International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

PhishAlert: An Efficient Phishing URL Detection via Hybrid Methodology

Bhawna Sharma, Parvinder Singh

Abstract: ***In spite of various research endeavors, phishing assaults stay common and exceedingly successful in attracting clueless clients to uncover delicate data, including account details and government managed savings numbers. Misfortunes due to phishing are developing consistently. A solitary methodology isn't effective for distinguishing a wide range of phishing assaults. So we propose a hybrid approach to deal with the classification of URLs as phishing or real. The investigation aftereffects of our proposed methodology, in view of a dataset gathered from phishing and legitimate URLs, have demonstrated that PhishAlert framework can successfully counteract phishing assaults and can thus ensure system security.***

Keywords: Phishing, Whitelist, Heuristics, Style Similarity, Hybrid Approach

I. INTRODUCTION

Phishing is a crime that takes sensitive information of users by misdirecting messages or phony sites [2], [5], [6]. "Phishing" is word was taken from the word "fishing". The term phishing came into record in 1996 in America. Fishers (for example assailants) utilize a snare (for example counterfeit email enticing clients for entering touchy data) to get a fish (for example to befool a client). Telephone Phreaking was the oldest type of hacking. So the character "f" of fishing was substituted by "ph" and the expression "phishing" came into picture. Online clients can be effectively beguiled into entering their own data in light of the fact that phishing sites are exceedingly like genuine ones.

Perniciously, by making phishing locales, "phishers" utilize various systems to trick their targeted people, including email messages, texts, discussion posts, telephone calls, and long range informal communication data. Phishing brings about serious monetary misfortune everywhere throughout the world, and phishing destinations are additionally developing quickly in amount and multifaceted nature [3]. The phishing assaults can be completed from various perspectives like email, SMS, voice, malware, website etc.

Revised Manuscript Received on October 30, 2019.
***** Correspondence Author
Bhawna Sharma*, Research Scholar, Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Sonepat-131039, Haryana, India, Email: bhawnash024@gmail.com
Dr. Parvinder Singh, Professor, Department of Computer Science & Engineering, Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Sonepat -131039, Haryana, India, Email: parvinder23@rediffmail.com

In this work, PhishAlert algorithm is used to identify phishing URLs. As indicated by the Anti-Phishing Working Group (APWG), the number of phishing locales recognized in first quarter of the year 2019 was 1.81lac,

Which was up outstandingly from the 1.38 lac seen in fourth quarter of the year 2018, and from the 1.51 lac seen in the third quarter of the year 2018. Phishing that focused mail services and Software-as-a-Service (SaaS) turned into the greatest classification of phishing. At 36 percent of all phishing assaults, it obscured phishing against the payment services category for the first time [39].

Some of the widely used anti-phishing techniques are listed below [7]:

- **List-based techniques-** Majority of the web-browsers (e.g. Google Chrome, Internet Explorer, Mozilla Firefox etc) utilize list based techniques. Blacklist and whitelist are two main types of list-based techniques. Whitelist is a list of all genuine URLs [1]. If the input URL is available in the whitelist, then the user can safely visit the URL. Because of this conduct, even the genuine sites which are not present in the whitelist are likewise choked bringing about high False Positive Rate. The blacklist contains phishing URLs which are hindered by the web browsers. Because of this conduct, the phishing URLs which are not available in the blacklist are allowed to visit by the user. This leads to high False Negative Rate. List based techniques are prone to zero-hour attacks since the most recent phishing site or the most recent legitimate site takes time for getting updated in blacklist or whitelist. Whitelists and blacklists are effective only if they are updated timely.
- **Heuristic-based technique-** This technique extracts features from the phishing site and uses these features to detect phishing attack [9], [12], [13], [14], [16], [23]. Heuristic method can detect zero-day phishing attacks [8]. Low false negatives and low false positives are achieved by using this technique. But this technique has less classification accuracy because all phishing sites do not have common features.
- **Visual Similarity-based Technique-** The user is tricked by the attacker by creating a website that has an analogous look as that of the genuine website [10], [11], [26]. This technique compares the image of the suspicious webpage with all the images present in the database containing legitimate webpages. The suspicious site is classified as phishing site when the parameter describing site-similarity comes out to be greater than threshold; otherwise the website is declared as legitimate. The comparison of images is a time consuming process. The consumption of storage becomes high due to the creation of a database containing images of legitimate webpages.

