An Intelligent System For Detecting Phishing Website Using Network Techniques

S.Sri Vidhya , Dr.M.Usha Devi

Department of Computer Science, Rathinam College of Arts and Science (Autonomous), Coimbatore, Tamilnadu, India.

Abstract *- Phishing attacks have become a serious cybersecurity threat, using deceptive websites and malicious URLs to steal sensitive information from individuals and organizations. Early detection is essential to protect user data and maintain secure network communication, but traditional rule-based methods often fail to identify new and evolving phishing techniques. This project proposes an intelligent system for detecting phishing websites using network-based techniques combined with machine learning. The system analyzes web URLs by extracting lexical, structural, statistical, and network-related features, and then classifies them as phishing or legitimate using trained models. Compared to traditional approaches, this method provides higher accuracy, better adaptability, and more efficient detection of modern phishing strategies. Therefore, it serves as a reliable and scalable solution for real-time cybersecurity applications.*

Keywords *– Phishing Detection, Intelligent System, Network Techniques, URL Analysis, Cybersecurity, Feature Extraction, Website Security, Malicious URLs, Real-Time Detection, Network Traffic Analysis*.

1. Introduction

Phishing websites have become a major concern in modern cybersecurity, as they attempt to steal sensitive information such as login credentials, banking details, and personal data. These malicious websites often imitate legitimate platforms, making it difficult for users to identify them. With the increasing use of online services, phishing attacks have grown more sophisticated and pose a serious threat to individuals and organizations. Detecting such websites at an early stage is essential to prevent financial loss and protect user privacy. However, traditional detection methods rely heavily on manual inspection and rule-based techniques, which are time-consuming and often ineffective against evolving threats.

Phishing detection primarily involves analyzing website URLs and network-related information to identify suspicious patterns. Characteristics such as URL structure, domain name variations, use of special characters, and redirection behavior are commonly used indicators of phishing websites. In addition, network-based attributes such as DNS records, IP address behavior, and server response patterns provide valuable insights into the authenticity of a website. By combining these features, intelligent systems can effectively distinguish between legitimate and malicious websites.

Several researchers have proposed different techniques for phishing detection using URL analysis and network-based approaches. Some methods focus on preprocessing URL data to remove noise and extract meaningful patterns, followed by analyzing domain characteristics and redirection chains. Others utilize techniques such as DNS analysis and IP verification to identify abnormal network behavior associated with phishing websites.

Despite these advancements, existing methods still face challenges such as detecting zero-day phishing attacks, handling large-scale data, and maintaining real-time performance. Some approaches may produce false positives or require significant computational resources. Therefore, there is a need for efficient and scalable systems that can quickly adapt to new phishing techniques. By integrating URL inspection with network analysis, the proposed system aims to overcome these limitations and provide accurate and reliable phishing detection.

Phishing websites have become a significant cybersecurity threat in the modern digital environment, targeting users to steal sensitive information such as login credentials, banking details, and personal data. These attacks often rely on deceptive URLs and fake websites that closely resemble legitimate platforms, making them difficult to detect. With the increasing use of online services, phishing techniques continue to evolve, reducing the effectiveness of traditional rule-based detection methods. To address this challenge, an intelligent system for detecting phishing websites using network techniques is proposed. The system analyzes URL features along with network-level information such as DNS records, IP address behavior, and traffic patterns to identify suspicious activities. By combining these approaches, the system can accurately distinguish between legitimate and phishing websites, providing a reliable and scalable solution for real-time cybersecurity protection , enhances early detection capabilities,This system enhances early detection capabilities, reducing the risk of data breaches and ensuring safer online interactions for users, monitoring of network behavior to effectively identify and prevent emerging phishing threats.

