ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 10, Issue 3, March 2025

Decentralized Anti-Spam OTP Network Using Block Chain Based Fraud Detection

[1] Manivel, [2] S.Jagadish, [3] S.Sudharsan, [4] A.M.Vignesh
[1] Assistant Professor, [2] [3] [4] Students, Department of Information Technology
Muthayammal Engineering College (Autonomous), Rasipuram-637408, TamilNadu, India
[1] manivelmecittzof@gmail.com, [2] Jagadishmecit@gmail.com, [3] Sudharasanmecit@gmail.com,
[4] vigneshamitmec@gmail.com

Abstract: ***Cyber criminals utilize the phishing tactic to seem to be reputable websites in order to collect personal data. The unique light weight phishing detection method presented is entirely based on the URL (uniform resource location). The SVM (support vector machine) evaluated record of phishing URLs, generates a fairly satisfactory recognition rate. Numerous books in the literature addressed the phishing assault. However, because to their complicated computation and high energy consumption, those systems are not the best for smartphones and other embedded devices. Just six URL features are needed OTP by the Suggested algorithm to complete the recognition. The features that are specified include the size of the URL, the number of dots, hyphens, and numeric characters, along with a corresponding discrete variable to the IP address contained in the URL, and lastly, the similarity index. The similarity index, a feature we present for the first time as input to the phishing detection systems, enhances the overall prediction rate, as demonstrated by the study's result.***

Keywords: Fake Review Products, Detections, cyber security, community computing, cybercrime

I. INTRODUCTION

In recent years, Due to its The project entitled “three-factor-based data transmission authentication scheme (TDTAS)” Most of the people requires 5G Communications information about the online product but the security of the next-generation wireless mobile communication technology, namely, 5G with IoT, has been a field of much interest among researchers in the last several years Before spending their economy on particular product can analyses the various blogs or web link Online purchases . Every day banks need to perform many activities related to users which needs huge fake phone call spam fake opt infrastructure with more online web applications etc. Almost everybody faces issues with online shopping, primarily because you need to make a purchase decision without actually looking at the product or trying it out. That admin providing for system needs to be more secure and reliable because each and every task performed is related to customers or End User money in online Transactions. Especially authentication and validation of user access is the major task in the online Product security systems. In this project,5G the system has provided the secure transactions find fake distributed user to the respective web server. This system will be accessible to all users who have a valid Secure transactions unfind out product. The user can view the Product detail when find & the fake address. The main objective is to create a secure online product purchase security providing the main server System and increase the users secure.as web usage has continued to grow, we have become more and more dependent on email services. In particular, the bulk delivery of unwanted messages—mostly business-related, but also containing harmful content or phony intentions—has become the main problem with email services for Internet service providers (ISPs), corporations, and individual users. Continuous summaries revealed that Phishing accounts for nearly 60% of all Email traffic. Phishing leads to email systems experiencing excessive transmission capacity and server storage limits, increasing the annual cost to businesses by several billions of dollars. Furthermore, because phishing communications attempt to obtain client data, they pose a serious threat to their security asking them to give over personal information such as their stick number and record number, using spoof communications that seem to be coming from reliable websites, such financial foundations. Phishing or non-Phishing messages are both possible. Phishing emails are also known as unsolicited or junk mail, even if non-phishing emails are legitimate and meant for a specific person or purpose. Devices and computations to handle content records in their information vector frame are provided by data recovery. The quantity of phishing statistics is rising. Extreme problems arise from Phishing communications, such as time wastage, system asset (data transfer capacity) waste, PC damage from malware, and moral dilemmas, such as Phishing communications promoting pornographic locations that are harmful to young people.

1.1 Problem Statement

Much research has been done on phishing detection strategies. The heuristic-based approach and the blacklist-based detection method are common phishing detection strategies. A consistent list of websites that are flagged as phishing sites is kept up to date using the blacklist-based method; if a user requests a page and it appears in the list, the connection is refused. This method is widely employed and has a low false-positive rate; yet, the quality of the list that is kept determines how accurate it is. As such, one of its drawbacks is that it can't identify transient phishing websites. The heuristic-based detection method uses information gleaned from the analysis and extraction of phishing site attributes to identify phishing sites. To suggest a fresh approach to phishing detection based on heuristics that addresses the drawbacks of the blacklist-based method. We put the suggested method into practice and evaluated its performance through experimentation. The suggested method determines whether a requested site is a phishing site by extracting features from the URLs of pages that users request and applying those features. This method can help lessen the harm caused by phishing assaults since it can identify phishing websites that blacklist-based methods are unable to identify.

