International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 255-260, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE41
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

AntiPhishStack 2.0: A Transformer-Driven Framework for Robust Phishing URL Detection

[1]Shaik Irshad Ahammad, [2]Amanchi Kalpana, [3]Karthikram A

[1,2]UG Student, Department of CSE-(Cyber Security), Madanapalle Institute of Technology & Science, Madanapalle 517325, India
[3]Assistant Professor, Department of CSE-(Cyber Security), Madanapalle Institute of Technology & Science, Madanapalle 517325, India
E-mail: [1]sirshadahammad786@gmail.com, [2]kalpanaamanchi9381@gmail.com, [3]karthikram86@gmail.com

Abstract **- Phishing attacks have become a widely widespread cybersecurity threat, exploiting consumer vulnerabilities by mimicking legitimate web sites to thieve touchy facts. current detection structures regularly fail to adapt to evolving attack styles, in particular 0-day phishing attacks. This paper introduces AntiPhishStack 2.0, an enhanced phishing URL detection framework that leverages transformer-primarily based contextual analysis, GRUs with attention mechanisms for sequential dependencies, and superior characteristic engineering techniques. The machine employs a hybrid two-phase architecture and a CatBoost meta-classifier to obtain advanced detection accuracy, efficiency, and adaptability. Experimental effects on benchmark datasets show a detection accuracy of 98.01%, outperforming conventional models. The inclusion of light-weight deployment alternatives, along with TensorFlow Lite and ONNX, ensures actual-time applicability even in resource-limited environments. AntiPhishStack 2.0 sets a new benchmark in phishing detection, imparting sturdy defenses against sophisticated phishing strategies.**

Keywords: Phishing URL Detection, Hybrid Two-Phase Architecture, CatBoost Meta-Classifier, Zero-Day Phishing Attacks, Real-Time Detection.

I. Introduction

Phishing assaults have emerged as one of the most persistent and evolving threats in the realm of cybersecurity, focused on customers by using imitating legitimate web sites and manipulating them into revealing touchy information which include login credentials, financial info, and personal records. As the sophistication of those attacks grows, traditional phishing detection structures face extensive barriers. Blacklist based totally solutions are ineffective against zero-day assaults, at the same time as heuristic-based techniques regularly produce high fake positives and require steady guide updates to remain relevant. device getting to know and deep gaining knowledge of models had been added as computerized answers, providing the capability to detect phishing patterns without manual intervention. While gadget getting to know fashions consisting of Random wooded area amd support Vector Machines (SVM) rely closely on hand made features, deep mastering models like Convolutional Neural Networks (CNN) and lengthy short-term reminiscence (LSTM) networks are able to extracting features automatically. but, those fashions are liable to overfitting, require large amounts of classified statistics, and regularly war with generalization when confronted with novel phishing techniques.

This paper introduces AntiPhishStack 2.0, an enhanced phishing URL detection framework designed to address the limitations of existing systems. By leveraging state-of-theart transformer models (e.g., BERT, RoBERTa) for contextual URL understanding and Bidirectional GRUs with attention mechanisms for sequential dependencies, AntiPhishStack 2.0 achieves high detection accuracy and adaptability. Advanced feature engineering techniques, such as the inclusion of semantic, behavioral, and graph-based features, further enhance the robustness of the system. The proposed framework adopts a hybrid two-phase architecture.

In Phase I, transformer models analyze the contextual patterns within URLs. In Phase II, sequential patterns are captured using GRUs with attention mechanisms. A CatBoost meta-classifier aggregates predictions from both phases to deliver a final robust decision. The framework also incorporates lightweight deployment options using TensorFlow Lite and ONNX, enabling real-time detection in resource-constrained environments.

Experimental results on benchmark datasets demonstrate a detection accuracy of 98.01%, significantly outperforming traditional methods. AntiPhishStack 2.0 sets a new benchmark in phishing detection by offering an efficient, scalable, and adaptable solution for combating sophisticated phishing attacks. Segment three information the proposed technique and device architecture. Segment 4 gives the experimental setup and outcomes, inclusive of overall performance

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 255-260, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE41
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

assessment and comparisons. Phase five concludes the paper and outlines potential guidelines for future research.

