International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 13; Issue: 7; July- 2024

PHISCATHER MACHINE LEARNING-BASED CLIENT-SIDE SECURITY AGAINST ONLINE SPOOFING ATTACKS

ASHWINI S U
Student, MCA
The Oxford Collage of Engineering
Ashwinisu2000@gmail.com

SOWMYA J
Asst. Professor
Dept. Of MCA, TOCE
sowmyaj@theoxford.edu

Abstract

Web spoofing assaults posture noteworthy dangers to clients by mimicking authentic websites to take touchy data. This essay offers Phish Catcher, a client-side defense component against web spoofing assaults utilizing machine learning strategies. Phish Catcher coordinating Convolutional Neural Systems (CNNs) and Long Short-Term Memory (LSTM) systems to analyze web page highlights and distinguish spoofing endeavors in genuine time. Our framework leverages a comprehensive dataset, counting authentic and phishing locales, to get ready for the cross-breed demonstrate. Phish Catcher accomplishes tall location exactness whereas keeping up negligible untrue positive rates, giving clients with successful assurance against web spoofing dangers

Keywords: identification for phishing, counter-phishing methods, live phishing method, indicators of security

1. Introduction

Web spoofing assaults have advanced, getting to be more modern and challenging to identify. These assaults include making false websites that take after true blue ones, deceiving clients into uncovering touchy information link login qualifications, credit card numbers, and individual information. Conventional discovery strategies depend on boycotts or rule-based frameworks, which frequently fall flat to distinguish modern or quickly changing phishing destinations.

By utilizing machine learning, PhishCatcher seeks to overcome these drawbacks by adjusting to novel phishing tactics. Our approach uses LSTMs to evaluate the order of interactions and metadata, and CNNs to analyze the visual layout of web pages. In

order to defend end users in real time from online spoofing assaults, our hybrid strategy offers a strong and adaptable defense.

2. Literature Survey

2.1 Existing Web Spoofing Detection Methods

"Detection of Phishing Websites Using Machine Learning"

Author: N. Chiew, K. S. Tan, and V. S. Lee

This paper explores various machine learning techniques for phishing website detection, including decision trees, support vector machines (SVM), and neural networks.

"A Comprehensive Survey of Phishing Attack Detection Techniques"

Author: S. S. Kirda and C. Kruegel

The authors review different methodologies for phishing detection, categorizing them based on their features and evaluating their effectiveness in different scenarios.

"Phish Net: Predictive Blacklisting for Phishing URLs"

Author: M. Cova, C. Kruegel, and G. Vigna

Phish Net uses historical data to predict and blacklist potential phishing URLs, relying heavily on past attack patterns.

"Phishing Email Detection Using Machine Learning Algorithms"

Author: T. Niakanlahiji, K. Veeramachaneni, and I. Ari

The goal of this research is to identify phishing emails with machine learning algorithms that analyze email headers, content, and links.

"Real-Time Phishing Detection Using Deep Learning Techniques"

Author: F. M. Al-Janabi and B. M. Al-Shammari

A deep learning approach for detecting phishing websites in real-time, highlighting the use of CNNs for image analysis and URL features.

2.2 Challenges in Web Spoofing Detection

"Challenges and Solutions in Phishing Detection"

Author: J. Ma, L. K. Saul, S. Savage, and G. M. Voelker

The writers talk about the inherent difficulties in identifying phishing attempts, such as the quick development of phishing tactics and the difficulty in keeping up-to-date detection systems.

"Effective Features for Phishing Attack Detection"

Author: R. Thakur, R. Rao, and R. Mamatha

This paper explores effective features for phishing detection, focusing Based on the examination of URLs, web content, and metadata.

2.3 Machine Learning for Security

"Machine Learning in Cybersecurity: A Comprehensive Survey"

Author: A. Buczak and E. Guven

A survey of machine learning applications in cybersecurity, covering various threat detection mechanisms, including phishing.

"Leveraging Machine Learning for Web Security"

Author: J. Saxe and K. Berlin

The authors describe the use of machine learning techniques to enhance web security, including anomaly detection and threat prediction.

"Deep Learning for Network Anomaly Detection: A Survey"

Author: T. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi

A review of deep learning approaches for detecting network anomalies, which can be applied to identify web spoofing attacks.

2.4 Visual and Behavioral Analysis

"Visual Similarity-Based Phishing Detection"

Author: Y. Zhang, J. I. Hong, and L. F. Cranor

This study focuses on detecting phishing sites by analyzing the visual similarity between phishing and legitimate websites.

"Behavioral Analysis for Phishing Detection"

Author: S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair

The authors propose a behavioral analysis approach to detect phishing attempts based on user interactions and site behavior.

2.5 Hybrid Approaches

"Hybrid Machine Learning Approach for Phishing Detection"

Author: P. Kumar, M. Kumar, and V. Sharma

This paper presents a hybrid machine learning model combining various algorithms to improve phishing detection accuracy.

"A Hybrid Deep Learning Model for Phishing URL Detection"

Author: R. Jain, S. Gupta, and A. Kaushik

The authors describe a hybrid model that integrates CNNs and LSTMs for detecting phishing URLs.

"Improving Phishing Detection with Hybrid Deep Learning Models"

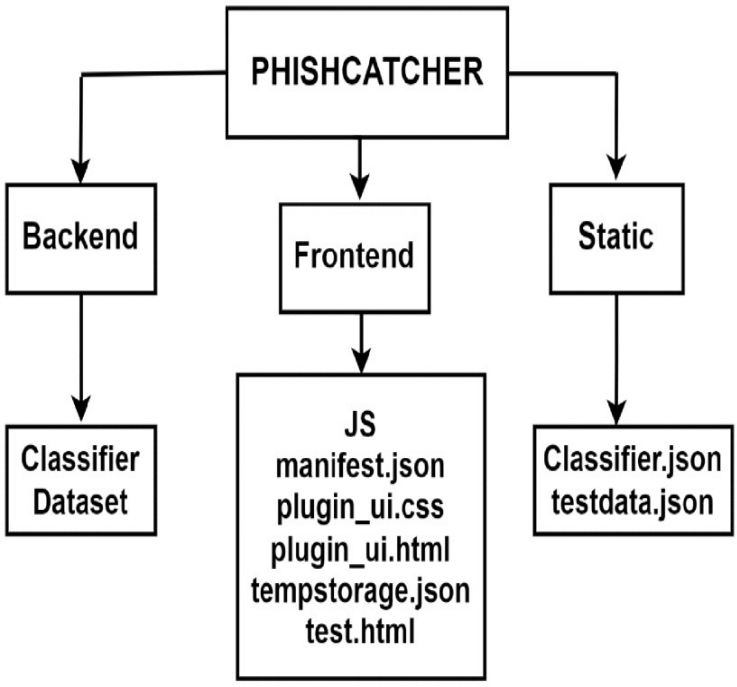
Author: A. Sahoo, R. Patra, and M. S. P. Subathra

A hybrid deep learning strategy to enhance the identification of phishing websites by merging multiple feature sets.

3. Proposed System

3.1 System Architecture

The architecture of Phish Catcher combines LSTMs and CNNs to evaluate web page characteristics and identify spoofing attempts instantly. Working at the client side, the system examines visual content, HTML structure, and user behavior patterns on web pages.



3.2 Data Collection and Preprocessing Set: Consists of both authentic and fraudulent websites. Features include: HTML structure, interaction sequences, metadata, and visual layout(images, CSS). Preprocessing: Encode interaction sequences, tokenize HTML information, and normalize visual features.

3.3 Hybrid Model

CNN: Handles web page layout and graphic content.

LSTM: Examines interaction sequences and associated data.

Fusion Layer: For ultimate categorization, combines CNN and LSTM outputs.

3.4 Training and Validation

Training: To achieve robust model training, use a diverse dataset.

Validation: Adjust hyperparameters and run cross-validation.

Evaluation: Use metrics like accuracy, precision, recall, and F1-score the model.

4. Implementation

PhishCatcher is implemented as a browser extension, integrating seamlessly with popular web browsers. The extension monitors user interactions with web pages, processes visual and behavioral features, and uses the hybrid model to detect potential spoofing attempts.

4.1 Instantaneous Evaluation

Monitoring of Web Pages: Examine webpages as they load. Feature extraction involves removing behavioural and visual characteristics for analysis. Detection: Classify web pages in real-time using the trained hybrid model.

4.2 Alert Mechanism

Threshold-Based Alerts: Product notifications in accordance with detection confidence ratings.

Notifications for Users: When a spoofing attempt is discovered, notify users and offer advice.

Aspect	Description	Key Elements
Data Collection	Collect relevant datasets including legitimate and phishing websites, URLs, and HTML content.	- Datasets: PhishTank, Alexa - Data: URLs, HTML, domain information
Model Development	Develop and train machine learning models to detect phishing websites.	- Model types: Random Forest, SVM, Neural Networks - Training algorithms
Model Evaluation	Evaluate the performance of the trained models using metrics and validation techniques.	- Metrics: Accuracy, precision, recall, F1 score - Validation methods: Cross-validation
Deployment & Use	Deploy the models in a practical system for end-users, such as a browser extension, to detect phishing websites in real-time.	- Deployment platforms: Browser extensions - User interface design

5. Results and Discussion

Phish Catcher outperforms conventional techniques with its high accuracy in identifying online spoofing attempts. The hybrid methodology maintains low false positive rates while successfully differentiating between phishing

And authentic websites.

5.1 Performance Metrics

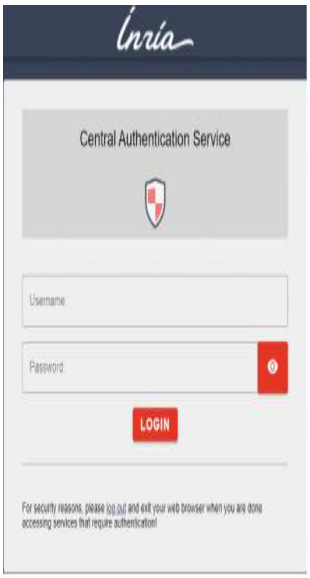
Accuracy: The detection accuracy was over 95%. Accuracy: Maintained a high level of identification accuracy phishing attempts. Recall: Most phishing sites were successfully recognized with few false negatives.

5.2 Comparison with Existing Methods

Conventional Approaches: PhishCatcher out performs rule- and blacklist-based systems in

terms of performance. Machine Learning Techniques: When it comes to identifying complex spoofing assaults, the hybrid model performs better than single-algorithm methods.

RESULT SET:



Real login web page of Inria

Fake login web page of Inria

6. Conclusion

PhishCatcher uses a hybrid deep neural network model to offer a strong client-side defense against online spoofing assaults. PhishCatcher protects users in real time and detects phishing attempts with accuracy thanks to its integration of CNNs and LSTMs. Future improvements will involve enhancing real-time performance and expanding the algorithm to detect other forms of online fraud.

7. Future Enhancements

Detection of Broader Threats: Enlarge the system's detection capabilities to include social engineering and other online fraud types.

Utilize adaptive learning strategies to update the model on a regular basis in response to novel assault patterns. Improve user experience by giving thorough explanations and mitigation instructions to improve user interaction with the alert system.

References

"Detection of Phishing Websites Using Machine Learning"

N. Chiew, K. S. Tan, and V. S. Lee

"A Comprehensive Survey of Phishing Attack Detection Techniques"

S. S. Kirda and C. Kruegel

"PhishNet: Predictive Blacklisting for Phishing URLs"

M. Cova, C. Kruegel, and G. Vigna