PhishNinja

12/11/2023

20201CCS0131	Umraz Khan
20201CCCS0063	Jonathan R
20201CCS0132	Lenin E
20201CCS070	Samrudh D Yash
20201CCS0046	K Rehan

Abstract

As the prevalence of online communication and transactions continues to rise, the threat of phishing attacks has become a critical concern for individuals and organizations alike. Phishing attacks exploit human vulnerabilities by tricking users into divulging sensitive information, such as login credentials and financial details. This paper introduces an advanced phishing detection algorithm designed to enhance cybersecurity measures and mitigate the risks associated with phishing.

The proposed algorithm combines machine learning techniques with feature engineering to create a robust and adaptive system for detecting phishing attempts. It leverages a diverse set of features, including URL analysis, content inspection, and behavioral patterns, to comprehensively evaluate the legitimacy of a given webpage. The machine learning model is trained on a large dataset of known phishing and legitimate websites, enabling it to learn and generalize patterns indicative of phishing behavior.

Key components of the algorithm include:

- **URL Analysis**: The algorithm evaluates the structure and components of URLs to identify suspicious patterns, such as misspelled domain names or the presence of subdomains commonly associated with phishing.

- **Content Inspection**: Deep content analysis is performed to assess the webpage's textual and visual elements. Natural Language Processing (NLP) techniques are employed to identify phishing indicators in the page content.

- **Behavioral Patterns**: The algorithm monitors user interactions with web pages, including mouse movements and click patterns, to detect anomalous behavior that may be indicative of a phishing attempt.

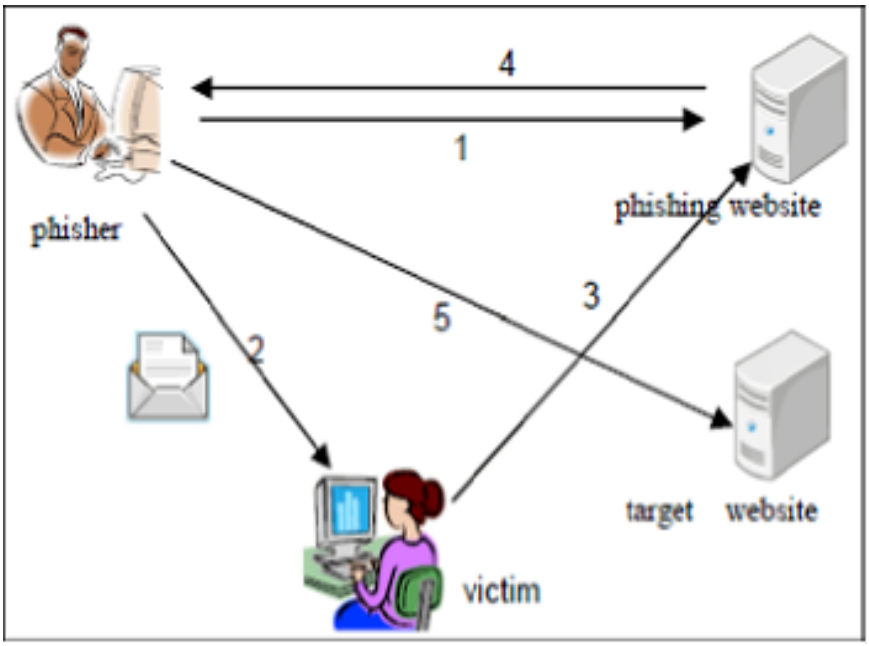
- **Machine Learning Model**: A supervised machine learning model, trained on a diverse and up-to-date dataset, classifies web pages as either phishing or legitimate based on the extracted features. The model is continuously updated to adapt to evolving phishing techniques.

Literature Survey

Phishing attacks are insidious cyber threats that rely on deceit and trickery to compromise personal and sensitive information. These malicious campaigns typically involve fraudulent emails, websites, or messages that impersonate trustworthy sources, enticing recipients to divulge sensitive data like passwords, credit card details, or personal information. Phishing attacks exploit human psychology and often lead to dire consequences, from financial loss to identity theft. As a prevalent and constantly evolving cybersecurity menace, understanding the basics of phishing is essential for individuals and organizations alike to bolster their defenses against this ever-present online danger.

Phishing Mechanism

The mechanism with which phishing works fairly simple, manipulate the victim into providing confidential information about him/her.[1]
To perform such an attack, the attacker or phisher mimics a legitimate website. The phishing website would gather all the information on the target and provide it to the attacker.



Types of Phishing Attacks

The various types of phishing attacks are [2]:

- Deceptive Phishing
  Deceptive Phishing is the most prevalent type of phishing attack. It involves imitating a legitimate website and sending an email to the target appearing it to be genuine.

- Spear Phishing
  This variety of phishing is nearly identical to deceptive phishing.
  The only difference is the target. Unlike deceptive phishing, spear phishing targets one individual only.

- Whaling
  Whaling attacks occur when the phisher targets an individual at an executive position like CEO. The attacker would be profiling the victim for a considerable period before performing the attack.

- Pharming
  Pharming is another variation of phishing. Unlike, the other techniques, it is not necessary to target individuals. The attack can victimize a large number of people without having to be targeted individually.

Anti-Phishing Techniques

In [3], the authors proposed a solution to defend phishing attacks using a combination of visual similarity based techniques and white list. The Computer Vision (CV) tool called Speed up Robust Features (SURF) detector. This detector uses square shaped filters for extracting discriminative key point features.

In [4], a different solution was proposed by the use of Support Vector Machines (SVM) to detect if the mail is malicious or not. The SVM extracted common characteristics of the mail such as language used, layout of the mail, structure of the mail, etc. It then compares the extracted details with the details present in the system to check the similarity accuracy.

A more advanced technique of filtering and classification was used in [5]. In this paper, the authors tested the URLs and verified whether it was malicious or not. They used an automated approach for detecting phishing. It had two phases- Pre-filtering and Classification phase. In the pre-filtering phase, the URL was compared against a black list using the domain part of the URL. If the URL was present in that list then it was classified as malicious and would not be proceeding to the Classification Phase.

The following methods can be combined to come up with a method to detect phishing:

- **Domain Reputation Services:**
  Utilize domain reputation services like Google Safe Browsing, and IBM X-Force to check if a domain is known for hosting phishing content.

- **WHOIS Data Analysis:**
  Analyze WHOIS registration data for domains and look for suspicious information

- **DNS Analysis:**
  Analyze DNS records for domains and look for anomalies like unusual IP addresses, multiple domains resolving to the same IP.

- **Keyword Analysis:**
  Use keyword analysis to detect suspicious words or phrases in domain names. Phishing domains often contain misspelled versions of legitimate domain names or keywords related to popular brands.

- **Certificate Transparency Logs:**
  Monitor Certificate Transparency logs to detect newly registered SSL certificates for domains. Phishers often use SSL certificates to make their websites appear legitimate.

Objectives

- **Chrome Extension Development**: Create a Chrome extension designed to enhance the security of online/mobile wallets and net banking.

- **Phishing Attack Prevention**: Develop features within the extension aimed at preventing phishing attacks, including real-time analysis of web content and URLs for potential threats.

- **User Awareness and Education**: Incorporate educational components into the extension, such as pop-up notifications, tooltips, and guides, to increase user awareness regarding phishing threats and safe online banking practices.

- **Reporting and Response Features**: Integrate a user-friendly reporting mechanism within the extension, allowing users to report potential phishing incidents directly. Implement a quick response protocol to assist affected users.

- **Phishing Incident Analytics**: Collect and analyze data related to phishing incidents through the extension, including attack vectors and trends, to continually improve anti-phishing strategies.

- **Continuous User Training**: Offer in-extension training materials, tutorials, and webinars to keep users informed about evolving phishing tactics and security best practices.

Existing Methods-Drawbacks

Email Filtering:
Strengths: Automated systems can analyze incoming emails and filter out potential phishing messages based on known patterns.
Drawbacks: False positives are common, and legitimate emails may be flagged. Sophisticated phishing emails that mimic legitimate communication may still slip through.

URL Filtering:
Strengths: Blocking or warning users about suspicious links in emails or websites.
Drawbacks: Advanced phishing attacks may use URLs that haven't been blacklisted yet. Legitimate websites with compromised content might also be flagged.

Multi-Factor Authentication (MFA):
Strengths: Adds an extra layer of security by requiring multiple forms of identification.
Drawbacks: Not foolproof; some phishing attacks can still trick users into providing all the necessary authentication factors. Additionally, implementing MFA can be inconvenient for users.

User Education and Awareness:
Strengths: Training users to recognize phishing attempts can be effective in preventing attacks.
Drawbacks: Human error is still a significant factor. Users may still fall for well-crafted and convincing phishing schemes, especially as tactics evolve.

Web Browsing Protection:
Strengths: Browser tools and extensions that warn users about potentially malicious websites.
Drawbacks: Limited to online activities and may not protect against phishing attempts outside of web browsing. Users might also ignore or override warnings.

Behavioral Analysis:
Strengths: Analyzing user behavior to detect anomalies that could indicate a phishing attack.
Drawbacks: False positives are possible, and the system may not detect new, sophisticated attacks. It might also be challenging to distinguish between normal and malicious behavior accurately.

Proposed Method

The following methods can be combined to come up with a method to detect phishing:

Domain Reputation Services:
Utilize domain reputation services like Google Safe Browsing, and IBM X-Force to check if a domain is known for hosting phishing content.

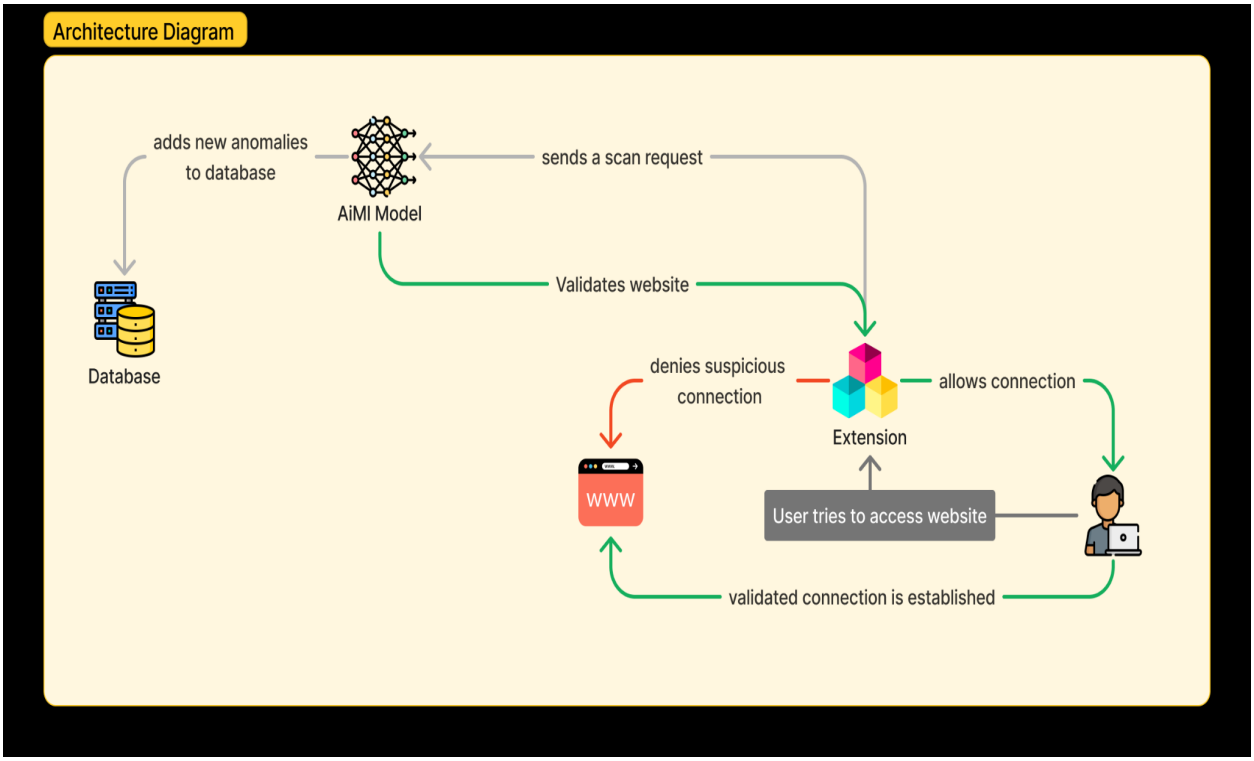
WHOIS Data Analysis:
Analyze WHOIS registration data for domains and look for suspicious information

DNS Analysis:
Analyze DNS records for domains and look for anomalies like unusual IP addresses, multiple domains resolving to the same IP.

Keyword Analysis:
Use keyword analysis to detect suspicious words or phrases in domain names. Phishing domains often contain misspelled versions of legitimate domain names or keywords related to popular brands.

Certificate Transparency Logs:
Monitor Certificate Transparency logs to detect newly registered SSL certificates for domains. Phishers often use SSL certificates to make their websites appear legitimate.

Architecture Diagram



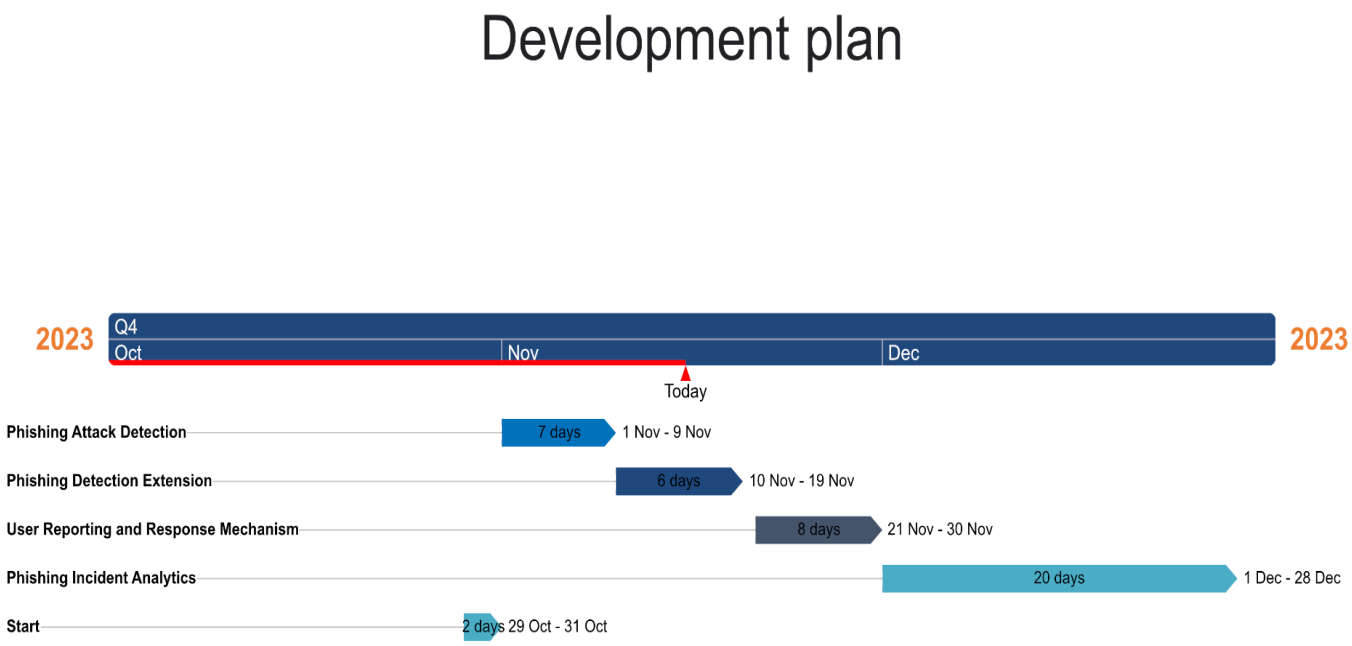
Modules

- Phishing Attack Detection System
- Phishing Detection Extension
- User Reporting and Response Mechanisms
- Phishing Incident Analytics

Hardware and Software Details

- Python
- Javascript
- REST api

Timeline by Gantt Chart



References

1. Akarshita Shankar, Ramesh Shetty and Badari Nath - A Review on Phishing Attacks, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 9 (2019) pp. 2171-2175.

2. Gaurav, Madhuresh Mishra, Anurag Jain - Anti- Phishing Techniques: A Review, International Journal of Engineering Research and Applications (IJERA), vol. 2, pp. 350-355, April – 2012.

3. Routhu Srinivasa Rao and Syed Taqi Ali - A Computer Vision Technique to Detect Phishing Attacks, 5th International Conference on Communication Systems and Network Technologies, IEEE, October 2015.

4. Madhusudhanan Chandrasekaran, Krishnan Narayanan and Shambhu Upadhyaya - Phishing E- mail Detection based on Structural Properties, IEEE, November 2015.

5. Yi-Shin Chen, Huei-Sin Liu, Yi-Hsuan Yu and PangChieh Wang, Detect Phishing by Checking Content Consistency, IEEE, 2017.

Guide Signature **Mr. Praveen Pawaskar**	**Reviewer Signature** **Dr. Mohana S D**