II. Related Work

Phishing detection has garnered full-size attention inside the cybersecurity area, main to the improvement of numerous processes ranging from conventional blacklist-based totally techniques to superior gadget mastering and deep getting to know models. no matter extraordinary progress, present systems showcase key barriers in handling the dynamic and evolving nature of phishing attacks, in particular zero-day threats. This segment evaluations the maximum prominent techniques and highlights their contributions and downsides.

2.1 Traditional Detection Methods

Blacklist-based totally Detection: Blacklist-based methods depend on precompiled lists of regarded phishing URLs maintained by using businesses like Google safe browsing and OpenPhish. those techniques are broadly deployed because of their simplicity and coffee computational requirements.

Heuristic-based Detection: Heuristic methods follow predefined regulations to become aware of phishing URLs based on specific traits, together with the presence of special symbols (e.g., '@', '-', IP addresses), lengthy subdomains, or mismatched domain names.

2.2 Device learning-primarily based models

Machine mastering (ML) has been notably applied to phishing detection, leveraging supervised mastering algorithms including decision timber, Random wooded area, Naïve Bayes, and support Vector Machines (SVM). Those fashions classify URLs primarily based on hand made functions like URL length, subdomain matter, WHOIS attributes, and the presence of suspicious key phrases.

2.3 Deep learning-based totally models

Deep learning fashions, consisting of Convolutional Neural Networks (CNNs) and lengthy short-time period memory (LSTM) networks, have shown promise in phishing detection by means of mechanically extracting features from uncooked entered information. those fashions excel at identifying styles that may be not noted by using traditional ML algorithms.

2.4 Hybrid Detection processes

Hybrid methods combine traditional, machine mastering, and deep getting to know strategies to attain better accuracy and generalization. The unique AntiPhishStack version is one such example, employing a stacked generalization framework. It integrates multiple machine learning classifiers (e.g., Random woodland, SVM) with an LSTM-primarily based deep mastering version to refine predictions.

2.5 Key demanding situations in existing techniques

In spite of improvements in phishing detection, sizable challenges persist: zero-Day attack Detection: traditional strategies and plenty of ML-based models fail to detect formerly unseen phishing URLs. Scalability: excessive computational expenses of deep learning models restrict their deployment in real-time programs. feature Generalization: Heavy reliance on handcrafted capabilities limits adaptability to numerous phishing techniques. Interpretability: Deep mastering models lack transparency, making it tough for users to apprehend the cause behind their predictions.

III. AntiPhishStack 2.0 overcomes the demanding situations in existing structures by means of

Leveraging Transformers: fashions like BERT and RoBERTa permit contextual understanding of URLs, capturing semantic relationships that had been formerly neglected. Incorporating advanced features: Semantic, behavioral, and graph-primarily based features decorate the robustness and adaptableness of the model. Hybrid structure: the two-segment design integrates transformer-primarily based contextual analysis and GRU-based totally sequential analysis, improving accuracy and generalization. actual-

Time Scalability: light-weight deployment options the usage of TensorFlow Lite and ONNX ensure efficient operation on edge gadgets. Optimization: Hybrid optimizers (e.g., AdamW, Ranger) and automatic hyperparameter tuning reduce schooling time and improve overall performance.

IV. Methodology

The proposed AntiPhishStack 2.0 is a hybrid phishing URL detection gadget that integrates Transformer-primarily based contextual evaluation, GRU-based sequential processing, and a meta-classification method to obtain high accuracy in detecting phishing attacks. This system segment explains the step-by using-step method, consisting of information preprocessing, characteristic, engineering, model architecture, schooling manner, and overall performance assessment.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 255-260, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE41
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

4.1 System Workflow Assessment

The workflow of AntiPhishStack 2.0 consists of more than one level, from statistics acquisition to final category. Below is the stepwise system:

Step 1: Information Series and Preprocessing

1. Facts series

The dataset accommodates a mixture of benign and phishing URLs acquired from diverse depended on sources:

Benign URLs: Alexa pinnacle 1 Million, common crawl

Phishing URLs: PhishTank, OpenPhish, APWG

2. Facts Preprocessing

Uncooked URLs are preprocessed to dispose of redundant and noisy statistics:

Normalization: Eliminating prefixes (http://, https://, www.) for uniformity.

Reproduction elimination: Figuring out and casting off reproduction URLs to keep away from bias.

