL can efficiently defend against pharming attacks, because AIWL will alert the user when the legitimate IP is maliciously changed. Naïve Bayesian classifier is used to automatically maintain the white-list in AIWL.

Method proposed by Purnima Singh, Manoj D. Patil in [4] first tries to identify whether a particular web page is phishing or not based on a large set of heuristics extracted from related. The most probable phishing target of that web page is obtained using Google Search API if the web page is found to be phishing. The proposed paper suggests extraction of many features that identifies legitimate web pages. The process starts from examining whether the web page contains text fields. This step proves useful as phishing web pages ask for user inputs through text fields. On detection of minimum one text field the process goes on to explore the web page further.

In order to tackle the semantic attacks launched by the phishers, this paper [5] presents a method to detect the phishing webpages semantically. According to the linguistic characteristics that appear in the phishing pages, the phishing domain ontology is proposed and then its corresponding description model, Phishing Descriptive Model (PDM) is proposed. After that, the mechanism for the detection of the phishing pages based on the PDM is introduced.

Another noteworthy approach [6] tries to demonstrate that the source of phishing URLs and the freshness of the URLs tested can significantly impact the results of anti-phishing tool testing. It also demonstrates that many of the tools tested are vulnerable to easy exploits.

There also exists a system [7] that exploits algorithms and underlying principles. All existing algorithms are categorized into three streams based on the type of information they use: content-based methods, link-based methods and methods based on non-traditional data such as user behavior, clicks and HTTP sessions.

Jian Mao et al suggest a new solution [8], BaitAlarm that presents an algorithm to quantify the

suspicious ratings of web pages based on similarity of visual appearance between the web pages.

A newly discovered "Offpath TCP sequence number inference", by Zhiyun Qian et al [9] allows an off-path attacker to attack enabled by firewall middleboxes is reported hijack a TCP connection and inject malicious content, effectively granting the attacker write-only permission on the connection.

The paper by Zhen Chen et al [10], based on wellknown TimeMachine, presents TIFAflow, the design and implementation of a novel system for archiving and querying network flows.

It allows an off-path attacker to attack enabled by firewall middleboxes is reported hijack a TCP connection and inject malicious content, effectively granting the attacker write-only permission on the connection.

The paper by Zhen Chen et al [10], based on wellknown TimeMachine, presents TIFAflow, the design and implementation of a novel system for archiving and querying network flows.

System architecture in Fig. II describes the precise view of the current system. The overall architecture summarizes the working of the proposed system as in how the fraudulent website is been identified and notified to the user. This anti-phishing technique helps to protect the private information of the user. The system provides a CSS based comparison strategy to pick the genuine emails. This strategy focuses on certain rules which involve tests on page layout of the webpage such as CSS domain name, URL of the site, title, CSS content of the site etc. It filters the legitimate emails through these tests thereby protecting the users from the phishers.

III. CONCLUSION

Life, in almost all its forms, revolves enormously around use of Internet. The fast –track world demands speedy interactions between communicating parties. The Internet provides us this required pace of life. But with convenience in sight, security is often overlooked. With phishers constantly on the loose, exploiting every possible loop-hole they can lay their hands upon, the situation demands that we arrest the situation before it further worsens. Some techniques rely on white-lists and black-lists and some go on to propose newer systems to tackle the menace of phishing. But constraints such as dependency on textual data in Webpage, cloud storages, additional resource requirements, white-lists and black-lists (that need manual updating) clearly indicate their scope of improvement. CSS-driven technique appears to be most promising amongst these technologies as it works directly on the basic structure of Webpages that are the page layouts. Combining CSS-based technique with URL-driven [12] approach to further improvise on efficiency will be a contribution that can be worked on in the future.

ACKNOWLEDGEMENTS

We would like to thank our project guides Asst. Prof. Shanthi K. Guru and Asst. Prof. Pallavi Khude for their utmost valuable guidance. We wish to thank them for their unending support and their promptness to help us with the most diverse of problems that we have encountered along the way right from when the idea was conceived. We also would like to express our sincere gratitude towards all our staff and colleagues who have helped us directly or indirectly in the course of writing this paper.

REFERENCES

[1]. Adele E. Howe, Indrajit Ray, Mark Roberts, Malgorzata Urbanska, "The Psychology of Security for the Home Computer User," IEEE Symposium on Security and Privacy 2012.

[2]. Jake Drew, Tyler Moore, "Automatic Identification of Replicated Criminal Websites Using Combined Clustering", 2014 IEEE Security and Privacy Workshops.

[3]. Ye Cao, Weili Han, Yueran Le, "Anti-phishing Based on Automated Individual White- List," Fairfax, Virginia, USA, ACM, October 31, 2008.

[4]. Purnima Singh, Manoj D. Patil, "Identification of Phishing Web Pages and Target Detection", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 2, February 2014.

[5]. Jianyi Zhang, Qi Li, Qian Wang, Tao Geng, Xi Ouyang, Yang Xin, **"**Parsing and Detecting Phishing Pages Based on Semantic Understanding of Text", Journal of Information & Computational Science 9: 6 (2012) 1521–1534.

[6]. Yue Zhang, Serge Egelman, Lorrie Cranor, and Jason Hong, "Phinding Phish: Evaluating Anti-

Phishing Tools," Research showcase @ CMU, Human-Computer Interaction Institute School of Computer Science, 2006.

[7]. Nikita Spirin, Jiawei Han, "Survey on Web Spam Detection: Principles and Algorithms," SIGKDD Explorations Volume 13, Issue 2.

[8]. Jian Mao, Pei Li, Kun Li, Tao Wei, and Zhenkai Liang, "BaitAlarm: Detecting Phishing Sites Using Similarity in Fundamental Visual", Features, 2013 5th International Conference on Intelligent Networking and Collaborative Systems.

[9]. Zhiyun Qian, Z. Morley Mao, "Off-Path TCP Sequence Number Inference Attack: How Firewall Middleboxes Reduce Security", 2012 IEEE Symposium on Security and Privacy.

[10]. Zhen Chen, Lingyun Ruan, Junwei Cao, Yifan Yu, and Xin Jiang, "TIFAflow: Enhancing Traffic Archiving System with Flow Granularity for Forensic Analysis in Network Security", Tsinghua Science and Technology, August 2013, 18(4): 406-417.

[11]. , "Phishing Activity Trends Report," Unifying the Global Response To Cyber Crime, October-December 2014. [Online]. Available:http://docs.apwg.org/reports/apwg_trends_report_q4_2014.pdf.

[12]. Samuel Marchal, Jérôme François, Radu State, and Thomas Engel, "PhishStorm: Detecting Phishing With Streaming Analytics", IEEE transactions on network and service management, vol. 11, no. 4, December 2014.