ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

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 Ms. Pallavi Khude

Asst. Prof. DYPCOE, Akurdi. Asst. Prof. DYPCOE, Akurdi.

Student, Computer Engineering, DYPCOE, Pune, India [1 2 3]

Abstract*:* Phishing implies an activity where an individual or a group of individuals attempt to obtain sensitive user information thus, threatening the financial or identity-related security of other individuals. Few of the preventive measures currently being taken in the said direction are: spam filtering, better authentication and detecting infringed domain-names. Better mutual authentication needs awareness on the user's part and it requires exhaustive efforts from the user's and the respective organization's side. The proposed paper aims at using a hybrid approach as the way to find web page similarities. As CSS is a commonly used technology used to define visual appearances of web pages. This paper exploits it as a way to compare genuine websites against phishing ones in order to warn the user against such attacks. Besides this, the other algorithms such as Ob-URL detection algorithm and Google safe browsing are used for enhancing the accuracy of detecting the fraudulent websites. This paper takes us through the insides of the above mentioned approach.

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

I. INTRODUCTION

As security expert Chuck Wade of Interisle Group states, "Technology is the rising tide that lifts all ships—including pirate ships." Deception in tandem with automation over the World Wide Web, done in order to procure authentication information like passwords and account numbers for malicious intents or "phishing" is on the rise. Unaware and credulous users are the easiest targets. Users falsely enter their private information on any site without testing Perpetrators intercept the exchange of information. There is loss of private information. In day to day life most of the people choose to transfer money online. But because of phishing it becomes risky to do such transactions.

Phishing is largely dependent on emails. Rogue URLs asking for authorization details are sent to a targeted individual or a community. Using users private information user accounts are hacked.User - awareness regarding phishing activities is the highest-priority preventive measure. But creating that awareness will need conscious, continuous efforts and cannot be achieved in a day. There are many techniques present to control phishing. Users falsely enter their private information on any site without testing .That information may use by anyone for bad purpose. So prevention is must. To detect the fraud the whole site must go through the tests so that user will come to know about fraud sites. There are many techniques present to control phishing

Our aim is to use fundamental visual features of a web page's appearance as the basis of detecting page similarities. There are many similarities in visualizing fraud sites and original site in web pages. So it is difficult to detect the original site. So we are using CSS detection algorithm to find out the similarity in web pages. Also another algorithm is ObURL algorithm. In this algorithm we are performing multiple tests such as IP address Test, Shorten URL Test, Black and White list Test, multiple recipient ,spam word's in the email .Blacklisting of phishing websites, filtering out suspicious mails, better authentication processes adopted by

All Rights Reserved © 2016 IJARTET

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

organizations are the steps being taken in this direction. For blacklisted phishing sites, this paper also uses Google Safe Browsing. Google Safe Browser is a service provided by Google for list of already detected phishing or fraud site. We are using Google chrome extension to detect the web pages as genuine or not.

Statistics provided by the Anti-Phishing Working Group (APWG) presents some noteworthy facts. The excerpts of the fourth quarterly APWG report [11] of 2014 are:

- Average number of detected malware variants - 255,000 new threats each day
- An increase of 18% in unique phishing reports submitted to APWG during Q4 as compared to Q3 of 2014
- 46,824 phishes observed in Q4
- A total of 437 brands targeted by phishers in Q4.

Fig. I sheds some light on the extent to which phishing impacts the world.

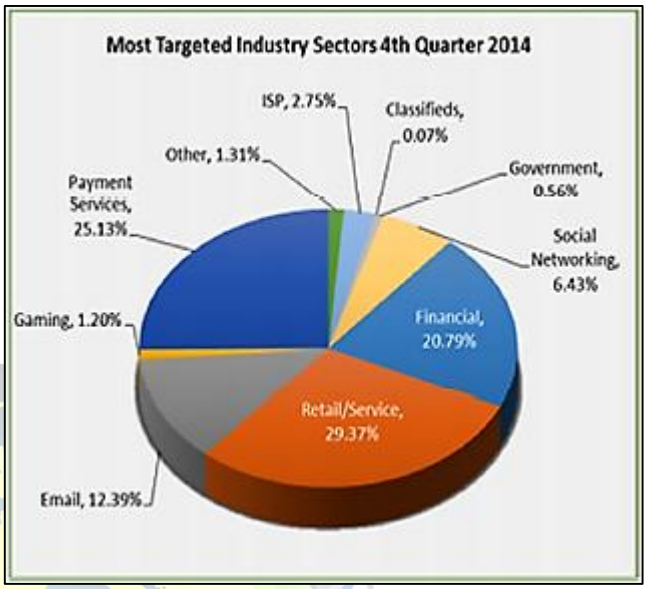

Fig. 1. Most Targeted Industry Sectors

II. LITERATURE SURVEY

Home computer users always make a trade-off in favour of convenience as opposed to safe surfing. In the process, they often overlook potential threats making them the weakest links in computer security. Paper [1] reviews factors that have an effect on the security related decisions of computer users.

Another paper [2] presents a combined clustering method that links together replicate scam websites, even when a criminal has attempted to hide connections.

Another noteworthy approach [6] suggests that the source of phishing URLs and freshness of the tested URLs can significantly impact the results of anti-phishing tool testing. There also exists a system [7] that exploits algorithm-intensive approach for phishing detection. On the other hand, Jian Mao et al. have come up with a novel possible solution in their paper [8], Bait Alarm that presents an algorithm to quantify the suspiciousness rating of web pages based on similarity of visual appearance between them.

"Offpath TCP sequence number inference", by Zhiyun Qian et al [9]makes use of a TCP connection in an attempt to reach an efficient solution whereas the paper by Zhen Chen et al [10], presents TIFA flow, design and implementation of a novel approach for querying and archiving network flows.

III. PROPOSED WORK

The proposed system will be a Phishing Detector aimed at web users who have Internet connection at their disposal. The system will be designed so as to minimize users' exposure to potentially risky webpages that threaten to misuse their sensitive information. The focus will be email system domain, where rogue links arrive in the users' mailboxes urging them to provide their credentials.

More specifically, this system will be designed to prompt Internet users when they arrive at phishing

All Rights Reserved © 2016 IJARTET

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

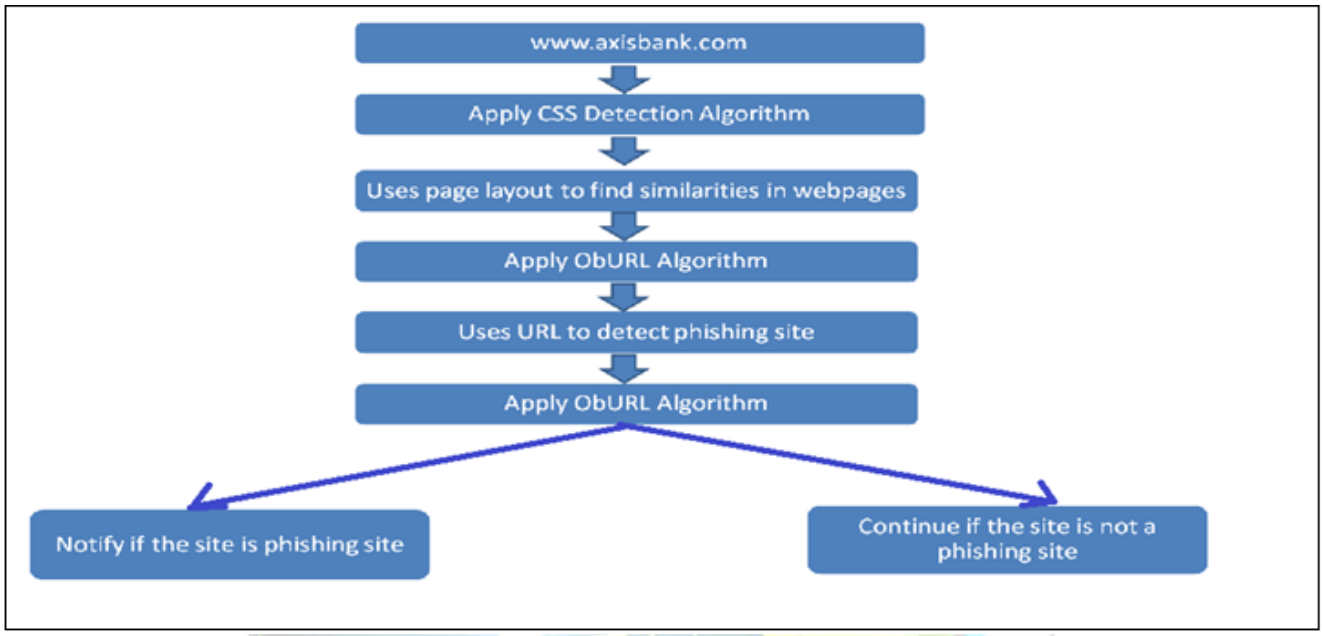

Fig. 2. System Flow Diagram

websites by means of pop-up messages. Previously encountered phishing websites will be stored into the databases along with lists of commonly known phishing websites that are already published on the World Wide Web. Fig.2. gives a clear picture of the proposed system flow.

