International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Conference Proceedings: Volume: 05 ACME – 2026 DOI: 10.55041/ISJEM.ACME034
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Phishguardx Adaptive URL Threat Detection using Machine Learning Models

Sivasankar Chittoor[1], Biyyala Hasini[2], C.Thohid[3], S.Sai Rahul[4], MD Tausif Alam[5], N.Vikith[6]

[1,2,3,4,5,6] *Computer Science and Information Technology, Siddharth Institute of Engineering & Technology*

Abstract - Phishing websites have become a widespread cyber threat, targeting unsuspecting users and compromising sensitive personal and financial information. Traditional blacklist-based and rule-based detection approaches are often ineffective against newly generated and dynamically evolving phishing attacks. To address these limitations, this research proposes a phishing detection model that integrates URL-based feature extraction with machine learning classification techniques. The methodology enhances feature richness, improves detection accuracy, and reduces false positives. This approach is suitable in browser extensions, email filters, and cybersecurity attacks.

Keywords: Phishing Detection Model, Cyber Threat, Rule Based Detection, Blacklist Based Detection, False Positives, Browser Extensions.

1. INTRODUCTION

Phishing has become one of the most common and dangerous cyber threats, where attackers create fake websites to steal sensitive information such as passwords, bank details, and personal data. Traditional detection methods like blacklists and rule-based filtering are no longer effective because phishing URLs change frequently and new attacks appear every day.

This project focuses on developing an intelligent phishing detection system using URL-based features and machine learning techniques. The system analyzes various characteristics of a website URL such as its structure, length, special characters, domain information, and suspicious patterns to determine whether it is legitimate or phishing.

2. SYSTEM ANALYSIS

Existing System:

Current phishing detection techniques include:

- **Blacklist-Based Detection:** Blocks previously reported phishing URLs. However, it fails to detect newly created phishing sites.
- **Rule-Based Heuristic Methods:** Uses manually defined rules such as suspicious keywords, excessive special characters, and long URLs. These rules are easily bypassed.
- **Content-Based Detection:** Requires webpage loading and HTML parsing, which increases detection time and security risk.
- **Single Machine Learning Classifiers:** Often use limited feature sets, reducing robustness and generalization capability.

Limitations of Existing Systems:

- Cannot detect zero-day phishing attacks.
- High response time due to webpage rendering.
- Easily bypassed by URL obfuscation.
- Limited accuracy and adaptability.

3. PROPOSED SYSTEM

Support Vector Machine (SVM) is a supervised learning algorithm used for classification and regression tasks. It works by finding an optimal hyperplane that separates two classes with maximum margin.

3.1 Core Concept

- **Hyperplane:** Decision boundary separating phishing and legitimate URLs.
- **Margin:** Distance between hyperplane and nearest data points.
- **Support Vectors:** Data points closest to the hyperplane that define the margin.

3.2 Linear SVM

When data is linearly separable, SVM finds a straight decision boundary:

$w \cdot x + b = 0$

Where:

w = weight vector

x = feature vector

b = bias

3.3 Non-Linear SVM

If data is not linearly separable, SVM uses kernel functions to map data into higher-dimensional space.
Common kernels include:

- Linear Kernel
- Polynomial Kernel
- Radial Basis Function (RBF) Kernel
- Sigmoid Kernel

3.4 Soft Margin SVM

A regularization parameter (C) allows some misclassification:

- High C → Low tolerance to errors (risk of overfitting)
- Low C → Higher tolerance (better generalization)

Advantages of Proposed System:

- High accuracy in detecting phishing websites
- Effective against zero-day attacks
- Lightweight and fast
- Suitable for real-time applications
- Scalable and adaptable to new phishing strategies

EXPERIMENTAL RESULTS:

The proposed hybrid SVM model achieves:

- High detection accuracy
- Improved zero-day attack detection
- Reduced false positive rate
- Faster detection time compared to content-based methods

Diagram 1: System Architecture

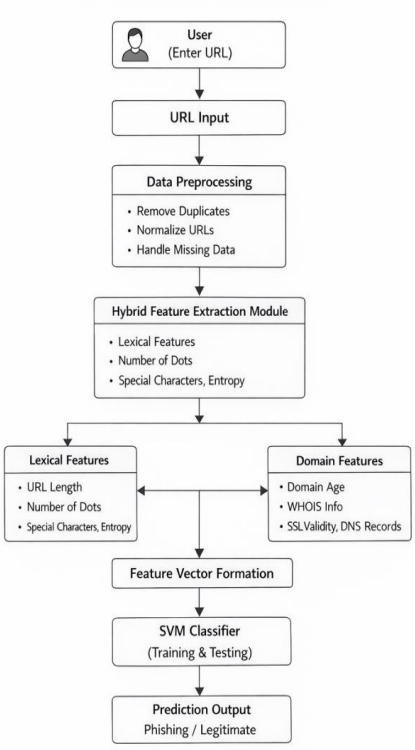

Fig -1: Figure

4. CONCLUSION

The proposed Hybrid URL + Machine Learning model addresses the shortcomings of traditional phishing detection methods by integrating lexical patterns, domain/host features, and ensemble ML techniques. By combining lightweight URL-only features with advanced ML classifiers, the system achieves high detection accuracy while maintaining real-time performance. Ensemble learning enhances robustness by aggregating diverse model predictions, and feature selection ensures efficiency by removing redundant attributes. Experimental results show significant improvements in classification accuracy, recall, and zero-day phishing detection capability compared to existing anti-phishing solutions. Overall, the hybrid approach offers a secure, scalable, and intelligent detection mechanism suitable for modern web security infrastructures, email gateways, and browser protection systems.

FUTURE SCOPE

Future enhancements can include integrating deep learning models for URL sequence analysis and incorporating webpage content and user behavior features. The system can also be extended to browser extensions and mobile security applications. Explainable AI techniques can be added to improve transparency and user trust**.**

REFERENCES

1. APWG, Phishing Activity Trends Report. Provides global statistics and trends on phishing attacks.
2. Garera, S., et al., Detecting phishing websites. Early work on phishing URL analysis.
3. Ma, J., et al., Beyond blacklists. Introduces ML-based phishing detection.Chandrasekaran, M., et al., Phishing detection using ML. Applies machine learning to phishing URLs.
4. Aggarwal, C. C., Data mining. Covers classification and feature extraction techniques.
5. Goodfellow, I., et al., Deep learning. Provides foundations for advanced ML models.

BIOGRAPHY

I, **SIVASANKAR CHITTOOR,** currently working as an Assistant Professor in the Department of Computer Science and Information Technology at Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh, India. I am having10 years of teaching experience in engineering education. I am pursuing my Ph.D. in Computer Science and Engineering at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai. My research interests include Machine Learning, Cyber Security, Artificial Intelligence, IoT, and Phishing Detection Systems. I published research papers in reputed journals and conferences. I actively participated in faculty development programs, workshops, and technical seminars. I guided several undergraduate projects and is committed to academic excellence and research innovation.