2.Related Works

Several approaches have been proposed for detecting phishing websites using different cybersecurity techniques. Initially, blacklist-based methods were used to compare user-requested URLs with known phishing sites, but these methods fail to detect new attacks. To address this, heuristic-based techniques analyze URL features such as length, special characters, and redirection behavior, though they are not fully effective against advanced phishing strategies. In recent years, machine learning algorithms such as Decision Tree, Naïve Bayes, Support Vector Machine (SVM), and Random Forest have been widely used to classify websites based on features extracted from URLs, webpage content, and domain information. These methods provide better accuracy and adaptability. Further improvements have been achieved using deep learning techniques like Convolutional Neural Networks (CNN), which automatically learn complex patterns from data but require higher computational resources. Additionally, network-based techniques analyze DNS records, IP behavior, and traffic patterns to detect suspicious activities, especially in zero-day attacks. Hybrid approaches combining multiple techniques have shown improved performance; however, challenges such as false positives and real-time detection still exist, necessitating more efficient intelligent systems.

Several approaches have been developed to detect phishing websites in cybersecurity. Initially, blacklist-based methods were used to identify known phishing URLs, but they fail to detect new attacks. Heuristic-based techniques were later introduced to analyze URL features such as length, special characters, and redirection patterns. However, these methods are not effective against sophisticated phishing strategies. Machine learning algorithms like Decision Tree, Naïve Bayes, Support Vector Machine (SVM), and Random Forest have improved detection accuracy by analyzing multiple features. These models classify websites based on URL, domain, and content characteristics. Deep learning techniques such as Convolutional Neural Networks (CNN) further enhance performance by automatically learning patterns from data. Network-based methods analyze DNS records, IP behavior, and traffic patterns to detect suspicious activities. Hybrid approaches combining multiple techniques provide better results. Despite these advancements, challenges like zero-day attacks and false positives still exist.

Several approaches have been proposed for detecting phishing websites using network-based techniques. Early methods relied on blacklist systems, which compare URLs with known phishing databases but fail to detect new threats. Heuristic techniques analyze URL structures and redirection behavior to identify suspicious patterns. Network-based approaches focus on DNS records, IP address tracking, and traffic flow analysis to detect malicious activities. Some systems also examine server behavior and domain registration details for better identification.

3.System Design

The background research indicates a growing need for intelligent systems in detecting phishing websites using network-based techniques. This study introduces an optimized approach for identifying phishing websites by analyzing URL characteristics and network-level information to achieve higher detection accuracy. The schematic representation of the proposed system is illustrated.The primary contributions of this study are outlined as follows:

(i)Implementing effective URL preprocessing techniques to ensure accurate and clean input data for analysis.
(ii)Utilizing feature extraction methods to identify significant URL and network features that indicate phishing behavior.
(iii)Rule-based detection techniques is to classify websites as legitimate or phishing based on analyzed data.
(iv) Analyzing network-related attributes such as DNS records, IP address behavior, and traffic patterns to detect suspicious activities.

3.1 URL Input & Preprocessing

This process begins with the user entering a website URL into the system. The system validates the input to ensure it is correctly formatted and usable. Any unnecessary characters or errors are removed during preprocessing. The cleaned and standardized URL is then passed to the next stage for further analysis.

3.2. Feature Extraction

In this process, important features are extracted from the given URL. These include lexical features such as URL length and special characters, as well as structural features like domain information. The extraction helps in identifying patterns that may indicate phishing. The processed feature set is then forwarded for deeper analysis.

3.3. Network Analysis

This process focuses on analyzing network-related information of the website. It examines DNS records, IP address details, and server behavior to detect suspicious activity. Traffic patterns and response characteristics are also considered. The analyzed network data plays a key role in identifying potential phishing threats.

3.4. Technique Selection

The research focuses on analyzing URL characteristics and network-related data for phishing website detection. To identify the most suitable approach, various network-based techniques were studied and selected based on their effectiveness in detecting suspicious activities. These techniques are widely used in cybersecurity for identifying

malicious websites through communication and domain analysis. The selected techniques include:

- DNS Analysis
- IP Address Verification
- URL Inspection

Considering the nature of the dataset and the objective of detecting phishing websites, these techniques were chosen for their ability to identify abnormal patterns and improve detection accuracy.

3.4.1. DNS Analysis

DNS (Domain Name System) Analysis is an important technique used to examine domain-related information of a website. It helps in identifying suspicious domains by analyzing factors such as domain age, registration details, and DNS record consistency. Phishing websites often use newly registered domains or frequently change their DNS records to avoid detection. This technique also checks mismatches between domain names and their associated records. By monitoring DNS behavior, abnormal patterns can be detected effectively. DNS analysis plays a key role in identifying hidden malicious activities. It improves the reliability of phishing detection systems.

