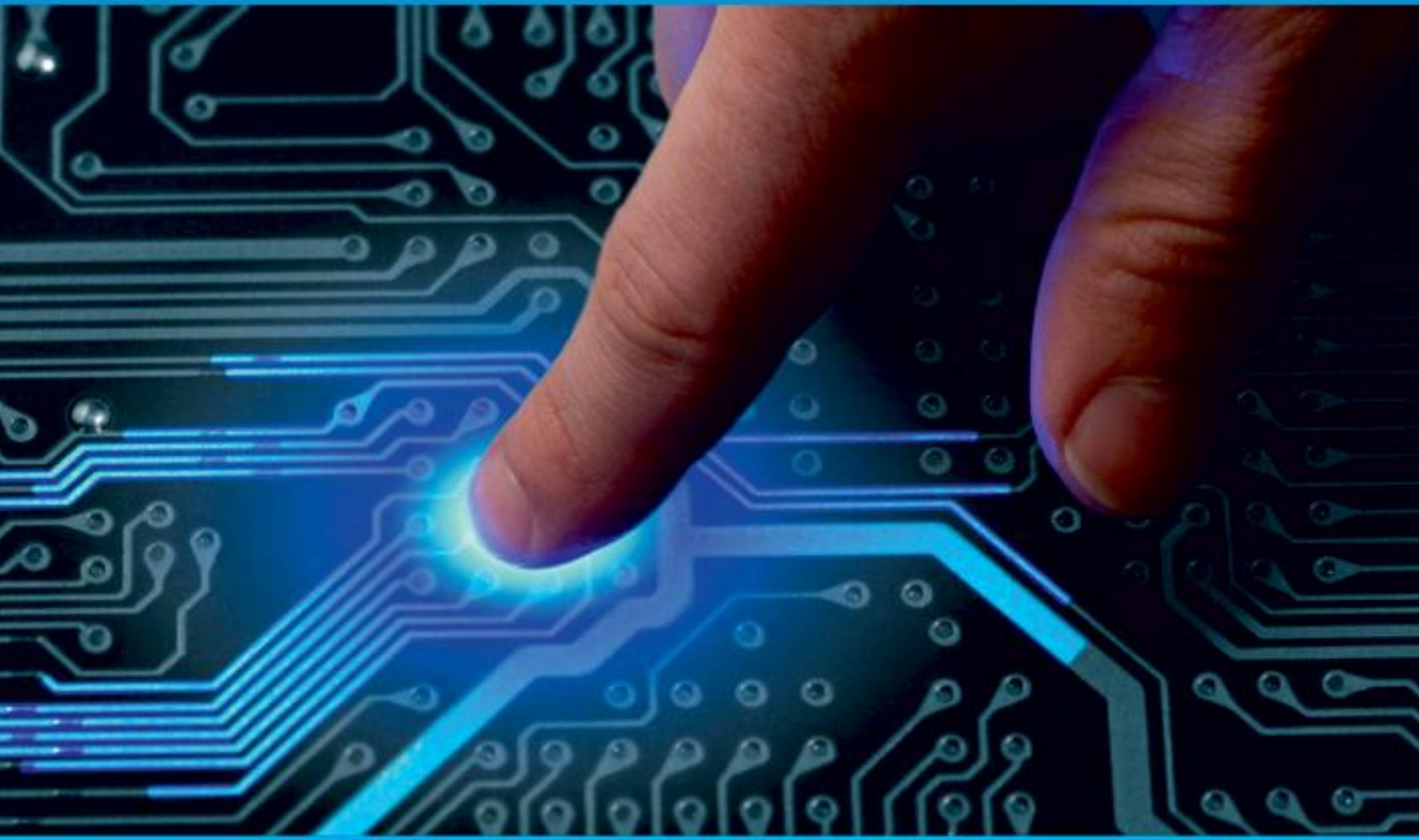
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379

Scam Link Detection

Ranjini S[1], Muhammed Ajas K[2], Sisira S[3], Anusree S[4], Aju George[5]

Assistant Professor, Department of Computer Science, Nethaji Memorial Arts & Science College, Nemmara, Palakkad, Kerala, India[1]

Student, Department of Computer Science, Nethaji Memorial Arts & Science College, Nemmara, Palakkad, Kerala, India [2 3 4 5]

ABSTRACT: A Scam attempt link or phishing site link is a fraudulent email or website that tricks you into giving out personal information like passwords, usernames, and credit/debit card details. The phishing site will appear just as legitimate sites do but instead direct the user to a page where they are asked for their sensitive info. There's no way of telling which pages on the web aren't legit until it's too late so make sure your browser has anti-phishing features. This approach provides the highest level of accuracy in predicting phishing attempts by identifying specific traits. This method is used to detect phishing site URLs that have these specific characteristics, which we identified in order to keep up with our times.

KEYWORDS: Scan, Phishing attack, Smishing, vishing, Mishing, websites

I. INTRODUCTION

The aim of scam link detection is to identify and prevent the dissemination of malicious or fraudulent links on the internet. Scam links often lead users to phishing websites, malware downloads, or other online scams designed to deceive individuals and compromise their personal information, financial data, or security.

Our system accurately predicts URL based scam attacks while also being able to protect the user's identity. There are many tools and algorithms that can help in making decisions, predictions, etc., one of which is machine learning. Because the algorithm builds a model from historical data without biasing what will happen next, resulting in more accurate overall results, it enables for better decision-making with less risk than other alternatives like prediction markets or expert systems. Our system is designed to detect scam link attacks by using different machine learning algorithms One technique in use, a hybrid algorithm approach that combines several different machine learning algorithms, improves accuracy and detection rates while retaining simplicity in comparison to more intricate models.

Online procedures, online business or trading, or exposure, so the online systems already in place at that time faced little threat. However, in the past five years, the world has experienced a big boom in the IT sector, resulting in most of the daily operations going online. From shopping to banking the term "phishing" was coined in 1996 by his hacker, who stole the America Online account by stealing passwords from unsuspecting AOL users. The word phishing comes from the phrase "website phishing" and is his variation of the word "phishing". The idea is that, like a fish, it casts the bait in hopes that the user grabs it and bite. In most cases, bait is either an e-mail or an instant messaging site, which will take the user to hostile phishing websites.

Over the years, phishing attacks grew in number and intensity too. Phishing attacks now target users of online banking, payment services such as PayPal, and online e-commerce sites. There are different modes through which phishing can be carried out and hence there are various types of phishing like Vising (voice over phishing), Smishing (Phishing via SMS), whaling, Mishing (mobile phishing), social engineering, spear phishing, etc. A typical phishing attack typically consists of four phases: preparation, widespread broadcast, maturity, and account hijacking.

II. SYSTEM ANALYSIS

Existing System:
Various spam link detection techniques are employed to identify and mitigate the impact of spam and malicious links on the internet. These techniques often leverage a combination of heuristics, machine learning, and other methods to analyze patterns and characteristics associated with spam.The existing scam link detection systems are in-adequate to meet the security requirements. The existing system mainly uses datasets to examine dependencies between URL

patterns. However they are not sure to obtain accuracy above 80 %. Therefore a more trustable system needs to be proposed for predicting scam links.

Proposed System:

This project's primary goal is to find the best method for identifying phishing websites, which have recently grown to be a serious concern. There are several ways to acquire this, but none of them are responsive to the rapidly advancing technologies. In order to overcome the adaptiveness, we are using Machine Learning and pattern matching to find an optimal approach. The proposed system uses pattern matching features to identify scam links. The properties of each URLs are analyzed by the software and the system predicts the website link as either scam or normal. The system checks many features including

- The url’s protocol -like http or https
- The websites Ranking and credibility score
- The website’s signature (Extracted using feature extraction algorithms)
- Any complaints/allegations against the website header in nearest history.

III. FUTURE ENHANCEMENT

To the best of our knowledge, this study is the first to incorporate the results of all previous analyses into the analysis of machine learning approaches for phishing website identification. The phishing research that has been proposed uses a categorical paradigm, in which phishing websites are assumed to automatically categorize websites into a specific range of sophisticated values based on the grandeur variable and a number of other criteria. ML-based phishing techniques leverage website functionality to gather data that may be utilized to categorize websites in order to identify phishing sites. Developing focused anti-phishing approaches and methods as well as minimizing their inconvenience are two ways to prevent phishing.

One of the main issues facing security experts nowadays is the rise in phishing attempts. The traditional tools for identifying phishing websites use signature-based approaches which are not able to detect newly created phishing webpage. Thus, researchers are coming up with machine learning-based methods which are capable to detect and classify the phishing webpages with high accuracy if a large and variety of features is considered. The timely identification of phishing websites is hampered by the time-consuming nature of developing a classification model employing a large number of characteristics. In order to create high-performance classification models faster, it is necessary to use a feature selection method to shortlist a set of features. In this chapter, we study the role of feature selection methods in detecting phishing WebPages efficiently and effectively.

A comparative analysis of machine learning algorithms is carried out on the basis of their performance without and with feature selection. Experiments are conducted on a phishing dataset with 30 features containing 4898 phishing and 6157 benign WebPages. Various machine learning techniques are employed to get optimal outcomes. The models are then made more efficient by using a feature selection technique. Random forest yields the best accuracy both before and after feature selection, while also significantly reducing the time required to create the model. The tests show that the accuracy of phishing detection classification models may be maintained while building faster by utilizing a feature selection strategy in conjunction with machine learning methods.

IV. CONCLUSION

This project aims to solve the issue of unawareness of the nature of websites and identify it by using datasets created specifically for this purpose and to train machine learning models to detect phishing websites. The dataset, which contains both phishing and benevolent URLs of websites, is used to generate the necessary URL and website content-based functionality. Every model's performance standard is calculated and contrasted. The Implementation of this proposed system will avoid the phishing site to compromise a user by identifying the attack and notifying the same.

From this point on, preventing such an attack is considered a difficult task in the field of system security. Phishing assaults can now be identified via a reliable identification technique with fewer false positives. Data analysis and heuristics, machine learning, and deep learning algorithms are some of the guiding techniques discussed in this article. While heuristic and data analysis methods have low False Positive rates and high computational costs, they are better at identifying phishing attacks. As opposed to other approaches, ML procedures provide the most straightforward outcomes. Some machine learning algorithms can detect TP up to 91 % of the time. Since malicious URLs are created every day, attackers use strategies to trick users and alter URLs in order to attack, the web Plug-in is an astonishing method that indicates a warning message to the user and avoids the attack.

REFERENCES

1. Muhammet Baykara, ZahitZiyaGure, NC “Detection of Phishing Attacks” 2018 6th International Symposium on Digital Forensic and Security (ISDFS) Firat University, Elazig, Turkey.
2. Athulya A.A, Praveen K. “Towards the Detection of Phishing Attacks” 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) 3) More akin Adebowale, KhinT.Lwin, M.A.Hossain “Intelligent phishing detection scheme using deep learning” (convolutional neural network (CNN) and the long short-term memory (LSTM)). (04 Aug 2018)
3. M somewhat “Efficient Deep Learning Technique for the Detection of Phishing Website” (27 June 2020)
4. Krishnamurthy“Chrome anti-phishing 43, Firefox anti-phishing 75”
5. Anu Yadav and Jatin Gemini “The Security threat in Cyber World – cybercrime as PHISHING” p- ISSN: 2393-9907; eISSN: 2393-9915 (April-June, 2017).
6. Phirashisha, Syiemlieh, Golden, Mary Khongsit1, Usha Mary Sharma, Bobby Sharma “Phishing-An Analysis on the Types, Causes, Preventive Measures And Case Studies in the Current Situation” e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 01-08 IOSR Journal of Computer Engineering (IOSR-JCE).
7. Ike Vayansky and Sathish Kumar “Phishing – challenges and solutions “ (January 2018).
8. Ram Basnet, Srinivas Mukkamala, and Andrew H. Sung ”Detection of Phishing Attacks: A Machine Learning Approach” New Mexico Tech, New Mexico 87801, USA
9. Amruta Deshmukh1, Sachin Mahabale2, Kalyani Ghanwat3, AsiyaSayyed “WEB PHISH DETECTION AN EVOLUTIONARY APPROACH” Department Of Computer Science and Engineering, Zeal Education Society’s, DCOER, Pune, Maharashtra, India. 63
10. Abdulghani Ali Ahmed, Nik QuosthoniSunaidi “Malicious Website Detection: A Review” Faculty of Computer Systems & Software Engineering University, Pahang, Malaysia (February 01, 2018)
11. Moruf akin Adebowale, KhinT.Lwin, M.A.Hossain “Intelligent phishing detection scheme using the deep learning” (convolutional neural network (CNN) and the long shortterm memory (LSTM)). (04 Aug 2018)
12. JyotiChhikara, RituDahiya, Neha Garg, Monika Rani “Phishing & Anti-Phishing Techniques” ISSN: 2277128X
13. PriyaSaravanana, Selvakumar Subramanian. “A Framework for Detecting Phishing Websites using GA-based Feature Selection and ARTMAP based Website Classification”.
14. Ayam el Assael, Shahryar Baki, Avishai Das, Rakesh M Verma “An In-Depth Benchmarking and Evaluation of Phishing Detection Research for Security Needs.”
15. M. Noushad Rahim and K.P. Mohamed Basheer “A survey on anti-phishing techniques: From conventional methods to machine learning.”
16. Victor E. Adeyemo, Abdullateef O. Balogun.“Ensemble-Based Logistic Model Trees for Website Phishing Detection”
17. Mehmet Korkmaz, EmreKocyigit, OzgurKoraySahingoz, and BanuDiri “Deep Neural Network-based Phishing Classification on a High-Risk URL Dataset”
18. Ms. SophiyaShikalgar Dr. S. D. SawarkarMrs.SwatiNarwane. “Detection of URL-based Phishing Attacks using Machine Learning”
19. Ms. Amrita Mitra “Phishing: Detection, Analysis And Prevention”

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

☎ 9940 572 462 6381 907 438 ✉ ijircce@gmail.com

www.ijircce.com

Scan to save the contact details