IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-Enhanced Phishing Defense & Real-Time Malicious URL Detection

[1]AMRUTGOUDA M PATIL, [2]MANGALA H S, [3]KARTHIK K S, [4]AYUSH BHATT, [5] BHARATH K

[1]Student, [2]Assistant Professor Author, [3]Student, [4]Student, [5]Student

[1] Computer Science and Engineering,

[1] Dayananda Sagar Academy of Technology& Management, Bengaluru, India

Abstract: The goal of this project is an AI-based development of the phishing defense and malicious URL recognition system that operates in real time. The purpose of this system employing machine learning and natural language processing techniques is to discriminate phishing web pages and URL links as harmful based on the domain name, content on the web page and users' expectations and patterns. The approach combines deep learning models that self-update based on new phishing approaches and thus are able to provide high performance and protection at scale. It supports real-time detection and alerting and can be integrated with other security tools, which improves the security of users and businesses. The objective of the project is to combat phishing and improve safety on the net.

Key Words - AI-Enhanced Phishing Defense, Malicious URL Detection,

I. INTRODUCTION

This project focuses on developing an AI-enhanced system for real-time phishing defense and malicious URL detection. Leveraging techniques in machine learning (ML) and natural language processing (NLP), it aims to accurately classify and detect phishing websites and malicious URLs based on features that include domain structure, web content, and user behavior. The solution integrates deep learning models that continuously update themselves to keep track of evolutions in phishing tactics; it is scalable and protects at high performance. It provides real-time detection, alerts, and seamless integration with existing security platforms, thereby enhancing cybersecurity for individuals and enterprises alike. The project aims to reduce phishing risks and improve overall online safety.

In this context,[1] AI would appear like a very potent tool that would enhance the ability to detect malicious URLs and phishing attacks and also help prevent them. Systems run by AI apply such advanced techniques that are parts of ML, NLP, and pattern recognition in performing real-time analysis on enormous datasets. These can catch even the subtlest patterns of anomalies or deceptive tricks that would otherwise go unnoted or bypassed under the old approach to detection.

This approach is based on the use of both supervised and unsupervised learning models for real-time adaptation towards higher accuracy without false positives. More deep learning techniques are involved in the complex feature analysis such as visual similarity in webpage design, linguistic patterns in emails, metadata from network traffic.

This new threat landscape covers problems of zero-day attacks, polymorphic threats, and sheer volume cyberattacks and provides insights into AI adoption, computed resources demanded, ethical considerations, and even adversarial attacks on AI models.

In fact, this system can be integrated into any of the existing web browsers across the globe, which provides live phishing protection and adaptability to make sure its effectiveness is implemented across all the most critical sectors, such as finance, health, and commerce, where phishing attacks' stake is at all high levels. This ability of the system to learn in real-time new attempts makes it efficient even if there is a continuous change from those malicious activities. At the heart of our method lies the Knowledge Distilled ELECTRA model, meticulously crafted and optimized for the task of URL classification. This is a break with the previous state and changes the nature of prevention:-Using the Knowledge Distilled ELECTRA model, thoroughly curated for URL classification, adds strength to the detection of our solution:-Our solution was made user-friendly as presented through browser extensions, with our approach allowing users the capability to make informed choices and minimize the chance that users will become victims:-Our paper contributes to how one can look at methods while assessing solutions critically while at it.

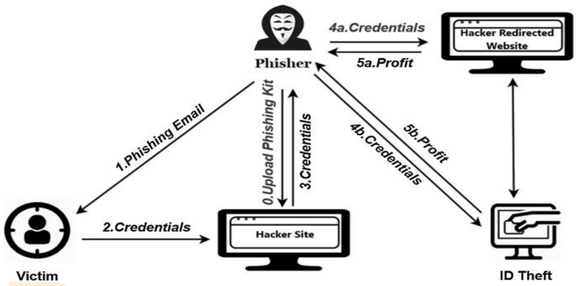

Fig 1

II. Literature Survey

1. Machine Learning-based Phishing Detection Authors: Zhang et al. Year: 2018

Zhang et al. has proposed a phishing detection system based on static features, which are the URL length and domain age along with the appearance of special characters for deciding the URLs. The research paper utilized classifiers like decision trees and random forests, to obtain excellent results on known phishing datasets. [2] However, it was not very adaptable for novel phishing patterns because it relies heavily on static features.

2. Deep Learning Breakthrough

Authors and Year: Bahnsen et al. (2020)

Bahnsen et al. introduced a deep learning model implemented on RNNs to predict sequences in URLs. [3] Its LSTM-based method increased the sensitivity of zero-day phishing detection by a wide margin with capturing context within URL strings. This work demonstrated that deep models can be versatile against legacy models.

3. Hybrids Detection Models

Authors and Year: Rao and Pais (2019)

Rao and Pais suggested a hybrid phishing detection framework combining heuristic-based blacklist checks and machine learning classifiers.[4] The mechanism proposed here may be more effective and efficient-the speed of heuristics for known threats and the adaptability of machine learning for new ones.

4. Natural Language Processing in Phishing Email Analysis

Authors and Year: Aburrous et al. (2021)

Aburrous et al. studied NLP for the analysis of content in phishing emails. The authors used Word2Vec, a technique that could determine word embeddings, thereby extracting patters of deceptive language and social engineering cues. [5] They have recently combined transformer-based models, such as BERT, to capture the subtlety of language for more accuracy in detection

Summary of recent phishing detection methods from 2019-24

Table 1

Authors	Used Algorithm	Challenges/ Limitations
Odeh	CatBoost, XGBoost, and LightGBM	Bias can occur in manual feature selection
Qasem	SNN, DT	Extracting 111 features from a real-time URL is not feasible.
Qasem	Bagging trees, KNN, Boosted decision trees, subspace discriminator, PSO-XG Boost	Used 59 features from a URL.
Sheikhi	-	Only 36000 URLs used.
Hannousse	SVM, DT, LR, RF, Naïve Bayes	Manual feature selection used.
Basit	RF, K-NN, DT, and ANN	Used an open-source dataset..
Rashid	SVM	Small dataset used.
Bu and Cho	CNN	Used character-level features.
Korkmaz	CNN	Low accuracy.
Feng and Yue	RNN	They used 17 features for classification .

III. METHODOLOGY

It passes through several stages from data collection to model evaluation for an effective phishing defense and malicious URL detection system driven by AI. This is especially designed for the challenge of real-time detection of phishing attempts and malicious URLs.

1. Data Collection and Preprocessing
1.1 Sources of Data
Available public datasets include PhishTank, OpenPhish, and top websites of Alexa for benign URLs.
[6] Real, from email and network traffic logs that would come with an assurance of privacy and observance.
Generated synthetic data for mimicking the effect of zero-day phishing.
1.2 Preprocessing:
Extract all features by pulling length of URL, whether the special characters exist or do not exist, existence of the subdomains, and domain's age.

Delete the duplicate records, incoherent, and irrelevant information. Only for text feature, split the token at URL and email address, splitting the text.

2. Feature Engineering
2.1 Static Features:
Domain-specific features such as WHOIS details, SSL certificate information.
URL structure analysis such as query parameters and path depth.
Dynamic Features:
Simulate runtime behavioral features by simulating the execution of URLs.
2.2 Visual Features:
Classify Web Page Screenshots based on Similarity of Websites by Processing Images
Natural Language Features:
As per NLP,[7] extract the semantic and syntactic feature from the text of a phishing email or the web pages

3. Modeling
3.1 Machine learning models:
Train a classical model with the inclusions of Random Forest SVM Gradient Boosting for a pilot run
3.2 Deep Learning Models:
Deep learning models applied on sequential patterns such as URL or content of the email. The visual pattern alongside the website screen
The transformer-based models, in this case,[8] BERT, can be utilized in deep content analysis for NLP while phishing email detection is the key concern of focus
Use of heuristics techniques such as blacklist or whitelist inspection can be integrated together with AI models to come up with more accuracy available.

4. Adaptive Learning for Zero-Day Attacks
Online learning and semi-supervised adaptation towards new patterns in the phishing attacks
[9]Unlabeled data identification mechanisms of Suspicious URL or behavior.

5. Real-Time Systems Design
5.1 Architecture:
Data ingestion, feature extraction along with AI classification module-based architecture.
5.2 Processing Pipeline:
Apply a framework like Apache Kafka or Spark Streaming for real-time analysis of URLs.
5.3 Latency Optimisation:
Utilize lightweight models or model compression for quicker inference.
Actionable Responses.
Automation Mechanisms.
[10]Automation mechanism responds as a blocker of suspicious URLs by alerting the administrators.

6. Testing of Robustness
Testing robustness with a model against evasion and data poisoning attacks
[11]Adversarial training renders the AI models resilient.

7. Model Evaluation and Validation
7.1 Metrics:
Precision, recall, F1-score and ROC-AUC for drawing the inference of performance.
[13]03False positive rate and false negatives for assessing the reliability.
7.2 Cross-validation:
Cross- validate models in a generalizable fashion by means of k-fold cross-validation
7.3 Comparison:
[12]Compare the AI model in contrast to traditional detection system where improved performance is provided by it.

8. Deploy and Scalability
Deploys on a cloud-based or an enterprise level with scalable architecture. Shall deploy using containerization techniques and technologies like Docker and Kubernetes.
Integrations within all existing cybersecurity frameworks; say, SIEMs security information and event management[13].

9. Continuous monitoring update
It needs to monitor its performance periodically and update its models. Feedback loops for self-improvement of the system should be implemented[14].

IV. Conclusion

The threat to an individual, organization, or a global cybersecurity framework is continued attacks by phishing and malicious URLs. However, leveraging AI technologies is transforming and turning the battle against these threats into one that can be fought with effectiveness and in real-time. Hybrid approaches that combine AI with heuristic techniques enhance the detection precision and speed of the system and are therefore very suitable for real-world deployments[15].
The real-time capability of AI systems has been crucial in shortening the window of vulnerability while ensuring timely responses to emerging threats. The developments in AI will help reduce the risks of phishing and malicious URLs because such systems can be proactive and dynamic in their defense. Issues of deployment challenges and ethics in AI practices will help these systems significantly enhance cybersecurity resilience on the global scale

References

[1] Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing Detection: A Literature Survey. *IEEE Communications Surveys & Tutorials*, 15(4), 2091–2121.
[2] Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2011). Learning to Detect Malicious URLs. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 1–24.
[3] Le, T. T., Malhi, K., & Lee, S. (2020). Deep Learning Models for Phishing Detection: A Survey. *IEEE Access*, 8, 129847–129864.
[4] Verma, R., & Hossain, N. (2019). Hybrid PhishNet: A Hybrid Deep Learning Framework for Phishing Detection. *Journal of Cybersecurity*, 5(1), 1–18.
[5] Lin, C. H., & Chen, M. Y. (2021). Natural Language Processing Techniques for Phishing Email Detection. *Computers & Security*, 101, 102110.
[6] Sahoo, D., Liu, C., & Hoi, S. C. H. (2017). Malicious URL Detection Using Machine Learning: A Survey. *arXiv preprint arXiv:1701.07179.*
[7] Marchal, S., Saari, K., Singh, N., & Asokan, N. (2017). Know Your Phish: Novel Techniques for Detecting Phishing Sites and Their Targets. *Proceedings of the 36th IEEE Symposium on Security and Privacy (S&P)*, 469–486.
[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*, 5998–6008.
[9] Shokri R., & Shmatikov V. (2015). Privacy-Preserving Machine Learning as a Service. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 1310–1321.
[10] Chen, S., Liu, J., & Zhai, E. (2019). Scalable Real-Time Phishing Detection System Using Apache Kafka. *ACM Transactions on Internet Technology (TOIT)*, 19(3), 1–20.
[11] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and Harnessing Adversarial Examples. *International Conference on Learning Representations (ICLR).*
[12] Dietterich,T.G.(2000).Ensemble Methods in Machine Learning. *International Workshop on Multiple Classifier Systems*, 1–15.
[13] Merkel, D. (2014). Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal*, 2014(239), 2.
[14] Aggarwal, C. C. (2013). Outlier Analysis. *Springer*.
[15] M. Mathapati, P. Nandihal, P. Mishra and V. Kotagi, "Improvisation of QoS in SDNFrame Work for UAV Networks Using Dijkstra Shortest Path Routing Algorithm," 2023 Electronics (AIKIIE), Ballari, India, 2023, pp. 1-7, doi: