Survey On Phisnet AI - Proactive Defense Against Malicious Url's

A.Nandhini[1], G.Gokulkannan[2], M.Sabeena[3], J.Saranya[4] , J.Shehara Banu[5]
[1]Assist prof, Dept of CSE
[2, 3, 4, 5] Dept of Information Technology
[1, 2, 3, 4, 5] Varuvan Vadivelan Institute Of Technology, Dharmapuri-636701

Abstract- *This project presents PhishNet AI, a proactive and intelligent mobile-based security framework designed to detect and prevent phishing attacks caused by malicious URLs. With the rapid increase in digital communication through SMS, emails, and social media platforms, phishing attacks have become a major cybersecurity threat.*

The proposed system analyzes URL-based, HTML-based, and derived features using advanced feature engineering and machine learning models to classify URLs as legitimate or malicious before user interaction. Unlike traditional systems that rely on reactive detection, PhishNet AI performs real-time analysis and blocks harmful links at an early stage.

The system integrates multiple security mechanisms such as Web Application Firewall (WAF), Intrusion Detection and Prevention Systems (IDS/IPS), bot detection, and secure communication using TLS and OAuth/JWT authentication. It is optimized for mobile devices, ensuring low latency, minimal resource consumption, and high accuracy.

Keywords: Phishing URL Detection, Machine Learning, WAF, Bot Detection, IDS/IPS, OAuth/JWT, TLS Encryption

I. INTRODUCTION

Phishing is one of the most critical cybersecurity threats in today's digital environment, especially with the widespread use of smartphones and online communication platforms. Users frequently receive links through SMS, emails, and social media, often accessing them without proper verification. These malicious URLs redirect users to fraudulent websites that mimic legitimate services such as banking, e-commerce, and payment systems.

Traditional security solutions are mostly reactive and depend on third-party databases such as blacklists and domain reputation services. These approaches are not effective in detecting newly generated phishing attacks and are not optimized for mobile environments.

To address these challenges, PhishNet AI is proposed as a real-time, proactive phishing detection system that uses machine learning and multi-layer security mechanisms. The system ensures early detection and prevention of malicious URLs before user interaction, providing enhanced security for mobile users.

II. PROBLEM STATEMENT

Existing phishing detection systems, as presented in the base paper, are primarily developed as web-based applications. While these systems are effective in controlled environments, they are not suitable for real-time usage in mobile platforms where most phishing attacks occur.

Users today frequently receive malicious links through SMS, WhatsApp, and other mobile applications. However, web-based systems require manual input of URLs and do not provide continuous monitoring, making them ineffective in preventing real-time attacks on mobile devices.

Additionally, these systems depend heavily on third-party services such as blacklists and external databases, which may be outdated and unable to detect zero-day phishing attacks. They also lack support for mobile-specific threats such as QR code phishing and background link monitoring.

Furthermore, most existing solutions operate reactively and do not provide immediate protection before user interaction. This creates a major security gap, especially in mobile environments.

To address these limitations, this project develops a mobile application that is deployed on Android devices, capable of real-time URL monitoring, proactive phishing detection, and automatic blocking of malicious links. This ensures improved security, usability, and accessibility compared to traditional web-based solutions.

If a URL is detected as malicious, the system immediately blocks access and prevents the webpage from loading. A warning notification is displayed to alert the user.

III. RELATED WORK

Several research works have been carried out in the field of phishing detection using different approaches such as signature-based methods, heuristic techniques, and machine learning models.

Signature-based or blacklist approaches are commonly used in many systems to detect phishing URLs by comparing them with a database of known malicious links. These methods are fast and efficient for detecting known threats, but they fail to identify new or zero-day attacks since unknown URLs are not present in the database.

Heuristic or rule-based methods analyze the structure and characteristics of URLs to identify suspicious patterns such as abnormal length, presence of special characters, and lack of secure protocols (HTTPS). These techniques can detect new attacks but may produce false positives due to rigid rule definitions.

Machine learning-based approaches have shown significant improvement in phishing detection. These systems extract features such as URL structure, domain information, and HTML content, and use models like Decision Trees, Random Forest, and Support Vector Machines to classify URLs. According to previous studies, these models can achieve high accuracy and adapt to new phishing patterns.

The base paper primarily focuses on implementing a phishing detection system as a web-based solution using machine learning and feature extraction techniques. While it provides high accuracy, it lacks real-time monitoring and is not optimized for mobile environments.

Recent research has also explored hybrid systems that combine multiple detection techniques to improve performance. These systems integrate signature-based, heuristic, and machine learning methods to provide better accuracy and reduce false positives.

However, most existing systems are limited to web-based platforms and do not provide real-time protection in mobile applications. To overcome these limitations, this project proposes a mobile-based application that performs real-time phishing detection, continuous monitoring, and proactive blocking of malicious URLs.

IV. METHODOLOGY

The proposed system, PhishNet AI, is developed as a mobile application that performs real-time phishing URL detection using a combination of feature extraction, machine learning, and multi-layer security mechanisms.

The methodology consists of the following steps:

- URL Monitoring Module
- URL Preprocessing and Normalization
- Feature Extraction Module
- Feature Selection
- Machine Learning Model
- Multi-Layer Security Mechanism
- Decision and Risk Scoring
- Blocking and Alert Mechanism

Url monitoring module: The system continuously monitors incoming URLs from various sources such as SMS, email, and social media platforms. This module operates in the background and captures URLs without interrupting normal user activity.

Url, preprocessing and normalization: Once a URL is collected, it is passed to the preprocessing module. In this stage, the URL is cleaned by removing unwanted characters, encoding, and obfuscation techniques.The URL is then divided into components such as protocol, domain name, subdomain, path, and query parameters. This structured format helps in accurate analysis and feature extraction.

Feature extraction module: Feature extraction is a key component of the system. The system extracts three types of features:

URL-based features: URL length, number of special characters, subdomains, and HTTPS usage

HTML-based features: forms, scripts, iframes, and login elements

Derived features: complexity score, suspicious patterns, and interaction behavior

These features help in identifying phishing characteristics effectively.

Feature selection: After extraction, the system performs optimized feature selection to remove unnecessary and redundant features. This reduces computational cost and improves real-time performance on mobile devices.

Machine learning model: The selected features are passed to machine learning models trained on labeled datasets. The

models classify URLs as legitimate or phishing based on learned patterns. The system also generates probability scores to represent the confidence level of prediction.

Multi-layer security mechanism: In addition to machine learning, the system uses:

- Signature-based detection
- Heuristic rule-based analysis
- Security components like IDS/IPS and anomaly detection

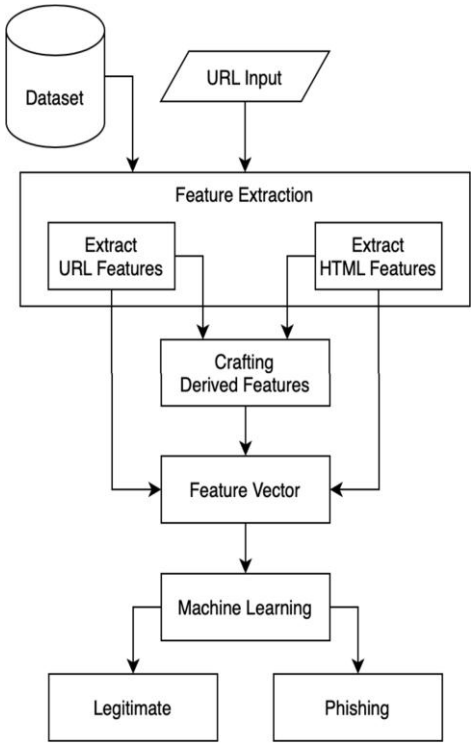
This multi-layer approach improves detection accuracy and handles both known and unknown threats.

Decision and risk scoring: The outputs from different detection layers are combined to calculate a risk score. Based on this score, the URL is classified as safe, suspicious, or malicious.

FUNCTIONAL REQUIREMENTS

The system must provide the following core functionalities:

- **Real-Time URL Analysis:** Capture and analyse URLs instantly when accessed through browsers, applications, or QR code scanners.
- **Multi-Layer Detection Mechanism:** Signature-based detection, heuristic analysis, and machine learning models to identify both known and unknown phishing threats.
- **QR Code Processing:** Decode QR codes and analyse the extracted content as a URL before allowing redirection.
- **Phishing Detection and Blocking:** Block malicious or suspicious URLs before the webpage is loaded, preventing user interaction with harmful content.
- **Typo squatting Detection:** Identify look-alike domains by comparing them with known legitimate domains using similarity analysis.
- **SQL Injection Detection in URLs:** Detect malicious query patterns in URLs, such as SQL injection attempts, following guidelines from OWASP.



- **User Alert Mechanism:** Notify users only when a confirmed threat is detected, minimizing unnecessary interruptions. Users should be able to customize trusted (whitelist) and blocked (blacklist) URLs for personalized security control.

SECTION	DESCRIPTION
Concept	The project focuses on developing a phishing detection system using machine learning techniques to identify malicious URLs and protect users from cyber threats.
Problem Statement	Existing phishing detection systems rely heavily on third-party services like blacklists and WHOIS, which are often outdated, slow, and ineffective against newly generated phishing attacks.
Solution	The proposed system is implemented as a mobile application that uses machine learning algorithms to analyze URL features and classify websites as legitimate or phishing in real-time without relying on external services.
Methodology/ Components	• Data Collection (phishing & legitimate URLs) • Feature Extraction (URL length, symbols, domain info, etc.) • Model Training (ML algorithms like Random Forest, SVM) • Model Evaluation (accuracy, precision) • Deployment (Web/App interface for detection)
Pros	• Real-time detection • Reduced dependency on third-party services • High accuracy using ML models • Scalable and adaptable
Cons	• Requires quality dataset • May give false positives/negatives • Model needs periodic updates • Performance depends on feature selection

V. CONCLUSION

The proposed phishing detection system successfully demonstrates the use of machine learning techniques within a mobile application to identify malicious URLs. By analyzing various URL-based features, the application can effectively

classify websites as legitimate or phishing in real-time without relying on third-party services. The mobile app implementation enhances usability, portability, and accessibility, allowing users to verify URLs instantly and improve their online safety. The system reduces dependency on traditional blacklist-based approaches and provides faster and more efficient detection of phishing attacks. Although the system achieves good accuracy, its performance depends on the quality of the dataset and feature selection. Future improvements can include integrating advanced deep learning models, real-time threat intelligence updates, and expanding feature analysis for better accuracy. Overall, the project provides a practical, scalable, and user-friendly solution for phishing detection, contributing to improved cyber security awareness and protection.

REFERENCES

[1] W. Li, S. Manickam, S. U. A. Laghari, and Y.-W. Chong, "Uncovering the cloak: A systematic review of techniques used to conceal phishing websites," IEEE Access, vol. 11, pp. 71925–71939, 2023, doi: 10.1109/ACCESS.2023.3293063.

[2] H. Cui, Y. Zhou, C. Wang, X. Wang, Y. Du, and Q. Wang,"PPSB: An open and flexible platform for privacy-preserving safe browsing,"IEEE Trans. Dependable Secure Comput., vol. 18, no. 4, pp. 1762–1778, Jul. 2021, doi: 10.1109/TDSC.2019.2937783.

[3] P. Zhang et al.,"CrawlPhish: Large-scale analysis of client-side cloaking techniques in phishing,"in Proc. IEEE Symp. Security and Privacy (SP), May 2021, pp. 1109–1124, doi: 10.1109/SP40001.2021.00021.

[4] APWG,"APWG Q1 Report: Phone-Based Phishing Grows Explosively, Shifting the Cybercrime Threatscape,"[Online]. Available: https://apwg.org/apwg q1-report-phone-based-phishing-grows explosively-shifting-the-cybercrime-threatscape/

[5] J. Spaulding, S. Upadhyaya, and A. Mohaisen,"You've been tricked! A user study of the effectiveness of typosquattingtechniques,"in Proc. IEEE ICDCS, Jun. 2017, pp. 2593–2596, doi: 10.1109/ICDCS.2017.221.

[6] P. Kintiset al.,"Hiding in plain sight: A longitudinal study of combosquattingabuse,"in Proc. ACM SIGSAC CCS, Oct. 2017, pp. 569–586, doi: 10.1145/3133956.3134002.

[7] A. Prasad et al.,"PermGuard: A scalable framework for Android malware detection using permission-to-exploitation mapping,"IEEE Access, vol. 13, pp. 507–528, 2025, doi: 10.1109/ACCESS.2024.3523629.

[8] A. Prasad and S. Chandra,"BotDefender: A collaborative defense framework against botnet attacks using network traffic analysis and machine learning,"Arabian Journal for Science and Engineering, vol. 49, no. 3, pp. 3313–3329, Mar. 2024.

[9] A. Prasad et al., "A collaborative prediction approach to defend against amplified reflection and exploitation attacks," Electronic Research Archive, vol. 31, no. 10, pp. 6045–6070, 2023.

[10] A. Safi and S. Singh, "A systematic literature review on phishing website detection techniques," Journal of King Saud University – Computer and Information Sciences, vol. 35, no. 2, pp. 590–611, Feb. 2023.

[11] N. A. B. M. Zin et al.,"Machine learning technique for phishing website detection,"in Proc. IEEE ICSECS, Aug. 2023, pp. 235–239.

[12] O. K. Sahingoz, E. Buber, O. Demir, and B. Diri,"Machine learning based phishing detection from URLs,"ExpertSystems with Applications, vol. 117, pp. 345–357, Mar. 2019.

[13] B. B. Gupta et al.,"A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment," Computer Communications, vol. 175, pp. 47–57, Jul. 2021.

[14] A. N. S. Charan, Y.-H. Chen, and J.-L. Chen, "Phishing websites detection using machine learning with URL analysis," in Proc. IEEE AIC, Jun. 2022, pp. 808–812.