3.4.2. IP Address Verification

IP Address Verification focuses on analyzing the IP address associated with a website. It helps in determining whether the IP address is trustworthy or linked to malicious activities. Phishing websites often use suspicious, shared, or dynamically changing IP addresses. This technique checks if the IP address is blacklisted or associated with multiple domains. It also verifies consistency between the domain name and its IP location. Any mismatch or unusual pattern may indicate a phishing attempt. This method enhances detection accuracy by validating network-level information.

3.4.3. URL Inspection

URL Inspection is used to analyze the structure and components of a website URL. It examines features such as URL length, use of special characters, presence of misleading words, and abnormal patterns. Phishing URLs often mimic legitimate websites but include slight variations that are difficult to notice. This technique also checks for multiple redirections and suspicious domain names. By identifying irregular patterns, the system can detect malicious intent. URL inspection is simple yet highly effective in phishing detection. It acts as a primary step in identifying suspicious websites.

3.4.4. System Evaluation

System evaluation is carried out to measure the performance of the phishing detection system. Various metrics such as accuracy, precision, recall, and F1 score are used for evaluation. These metrics help in understanding how well the system identifies phishing websites and avoids false detections. A good system should have high accuracy and low false positive rates. Evaluation is performed using both known and unseen data to ensure reliability. This process helps in improving the system's effectiveness. It ensures that the system performs efficiently in real-time scenarios.

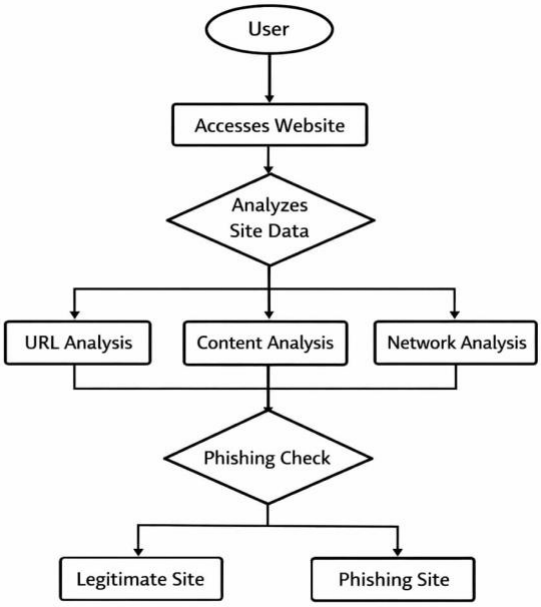

Fig 1. Dataflow Diagram

4. Object and Scope

The objective of this research is to develop an intelligent system for detecting phishing websites using network-based techniques. The primary goal is to accurately identify malicious URLs by analyzing their features and network-related patterns, thereby protecting users from cyber threats and data breaches. This study aims to improve detection accuracy, reduce false positives, and enhance the system's ability to identify newly emerging phishing attacks. By utilizing advanced analysis of network behavior and URL characteristics, the research seeks to provide a reliable and efficient solution for real-time phishing detection. The scope of this research includes a detailed analysis of various characteristics of web URLs, such as lexical, structural, and statistical features, along with network-level attributes to distinguish between legitimate and phishing websites. The study focuses on designing and evaluating an intelligent detection system using collected datasets and comparing its performance in terms of accuracy and efficiency.

5.Literature Review

Phishing website detection has gained significant attention in recent years due to the rapid increase in cyber threats and the growing dependence on online platforms. The rise of sophisticated phishing attacks, which exploit deceptive URLs and closely mimic legitimate websites, has created a strong need for more advanced and intelligent detection systems. Traditional security methods based on fixed rules and blacklists are no longer sufficient to handle evolving attack strategies, especially with the emergence of zero-day phishing attacks. As a result, recent research has increasingly focused on network-based techniques and intelligent analysis methods to improve the accuracy and efficiency of phishing detection systems.

A review of recent studies highlights the effectiveness of network-based approaches in detecting phishing websites by analyzing communication patterns and domain-related information. