Malicious URL Detection (Extension)

Mr. Pratik Suryabhan Patole[1], Mr. Tushar Ashok Khatale[2], Mr. Shubham Bapu Panchave[3], Miss. Sakshi Rajendra Mahatme[4], Miss. Suvarna Sujit Wakchaure[5]

Sir Visvesvaraya Institute Of Technology, Nashik, A/p. Chincholi, Tal. Sinnar, Dist. Nashik - 422102 (MS), India

Abstract

As cyber security continues to evolve, traditional methods for identifying harmful URLs—such as relying on blacklists and pattern recognition—are becoming less effective against new and sophisticated threats. Cybercriminals are constantly developing new tactics to evade detection, making it crucial to explore innovative solutions. This project aims to create a comprehensive system that detects malicious URLs by analyzing their fundamental features, rather than depending solely on historical data.

The proposed system will leverage blockchain technology to securely record and share information about URLs. By creating a decentralized and immutable ledger, the system ensures that once a URL is classified as malicious or benign, this information cannot be altered or tampered with. This secure approach not only strengthens the trustworthiness of the data but also facilitates collaboration among organizations, allowing them to share insights about potential threats more effectively.

Additionally, the system will employ encoding techniques to obscure URLs during the analysis process. This added layer of security helps protect sensitive information, making it more challenging for attackers to gain insights into the URLs being evaluated. By disguising URLs while still allowing for effective analysis, the system can better identify malicious threats without exposing users to unnecessary risks.

By focusing on the essential characteristics of URLs and integrating these advanced technologies, this project aims to significantly enhance organizations' ability to defend against cyber threats. Ultimately, the goal is to create a safer online experience for everyone, ensuring that users can navigate the Digital landscape with greater confidence and security.

Keywords: Cyber Security, Malicious Urls, URL Detection, Blockchain Technology, Decentralized Ledger, Threat Intelligence, Encoding Techniques, Data Integrity, Pattern Recognition, Blacklists, Static Feature Analysis, Cyber Threats, Online Safety

INTRODUCTION

As the digital landscape grows increasingly complex, the need for robust cybersecurity measures has never been more critical. Traditional methods for detecting malicious URLs, such as blacklists and pattern recognition, are becoming inadequate in the face of evolving cyber threats. Cybercriminals are constantly developing sophisticated tactics to bypass conventional defenses, making it essential to explore new and innovative approaches to URL detection.

This project proposes a cutting-edge system that focuses on analyzing the fundamental features of URLs to identify potential threats more effectively. By leveraging blockchain technology, the system ensures secure and immutable recording of URL classifications, enhancing trust and facilitating collaborative threat intelligence among organizations. Additionally, encoding techniques will be employed to obscure URLs during analysis, providing an extra layer of protection against potential exploits.Ultimately, this initiative aims to empower organizations to better defend against cyber threats, creating a safer online environment for all users. By prioritizing the inherent characteristics of URLs and integrating advanced technologies, we can enhance cyber security measures and foster greater confidence in the digital realm.

LITERATURE SURVEY

"Understanding the Threat Landscape of Malicious URLs: Impacts and Mitigation Strategies" This paper discusses the general risks of malicious URLs used for malware and phishing.:This paper talks about how malicious URLs are a big problem on the internet. These URLs are used by attackers to spread viruses (malware) or trick people into giving up personal information (phishing). The authors highlight how dangerous these links are for both regular users and organizations. [1]

"The Rising Threat of Cyberattacks in an Increasingly Connected World" Focuses on how the internet has become essential, and cyberattacks are growing due to smarter attackers.The authors of this paper discuss how the internet has become essential to our lives, but at the same time, cyberattacks are increasing because hackers are becoming smarter. This paper stresses the growing risk of cyber threats and the importance of better protection strategies. [2]

A Survey on Malicious URL Detection: Challenges and Future Directions" Covers the dangers of malicious URLs and the need for advanced detection methods.This paper also focuses on the threat of malicious URLs. It explains how these dangerous links are responsible for serious online crimes like phishing and malware spreading. The authors emphasize the need for strong detection systems to protect users from such threats. [3]

"Phishing URL Detection Using String-Based Features: A Traditional Approach Revisited" Highlights how phishing URLs can be detected through URL string characteristics.In recent years, phishing events occur frequently, and the detection of phishing URL has become a common concern in the field of network security. In previous studies, researchers distinguish phishing URLs from normal URLs by the string characteristics. This study looks specifically at phishing URLs. It points out that phishing is becoming more common, and one way to detect phishing is by analyzing the structure and patterns of the URLs. Earlier methods looked at how the URL "looks" (its characters and format) to tell if it's suspicious.[4]

Deep Learning for Phishing URL Detection: Transformer Models and Contextual Embeddings" Covers the use of transformers for improved phishing URL detection.In 2024, researchers have focused on enhancing phishing URL detection using deep learning techniques. Recent studies have explored transformer-based models to analyze sequential patterns in URLs, significantly improving classification accuracy compared to traditional machine learning approaches. These methods leverage contextual embeddings to identify subtle phishing characteristics more effectively.This recent paper (2024) talks about using deep learning, especially transformers (a type of AI model), to detect phishing URLs. These models can understand patterns in how URLs are written, helping identify tricky or hidden signs of phishing better than older methods.[5]

METHODOLOGY

1. Gather a diverse dataset of URLs, including both benign and malicious samples from various sources to ensure a wide representation of URL types.

2. Analyze the collected URLs to identify and extract important static features such as domain names, path structures, query parameters, and lengths.

3. Clean the dataset by removing duplicates and irrelevant entries.Normalize the extracted features to ensure uniformity and handle any missing values appropriately.

4. Utilize the preprocessed dataset to develop machine learning models for URL classification based on the extracted features. Experiment with various algorithms to identify the most effective approach for detecting malicious URLs.

5. Assess the performance of the trained models using key metrics such as accuracy, precision, recall, and F1-score to ensure reliability in threat detection.

6. Implement the developed system as a user-friendly interface or API, allowing for easy integration into existing cybersecurity frameworks and tools.

7. Establish a mechanism for continuous improvement by incorporating user feedback and new data into the model, ensuring adaptability to evolving threats.

OBJECTIVE

1. Create a model that enhances the accuracy of malicious URL detection by utilizing static feature classification techniques.

2. Focus on recognizing emerging and obfuscated threats that traditional detection methods may overlook.

3. Aim to reduce false positives, ensuring that legitimate URLs are not incorrectly flagged as malicious.

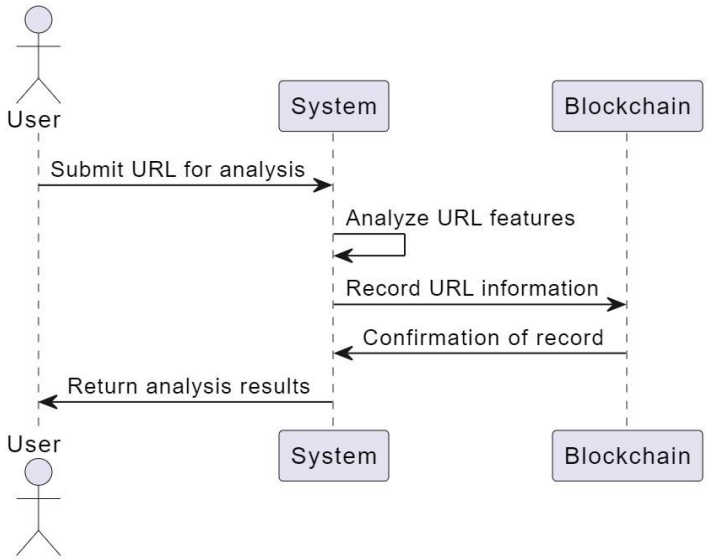
4. Deliver a scalable tool that organizations can implement to strengthen their cybersecurity measures against evolving threats.

PROBLEM DEFINATIONS

As cyber threats continue to evolve, traditional methods for detecting malicious URLs—such as blacklists and heuristic analysis—are becoming increasingly ineffective. These conventional approaches often fail to identify emerging and obfuscated threats that do not match established patterns, leaving organizations vulnerable to cyber-attacks. Furthermore, the reliance on historical data can lead to high false positive rates, causing legitimate URLs to be flagged incorrectly and disrupting business operations.

This project seeks to address these challenges by developing a more effective detection system that focuses on the static features of URLs. By enhancing the accuracy of malicious URL detection and minimizing false positives, the system aims to provide a robust solution for organizations facing an ever-changing landscape of cyber threats. The integration of blockchain technology and encoding mechanisms will further enhance the security and reliability of the detection process, ensuring that organizations can safeguard their digital environments effectively.

SEQUENCE DIAGRAM



NON FUCTIONAL REQUIREMENTS

1. The system must process and classify URLs within a specified time frame (e.g., under 2 seconds per URL).

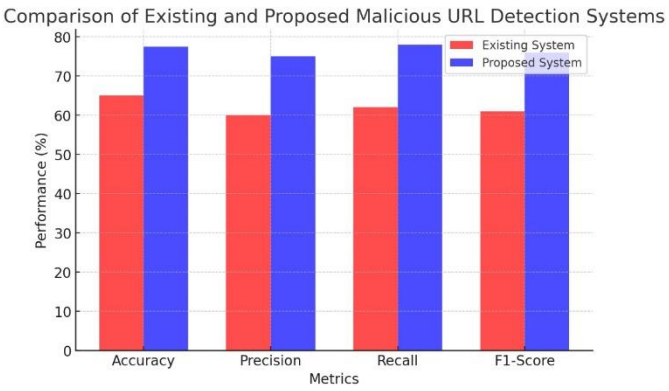
2. The system must handle an increasing number of URL classifications without a decline in performance, supporting growth in data volume.

3. The system must ensure data protection through encryption and secure access controls, safeguarding both user data and classification results.

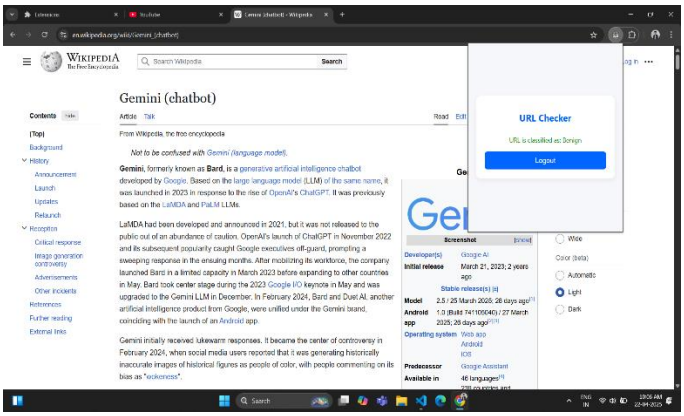
4. The system must demonstrate high availability, with minimal downtime and the ability to recover quickly from failures.

5. The user interface must be intuitive and easy to navigate, allowing users to quickly understand and use the system without extensive training.

RESULT



SNAPSHOTS



CONCLUSION

In an era where cyber threats are constantly evolving, the need for advanced methods to detect malicious URLs has become paramount. This project addresses the limitations of traditional detection techniques by proposing a novel framework that emphasizes static feature classification, blockchain integration, and encoding mechanisms. By focusing on the inherent characteristics of URLs, the system aims to accurately identify emerging and obfuscated threats that may evade conventional defenses.

The integration of blockchain technology ensures the security and integrity of URL classifications, while encoding techniques provide an additional layer of protection. Furthermore, the development of a user-friendly interface or API allows organizations to easily adopt this solution, enhancing their cyber security measures against increasingly sophisticated attacks.

Ultimately, this project aims not only to improve the accuracy and reliability of malicious URL detection but also to foster a safer online environment for all users. By equipping organizations with a scalable and robust tool, we contribute to the ongoing efforts to combat cybercrime and protect digital assets in an ever-changing landscape.

REFERENCES

1. A CNN-Based Model for Detecting Malicious URLs, 2023 RIVF International Conference on Computing and Communication Technologies (RIVF)
2. Machine Learning Supported Malicious URL Detection, 2023 4th IEEE Global Conference for Advancement in Technology (GCAT)
3. New Heuristics Method for Malicious URLs Detection Using Machine Learning, 2023 International Symposium on Networks, Computers and Communications (ISNCC)
4. A Malicious URL Detection Method Based on CNN, 2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS)