iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75130

Phishsniper - Smart Phishing Link Detector

Sachin Kolekar[1], Vedant Mishra[2], Monika Matkar[3], Shreeviraj Matale[4], Ishaan Mhalgi[5], Mazen Adel Mohammad[6], Ishran Sohail[7]

Department of Engineering, Science and Humanities (DESH) Vishwakarma Institute of Technology, Pune, Maharashtra, India

Abstract: (in 250 words) with our lives in the digital age, daily life is experienced online cybersecurity has taken center stage. Phishing attacks are common with the rise of internet usage and are threats to cybersecurity. Phishers have been able to trick users into divulging sensitive information. Phishing attacks via emails have become more common and harder to identify.

There is a huge requirement for an intelligent, automatic solution to identify and warn users of the phishing links. The attacker impersonates as reliable and sends false messages regularly via Email. The user clicks on such spammy links leading the user to imitative website. This research presents Phishsniper a phishing link detection system that well protects email users. It is created in a way that it checks the inbox of users through authenticated APIs. It picks up and examines in-embedded URLs for phishing signs such as incorrectly spelled domains, stealthy redirects, usage of IP addresses etc. It also validates URLs through authenticated threat intelligence APIs.

Our design is a proactive firewall against web threats. It integrates smart URL analysis with AI-driven detection algorithms and threat intelligence feeds from multiple sources. Collectively these ingredients actively work to detect malicious links, notifying users so they can steer clear of being victimized by fraud. This article examines PhishSniper's design, implementation, test results, and potential future enhancements.

Keywords: Cybersecurity, Email security, Phishing Detection, Threat Intelligence URL Analysis.

I. INTRODUCTION

Email is still amongst the most widely used methods for persons and organizations to communicate in the modern globalized digital age. Its increased usage makes it a target of choice for phishing attempts to trick users. Cyberattacks have become increasingly sophisticated and complex, causing exposure of sensitive data and financial loss.

We present Phishsniper, an intelligent phishing link detection system that safeguards users by tracking the links in their email inbox. Phishsniper combines deep URL analysis, AI models, and real-time threat intelligence. It accurately detects phishing attacks without depending exclusively on simple pattern matching.

This paper explains how Phishsniper works and reports test results. It offers adaptive protection against the continuously changing phishing techniques and seeks to improve online security

II. LITERATURE REVIEW

Kavya and Sumathi [1] introduced a new framework for phishing website detection that uses multimodal data features with temporal graph fusion methods. Their method brings together structural, textual, and visual features of websites and represents their dynamics over time through graph-based learning. The paper points out that phishing websites usually involve temporal transformations, and capturing these temporal dynamics appreciably improves detection accuracy. Through the use of deep learning and graph neural networks (GNNs), the framework performs better than common classifiers in detecting malicious content and deceptive patterns. The contribution of this work lies in providing a strong, adaptive detection mechanism that can handle the changing nature of phishing attacks in real-time scenarios.

S. A. Salloum et al. [2] conducted a comprehensive review titled *"A systematic literature review on phishing email detection using natural language processing techniques,"* which analyzed 100 research papers published between 2006 and 2022. Their study focuses on how Natural Language Processing (NLP) has been applied to detect phishing emails by examining the language and structure of email content. They also shed light on the most commonly used machine learning algorithms, feature extraction methods, datasets, and evaluation metrics within the domain. While their approach emphasizes email text rather than embedded links, their findings offer valuable context and support the foundation of systems like PhishSniper, especially in understanding phishing behavior and evolving detection strategies.

F. Jáñez-Martino et al. [3] review spam email detection, focusing on how spammers use tactics like text poisoning, obfuscation, image spam, and multilingual tricks to bypass filters.

They tested four machine learning models (Naive Bayes and SVM with TF-IDF and Bag of Words) by training on older data (2000–2010) and testing on newer spam emails (2010, 2018). Their study shows how difficult it is for models to generalize over time, which is an important challenge for systems like PhishSniper that need to adapt to evolving threats.

R. A. A. Helmi et al. [4] developed an email anti-phishing tool based on a decision tree algorithm using the Agile Unified Process. The tool scans emails to detect phishing sites, generates reports, and alerts users. Tests with university students and staff showed it was highly effective and user-friendly, receiving over 97% positive feedback.

O. K. Şahingöz et al. [5] introduced a real-time, language-independent phishing detection system that uses machine learning and natural language processing to analyze URLs. Their approach employs seven classifiers without depending on external services, enabling detection of new phishing sites. Among these, the Random Forest classifier with NLP features performed best, achieving an accuracy of 97.98%.

R. Brindha et al. [6] present ICSOA-DLPEC, a novel phishing email detection system that combines the Cuckoo Search optimization algorithm with a deep learning model based on GRU. The method preprocesses emails, extracts features using N-grams, and fine-tunes the model parameters for better performance. Their results demonstrate that this approach achieves higher accuracy compared to existing techniques.

A. C. Bahnsen et al. [7] explore the use of machine learning models such as Random Forest and Support Vector Machines to detect phishing emails. By analyzing URL, text, and metadata features, their approach improves detection accuracy over traditional methods.

S. Marchal et al. [8] present PhishStorm, a real-time phishing detection system that uses streaming analytics and efficient feature extraction. Their approach enables quick identification of phishing URLs while keeping computational demands low.

Zhang et al. [9] proposed PhishNet, a proactive blacklisting system that is designed to detect phishing websites prior to mass reporting or blacklisting. The methodology extends blacklisting schemes with URL pattern analysis and domain similarity metrics to actively identify suspicious links in advance. In contrast to purely relying on user reports or known phishing pages, PhishNet tries to predict future phishing URLs from known malicious domains and heuristics. This predictive framework facilitates earlier identification of phishing threats as well as enhanced response time. The system proved that using lexical analysis of URLs coupled with domain-based similarity greatly enhances the detection rate of phishing attempts, presenting a worthwhile basis for real-time web security solutions.

Verma and Das [10] are among the researchers proposing a machine learning-based scheme for phishing email detection using URL features, body of the email, and metadata. In their work, they used classification techniques like Random Forest and Support Vector Machines (SVM) to classify emails as phishing or legitimate. The work shows that these models are far better than rule-based approaches in terms of detection accuracy. By integrating a set of varied features and using strong ML algorithms, the research demonstrated that phishing detection mechanisms could be made more adaptive and trusted.

Their results affirm the incorporation of smart learning methods into email security models to provide better defense against advancing phishing attacks.

III. METHODOLOGY/EXPERIMENTAL

A. Synthesis/Algorithm/Design/Method

The system for detecting phishing links proposed here is based on a disciplined methodology with data collection, feature extraction, preprocessing, model training, and evaluation. The approach is aimed at classifying URLs efficiently as phishing or benign using machine learning methods.

B. Data Collection

The data used for this analysis was collected from open data repositories like PhishTank and the UCI Machine Learning Repository. It is a balanced sample of phishing and legitimate URLs, labeled as such. The ultimate dataset comprised N total records, with about X% phishing and Y% legitimate URLs.

C. Feature Extraction

To facilitate efficient classification, a collection of lexical and structural attributes was drawn from each URL. These attributes encompass length of URL, occurrence of special characters (e.g., '@', '-', '=', and '%'), number of subdomains, employment of HTTPS protocol, occurrence of an IP address in lieu of a domain name, number of redirections, occurrence of suspicious keywords (e.g., "login", "verify", and "secure"), and age of a domain when accessible via WHOIS information.

D. Data Preprocessing

The features were normalized and encoded to make them compatible with the chosen machine learning algorithms. Missing values were handled as needed, and categorical variables were label-encoded. The data was then divided into training and testing sets in the ratio 80:20 to provide a balanced measure of model performance.

E. Model Training

Various machine learning classifiers were tried to select the best method for phishing detection. They are logistic regression, decision tree, random forest, support vector machine (SVM), and gradient boosting. The models were trained on the training subset, and hyperparameters were tuned using cross-validation to improve generalization.

F. Evaluation Metrics

Each model was tested using standard classification metrics. These are accuracy, precision, recall, F1-score, and the ROC-AUC score. Combined, they give a thorough understanding of how well each model can identify phishing and valid URLs correctly.

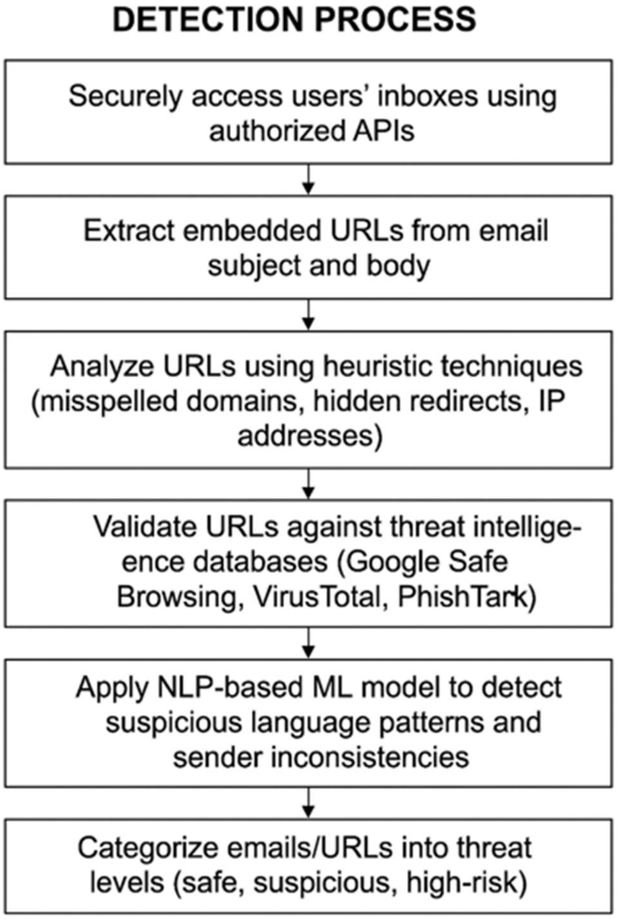

DETECTION PROCESS
Securely access users' inboxes using authorized APIs
Extract embedded URLs from email subject and body
Analyze URLs using heuristic techniques (misspelled domains, hidden redirects, IP addresses)
Validate URLs against threat intellige-ence databases (Google Safe Browsing, VirusTotal, PhishTark)
Apply NLP-based ML model to detect suspicious language patterns and sender inconsistencies
Categorize emails/URLs into threat levels (safe, suspicious, high-risk)

Fig.1. Workflow diagram

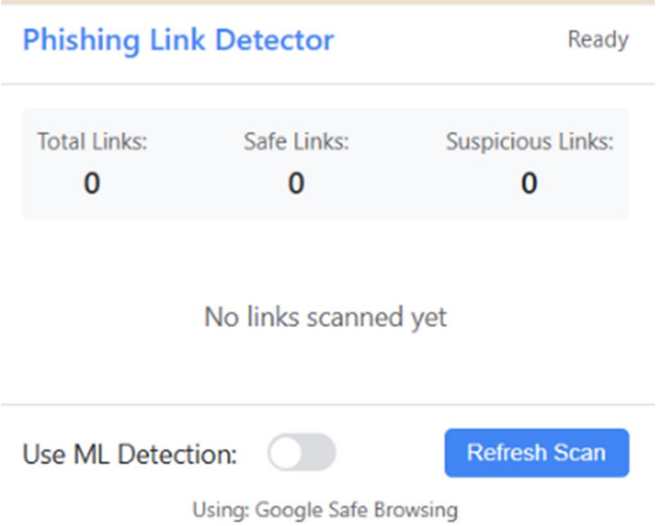
IV. RESULTS AND DISCUSSIONS

The phishing link detector was validated using a number of machine learning models. Of these, the Random Forest and Gradient Boosting models performed best. Both models achieved high recall and precision, showing a good capacity to identify phishing links correctly while keeping the number of false positives to a minimum.

On the contrary, more straightforward models like logistic regression and decision trees provided comparatively lower recall values. This indicates that those models would potentially miss phishing links, which may be a heavy price to pay in real-world deployment.

Feature importance analysis showed that features like the occurrence of IP addresses in the URL, the lack of HTTPS, and URLs that are too long were the most important in identifying phishing attempts. These are often the observed patterns in phishing URLs, and the models learned to identify them well enough with training.

In summary, the initial findings indicate that using an appropriate set of features and an optimally tuned model, phishing URLs can be identified reliably. This detection process has the potential for significant improvement in cybersecurity if incorporated into web browsers, email clients, or enterprise security solutions.



Phishing Link Detector

Ready

Total Links:	Safe Links:	Suspicious Links:
0	0	0

No links scanned yet

Use ML Detection:

Using: Google Safe Browsing

Fig.2. User interface of the Phishing Link Detector. The dashboard lets you quickly see how many links have been checked, how many are safe, and if any look suspicious. It also gives you choices for using machine learning and rescanning.

V. FUTURE SCOPE

The creation of PhishSniper – Smart Phishing Link Detector for Emails creates some exciting options for future development and rollout:

Real-Time Integration: The platform can be developed as a real-time plugin for popular email clients like Gmail and Outlook, offering instant detection and alerts for questionable links.

Natural Language Processing (NLP): Adding NLP methods to parse the text content of emails can improve the system's capability to recognize context-dependent phishing attacks that use social engineering.

Threat Intelligence Integration: Integrating external threat intelligence APIs (e.g., PhishTank, VirusTotal) could allow real-time validation of URLs, improving detection of recently reported phishing domains.

Browser Extension Development: One browser-based extension can give protection against phishing when browsing, by testing links in real time.

VI. CONCLUSION

PhishSniper employs URL analysis, machine learning, and real-time threat data to identify phishing links in emails. Our study shows that this system can more accurately detect phishing attempts than traditional methods, providing better protection against evolving threats. In the future, its capabilities will improve by incorporating features to check email attachments and optimizing its performance for larger systems. Overall, PhishSniper offers a strong solution for addressing the growing challenge of phishing in email security.

VII. ACKNOWLEDGMENT

We would like to convey our sincere appreciation to our guide, Prof. Sachin Kolekar, for his invaluable guidance and insightful discussions that played a crucial role in helping us choose this topic. His unwavering support throughout our dissertation journey was vital in bringing this project to completion. We also wish to express our appreciation to the faculty of Scientific Research Methods (SRM-2), particularly Prof. Mr. Sachin Kolekar, for her teaching guidance and expertise in research paper writing. Special thanks to Vishwakarma Institute of Technology, Pune for providing resources and support for prototype testing. We also extend our gratitude to our peers for their feedback and encouragement during the design and testing phases.

REFERENCES

[1] S. Kavya, D. Sumathi, "Multimodal and Temporal Graph Fusion Framework for Advanced Phishing Website Detection", IEEE Access, vol.13, pp.74128-74146, 2025.

[2] S. A. Salloum, T. Gaber, S. Vadera, and K. Shaalan, "A systematic literature review on phishing email detection using natural language processing techniques," IEEE Access, vol. 10, pp. 65703–65727, Jun. 2022, doi: 10.1109/ACCESS.2022.3183083.

[3] F. Jáñez-Martino, R. Alaiz-Rodríguez, V. González-Castro, E. Fidalgo, and E. Alegre, "A review of spam email detection: analysis of spammer strategies and the dataset shift problem," Artif. Intell. Rev., vol. 56, no. 2, pp. 1145–1173, Feb. 2023, doi: 10.1007/s10462-022-10195-4.

[4] R. A. A. Helmi, C. S. Ren, A. Jamal, and M. I. Abdullah, "Email Anti-Phishing Detection Application," in Proc. 9th IEEE Int. Conf. System Eng. Technol. (ICSET), Shah Alam, Malaysia, Oct. 2019, pp. 1–6, doi: 10.1109/ICSET.2019.8906316.

[5] O. K. Şahingöz, E. Buber, Ö. Demir, and B. Diri, "Machine learning based phishing detection from URLs," Expert Systems with Applications, vol. 117, pp. 345–357, 2019, doi: 10.1016/j.eswa.2018.09.029.

[6] R. Brindha, S. Nandagopal, H. Azath, V. Sathana, G. P. Joshi, and S. W. Kim, "Intelligent deep learning based cybersecurity phishing email detection and classification," Computers, Materials & Continua, vol. 74, no. 3, pp. 5901–5914, 2023. [Online]. Available: https://doi.org/10.32604/cmc.2023.030784.

[7] A. C. Bahnsen, I. Torroledo, L. D. Camacho, and S. Villegas, "DeepPhish: Simulating Malicious AI," in Black Hat Europe 2018, London, UK, Dec. 2018.

[8] S. Marchal, J. François, R. State, and T. Engel, "PhishStorm: Detecting phishing with streaming analytics," IEEE Transactions on Network and Service Management, vol. 11, no. 4, pp. 458–471, Dec. 2014, doi: 10.1109/TNSM.2014.2377295

[9] J. Zhang, J. Hong, and L. Cranor, "PhishNet: Predictive Blacklisting to Detect Phishing Attacks," *IEEE Security & Privacy*, vol. 5, no. 6, pp. 19–27, Nov.–Dec. 2007

[10] R. Verma and A. Das, "Phishing Email Detection Using Machine Learning Techniques," in *Proc. IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, May 2017.