II. SURVEY OF DETECTIONS:

2.1. Survey of review Phishing detection using machine learning techniques- Online reviews are a great source of information that can be used to ascertain the general public's opinion on items or services, and they are frequently the main deciding factor for customers when making a purchase. Manufacturers and retailers are very worried about customer feedback and reviews because of their impact. A dependence on internet evaluations raises the possibility that dishonest people would fabricate reviews in order to fraudulently promote or minimize goods and services. Opinion (review) phishing is the activity of manipulating and poisoning reviews (i.e., creating fictitious, dishonest, or misleading evaluations) for financial advantage. It's critical to have methods for spotting review phishing as not all internet reviews are reliable and truthful. By obtaining significant characteristics from the text using Natural Language Processing (NLP), a review of Phishing detection can be carried out with different machine learning methods. Aside from the content itself, reviewer information can also be utilized to help in this process. In this work, we examine the popular machine learning methods that have been put out to address the issue of review phishing detection as well as the effectiveness of various strategies for review phishing classification and detection. Most recent work has concentrated on supervised learning techniques, which necessitate labeled data, which is hard to get by in online review phishing. Given the millions of online evaluations that exist and the millions more that are created every day, research on Big Data techniques is interesting. We have not yet located any papers that investigate how big data analytics might be used to review phishing detection. This paper's main objective is to present a thorough and robust comparison of recent studies on the detection of review phishing using different machine learning approaches and to develop a methodology for carrying out additional research.

2.2 Fast and effective clustering of Phishing URL's based on structural similarity-

Phishing URLs cost businesses and individual users a great deal of money, time, and storage space every year. Locating and prosecuting Phishing URL's perpetrators as well as its eventual stakeholders should enable direct attack of the issue's core cause. In this research, we offer a methodology to quickly and effectively partition vast amounts of Phishing URLs into homogeneous campaigns using classification. This will help facilitate a challenging analysis that needs to be performed on big quantities of unclassified raw URLs structural resemblance. The framework makes use of the category Clustering Tree (CCTree), a revolutionary category clustering algorithm, and a set of 21 attributes that are typical of the email structure. The approach is assessed and verified using common tests carried out on three datasets containing more than 200k authentic, current phishing URLs.

2.3. Cosdes: A collaborative Phishing detection system with a novel e-mail abstraction scheme.- hese days, email communication is essential, yet the issue of email phishing keeps getting worse. The major goal of the similarity matching method for phishing detection is to prevent phishing attempts by keeping track of known phishing sites created through user feedback. Previous works mostly portray each email by a brief abstraction taken from the body of an email. Nevertheless, these email abstractions are insufficiently effective in near-duplicate detection because they fail to capture the dynamic nature of phishing attempts. In this work, we suggest a unique email abstraction method that uses the structure of emails as a representation of emails. We provide a process to create the email abstraction from HTML content in emails, and this newly created abstraction is better able to represent the Phishing's near-duplicate phenomenon. Additionally, we create a comprehensive Phishing detection system called COsdes (Collaborative Phishing Detection System), which has a progressive update strategy and an effective near-duplicate matching technique. The system Cosdes are able to maintain the most recent

data for near-duplicate detection thanks to the progressive updating scheme. We assess Cosdes using real-time data data gathered from an actual email server and demonstrate how our system performs better in real-world applications and detection results than previous methods.

2.4. Apache Mahout: Scalable machine learning and data mining.-Building scalable machine learning libraries is Mahout's mission. Scalable to reasonably large data sets is what we mean when we say scalable. We use the map/reduce paradigm to construct our key algorithms for batch-based collaborative filtering, clustering, and classification on top of Apache Hadoop. We do not, however, limit contributions to Hadoop-based implementations; contributions running on a single node or on a cluster that is not based on Hadoop are also welcome. The core libraries have undergone extensive optimization to enable strong performance even with non-distributed algorithms. scalable to back up your commercial argument. * Scalable: Mahout is offered under an Apache Software license that is beneficial to businesses community. Building a dynamic, responsive, and diverse community is Mahout's aim in order to promote conversations about possible use cases as well as the project itself. Visit the mailing lists for additional information.

III. EXISTING SYSTEM:

The likelihood that each word in an email's priority value indicates that it is phishing is calculated using existing email categorization methods. However, in the actual situation, the likelihood of phishing any given word is independent of the likelihood of any other word, and the likelihood of phishing a pair of words is independent of the likelihood of phishing any one of the individual words. For instance, the terms "Bumper" and "Prize" are both ham words; yet, when they are combined, "Bumper Prize" will result in phishing, which is not assessed under the current criteria. Our Phishing Detection system can discriminate between Phishing and non-Phishing URLs based on a self-learning algorithm in accordance with the principles of memory, which makes the process of Phishing detection similar to how memory is generated in our brains developing. These phishing messages can be used for other attacks in addition to increasing memory capacity and network communication. The assault has the ability to either destroy the user's data or expose their identity.

3.1 Drawbacks:

- A Phishingmer may transmit more than 100,000 bulk URLs in an hour with very little money.
- Transmission and storage bandwidth are wasted by junk mail.
- The reason phishing is problematic is that we, the receiver, are made to bear the expense.
- Phishing URLs will hog disk space.
- Waste time, generate malicious virus, and have a major negative impact on users' phishing links.

IV. PROPOSED SYSTEM:

It is more difficult to handle electronic phishing when dealing with a large number of URLs in the recipient's inbox and shielding them from phishing URL attacks. It depends on how each recipient interprets the communication and how they plan to use email exchanges. An official or authoritative figure who used to take action against it can view a phishing attempt as a ham to the average person. Certain emails could also be considered phishing because they frequently utilize phrases associated with phishing, even if they are issued by the authorities in charge of control or with the noble intention of warning people against phishing.

To prevent these types of misclassifications and to rigorously guard against Phishing attacks with minimal training requirements The suggested approach is arrived at. This methodology will use the likelihood that multiple distinct terms will occur in an email and their likelihood of being phished to draw inferences about the email's legitimacy. The suggested methodology classifies emails using SVM classifiers in order to determine if they are phishing or legitimate. SVM primarily works to achieve two goals: first, it accurately classifies emails into ham and phishing URLs; second, it classifies emails based on the relative frequency of words that indicate ham or phishing, using an approach that ensures none of the recipient's healthy emails should be identified as phishing.

Generally speaking, SVM classifiers use training data to classify a group of objects to determine the type of data that falls into a particular category. It will classify it into the appropriate category if it discovers something similar throughout the testing process. To comprehend the underlying classification mechanism, the following is a description of the basic work function of such an NB classifier.

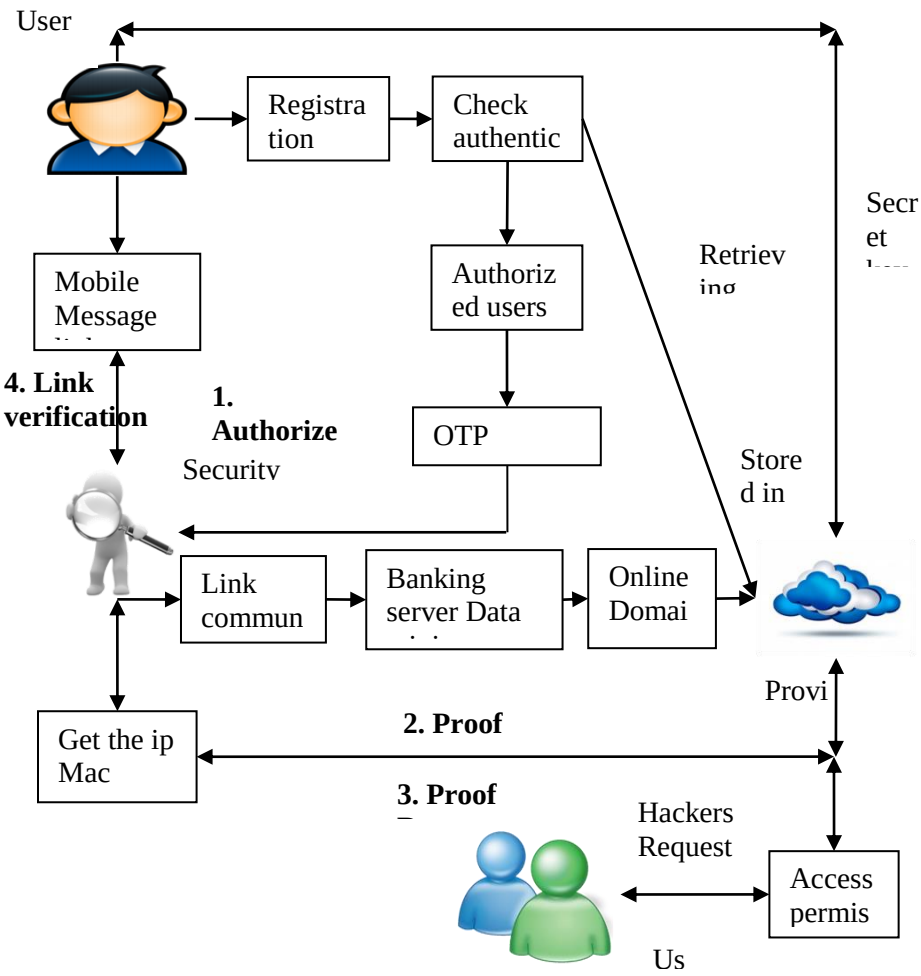
4.1 Advantages:

- Conserve storage and network bandwidth.
- Screen sent and received messages.

- Look for malware.

4.2 SYSTEM ARCHITECTURE

Phishing's are more dangerous and antagonistic for regular users. They also reduce system efficiency, slow down system transfer speeds, and cost businesses money. Therefore, every owner of a firm that uses email must process with the purpose of preventing Phishing from obtaining data through their email systems. Even while it might be challenging to stop every Phishing attempt, even a small amount of it can be stopped to lessen the harmful effects. With the ultimate goal of effectively sorting through spam and phishing emails, the suggested framework must be able to differentiate between typical phishing techniques and characteristics in order to distinguish phishing from legitimate emails. These methods are Best estimates and standards can be used to thwart these messages once they are known to the client. Since phishers are always improving their techniques, it's important to regularly implement new procedures to ensure that phishing is still effectively thwarted. Email headers and message body are the two areas of a message where phishing characteristics can be found.



V. MODULES

- Data set Acquisition
- Preprocessing
- Feature Selection
- Phishing Website Prediction

5.1 Data Set Acquisition

Please submit the datasets into this module. The phishing website is included in the dataset. Using a precompiled list of URLs from reputable and phishing websites, a classifier is created during the training phase.

5.2 Preprocessing

This module is used to remove noise, missing, or unnecessary data from the input. An essential phase in the data mining process is data pre-processing. The adage "garbage in, garbage out" is especially relevant to machine learning and data mining initiatives. A lot of the time, data collection techniques are not tightly controlled, which leads to missing values, impossible data combinations, and out-of-range numbers. Results from data analysis that hasn't been thoroughly checked for these issues may be deceptive.

5.3 Feature Selection

The feature extractor receives the gathered URLs and uses the predefined URL-based features to extract feature values. The taken out features are sent to the classifier generator, which uses the machine learning method and the stored features as input to create a classifier.

5.4 Phishing Website Prediction

The classifier ascertains if a requested website is a phishing site during the detection phase. A page request sends the requested site's URL to the feature extractor, which uses the predefined URL-based features to extract the feature values. The classifier receives certain feature values as input. Based on knowledge gained, the classifier decides if a new website is a phishing site. The person who requested the page is then informed of the classification outcome. The SVM algorithm is a straightforward probabilistic classifier that counts the frequency and combinations of values in a given set to determine a set of probabilities collection [4]. In this study, a text is represented as the bag of its words, and an SVM classifier uses these attributes to identify phishing emails. The bag of words is always utilized in document classification techniques, where the training classifier is trained using the frequency of occurrence of each word. The selected datasets have these bag of words attributes. The SVM approach was employed to ascertain the likelihood of phishing emails. Certain words are more likely to appear in non-phishing emails than in phishing emails. As an illustration, let's say we are certain that the word "free" will never appear in a legitimate email. We could then be certain that the email was phishing when we came across a message that contained this word. Bayesian Phishing Words like "free" and "viagra" have a very high likelihood of being phished, yet terms seen in non-phishing emails, including friend and family names, have a very low likelihood.

VI. CONCLUSION

SVM is a Phishing classifier that has a 99.5% classification accuracy on average. Additionally, it just needs a small amount of data—3.5 seconds—for training in order to achieve its standard performance. According to the study thus far, SVM's ability to relate the independent probabilities of terms inside an email's text suggests that it is a quick and accurate classifier. Combining independent probability of consecutive words in SVM offers a novel, moral method for classifying emails. in dataset while keeping the same accuracy will also aid in shortening the training dataset's development time. This study scheme lacks the functions of password change, inefficient login, and could not achieve user anonymity. Besides, the scheme could not guarantee session key security, which is vulnerable to several passive and active attacks. We proposed a three-party-based authentication scheme for 5G-enabled IoT environments along with a fuzzy extractor. The project title is “A Product link protection Schemes using for online purchase Security Systems” is a web based application. Every day online purchase need to perform many activities related to users which needs huge infrastructure with more End user ID etc. The security analysis results show TDTAS can resist most of the known attacks and security features. Unlike existing schemes, the formal security analysis of the proposed scheme has been proved under the RoR model. Moreover, the informal security analysis indicates that the scheme is secure and robust. The formal verification of our scheme has been done using a widely accepted AVISPA tool.

7.1 FUTURE ENHANCEMENT:

Obtaining precise categorization, with 0% of phishing emails being misclassified as ham emails and ham emails being misclassified as phishing emails. The attempts would be made to stop phishing emails, which are a greater cause for concern these days and carry phishing attacks. Additionally, the approach can be expanded to prevent Denial of Service (DoS) attacks, which are now known as Distributed Denial of Service Attacks (DDoS) because they occur in a distributed manner.

REFERENCE:

[1] R. Zhang, S. Cui, and C. Zhao, "A Three-Factor-Based Authentication Scheme of 5G Wireless Sensor Networks for IoT System," in Proc. Int. Conf. Commun. Signal Process. Syst., 2018, pp. 875–880.
[2] Y. Shi, Y. Zhao, R. Xie, and G. Han, "Designing a structural health monitoring system for the large-scale crane with narrow band IoT," in Proc. IEEE 23rd Int. Conf. Comput. Supported Cooper. Work Design (CSCWD), 2019, pp. 239–242.
[3] Y. Zhu, G. Jia, G. Han, Z. Zhou, and M. Guizani, "An NB-IoT-based smart trash can system for improved health in smart cities," in Proc. IEEE 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), 2019, pp. 763–768.
[4] D. He, S. Zeadally, B. Xu, and X. Huang, "An efficient identity-based conditional privacy-preserving authentication scheme for vehicular adhoc networks," IEEE Trans. Inf. Forensics Security, vol. 10, no. 12,
pp. 2681–2691, Dec. 2015.
[5] J. Zhang, J. Cui, H. Zhong, Z. Chen, and L. Liu, "PA-CRT: Chinese remainder theorem based conditional privacy-preserving authentication scheme in vehicular ad-hoc networks," IEEE Trans. Depend. Secure Comput., vol. 18, no. 2, pp. 722–735, Mar./Apr. 2021.
[6] P. Wang, C.-M. Chen, S. Kumari, M. Shojafar, R. Tafazolli, and Y.-N. Liu, "HDMA: Hybrid D2D message authentication scheme for 5G-enabled VANETs," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 8,
pp. 5071–5080, Aug. 2021.
[7] L. D. Xu, W. He, and S. Li, "Internet of Things in industries: A survey," IEEE Trans. Ind. Informat., vol. 10, no. 4, pp. 2233–2243, Nov. 2014.
[8] R. Amin, S. H. Islam, G. Biswas, M. K. Khan, L. Leng, and N. Kumar, "Design of an anonymity-preserving three-factor authenticated key exchange protocol for wireless sensor networks," Comput.
Netw., vol. 101, pp. 42–62, Jun. 2016.
[9] D. Abbasinezhad-Mood and M. Nikooghadam, "An anonymous ECCbased self-certified key distribution scheme for the smart grid," IEEE Trans. Ind. Electron., vol. 65, no. 10, pp. 7996–8004, Oct. 2018.
[10] A. K. Das, M. Wazid, N. Kumar, A. V. Vasilakos, and J. J. P. C. Rodrigues, "Biometrics-based privacy-preserving user authentication scheme for cloud-based Industrial Internet of Things
deployment," IEEE Internet Things J., vol. 5, no. 6, pp. 4900–4913, Dec. 2018.
[11] C. Stergiou, K. E. Psannis, B.-G. Kim, and B. Gupta, "Secure integration of IoT and cloud computing," Future Gener. Comput. Syst., vol. 78, pp. 964–975, Jan. 2018.
[12] C. Otto, A. Milenkovic, C. Sanders, and E. Jovanov, "System architecture of a wireless body area sensor network for ubiquitous health monitoring," J. Mobile Multimedia, vol. 1, no. 4, pp. 307–326, 2006.
[13] X. Li, J. Peng, M. S. Obaidat, F. Wu, M. K. Khan, and C. Chen, "A secure three-factor user authentication protocol with forward secrecy for wireless medical sensor network systems," IEEE Syst. J., vol. 14, no. 1, pp. 39–50, Mar. 2020.
[14] D. He, S. Zeadally, N. Kumar, and J. Lee, "Anonymous authentication for wireless body area networks with provable security," IEEE Syst. J.,vol. 11, no. 4, pp. 2590–2601, Dec. 2017.