IV. RESULTS

The proposed system is essentially based on the email system. But users can also manually copy-paste a suspicious URL, externally, to test its credibility. A user can create her/his login by supplying details that consist of a valid email ID and a username for the phishing detection system.

The login details will be safe-guarded with a password and will be stored in the linked database. Once the user is logged into the system, emails from the user's email account will be fetched for further processing. The system will evaluate the emails as "safe" or "unsafe" by referring to a combination of following factors:

- a database of typically used keywords(in emails) to lure unsuspecting web users
- an algorithm that checks for the authenticity of websites based on their visual elements
- an algorithm that checks for obfuscation of the website's URL
- a database of publicly announced white-listed websites

As an addition to Ob-URL and CSS detection algorithm, Google Safe Browsing provides a blacklist service which contains lists of URLs which are considered as fraudulent content. On cross-checking with the provided lists the further investigations based on CSS detection algorithm framework is applied to provide appropriate results.

This system loads the mails in the inbox of a particular user. The number of mails which can be fetched and analysed can be limited to increase the time efficiency.

a) ***UpdateEmailJob*** is a function to periodically update the emails in the system.

b) ***ReadRecentMail.java*** is a method used to update the system with the new incoming emails

c) ***StoreAllEmails*** fetches emails from gmail account and stores them in database table

All Rights Reserved © 2016 IJARTET

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

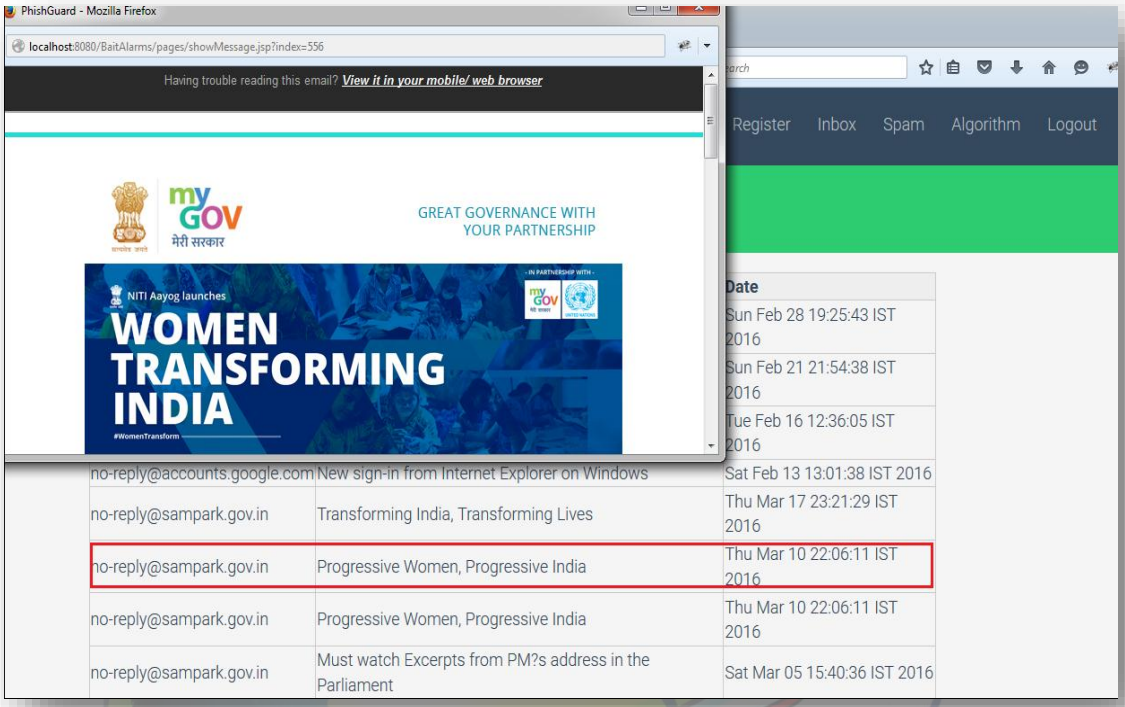

Fig.3. Successful navigation of a genuine UR

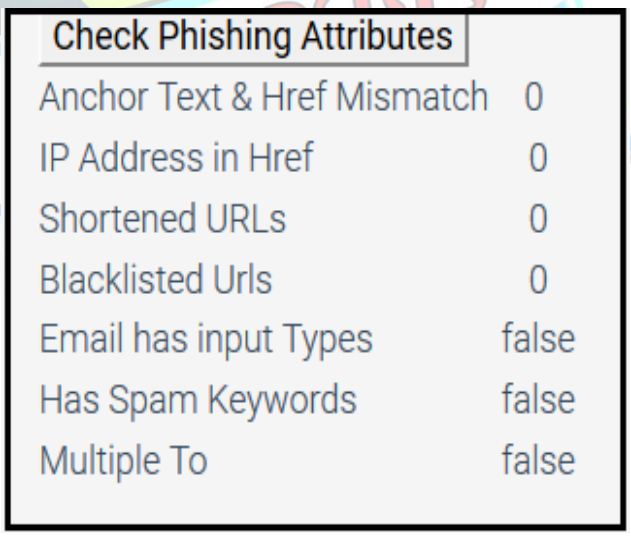

Fig.4. Check Phishing attributes

d) ***UpdatePhishingRules*** will update spam attributes for each Email

e) ***PhishMailGurad*** Function that checks email body for ob URL Detection Algorithm

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

If email fails in any of these checks it will be added to SPAM table. The following steps show how a mail is processed.

The function *processMessage* is called which performs as follows:

1. Email Content is parsed using JSOUP API
2. Content is searched for hyperlinks
3. ***getIpFromText*** – checks if url has an IP address and returns the IP address
4. ***getURL*** - check if given text contains a url or host name and returns it
5. ***getCompleteURL*** - check if url is shortened if yes return the original url (Used for shortened URL Test)
6. ***isWhitelist(url)*** - returns if the url is phishing, malware, virus
7. Check if email has an input type
8. Check if email is sent to multiple recipient

The ***Google safe browsing*** is applied to check the URL of the site and followed by the ***CSS detection algorithm.*** If the site is a genuine one then the system will navigate successfully on the given link else it will notify the user accordingly. Fig.3. shows the successful navigation of the given URL. Fig.4. shows checking of the attributes of the fetched email.

The hybrid application of the algorithms mentioned in this filters the fake emails efficiently as accurately. This enhances the performance of the system and performs better than the other approaches.

V. Conclusion

Life, in almost all its forms, revolves enormously around use of Internet. 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 limited 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.

VI. Future Work

The future work of this system consists of expanding the scope of the project. The system is restricted to email system only, which has a long way ahead to grow.

Nowadays, offline scams are also increasing resulting in the same effective threat that is caused by online ones. Offline phishing aims in avoiding clicking by the user on the hyperlinks which are provided in emails by the phishers. Phishers convince the user that the provided mail is a genuine one and encourages the user to submit the credentials and thereby achieving the illegal motives. Any organization such as banking sectors, financial institution, health care or government agency is as vulnerable to this form of phishing as hyperlink based phishing.

Hence, the only solution is to protect yourself in all kinds of such attacks and not being a victim of these online as well offline phishing activities. The future work will expand the scope of the project which will enhance the area of applications where the system is being implemented.

The phishing attacks through offline mode such as SMS, fax, advertisements etc, have the same impact and are proved to be one of the growing areas of scams as provided in the statistics of phishtank [13].

This future work will contribute in decreasing the threats and will help in making internet a safe and better place to live in.

VI. Acknowledgements

We would like to thank our project guides Asst. Prof. Shanthi K. Guru and Asst. Prof. Pallavi Khude for their utmost valuable guidance, 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.

References

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

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

All Rights Reserved © 2016 IJARTET

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Issue 3, March 2016

[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.

[13]. https://www.phishtank.com/stats.php

All Rights Reserved © 2016 IJARTET