Original Article

An Effective Secure Mechanism For Phishing Attacks Using Machine Learning Approach

Ms. U. Elamathi[1], Ms. A. V. M. B. Aruna[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur
Email: elamathiu@pmu.edu
[2]Assistant Professor, Department of Computer Science and Engineering, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur
Email: aruna1418@pmu.edu
DOI: 10.47750/pnr.2023.14.S02.320

Abstract

Phishing is one of the biggest crimes in this world in which it involves thefting the user's sensitive data. Usually, phishing websites targets individual's websites, organizations, sites for cloud storages, and government websites. Most of the users while surfing on internet are unaware of the phishing attacks. Many existing phishing approaches have been failed for providing a proper way to the issues facing for E-mails attacks. Currently hardware based phishing approaches have been used due to software attacks it rises a large factor. Due to rise of these kinds of problems, the proposed research work focused on three stage phishing series attack for precisely detecting the problems in a content based manner and the method was named as Phishing Attack Mechanism. Three input values had been taken such as Uniform Resource Locators, traffics and web content as input features by based on features of phishing attack and non-attack of phishing website technique features are implemented. To implement the experimental analysis proposed Phishing Attack Mechanism, dataset is collected from the recent phishing cases. In which it has been founded that the real phishing cases from proposed giving a higher accuracy on both zero-day phishing attack and in both phishing attack detections. Three different classifiers were used to find out the classification accuracy in order to detect the phishing, which yields the 95.18%, 85.45%, 78.89%, by NN, SVM, & RF classifications respectively. The results suggested and recommended is better for detecting the phishing by machine learning approach among the different approach.

Keywords: Phishing, Attack detection, Web Crawler, Heuristic analysis, Machine Learning Classification.

1. Introduction

In the modern era of network, development of several industries, people use the internet. While, different security attacks affect their business. One of main attack is called as phishing. In phishing threat, this can be done by e-mail spoofing and similar webpage functioning by this activity. phisher can perform attacks with the help of spoofed e-mails and copy of website design. Based on the internet, phisher can hack the user personal belongings.

Phishing is an offence scheme implementing by technical and social technology and hacking their user identity information and banking information. Since phishing threats happens because of user weakness and developing by phishers [1].

While working on internet users need to enter the useful data such as personal information, Banking information etc. This attacks are used to stole the user's personal data. Day-to-day economy phishing attacks are increasing says reporters. Phishing websites look like as same as the original website [2]. The main target of the phishers is to attack the victim's e-mails, messages and phone-calls. Many various kinds of phishing are there such as deceptive phishing attacker's focus on organization in which employees on the organization people information are stealed [3]. Figure 1. Represents the Phishing sites reported on 2020.

Figure 1. Phishing sites reported on 2020

1.1 Proposed Research Key Features

- Phishing attack is happening day-to-day life, based on the database features this research work mechanism evolved.
- The proposed research work focused on recent database and performance can be evaluated based on parameter.
- Three different classifiers were used to find out the classification accuracy in order to detect the phishing, which yields the 95.18%, 85.45%, 78.89%, by NN, SVM & RF classifications respectively.
  Figure 2. represents the Internet Usage per Year from 2014 to 2019. Figure 3. represents the Year-Wise Online Report of Phishing Attack Incident from 2014 to 2019. Figure 4. Illustrate the Phishing attacks in Industry.

