E-ISSN: 2583-9667
Indexed Journal
Peer Reviewed Journal
https://multiresearchjournal.theviews.in

Received: 16-01-2025
Accepted: 27-02-2025

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH IN MULTIDISCIPLINARY

Volume 3; Issue 2; 2025; Page No. 155-158

Phish shield: AI-powered browser extension for phishing detection

[1]Dr. Parameswari R and [2]S Akashkumar

[1]Professor and Head, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India
[2]Student, Department of Computer Science and Information Technology, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai, Tamil Nadu, India

DOI: https://doi.org/10.5281/zenodo.15575050

Corresponding Author: Dr. Parameswari R

Abstract

Phishing scams have increasingly become a significant cybersecurity concern, luring users into disclosing sensitive information through fake websites. Enter PhishShield, an innovative phishing detection system that utilizes machine learning, seamlessly integrated into a browser extension. At its core, PhishShield harnesses a trained Random Forest model to analyze URLs and classify them as either phishing attempts or legitimate sites based on a variety of extracted features. The backend, developed using Python and Flask, hosts the model and serves as the API for processing URL queries. On the other hand, the frontend operates as a Chrome extension that delivers real-time detection capabilities. With PhishShield, users can navigate the internet with greater confidence, as the tool actively works to prevent them from becoming victims of phishing attacks. To achieve accurate classifications, PhishShield employs a comprehensive Random Forest classifier trained on a vast dataset of both phishing and legitimate URLs. It extracts a multitude of features related to the URLs, such as their length, the inclusion of special characters, the use of IP addresses, domain age, and SSL certificate validity. These specially designed features allow the model to effectively pinpoint deceptive websites, steering clear of traditional blacklist methods. The backend effectively manages feature extraction and model inference, responding promptly to requests from the frontend. When a user visits a page, the Chrome extension captures the current URL and sends it for classification. If the site turns out to be suspicious, users receive an immediate warning, advising them not to proceed, thus providing an extra layer of security in their online activities.

Keywords: Phishing detection system, Random Forest model, URL classification, Feature extraction, Python, Flask, API, Real-time detection, User warning system, Blacklist methods, Online security, Frontend-backend integration

Introduction

Phishing attacks represent a significant and evolving cybersecurity threat in the digital age. These deceptive attempts often involve the creation of malicious websites that closely mimic legitimate platforms, aiming to trick users into divulging sensitive information such as login credentials, financial details, and personal data. The consequences of falling victim to phishing can be severe, ranging from identity theft and financial fraud to significant data breaches for organizations.

Traditional security mechanisms, such as static blocklists of known malicious URLs, struggle to keep pace with the rapid emergence of new phishing domains and sophisticated attack techniques. Attackers can easily bypass these lists by registering new domains or employing URL obfuscation methods. Furthermore, relying solely on user awareness is insufficient, as even tech-savvy individuals can be deceived by well-crafted phishing campaigns. Rule-based detection systems often lack the adaptability required to identify novel phishing strategies.

With the increasing number of cyberattacks, phishing has emerged as one of the most deceptive methods to steal credentials, financial details, and personal information. Traditional security solutions rely on blocklists, which can be bypassed by attackers using new domains**.** PhishShield overcomes this limitation by using machine learning algorithms to analyze URL structures and detect phishing attempts in real-time. The extension integrates a trained phishing detection model into a browser environment, offering users a seamless way to identify malicious websites before interacting with them.

PhishShield addresses these limitations by introducing an AI-powered, real-time phishing detection system seamlessly integrated into a user's browsing experience as a Chrome

extension. By leveraging machine learning, specifically a trained Random Forest Classifier, PhishShield analyzes the characteristics of URLs in real-time to predict whether they are indicative of a phishing attempt. The system comprises a Python-based Flask backend that hosts the trained model and a user-friendly Chrome extension that extracts the active URL and communicates with the backend for instant analysis. This proactive approach aims to significantly enhance online security by identifying and alerting users to potential phishing threats before they can interact with malicious websites.

Literature Review

Phishing detection has been a growing area of research due to the increasing sophistication and frequency of online phishing attacks. Traditional approaches primarily relied on blacklist-based detection, where known phishing URLs are stored in a database and checked against incoming URLs. While simple to implement, this approach is highly ineffective against zero-day attacks, as new phishing URLs can bypass the blacklist until reported and updated.

To overcome the limitations of static lists, researchers began to explore z which use predefined rules based on URL structures, domain patterns, and website behavior. However, these rule-based systems often suffer from high false positives and are difficult to maintain as phishing techniques evolve.

With the advent of machine learning (ML), the focus shifted toward intelligent phishing detection systems capable of learning from large datasets. Algorithms such as Support Vector Machines (SVM), Naive Bayes, Decision Trees, and Random Forest Classifier have been widely used for classifying URLs as phishing or legitimate based on features such as:

- Length of URL
- Use of HTTPS protocol
- Presence of IP addresses in the URL
- Age of the domain
- Use of suspicious keywords or subdomains

For instance, the study by Abdelhamid *et al*. (2014) [1] applied associative classification to phishing detection and achieved promising results by combining content-based and address-bar-based features. Similarly, Mohammad *et al*. (2015) [2] introduced a phishing detection method using machine learning models trained on 30+ features, reporting high accuracy using Random Forests and Decision Trees.