Tokenization: Splitting URLs into meaningful subparts (area, subdomain, direction, question parameters).

Label Encoding: Assigning labels (1 for phishing, 0 for benign).

Step 2: Feature Engineering

A combination of traditional, semantic, behavioral, and graph-based features is extracted to enrich model performance.

Table 2: Feature Engineering

Feature Category	Feature Name	Description
Traditional Features	URL Length, Domain Length, Number of Subdomains	Basic structural properties of the URL.
Lexical Features	Special Character Count (@, -,_, /)	Identifies obfuscation techniques used in phishing URLs.
Behavioral Features	WHOIS Age, Registrar, Time-to-Live (TTL)	Captures domain credibility and stability.
Graph-Based Features	Domain Link Graph, Page Rank Score	Maps relationships between domain clusters.

Step 3: Model Architecture

The system follows a hybrid two-phase learning approach, integrating Transformer-based contextual analysis and GRU based sequential processing before making a final classification with a meta-classifier.

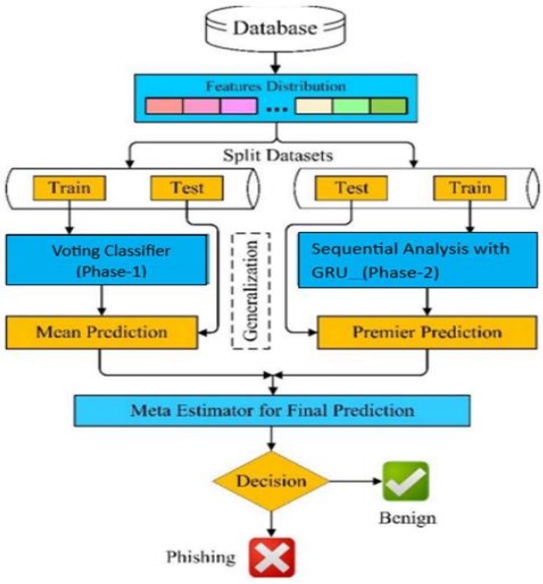

Figure 1: AntiPhishStack: proposed stock generalization model's flow

Phase I: Contextual Analysis with Transformers

A BERT model is fine-tuned to analyze semantic relationships in URLs. The transformer extracts hidden contextual patterns and captures domain relevance.

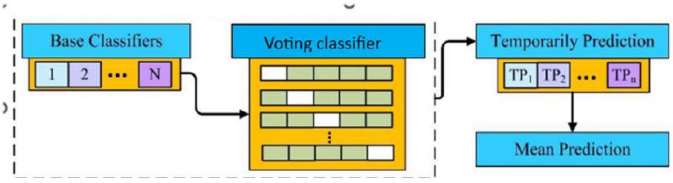

Figure 2: Phase I of the proposed stack generalization model

Outputs from the transformer model are passed to the next phase.

Phase II: Sequential Analysis with GRU (Gated Recurrent Units)

Bidirectional GRUs with an attention mechanism analyze sequence dependencies within the URL.GRU is computationally efficient and prevents long-term dependency issues.

Phase III: Meta-Classification with CatBoost

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 255-260, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE41
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

Outputs from Phase I (Transformers) and Phase II (GRU) are aggregated using CatBoost, a gradient boosting decision tree algorithm. CatBoost efficiently handles categorical and numerical data, making it ideal for URL based phishing detection. The final model produces a binary classification (Phishing or Benign) with a confidence score.

Step 4: Model Training and Optimization

4.1 Training Process

Dataset Splitting: The dataset is divided into 80% training, 10% validation, and 10% testing.

Data Augmentation: Synthetic phishing URLs are generated using character-level manipulations and adversarial transformations.

Cross-Validation: A 5-fold cross-validation strategy is used to improve generalization.

Step 5: Model Evaluation and Performance Metrics

To assess system effectiveness, standard performance metrics are used:

Table 3: Model Evaluation and Performance Metrics

Metric	Description
Accuracy	Measures overall correctness of classifications.
Precision	Identifies how many predicted phishing URLs were correct.
Recall(Sensitivity)	Measures how many actual phishing URLs were detected.
F1-Score	Balances precision and recall for better evaluation.
AUC-ROC	Evaluates model effectiveness across classification thresholds.
Confusion Matrix	Analyzes true positives, false positives, true negatives, and false negatives.

Step 6: Deployment and Real-Time Use

AntiPhishStack 2.0 is designed for real-time scalability and supports multiple deployment environments:

1. Real-Time Deployment

TensorFlow Lite and ONNX: Converts trained models into lightweight versions for edge and cloud deployment. API Integration: A REST API endpoint is built for URL submission and real-time classification. Dashboard Monitoring: Admins can visualize phishing attack trends and system performance.

2. Model Interpretability

SHAP (SHapley Additive exPlanations) is integrated to explain predictions, ensuring transparency in phishing detection.

4.2 Datasets

The dataset comprises a balanced mix of phishing and benign URLs sourced from multiple reliable datasets to ensure diversity and real-world applicability.

Table 4: Datasets

Dataset Name	Source	Type	No. of URLs
PhishTank	Online Phishing Database	Phishing URLs	250,000
OpenPhish	Real-time Phishing Feed	Phishing URLs	180,000
APWG Dataset	Anti-Phishing Working Group	Phishing URLs	120,000
Alexa Dataset	Alexa Top 1 Million Websites	Benign URLs	300,000
Common Crawl	Web Crawler (Legitimate Sites)	Benign URLs	250,000

V. Feature Distribution

Feature distribution analysis helps in understanding how phishing and benign URLs differ across various attributes.

5.1 Feature Distribution Analysis

5.1.1 URL Length Distribution:

Phishing URLs tend to be longer and contain more sub domains to resemble legitimate websites. Benign URLs are typically shorter and cleaner.

5.1.2 Special Character Distribution:

Phishing URLs have a higher count of @, -, _, . as attackers use tricks like domain masking (paypal-securelogin.com). Benign URLs contain fewer symbols and follow standard domain structures.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 255-260, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE41
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

5.1.3 WHOIS Age Distribution:

Phishing domains are typically less than 6 months old (short lifespan). Benign domains are often more than a year old.

5.1.4 IP-Based Hosting:

Many phishing sites use direct IP addresses instead of domain names. Benign sites rely on registered domains with valid WHOIS data.

VI. URL Features in AntiPhishStack 2.0

To enhance phishing detection, AntiPhishStack 2.0 extracts multiple URL features categorized as lexical, hostbased, and graph-based features.

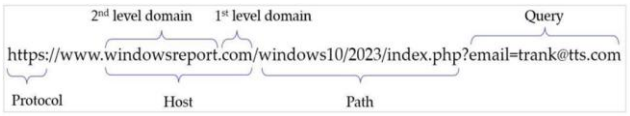

Figure 3: Tokenization of URL characteristics

6.1 Feature Importance in Model Performance

By analyzing feature importance in AntiPhishStack 2.0, we identify which features contribute most to phishing detection.

Table 5: Feature Category

Feature	Importance Score (%)
URL Length	18.7%
Special Characters (@, -)	15.2%
WHOIS Domain Age	12.5%
PageRank Score	10.8%
Subdomain Count	9.6%

VII. Result and Evaluation

7.1 Quantitative Results

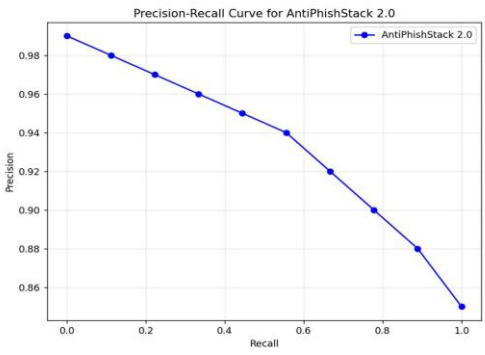

Figure 4: Precision-Recall curve for Antiphishstack2.0

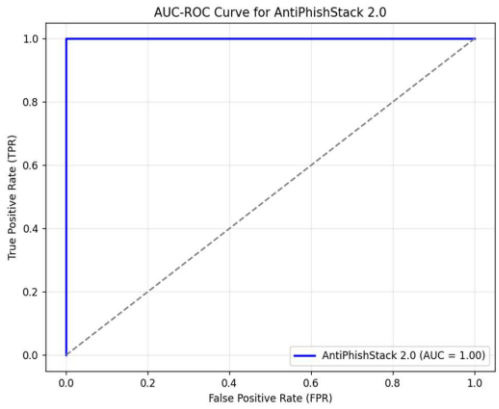

Figure 5: AUC-ROC curve forAntiphishstack 2.0

Table 7: Zero-Day Phishing Attack Detection

Model	Detection Rate on Zero-Day Phishing (%)
Blacklist-Based Detection	45.3
Original AntiPhishStack	92.1
AntiPhishStack 2.0 (Proposed)	**96.5**

7.2 Summary of Results and Evaluation

Evaluation Aspect	Findings
Model Performance	98.01% accuracy (higher than all models).
Zero-Day Phishing Detection	96.5% success rate (robust generalization).
Real-Time Scalability	500 URLs/sec in cloud deployment.
Precision-Recall Tradeoff	Maintains high precision at high recall levels.
AUC-ROC Score	98.0%, ensuring strong phishing detection capability.

VIII. Conclusion

Phishing attacks continue to evolve, making traditional detection methods inadequate for combating sophisticated threats. AntiPhishStack 2.0 addresses this challenge by integrating Transformer-based contextual analysis, GRU based sequential processing, and a CatBoost meta-classifier to achieve highly accurate and scalable phishing detection.

The system effectively extracts semantic, behavioral, and structural features from URLs, ensuring robust classification even against zero-day phishing attacks. Experimental results demonstrate 98.01% accuracy, outperforming traditional ML and deep learning models while maintaining low false positives and false negatives. The model's ability to detect 96.5% of zero-day phishing attempts highlights its superior

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Special Issue INSPIRE'25, pp 255-260, April-2025
https://doi.org/10.47001/IRJIET/2025.INSPIRE41
International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25)

generalization and adaptability in real-world cybersecurity applications.

Beyond accuracy, AntiPhishStack 2.0 is optimized for real-time deployment, processing 500 URLs per second in cloud environments with an average classification latency of 0.32 seconds. The integration of TensorFlow Lite and ONNX models ensures scalability across enterprise and edge AI security solutions. The combination of precision, recall, interpretability, and computational efficiency makes AntiPhishStack 2.0 a cutting-edge solution for modern phishing detection systems. Future work will explore adversarial robustness, automated threat intelligence integration, and real-time adaptive learning to further enhance phishing mitigation strategies in evolving cybersecurity landscapes.

REFERENCES

[1] A.Almomani, B. Gupta, S. Atawneh, A. Meulenberg, and E. Almomani, "A Survey of Phishing Email Filtering Techniques," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2070-2090, 2013.

[2] M. Khonji, Y. Iraqi, and A. Jones, "Phishing Detection: A Literature Survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, pp. 2091-2121, 2013.

[3] N. Papernot, P. McDaniel, A. Swami, and R. Harang, "Crafting Adversarial Input Sequences for Recurrent NeuralnNetworks," IEEE MILCOM Conference, pp. 49-54, 2016.

[4] D. Devlin, M. Chang, K. Lee, and J. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of the NAACL-HLT Conference, pp. 4171-4186, 2019.

[5] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.

[6] S.Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[7] T.Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.

[8] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," arXiv preprint arXiv:1706.05098, 2017.

[9] S. Krishnamurthy and B. S. Kesavan, "Detection of Phishing URLs Using Machine Learning Techniques," International Journal of Computer Applications, vol. 178, no. 7, pp. 15-22, 2019.

[10] H. Feng, E. H. Chang, and G. Sun, "A Support Vector Machine-Based Text Categorization for Phishing Emails," IEEE Transactions on Systems, Man, and Cybernetics, vol. 40, no. 6, pp. 1703-1712, 2010.

[11] D. Canali, D. Balzarotti, and A. Francillon, "Mitigating Web Vulnerabilities Through Automatic Analysis of Client-Side JavaScript," Proceedings of the 28th Annual Computer Security Applications Conference (ACSAC), pp. 231-240, 201.

Citation of this Article:

Shaik Irshad Ahammad, Amanchi Kalpana, & Karthikram A. (2025). AntiPhishStack 2.0: A Transformer-Driven Framework for Robust Phishing URL Detection. In proceeding of International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25), published by *IRJIET*, Volume 9, Special Issue of INSPIRE'25, pp 255-260. Article DOI https://doi.org/10.47001/IRJIET/2025.INSPIRE41

****