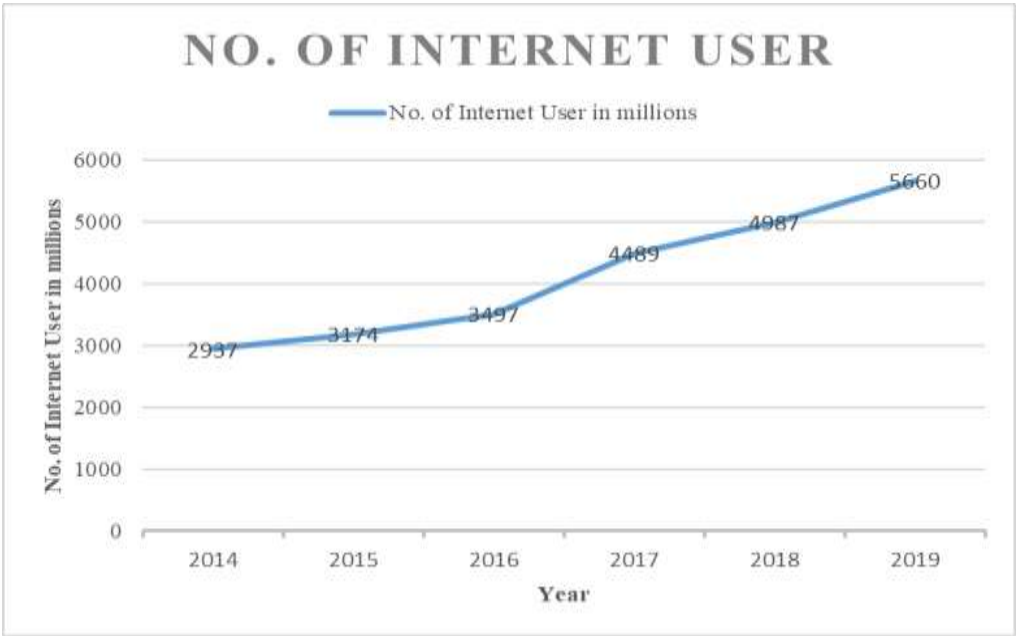

Figure 2. Internet Usage per Year from 2014 to 2019.

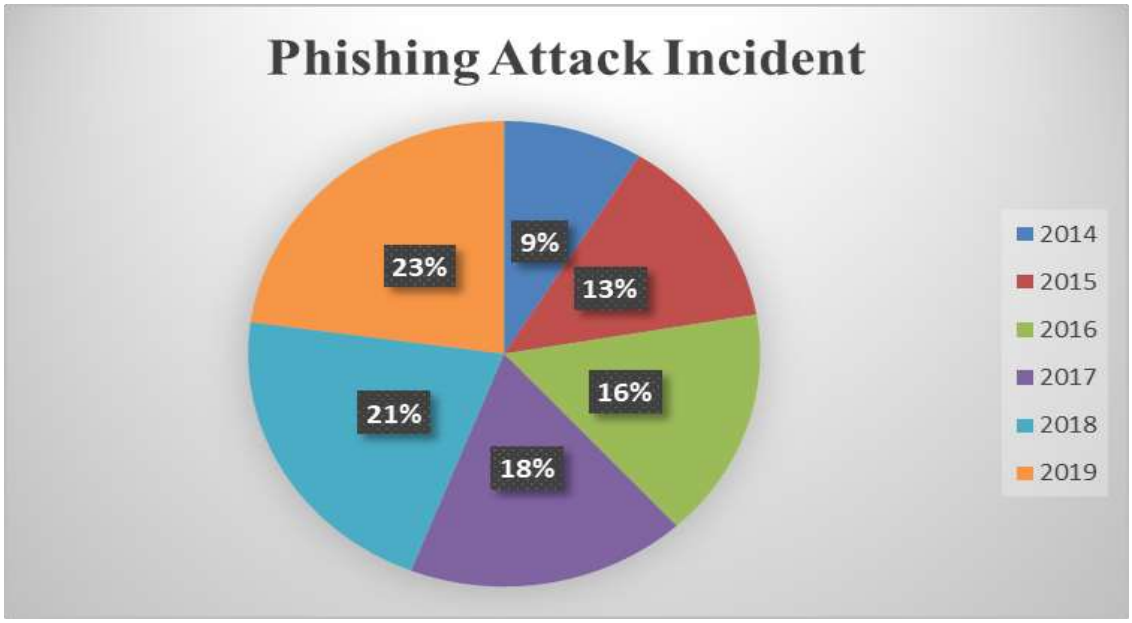

Figure 3. Year-Wise Online Report of Phishing Attack Incident from 2014 to 2019

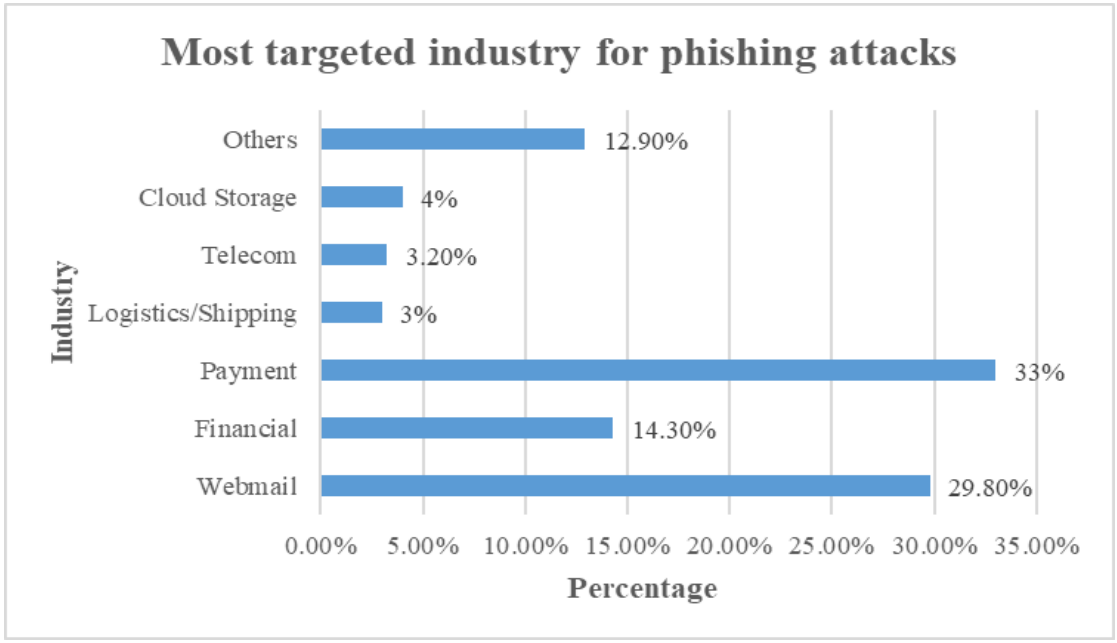

Figure 4. Phishing attacks in Industry

2. Literature survey

Literature survey which is related to the work adopted in this investigation is presented in this section. Table 1. Illustrate the Literature Survey for detecting the Phishing – A Brief Overview.

Table 1. Literature Survey for detecting the Phishing – A Brief Overview

S.No.	Author & Year	Algorithms	Techniques	Merits	Demerits

1.	Y.Zhang et al., (2007) [4]	Neural networks	Neural network has been classified with Monte Carlo algorithm.	Increases accuracy rate and stability detection.	Whole page has to be downloaded.
2.	M. Babagoli et al., (2018) [5]	Heuristic value based on extracting the parameters	Decision tree and wrapper is used for selecting the heuristic based nonlinear regression.	Using decision trees original dataset can be reduced.	11055 phishing and real web pages only used as it contains a less dataset.
3.	H. Kim et al., (2017) [6]	Features of Machine learning	Authentication on user and domain level.	Communication of security can be increased.	Same technology should be used on sender side and receiver side.
4.	T. Peng et al., (2018) [7]	Naive bayes classifier	Phishing email is identified by using naïve bayes classifier on machine learning and NLP techniques.	To detect the appropriateness of each words NLP is used.	To establish the virus pairs machine learning is used. Emails text analysis on rely.

3. Proposed Methodology

Phishing Attack Mechanism can be categorized in three categories such as

1) DNS blacklist,
2) Web crawler based approach,
3) URL analysis based approach.

These approaches are used for future purpose for phishing attacks in future extraction.

3.1 DNS blacklist & Web Crawler

DNS blacklist (Domain Name System blacklist) are used for generating a number of Internet Protocol address which can be easily mounted for programming on browser. It is built on the top source file on the internet [8]. This Domain Name System blacklist generates the Internet Protocol address for involving spam purpose. Information's are frequently updated on the DNS system. Web crawler starts to the websites interconnecting with pages and links. Crawling from one website by Phishing Attack Mechanism goes through all the links from web index. Proposes a Phishing Attack Mechanism crawl as creating a web crawler for each webpage in a website since attackers. Figure 6. represents the Crawler for Web Indexing.

3.2 URL analysis

URL partition is as follows. <protocol>://<sub domain> <Primary domain> <TLD>/<Path Domain>. Algorithm 1 explains the working module of URL based phishing detection.

Algorithm 1: URL based Phishing Detection	
Input	Primary domain –k are the features of URL, @, -dots, ID.
Output	Either phished or legal classification.

Step 1	if k is IP address then Condition= Phished else if turned up ('@','-','.');
Step 2	If '@' && '-' Condition= Phished.
Step 3	else if turned up ('.')>5 Condition= Phished end if
Step 4	else if ld<3 Condition= Phished end if

Step 5	if Condition is Phishing then
	"User notify"
	end if

4. Machine Learning Approach for Detecting the Attacks

The Data sets are collected from recent Alexa, Siri, Phish Tank and then processed into machine learning algorithms. Figure 5. Represents the Machine Learning Techniques for detecting Phishing Attacks. Figure 6. illustrate the Classifier Performance. Table 2. Represents the Accuracies in % of different classifiers against extracted features. Figure 7. Illustrates the Spam Mails detected by Proposed Methodology in E-mail. Figure 8. Illustrate the Fake website detected and blocked by Proposed Methodology in Netcraft. Figure 9. represents the Fake website detected by Proposed Methodology in Google Safe Browsing. Figure 10. Illustrate the Detection of Fake website by Google Safe browsing. Figure 11. Represents the High risk Fake website displayed but not blocked by Netcraft.

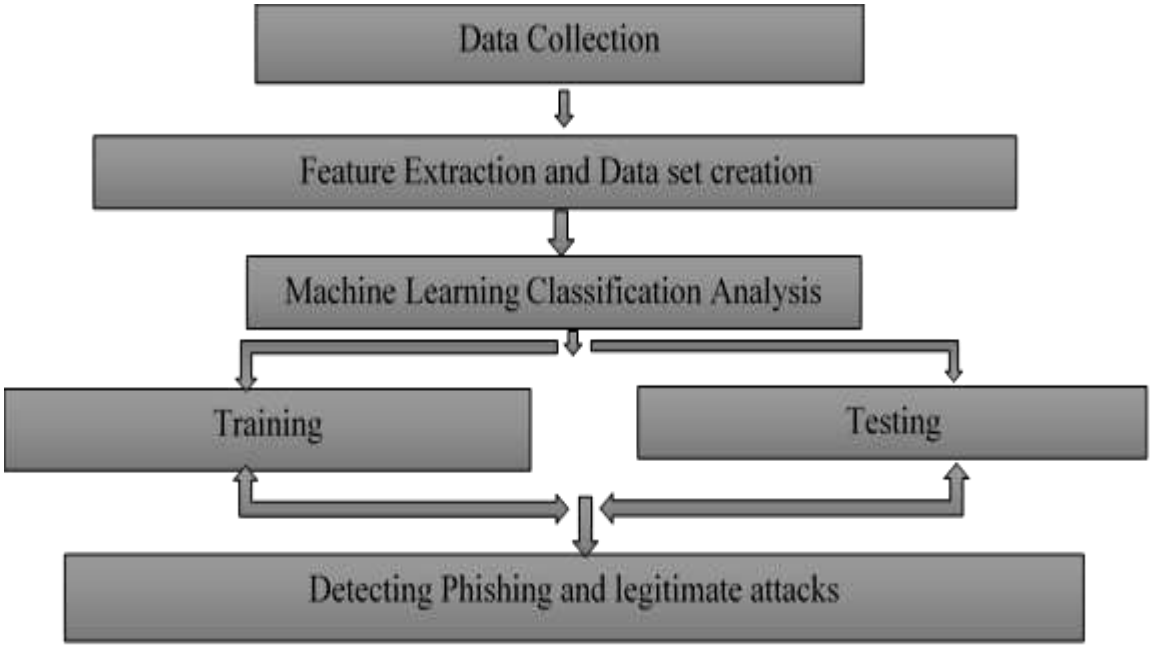

Figure 5. Machine Learning Techniques for detecting Phishing Attacks

Figure 6. Classifier Performance

Table 2. Accuracies in % of different classifiers against extracted features

Machine Learning Accuracy			
Support Vector Machine	**Random Forest**	**Neural Network**	**Total Extracted Parameter**
85.45%	78.89%	95.18%	12 Parameters

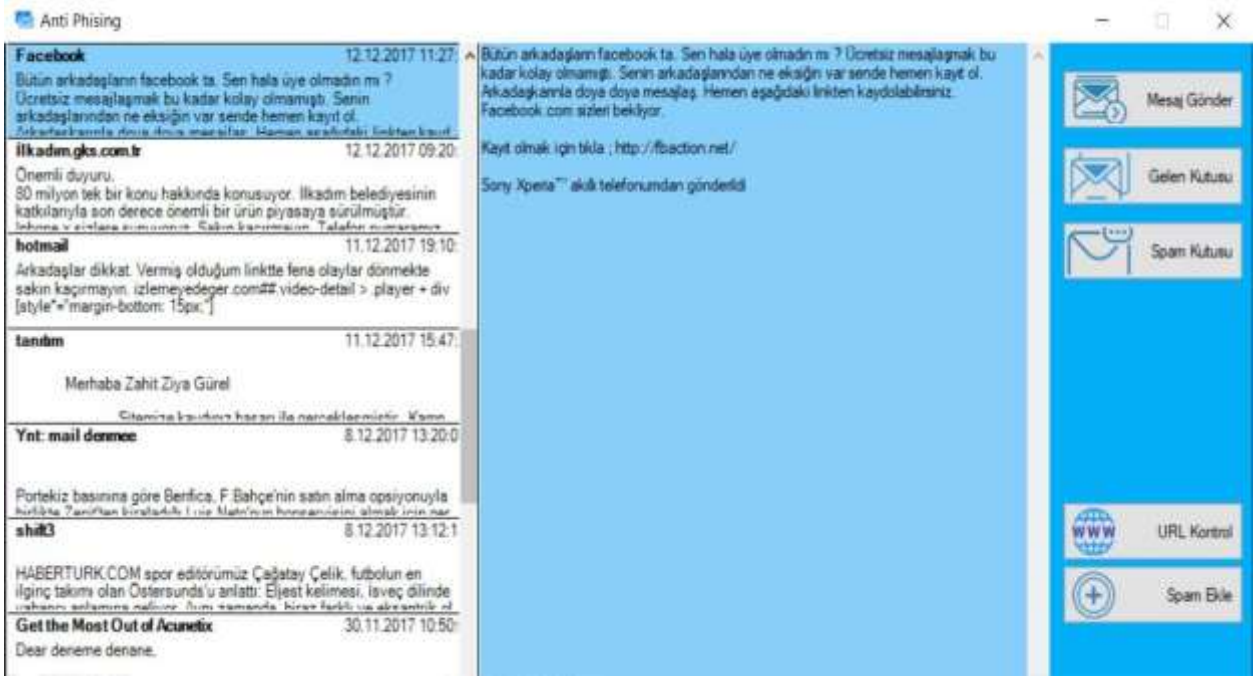

Figure 7. Spam Mails detected by Proposed Methodology in E-mail

Figure 8. Fake website detected and blocked by Proposed Methodology in Netcraft

Figure 9. Fake website detected by Proposed Methodology in Google Safe Browsing

Figure 10. Detection of Fake website by Google Safe browsing

Figure 11. High risk Fake website displayed but not blocked by Netcraft

5. Conclusion

Detecting the phishing attackers has been proposed by using Phishing Detection Mechanism. In which Phishing Detection Mechanism performing a three phases identifying mechanism, it is easy to easily identify the DNS blacklist testing for frequently using phishing IP. Using the web crawler and analysis phase phishing E-mail & sites are identified. Proposed experimental analysis had been done for Phishing Detection Mechanism and it is used for precisely detecting the websites which are phishing as Phishing Detection Mechanism produces a detecting best accuracy. Three different classifiers were used to find out the classification accuracy in order to detect the phishing, which yields the 95.18%, 85.45%, 78.89%, by NN, SVM, RF classifications respectively. The results suggested and recommended for detecting the phishing by machine learning approach.

References

1. K. D. Rajab, ‘‘New hybrid features selection method: A case study on Websites phishing,’’ Secur. Commun. Netw., vol. 2017, Mar. 2017.
2. Khonji, M., Iraqi, Y. and Jones, A. (2013) ‘Phishing detection: a literature survey’, IEEE Communications Surveys and Tutorials, Vol. 15, No. 4, pp.2091–2021.
3. S. Lee and J. Kim, ‘‘Warning bird: A near real-time detection system for suspicious URLs in Twitter stream,’’ IEEE Trans. Dependable Secure Comput., vol. 10, no. 3, pp. 183–195, May 2013.

4. Y.Zhang, J.I.Hong, andL.F.Cranor,"Cantina:a contentbased approach to detecting phishing websites,"in Proceedings of the 16th International World Wide Web Conference (WWW'07),pp. 639–648,Banff,Canda,May2007.
5. M. Babagoli, M. P. Aghababa, and V. Solouk, "Heuristic nonlinear regression strategy for detecting phishing websites," Soft Comput., pp. 1–13, 2018.
6. H. Kim and E. A. Lee, "Authentication and Authorization for the Internet of Things," IT Prof., vol. 19, no. 5, pp. 27–33, 2017.
7. T. Peng, I. Harris, and Y. Sawa, "Detecting Phishing Attacks Using Natural Language Processing and Machine Learning," Proc.-12th IEEE Int. Conf. Semant. Comput. ICSC 2018, vol. 2018– Janua, pp. 300–301, 2018.
8. M. Aburrous, M. A. Hossain, K. Dahal, and F. Thabtah, "Intelligent phishing detection system for e-banking using fuzzy data mining," Expert Syst. Appl., vol. 37, no. 12, pp. 7913–7921, 2010.