More recent research, like Marchal *et al*. (2016) [3], emphasized the importance of real-time detection through browser plugins and lightweight models, paving the way for practical deployments. These studies have shown that ML-based systems can significantly outperform traditional approaches in terms of accuracy, adaptability, and speed.

However, real-world adoption faces challenges such as

- Efficient feature extraction in real-time
- Minimizing latency in browser-based environments
- Ensuring model generalization on unseen data
- Providing a user-friendly interface for non-technical users

PhishShield builds upon this body of work by integrating a trained Random Forest model into a Chrome browser extension. Unlike previous work that focused solely on backend classification, this project emphasizes real-time usability, lightweight deployment, and seamless integration with the user's browsing experience. It bridges the gap between academic research and practical application by offering a proactive, intelligent defense system against phishing attacks.

Proposed Solution

Proactive and Real-Time Detection: PhishShield harnesses machine learning to analyze URLs in real-time, enabling it to identify emerging phishing threats that may not yet be on existing blocklists. Seamless Browser Integration: As a Chrome extension, PhishShield offers immediate protection directly within your browsing experience. AI-Driven Adaptability: The machine learning model can be continuously updated with fresh data, allowing it to adapt to new phishing techniques more effectively than traditional rule-based systems. Efficient Design: The goal is to conduct thorough analysis without noticeably slowing down your browser's performance. Clear User Alerts: Users receive instant and straightforward warnings through browser notifications, ensuring they stay informed. Room for Future Growth: The underlying architecture is designed to support more advanced capabilities down the line, such as content analysis and integration with cloud-based threat intelligence. The proposed system, PhishShield, is designed to detect and prevent phishing attacks in real-time using a machine learning-based approach integrated into a browser environment. Unlike traditional anti-phishing techniques that rely on static blacklists or user reports, PhishShield leverages the predictive power of a trained Random Forest classifier to analyze URL characteristics and identify suspicious behavior dynamically.

The system consists of two primary components: a Python-Flask backend and a Chrome browser extension frontend. The backend hosts the trained Random Forest model and is responsible for extracting features from URLs, performing classification, and returning results. These features include lexical attributes such as URL length, presence of special characters, use of IP addresses, subdomain count, HTTPS usage, and domain-related metadata like age and registrar information. This ensures that detection is not dependent on site content or user interaction, making it lightweight and fast.

The frontend, implemented as a browser extension, operates silently in the background as users navigate the web. Each time a new page is loaded, the extension captures the current URL and sends it to the backend via a secure API request. The backend processes the request, extracts the relevant features, and feeds them to the Random Forest model for classification. The model then determines whether the URL is likely to be phishing or legitimate. If the result is flagged as phishing, the extension instantly alerts the user with a warning notification and provides options to exit or proceed with caution, thereby minimizing the risk of falling victim to fraudulent websites.

PhishShield is designed with real-time performance and user transparency in mind. The system ensures minimal latency in communication between the browser and the backend, and its unobtrusive interface aims to enhance the user

experience without interfering with regular browsing. Additionally, the machine learning model can be retrained and updated periodically to adapt to emerging phishing techniques, offering continuous protection against evolving threats.

By integrating intelligent URL analysis with seamless browser-level interaction, PhishShield offers a proactive and practical solution to combat phishing attacks. Its architecture not only improves detection accuracy but also empowers users with immediate awareness, bridging the gap between machine intelligence and end-user cybersecurity, PhishShield, is a browser-embedded cybersecurity tool designed to detect phishing websites proactively using intelligent, machine learning-driven URL analysis. This system shifts the conventional approach of phishing detection from reactive, database-reliant solutions to a predictive model capable of identifying suspicious web activity in real time. By embedding this technology directly into a browser extension, PhishShield offers users seamless and continuous protection as they navigate the internet, without requiring manual intervention or prior knowledge of malicious websites.

At the heart of PhishShield lies a machine learning pipeline trained using a labeled dataset containing both phishing and legitimate URLs. The model, built using the Random Forest algorithm, is capable of learning complex patterns from a diverse set of URL features. These features are primarily drawn from the structural and semantic properties of the URL itself, such as the frequency of unusual characters, subdomain structure, use of HTTPS, domain entropy, and registration data. This allows the model to assess potential risks based on the intrinsic nature of the URL, rather than relying on known blacklisted domains.

The system architecture is divided into two synergistic components: the browser-based frontend and the Flask-powered backend. The browser extension operates as a real-time monitoring agent, capturing the active tab's URL and sending it to the backend for evaluation. Upon receiving the request, the backend processes the URL, extracts relevant features, and feeds them into the trained model. A decision is returned instantly, classifying the URL as either phishing or safe. In the case of a threat, the user is immediately presented with a warning, along with a recommended action to ensure their safety.

PhishShield emphasizes low-latency response and minimal system overhead, making it ideal for real-world deployment where speed and accuracy are paramount. Furthermore, its modular design allows for future enhancements, such as adding support for multilingual URL patterns, cloud-based model retraining, or behavioral analysis based on user interaction with web content. Overall, PhishShield offers a forward-thinking, automated defines system that strengthens user awareness and security without compromising the web browsing experience.

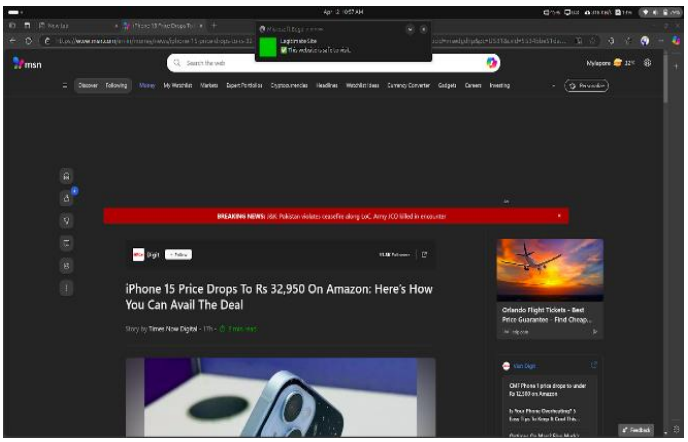

Fig 1

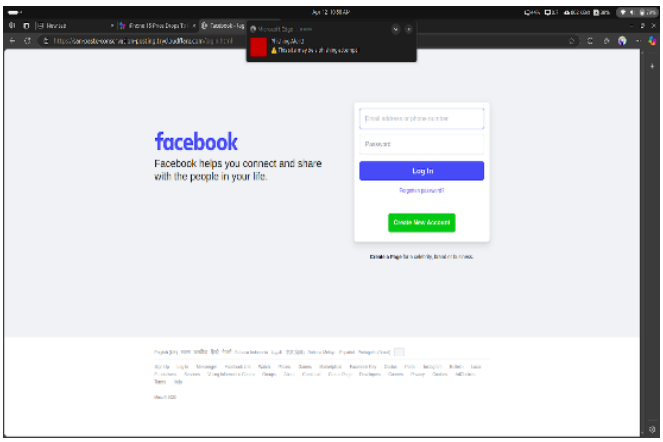

Fig 2

Conclusion

Phishing remains a persistent and evolving threat in today's digital landscape, targeting individuals and organizations alike. This project, PhishShield, offers a proactive defense mechanism by integrating machine learning into the web browsing experience, enabling real-time detection of phishing attempts. By leveraging a Random Forest classifier trained on key URL-based features, the system can accurately distinguish between malicious and legitimate websites without relying on outdated blacklists or manual user reports.

Through its dual-structured architecture-combining a lightweight Python-Flask backend with a responsive Chrome extension frontend-PhishShield delivers fast, efficient, and user-friendly protection. The system's ability to analyze URLs dynamically as users browse ensures a seamless and secure online experience. Additionally, the model's adaptability to retraining and updates positions it well for handling the continuously shifting tactics employed by attackers.

Overall, PhishShield demonstrates the effectiveness of intelligent, machine learning-driven cybersecurity tools in combating phishing threats. It not only enhances user safety but also sets the foundation for more advanced, scalable solutions that can evolve with the threat landscape. With further development, this system has the potential to become a comprehensive security layer for everyday internet users.

References

1. Abdelhamid N, Ayesh A, Thabtah F. Phishing detection based on associative classification data mining. Expert Systems with Applications. 2014;41(13):5948–5959. doi:10.1016/j.eswa.2014.03.019
2. Mohammad RM, Thabtah F, McCluskey L. Intelligent phishing detection system using rule induction techniques. Expert Systems with Applications.

2015;40(11):4697–4706. doi:10.1016/j.eswa.2012.12.073
3. Marchal S, Saari K, Singh N, Asokan N. Know your phish: Novel techniques for detecting phishing sites and their targets. In: 2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS); 2016 Jun 27–30; Nara, Japan. IEEE; c2016. p. 323–333. doi:10.1109/ICDCS.2016.48
4. Basnet R, Sung AH, Liu Q. Rule-based phishing attack detection. In: Proceedings of the 2012 International Conference on Security and Management (SAM); 2012 Jul 16–19; Las Vegas, NV, USA. CSREA Press; c2012. p. 1–7.
5. Rao RS, Pais AR. Detection of phishing websites using an efficient feature-based machine learning framework. Computers & Security. 2018;73:291–307. doi:10.1016/j.cose.2017.11.008
6. Jain AK, Gupta BB. A machine learning based approach for phishing detection using hyperlinks information. Journal of Ambient Intelligence and Humanized Computing. 2019;10(5):2015–2028. doi:10.1007/s12652-018-0871-9
7. Verma R, Das A. What's in a URL: Fast feature extraction and malicious URL detection. In: Proceedings of the 3rd ACM Conference on Data and Application Security and Privacy (CODASPY); 2017 Mar 22–24; Scottsdale, AZ, USA. ACM; c2017. p. 300–307. doi:10.1145/3029806.3029810