International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

PHISHCATCHER CLIENT-SIDE DEFENSE AGAINST WEB SPOOFING ATTACKS USING MACHINE LEARNING

[1]NIMMANAGOTI SHIVASAI,[2]MAREPALLY SUKESH REDDY,[3]MADIRA PAVAN RAJ,[4]GURRAM SAKETH,[5]B.SANKARAIAH

[1,2,3,4]Students, Department of computer Science And Engineering, Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100

[5]Assistant Professor, Department of computer Science And Engineering, Malla Reddy Engineering College (Autonomous),Hyderabad Telangana, India 500100

ABSTRACT

Cybersecurity faces a significant challenge in protecting the confidentiality and integrity of users' private information, such as passwords and PIN codes, against various forms of cyberattacks. Phishing and web spoofing attacks, where attackers create fake copies of legitimate web pages to steal sensitive information, are among the most common threats. These attacks can be initiated through phishing emails, malicious ads, malware, and other tactics, making it difficult for users to identify fraudulent websites. While several security strategies have been proposed, many face issues related to latency and accuracy. To address these challenges, we propose a client-side defense mechanism based on machine learning techniques to detect spoofed web pages and safeguard users from phishing attacks. We developed a Google Chrome extension called PhishCatcher, which utilizes a random forest classifier to assess whether a login page is legitimate or spoofed. The classifier uses four types of web features as input to make its determination. Our experimental results demonstrate that PhishCatcher achieves a remarkable accuracy and precision of 98.5% in classifying 400 legitimate and 400 phishing URLs. Additionally, the average response time of the extension was just 62.5 milliseconds, showcasing its efficiency in real-world applications.

Keywords: Phishing detection, web spoofing, cybersecurity, machine learning, random forest, browser extension, client-side defense, phishing URLs, web security, PhishCatcher, real-time protection, user authentication, spoofed websites.

I.INTRODUCTION

In October 2022, members of the National Institute for Research in Digital Science and Technology (Inria) in France received a phishing email written in French, asking users to confirm their webmail accounts by clicking a suspicious link. The link led to a fake login page resembling Inria's legitimate authentication page. When unsuspecting users entered their username and password on this fake site, the attacker could later submit this information to the real Inria login page, thereby compromising user credentials. This phishing attack exploited the similarity between the fake and real pages, making it difficult for users to discern the fraudulent nature of the website. This scenario highlights the growing threat of phishing and web spoofing attacks, where attackers aim to steal valuable information by tricking users into revealing sensitive data. The rise of online services such as e-commerce, online banking, and e-health has increased the risks associated with phishing attacks. Users typically create personalized accounts on websites by setting up usernames and passwords. However, when accessing these accounts, users face the risk of identity theft and data breaches, particularly if they fall victim to phishing attacks. In such attacks, users are directed to fake websites that mimic legitimate login pages, often via emails containing convincing messages. These sites can easily deceive users into submitting their credentials, putting their personal information at risk. As phishing techniques continue to evolve, attackers have found new ways to bypass traditional security measures like firewalls, digital certificates, and even two-factor authentication. Phishing attacks and web spoofing have become more sophisticated, with attackers now targeting high-value information such as national security data and intellectual property. These attacks often involve copying logos or HTML elements from legitimate websites to make the fake site appear genuine. Common phishing attack vectors include emails, trojan horses, keyloggers, and man-in-the-middle proxies. The primary targets for these attacks are online banking systems, third-party payment services, and e-commerce sites. Although cryptographic security protocols like SSL/TLS provide some protection, they are not foolproof. To effectively combat phishing, additional mechanisms need to be implemented on either the server-side, client-side, or both. While server-side solutions can identify spoofed sites, client-side solutions offer an alternative way to protect users without requiring modifications to the server. In response to the growing threat of phishing and spoofing, this paper introduces PhishCatcher, a stateless client-side tool designed to protect users from web spoofing attacks. PhishCatcher, a Google Chrome extension, leverages machine learning techniques, specifically the random forest algorithm, to determine whether a login page is legitimate or spoofed. The tool was evaluated on real-world applications, yielding promising results in terms of accuracy and efficiency. Key contributions of this research include the development of the PhishCatcher Chrome extension, careful selection of web features for the classifier algorithm, and experimental testing of the tool's effectiveness. This work provides a novel solution to phishing attacks by combining machine learning and client-side protection, offering users enhanced security against web spoofing.

II.LITERATURE SURVEY

Khan et al., 2021: This paper presents SpoofCatch, a client-side tool designed to protect users from phishing attacks. It highlights the effectiveness of detecting and mitigating phishing threats by leveraging browser-side detection mechanisms. The work emphasizes user-level defense against web-based threats, contributing to the broader fight against social engineering attacks.

Schneier, 2005: Schneier critiques the effectiveness of two-factor authentication (2FA), arguing that it often arrives too late in the security process. The paper discusses its limitations in the broader context of security vulnerabilities, suggesting that 2FA is insufficient on its own and should be part of a layered defense.

Garera et al., 2007: This framework provides a methodology for detecting and measuring phishing attacks, focusing on the analysis of web traffic and phishing behaviors. The work introduces strategies for identifying malicious websites and phishing URLs, providing a basis for future detection systems and strategies.

Oppliger and Gajek, 2005: This research examines effective defenses against phishing and web spoofing. It explores techniques such as secure communications protocols and user awareness strategies. The paper highlights the growing need for robust protection mechanisms to prevent users from falling victim to deceptive online practices.

Pietraszek and Berghe, 2005: This paper discusses defending against injection attacks by employing context-sensitive string evaluation. By evaluating strings in context, the authors propose a way to prevent malicious code injection, which is a common vulnerability in web applications.

Johns et al., 2011: The paper addresses session fixation attacks, offering methods to prevent such attacks in web applications. The authors propose secure mechanisms to validate and protect session IDs, reducing the risk of unauthorized session hijacking.

Bugliesi et al., 2014: This research focuses on enhancing client-side protection for cookie-based sessions, aiming to prevent session hijacking through automatic and robust security measures. The paper suggests a method of validating session cookies to improve the integrity of web sessions, particularly in the face of potential security breaches.

III.PROPOSED METHODOLOGY

The latest proposed method, PhishCatcher, for defending against web spoofing attacks using machine learning focuses on enhancing client-side protection by leveraging advanced machine learning models. PhishCatcher aims to detect phishing websites by analyzing features such as visual elements, domain names, and the behavior of the web page.

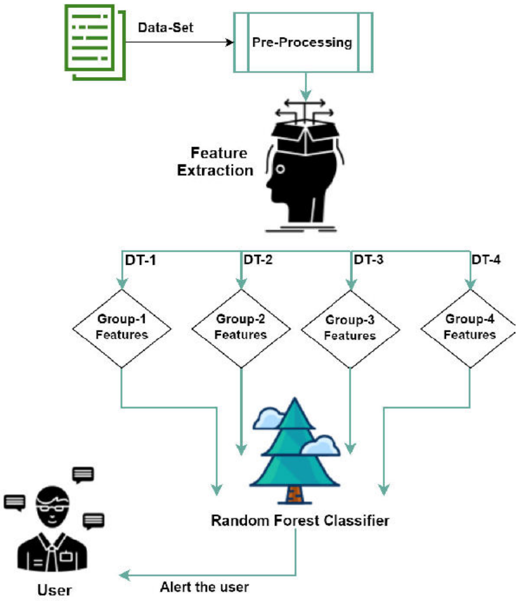
Feature Extraction: It extracts both content-based (e.g., URLs, HTML structure) and context-based features (e.g., page layout, styling, and JavaScript behavior). These features are crucial for identifying fraudulent sites designed to mimic legitimate ones.

Machine Learning Algorithms: PhishCatcher uses supervised learning algorithms, training on large datasets of legitimate and phishing websites. Commonly used algorithms could include decision trees, random forests, and neural networks, which allow for the detection of subtle patterns distinguishing phishing websites from legitimate ones.

Client-Side Deployment: Unlike server-side solutions, PhishCatcher is deployed directly on the client's browser as a plug-in or extension.

Real-Time Detection: The model continuously monitors and analyzes the web pages the user visits, flagging suspicious sites in real-time. This proactive defense minimizes the chances of users falling victim to phishing attacks before they can interact with a harmful site.

Accuracy and Adaptability: Machine learning models adapt over time, improving detection rates as more phishing patterns and tactics emerge. This dynamic adaptability enhances the system's ability to stay ahead of evolving phishing techniques.

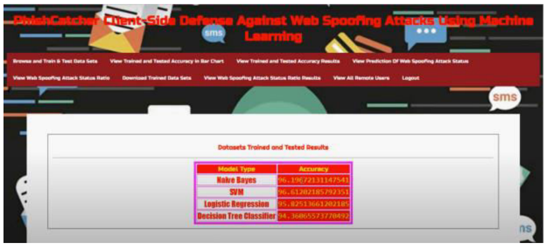


IV.WORKING METHODOLOGY

The content involves a web-based application designed to manage and detect web spoofing attacks using machine learning models, primarily focusing on phishing and web spoofing detection. The system is structured into two key components: the remote user interface and the service provider interface.

PhishCatcher Client-Side Defense Against Web Spoofing Attacks Using Machine Learning
Login Service Provider:
User Name
Password Admin
Login
User Login
sms
sms

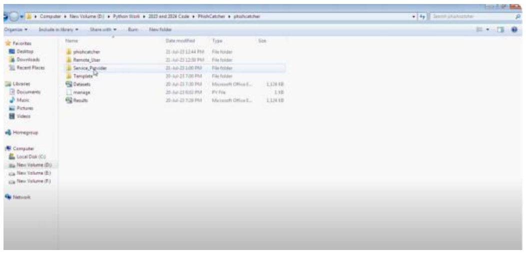
For the remote user, the process begins with user registration and login functionalities, allowing clients to create accounts and securely access their profiles.


PhishCatcher Client-Side Defense Against Web Spoofing Attacks Using Machine Learning
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Web Spoofing Attack Status
View Web Spoofing Attack Status Ratio
Download Trained Data Sets
View Web Spoofing Attack Status Ratio Results
View All Remote Users
Logout

Datasets Trained and Tested Results

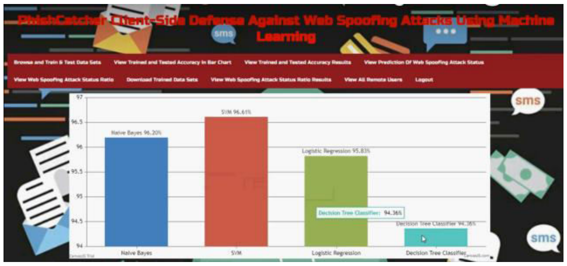
Model Type	Accuracy
Naive Bayes	96.19672131147541
SVM	96.61202185792351
Logistic Regression	95.82513661202185
Decision Tree Classifier	94.36065573770492

Once logged in, users can upload a dataset (in this case, URLs) and use the system to predict whether a URL is associated with a phishing or web spoofing attack.

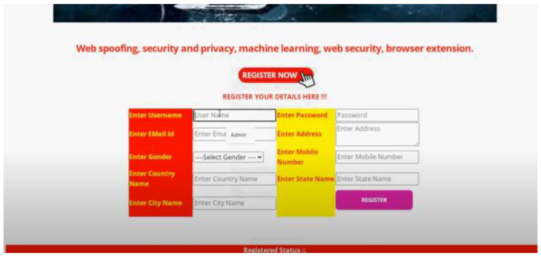


Name	Date modified	Type	Size
phishcatcher	21-Jul-23 12:44 PM	File folder	
Remote_User	21-Jul-23 12:30 PM	File folder	
Service_Provider	21-Jul-23 1:00 PM	File folder	
Template	20-Jul-23 7:00 PM	File folder	
Datasets	20-Jul-23 7:30 PM	Microsoft Office E...	1,124 KB
manage	20-Jul-23 6:02 PM	PY File	1 KB
Results	20-Jul-23 7:28 PM	Microsoft Office E...	1,124 KB

This prediction is based on machine learning models trained on historical data, which includes techniques such as Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and Decision Trees. The system processes the data using a CountVectorizer to transform the URL data into numerical vectors that can be used by the machine learning algorithms for training and prediction.

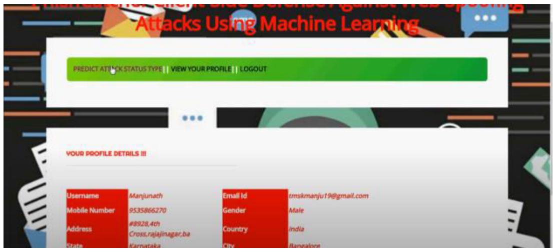

PhishCatcher Client-Side Defense Against Web Spoofing Attacks Using Machine Learning
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Web Spoofing Attack Status
View Web Spoofing Attack Status Ratio
Download Trained Data Sets
View Web Spoofing Attack Status Ratio Results
View All Remote Users
Logout
97
96.5
96
95.5
95
94.5
94
Naive Bayes 96.20%
SVM 96.61%
Logistic Regression 95.83%
Decision Tree Classifier: 94.36%
Decision Tree Classifier 94.36%
Naive Bayes
SVM
Logistic Regression
Decision Tree Classifier

Upon receiving the user input (URLs), the models predict the likelihood of an attack and provide the results. These predictions are stored in a database for further analysis.The service provider's role is to manage and monitor the performance of the web spoofing detection system. Service providers can view the statistics related to the attack prediction, such as the overall success rate and detection ratio. This is achieved by querying the attack_prediction model to calculate the ratio of phishing and non-phishing URLs in the system. Service providers also have access to charts displaying these ratios, as well as the accuracy of different machine learning models (e.g., Naive Bayes, SVM, Logistic Regression, and Decision Trees) used in the detection process.



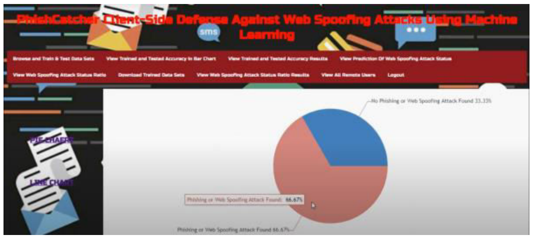
These models are retrained periodically to improve performance. The system also provides the ability to export the predicted dataset in Excel format for further use.

Additionally, the service provider can access detailed reports, such as confusion matrices, classification reports, and model accuracy scores.


Attacks Using Machine Learning
PREDICT ATTACK STATUS TYPE || VIEW YOUR PROFILE || LOGOUT
YOUR PROFILE DETAILS !!!

Username	Manjunath	Email Id	tmskmanju19@gmail.com
Mobile Number	9535866270	Gender	Male
Address	#8928,4th Cross,rajajinagar,ba	Country	India
State	Karnataka	City	Bangalore

These reports help in evaluating the effectiveness of each machine learning model used for attack detection. The training process involves cleaning and transforming the dataset (URLs), applying different classification algorithms, and recording the accuracy and performance of each model.


PhishCatcher Client-Side Defense Against Web Spoofing Attacks Using Machine Learning
Browse and Train & Test Data Sets
View Trained and Tested Accuracy in Bar Chart
View Trained and Tested Accuracy Results
View Prediction Of Web Spoofing Attack Status
View Web Spoofing Attack Status Ratio
Download Trained Data Sets
View Web Spoofing Attack Status Ratio Results
View All Remote Users
Logout
PIE CHART
LINE CHART
No Phishing or Web Spoofing Attack Found 33.33%
Phishing or Web Spoofing Attack Found: 66.67%
Phishing or Web Spoofing Attack Found 66.67%

This data is then presented to the service provider for decision-making. Lastly, the system allows for the download of datasets containing predictions, and it tracks detection accuracy and detection ratios for continuous monitoring and improvement.

V.CONCLUSION

Users have become dependent on the online applications as they provide significant quality of service in many domains i.e., online banking, e-commerce, social connectivity, digital libraries, online health services, virtual education, digital marketing and multi-player gaming applications. Commonly, an authentication procedure is followed by the users for the creation of their online account to access the private web content. The security and privacy of users is at stack amid highly sophisticated web spoofing attacks. Several research and commercial tools have been developed to fight against web spoofing attacks but most of them appear with a few lapses. We have developed an optimized user-friendly browser plug-in dubbed as Phish Catcher for the smart disclosure of phishing attacks based on supervised machine learning. Contrary to the traditional approaches, our scheme offers to run the classification in the browser itself. It addresses the loopholes in the existing web applications by fixing the latency issues and improving the efficiency of the tool. The user interface of our plug-in is made simple for the better understanding of the user. When a user enters a phished URL, it displays a phishing alert on the screen and highlights the corresponding phishing features of that URL in a drop-down menu.

The feature set contains thirty features, though, the addition of more automated features might be a great idea to improve the overall performance. Some other discriminative classifiers such as SVM can also be implemented for the prediction of fake or real URL by training larger data-sets. Evaluation metrics may also be evolved by using different tools for a better performance analysis.

VI.REFERENCES

[1] W. Khan, A. Ahmad, A. Qamar, M. Kamran, and M. Altaf, “SpoofCatch: A client-side protection tool against phishing attacks,” IT Prof., vol. 23, no. 2, pp. 65–74, Mar. 2021.

[2] B. Schneier, “Two-factor authentication: Too little, too late,” Commun. ACM, vol. 48, no. 4, p. 136, Apr. 2005.

[3] S. Garera, N. Provos, M. Chew, and A. D. Rubin, “A framework for detection and measurement of phishing attacks,” in Proc. ACM Workshop Recurring malcode, Nov. 2007, pp. 1–8.

[4] R. Oppliger and S. Gajek, “Effective protection against phishing and web spoofing,” in Proc. IFIP Int. Conf. Commun. Multimedia Secur. Cham, Switzerland: Springer, 2005, pp. 32–41.

[5] T. Pietraszek and C. V. Berghe, “Defending against injection attacks through context-sensitive string evaluation,” in Proc. Int. Workshop Recent Adv. Intrusion Detection. Cham, Switzerland: Springer, 2005, pp. 124–145.

[6] M. Johns, B. Braun, M. Schrank, and J. Posegga, “Reliable protection against session fixation attacks,” in Proc. ACM Symp. Appl. Comput., 2011, pp. 1531–1537.

[7] M. Bugliesi, S. Calzavara, R. Focardi, andW. Khan, “Automatic and robust client-side protection for cookie-based sessions,” in Proc. Int. Symp. Eng. Secure Softw. Syst. Cham, Switzerland: Springer, 2014, pp. 161–178.

[8] A. Herzberg and A. Gbara, ‘‘Protecting (even naıve) web users from spoofing and phishing attacks,’’ Cryptol. ePrint Arch., Dept. Comput. Sci. Eng., Univ. Connecticut, Storrs, CT, USA, Tech. Rep. 2004/155, 2004.

[9] N. Chou, R. Ledesma, Y. Teraguchi, and J. Mitchell, ‘‘Client-side defense against web-based identity theft,’’ in Proc. NDSS, 2004, 1–16.

[10] B. Hämmerli and R. Sommer, Detection of Intrusions and Malware, and Vulnerability Assessment: 4th International Conference, DIMVA 2007 Lucerne, Switzerland, July 12-13, 2007 Proceedings, vol. 4579. Cham, Switzerland: Springer, 2007.

[11] C. Yue and H. Wang, ‘‘BogusBiter: A transparent protection against phishing attacks,’’ ACM Trans. Internet Technol., vol. 10, no. 2, pp. 1–31, May 2010.

[12] W. Chu, B. B. Zhu, F. Xue, X. Guan, and Z. Cai, ‘‘Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs,’’ in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2013, pp. 1990–1994.

[13] Y. Zhang, J. I. Hong, and L. F. Cranor, ‘‘Cantina: A content-based approach to detecting phishing web sites,’’ in Proc. 16th Int. Conf. World Wide Web, May 2007, pp. 639–648. [14] D. Miyamoto, H. Hazeyama, and Y. Kadobayashi, ‘‘An evaluation of machine learning-based methods for detection of phishing sites,’’ in Proc. Int. Conf. Neural Inf. Process. Cham, Switzerland: Springer, 2008, pp. 539–546.

[15] E. Medvet, E. Kirda, and C. Kruegel, ‘‘Visual-similarity-based phishing detection,’’ in Proc. 4th Int. Conf. Secur. privacy Commun. Netowrks, Sep. 2008, pp. 1–6.