- **Machine learning-based techniques:** These days, majority of the analysts are utilizing Machine Learning (ML) algorithms [15], [19], [20], [38] for classifying sites as phishing or legitimate.

These systems are a blend of Machine Learning algorithms and heuristic strategies. Some ML methods are: AdaBoostM1, Sequential Minimum Optimization (SMO), Support Vector Machine (SVM), J48 Tree etc. These algorithms can detect zero-hour phishing attack. These algorithms are capable of handling large datasets efficiently. The performance of ML based techniques relies on the type of classifiers, training-data's size, features used and feature-set's size.

Phishing is complicated cyber theft, and it is using a variety of phishing attacks [27], [28]. A single approach can't identify all phishing sites efficiently [22], [30]. PhishAlert identifies phishing sites using a hybrid approach that combines whitelist, heuristics, content similarity and style similarity.

The primary objectives of this paper are:

- To put forward a vigorous solution to recognize phishing URLs rendering to a hybrid approach named PhishAlert which makes use of whitelist, heuristics, content similarity and style similarity.
- To prototype PhishAlert and to perform its evaluation on an enormous dataset.
- To compare PhishAlert with existing methods which are used for classifying sites as: phishing or legitimate.

The remaining paper is structured in this manner: Section II organizes the related work. Section III describes the proposed work. Section IV illustrates experimentation and results. The conclusion is deliberated in Section V.

II. RELATED WORK

Some investigations have tended to the issue of phishing lately. The key elements of every method are investigated in this section.

Blacklists are every now and again refreshed arrangements of recently distinguished phishing sites, locations of Internet Protocol (IP) or catchphrases [2, 7]. Google Safe Browsing API enables customer to receive warnings for phishing sites. It uses two blacklists namely goog-phishshavar (for phishing) and goog-malware-shavar (for malware).

Whitelists are arrangements of checked authentic URLs. Ankit et al. [1] proposed an automated whitelist that keeps up a whitelist of the features depicting trusted Login User-Interfaces (LUIs). The automated whitelist contains all trusted LUIs.

Wenyin Liu et al. [10] anticipated a method that utilized visual-layout based attributes to distinguish potential phishing websites and measure their similarity with legitimate sites. It developed a framework containing two procedures. The first procedure kept running on nearby mail servers and screened messages for watchwords and suspicious URLs. The second procedure looked at the phishing pages and detemined visual-layout based likenesses between both of them as far as key areas, page designs, and generally styles. Kuan-Ta Chen et al. [11] introduced an anti-phishing method based on discriminative keypoints. This method used Contrast Context Histogram (CCH) to calculate the visual-similarity between phishing site and legitimate site.

S.Carolin Jeeva [14] proposed an approach based upon Intelligent Phishing in URLs detection which extricated the important highlight that impact on authentic and phishing URLs. Further the identified highlights are exposed to apriori based algorithm and predictive apriori based algorithm in associative rule mining. The predominant highlights are considered. The proposed algorithm identified important feature i.e. security in Transport layer, absence of URL's Top Level Domain and availability of reserve-word in the URL.

Zhang et al. [12] proposed CANTINA algorithm, based upon HTML content technique to recognize phishing sites and also examines the website page code and utilizes TF-IDF to identify the keywords which frequently occurred. The domain name and the frequent keywords are combined to form a search engine query. If no links are returned by the search query then the website is declared as phishing. Xiang et al. proposed the CANTINA+ algorithm [13]. This algorithm was reformatted from CANTINA algorithm. CANITINA+ added new features to the feature set of CANTINA. CANTINA+ used a rich feature set and applied machine learning techniques to identify phishing sites.

Jail-Phish [18] method utilized URL as info and showed the result as genuine or phishing. This method works for all languages and is independent of web history. This method was tried on old phishing set and new phishing set. It considered a large size dataset and achieved True Positive Rate of 97.92%.

Routhu Srinivasa Rao et al. [19] proposed CatchPhish to identify phishing websites depending on the highlights extricated from the URL of a given site. CatchPhish utilized two types of feature sets. One was based upon URL while other was based upon TF-IDF. TF-IDF algorithm was based upon Term Frequency- Inverse Document Frequency. Random Forest algorithm was utilized to identify phishing websites. On an enormous dataset, Catch-Phish accomplished 94.26% accuracy.

Rao and Pais [20] anticipated a method to identify phishing sites utilizing heuristic-based approach. The highlights utilized in this model to identify phishing sites don't rely upon image database or web history. This technique detected the phishing websites using URL features (highlights). This model accomplished high accuracy of 99.55% utilizing Random Forest algorithm.

Phishing-Aware [21] used neuro-fuzzy approach. It structured an anti-phishing model, named Fi-NFN, to ensure security for nearby devices effectively and rapidly. Without expending numerous assets from nearby devices, Fi-NFN model straightforwardly secured clients progressively, yet in addition improved the nature of services at the edge of the system.

PhiDMA [22] planned by Sonowal and Kuppusamy, represented a Phishing Detection Model with Multi-layered Approach. Auto-updated whitelist, URL features-extractor, lexical-signature generator, string matching algorithm and accessibility-score comparator were used to identify phishing sites. From the trial results model could identify phishing websites effectively with an accuracy of 92.72%.

Phishing-Alarm [25] introduced a model to measure the suspiciousness of websites using the similitude of visual appearance between the websites.

This methodology utilized Cascading Style Sheets (CSSs) as the premise to precisely evaluate the visual similitude of each page component.

Erzhou Zhu [29] et al. proposed OFS-NN, a viable phishing sites recognition model dependent on the optimal feature assortment and neural-network based method. Over-fitting problem was resolved using OFS-NN.

Peng et al. [35] introduced a methodology which utilized natural language processing to detect phishing attacks.

Patil et al. [37] proposed three methodologies for identifying phishing sites. First methodology was to examine different highlights of URL. Second methodology was to check the authenticity of site by finding the place where the site is being facilitated and the people who are overseeing it. Third methodology utilized visual layout based examination to check validity of site.

III. PROPOSED APPROACH

PhishAlert is an algorithm that classifies URL as phishing (fake) or legitimate (genuine). It is proposed on the idea that a single approach cannot deal with all types of phishing attacks. So a hybrid approach is a better solution for the phishing problem. PhishAlert makes use of whitelist, heuristics, content similarity and style similarity. PhishAlert takes URL as input. The URL is checked in a whitelist. Whitelist is a list containing legitimate URLs. We initialize the list to GoogleIndex. If the input URL is available in GoogleIndex whitelist, at that moment it is stated as legitimate; else it is passed to the next step. Features of URL are extracted in the next step. Phishing features are listed in Table-I. If the URL comprises of phishing features; then exit; else forward it to the next step. In the next step, find the top 5 frequent terms in the given webpage. Feed the URL along-with top 5 frequent terms on Google. If Google does not find any result, at that point the URL is declared to be a phishing URL and exit. If Google is able to return one or more results, then pass the input URL to the next step. The input URL is compared with search engine results URL on the basis of styling rules. CSS (Cascading Style Sheets) are used for styling webpages. Cosine similarity can be used as a similarity measure which can be computed using equation 1.

$$\text{cosine} - \text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|.\|B\|} \qquad (1)$$

Here A and B are two vectors and they will be similar to each other if they have high value for cosine similarity; θ is the angle between A and B. If the resulting percentage of the style-similarity of all URLs with the input URL comes out to be less than threshold, then affirm the site as phishing and exit. Otherwise forward the URL to the next step. If the URL is Legitimate then update the white list with this URL. The main steps of PhishAlert algorithm are summarized below.

Algorithm **PhishAlert**
INPUT: URL
OUTPUT: Class Label of URL (Phishing or Legitimate)

Step1: Verify the URL by whitelist filter. If the input URL is available in the GoogleIndex whitelist, then it can be visited by the user; otherwise it is passed to the next step.

Step2: Extract URL features. If the input URL comprises of the features shown in table 1, then declare the URL as phishing and exit, otherwise forward the URL to the next step.

Step3: Apply text cleaning on the URL content. Extract top five frequent terms from the URL content. Feed the URL along-with top 5 frequent terms on any of the search engines (say Google). If Google does not show any link as result; at that point affirm the URL as phishing and exit. If Google is able to return any link; at that point pass the URL to the next step.

Step4: Measure the style similarity of the input URL alongwith all the URLs reverted by Google as result. If the style match of any one of the Google result URL with the given URL comes out to be larger than threshold, at that moment declare the URL as a phishing URL and exit. Otherwise pass the URL to the next step.

Step5: If the URL is Legitimate then update the white list with this URL.

Table-I: Phishing Features

S.No.	PHISHING FEATURES
1	HOST URL LENGTH>20
2	COUNT OF SLASHES IN URL>6
3	COUNT OF DOTS IN URL-HOST>4
4	COUNT OF TERMS IN URL-HOST >4
5	COUNT OF DOTS IN URL-PATH >2
6	COUNT OF HYPHENS IN URL-HOST >1
7	URL-LENGTH>75
8	USE OF IP ADDRESS AS A SUBSTITUTE OF DOMAIN NAME
9	PRESENCE OF HTTP PROTOCOL
10	NON-EXISTENCE OF TOP LEVEL DOMAIN

IV. EXPERIMENTATION AND RESULTS

A. Performance Measures

Here PP is number of Phishing instances that are classified as Phishing instances. LP is number of Legitimate instances classified as Phishing. PL is number of Phishing instances that are classified as Legitimate. LL is number of Legitimate instances that are classified as Legitimate.

Table-II presents the confusion matrix [7].

Table-II: Confusion Matrix

Predicted TRUE	Declared as Phishing?	Declared as Legitimate?
Actually Phishing?	PP	PL
Actually Legitimate?	LP	LL

Different performance measures are [7], [18]:

1. **False Positive Rate (FPR)-** It computes the ratio of all legitimate instances that are classified as phishing to the total number of legitimate instances. $FPR = LP/(LL + LP)$ (2)
2. **True Positive Rate (TPR)-** It computes the ratio of all phishing instances that are classified as phishing to the total number of phishing instances.
   $TPR = PP/(PP + PL)$ (3)
3. **False Negative Rate (FNR)-** It computes the ratio of all phishing instances that are classified as legitimate to total number of phishing instances.
   $FNR = PL/(PP + PL)$ (4)
4. **True Negative Rate (TNR)-** It computes the ratio of all legitimate instances that are classified as legitimate to the total number of legitimate instances.
   $TNR = LL/(LL + LP)$ (5)
5. **Accuracy (*ACC*)-** It quantifies the general rate of accurately recognized phishing and authentic occasions in connection to all examples.
   $ACC = (LL + PP)/(LL + LP + PL + PP)$ (6)
6. **Precision (*P*)-** It computes the ratio of all phishing instances that are classified as phishing to the total number instances that are identified as phishing.
   $P = PP/(LP + PP)$ (7)
7. **Recall (*R*)-** It is same as True Positive Rate
   $R = TP$ (8)
8. ***f*1-Score-** The Harmonic Mean of Precision and Recall is termed as f1-score.
   $f1 - Score = (2 * P * R)/(P + R)$ (9)

B. Results and Discussion

PhishAlert algorithm is used to classify every URL in the dataset as phishing or legitimate. PhishAlert algorithm has been implemented in Python.

An experiment was performed using a dataset containing 1000 URLs. Out of these, 500 URLs were legitimate and 500 were phishing. The legitimate URLs were collected from stuffgate database whereas the phishing URLs were collected from phishtank database as shown in Table-III.

Table-III: Database used

S.No.	Database	Links	Instances
1.	Stuffgate	http://stuffgate.com/stuff/website/top-sites	500
2.	phishtank	https://www.phishtank.com/	500

Table-IV shows the confusion matrix used in the experiment. From the confusion matrix it is clear that PhishAlert correctly classifies all the legitimate instances but incorrectly classifies 25 phishing instances. So TPR=0.95, FPR=0.05, TNR=1, FNR=0 and Accuracy=0.975. Table-V shows the values computed in the experiment for measuring precision, recall and f1-score. The average precision of phishing and legitimate instances comes out to be 0.975. The average recall value results into 0.975. The average f1-score becomes 0.975. The comparison of PhishAlert with existing approaches is presented in Table-VI. It can be concluded from Fig. 1 that the average value for all three performance measures (precision, recall and f1-measure) is greater than 0.9 which is quite satisfactory. PhishAlert gives highest value for all the three performance measures (precision, recall and f1-score) as compared to CANTINA [12] and CANTINA + [13]. Fig. 2 shows that PhishAlert is more efficient algorithm as compared to CANTINA [12] and CANTINA+ [13].

Table-IV: Confusion Matrix used in Experiment

n = 1000	Predicted: Legitimate	Predicted: Phishing
Actual : Legitimate	500	0
Actual: Phishing	25	475

Table -V: Precision Recall and f1-score for the experimental dataset

Instances	Precision	Recall	F1-score
Phishing	1	0.95	0.97
Whitelist	0.95	1	0.98
Average	0.975	0.975	0.975

Table-VI: Comparison of PhishAlert with CANTINA and CANTINA+

Approaches	Precision	Recall	F1-score
PhishAlert	0.98	0.97	0.97
CANTINA	0.942	0.97	0.956
CANTINA +	0.975	0.934	0.963

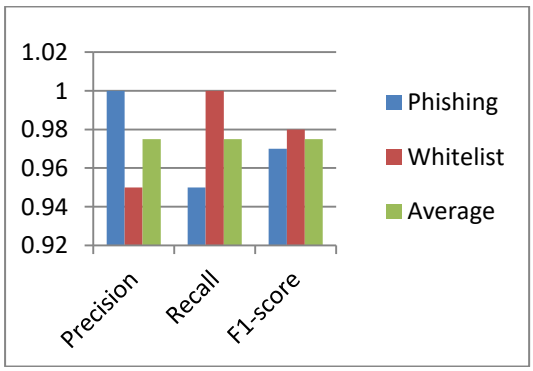

Fig. 1. Precision, Recall and f1-measure for PhishAlert

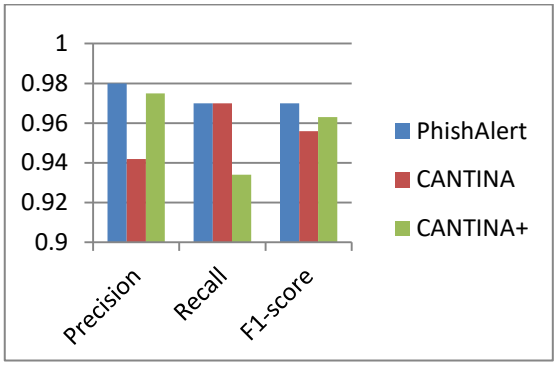

Fig. 2. Comparison of PhishAlert with CANTINA and CANTINA+ on the basis of precision, recall and f1-measure

V. CONCLUSION

Phishing is said to be a noteworthy cyber theft which has been making money related misfortune for the individual clients and associations. It is very complicated issue, and a solitary identification prototype is not able to differentiate all classes of phishing attacks. PhishAlert model was designed in this paper to recognize phishing i.e. fake URLs by utilizing hybrid approach which is a combination of various methodologies to be like whitelist, heuristics, content similarity and style similarity. PhishAlert performed more efficiently as compared to existing algorithms, namely, CANTINA and CANTINA +, on the experimental dataset that contained 500 phishing instances and 500 legitimate instances. PhishAlert was found to be 98% precise. This outcome demonstrates that the model could effectively distinguish all phishing URLs. PhishAlert model's performance decreases with increment in size of dataset. More features ought to be incorporated into the future work to the PhishAlert algorithm so as to achieve better classification rate.

REFERENCES

1. Ankit Kumar Jain and B. B. Gupta, "A novel approach to protect against phishing attacks at client side using auto-updated white-list", *EURASIP Journal on Information Security*, *Springer Open*, Vol. 2016, No. 1, pp. 1-11, Dec. 2016
2. Rami M. Mohammad , Fadi Thabtah and Lee McCluskey, "Tutorial and critical analysis of phishing websites methods", *Computer Science Review*, *Elsevier*, Vol. 17, No. 1, pp. 1-24, Aug. 2015
3. K. Nirmal, B. Janet and R. Kumar, "Phishing - The threat that still exists", *International Conference on Computing and Communications Technologies, IEEE*, pp. 139-143, 2015
4. Cristian Iuga, Jason R. C. Nurse and Arnau Erola, "Baiting the hook: factors impacting susceptibility to phishing attacks", *Human-centric Computing and Information Sciences, SpringerOpen*, Vol. 6, No. 1, pp. 1-20, Dec. 2016
5. Dr. M. Nazreen Banu S. Munawara Banu," A Comprehensive Study of Phishing Attacks", *International Journal of Computer Science and Information Technologies*, vol. 4 , no. 6, pp. 783-786, 2013
6. Elmer EH Lastdrager, "Achieving a consensual definition of phishing based on a systematic review of the literature", *Crime Science, Springer Open Journal*, pp. 1-10, 2014
7. Mahmoud Khonji, Youssef Iraqi, and Andrew Jones, "Phishing Detection: A Literature Survey", *IEEE Communications Surveys & Tutorials*, vol. 15, No. 4, pp. 2091-2121, 2013
8. Narendra. M. Shekokar, Chaitali Shah, Mrunal Mahajan and Shruti Rachh, "An Ideal Approach for Detection and Prevention of Phishing Attacks", *Proceedings of 4th International Conference on Advances in Computing, Communication and Control, Elsevier*, Vol. 49, pp. 82 -91, 2015
9. Weibo Chu, Bin B. Zhu, Feng Xue , Xiaohong Guan, Zhongmin Cai, "Protect Sensitive Sites from Phishing Attacks Using Features Extractable from Inaccessible Phishing URLs", *IEEE International Conference on Communications, IEEE Communication and Information Systems Security Symposium*, pp. 1990-1994, 2013
10. Wenyin Liu, Xiaotie Deng, Guanglin Huang, and Anthony Y. Fu, "An Antiphishing Strategy Based on Visual Similarity Assessment", *IEEE Internet Computing*, Vol. 10, No. 2, pp. 58-65, March-April, 2006
11. Kuan-Ta Chen, Chun-Rong Huang, and Chu-Song Chen, "Fighting Phishing with Discriminative Keypoint Features", *IEEE Internet Computing*, Vol. 13, No. 3, pp. 56-63, May-June, 2009
12. Yue Zhang, Jason Hong, Lorrie Cranor, "CANTINA: A Content-Based Approach to Detecting Phishing Web Sites", *ACM Proceedings of 16th international conference on World Wide Web*, pp. 639-648, May 2007
13. Guang Xiang, Jason Hong, Carolyn P. Rose, Lorrie Cranor, "CANTINA+: A Feature-rich Machine Learning Framework for Detecting Phishing Web Sites", *ACM Transactions on Information and System Security* , Vol. 14, No. 2, pp. 1-32, Sept. 2011
14. S.Carolin Jeeva and Elijah Blessing Rajsingh, "Intelligent Phishing URL Detection using Association Rule Mining", *Human Centric Computing and information Sciences, SpringerOpen* , Vol. 6, pp. 1-19, 2016
15. Mahmood Moghimi , Ali Yazdian Varjan, "New Rule-Based Phishing Detection Method", *Expert Systems with Applications, Elsevier*, Vol. 53, pp. 231-242, July 2016
16. Wa'el Hadi, Faisal Aburub, Samer Alhawari, "A New Fast Associative Classification Algorithm for Detecting Phishing Websites", *Applied Soft Computing, Elsevier*, Vol. 48, pp. 729-734, Nov. 2016
17. Hossein Shirazi, Kyle Haefner, Indrakshi Ray, "Fresh-Phish: A Framework for Auto-Detection of Phishing Websites", *IEEE International Conference on Information Reuse and Integration*, pp. 137-143, 2017
18. Routhu Srinivasa Rao, Alwyn Roshan Pais, "Jail-Phish: An Improved Search Engine Based Phishing Detection System", *Computers & Security, Elsevier*, Vol. 83, pp. 246-267, June 2019
19. Routhu Srinivasa Rao, Tatti Vaishnavi, Alwyn Roshan Pais, "CatchPhish: Detection of Phishing Websites by Inspecting URLs", *Journal of Ambient Intelligence and Humanized Computing, Springer*, Vol. 10, No. 5, pp. 1-13, May 2019
20. Routhu Srinivasa Rao, Alwyn Roshan Pais, "Detection of Phishing Websites Using an Efficient Feature-Based Machine Learning Framework", *Neural Computing and Applications, Springer*, pp. 1-23, Jan. 2018
21. Chuan Pham, Luong A. T. Nguyen, Nguyen H. Tran, Eui-Nam Huh, Choong Seon Hong, "Phishing-Aware: A Neuro-Fuzzy Approach for Anti-Phishing on Fog Networks", *IEEE Transactions on Network and Service Management*, Vol. 15, No. 3, pp. 1076 - 1089, Sept. 18
22. Gunikhan Sonowal, K.S. Kuppusamy, "PhiDMA - A Phishing Detection Model with Multi-filter Approach", *Journal of King Saud University - Computer and Information Sciences*, pp. 1-14, July 2017
23. Lew May Form, Kang Leng Chiew, San Nah Sze, Wei King Tiong, "Phishing Email Detection Technique by using Hybrid Features", *9th International Conference on IT in Asia*, pp. 1-5, Dec. 15
24. Ibrahim Waziri Jr., "Website Forgery: Understanding Phishing Attacks & Nontechnical Countermeasures", *IEEE 2nd International Conference on Cyber Security and Cloud Computing*, pp. 445-450, 2015
25. Jian Mao, Wenqian Tian, Pei Li, Tao Wei, Zhenkai Liang, "Phishing-Alarm: Robust and Efficient Phishing Detection via Page Component Similarity", *IEEE Access*, Vol. 5, pp. 17020 – 17030, Aug. 2017
26. S.Nisha, Dr.A.Neela Madheswari, "Prevention of Phishing Attacks in Voting System using Visual Cryptography", *International Conference on Emerging Trends in Engineering, Technology and Science*, pp. 1-4 , Oct. 2016
27. Dr. M. Nazreen Banu, S. Munawara Banu, "A Comprehensive Study of Phishing Attacks", *International Journal of Computer Science and Information Technologies*, Vol. 4, pp. 783-786, Jan. 2013
28. Srushti Patil, Sudhir Dhage, "A Methodical Overview on Phishing Detection along with an Organized Way to Construct an Anti-Phishing Framework", *5th International Conference on Advanced Computing & Communication Systems*, pp. 589-593, June 2019

29. Erzhou Zhu, Yuyang Chen, Chengcheng Ye, Xuejun Li, Feng Liu, "OFS-NN: An Effective Phishing Websites Detection Model Based on Optimal Feature Selection and Neural Network", *IEEE Access*, Vol. 7, pp. 73271 - 73284, June 2019
30. Muhammet Baykara, Zahit Ziya Gürel, "Detection of Phishing Attacks", *International Symposium on Digital Forensic and Security*, pp. 1-5, May 2018
31. Samar Muslah Albladi, George R. S. Weir, "User Characteristics that Influence Judgment of Social Engineering Attacks in Social Networks", *Human-centric Computing and Information Sciences, SpringerOpen*, Vol. 8, No. 1, pp. 1-24, Dec.2018
32. Tore Pedersen, Christian Johansen, Audun Josang, "Behavioural Computer Science: An Agenda for Combining Modelling of Human and System Behaviours", *Human-centric Computing and Information Sciences, SpringerOpen*, Vol. 8, No. 1, pp. 1-20, Dec. 2018
33. Wenbin Yao, Yuanhao Ding, Xiaoyong Li, "Deep Learning for Phishing Detection", *2018 IEEE International Conference on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications*, pp. 645-650, 2018
34. Jhen-Hao Li, Sheng-De Wang, "PhishBox: An Approach for Phishing Validation and Detection", *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress*, pp. 557-564, 2017
35. Tianrui Peng, Ian G. Harris, Yuki Sawa, "Detecting Phishing Attacks Using Natural Language Processing and Machine Learning", *12th IEEE International Conference on Semantic Computing*, pp. 300-301, 2018
36. Rana M. Amir Latif, Muhammad Umer, Tayyaba Tariq, Muhammad Farhan, Osama Rizwan, Ghazanfar Ali, "A Smart Methodology for Analyzing Secure EBanking and E-Commerce Websites", *16th International Bhurban Conference on Applied Sciences and Technology*, pp. 589-596, Mar. 2019
37. Vaibhav Patil, Pritesh Thakkar, Chirag Shah, Tushar Bhat, S. P. Godse, "Detection and Prevention of Phishing Websites using Machine Learning Approach", *Fourth International Conference on Computing Communication Control and Automation*, pp.1-5, 2018
38. Prasanta Kumar Sahoo, "Data mining a way to solve Phishing Attacks", *IEEE International Conference on Current Trends toward Converging Technologies*, pp. 1-5, 2017
39. https://docs.apwg.org/reports/apwg_trends_report_q1_2019.pdf

AUTHORS PROFILE

Bhawna Sharma is currently pursuing Ph.D from Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Sonepat, Haryana,India.. She received her M.Tech. degree in Computer Science & Engineering from ITM University, Gurgaon in 2013. Also, she holds the degree of B.Tech. in Computer Science & Engineering from PDM College of Engineering, Bahadurgarh affiliated to Maharishi Dayanand University, Rohtak. She is working as Assistant Professor in JMIT College, Radaur. Her research interests include cyber security and machine learning.
Email:bhawnash024@gmail.com
(Corresponding author)

Dr. Parvinder Singh is Dean and Chairperson in the department of Computer Science & Engineering at Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Sonepat, Haryana,India. He holds Ph.D degree from Maharishi Dyanand University, Rohtak. He pursued M.Tech(CSE) from Guru Jambheshwar Univeristy, Hisar . He pursued B.E.(Electronics) from Baba Saheb Ambedkar Marathwara University, Aurangabad (STB College of Engg, Tuljapur). He has published many research papers in journals of reputed publishers and also associated with review and editing work with many journals.
Email: parvinder23@rediffmail.com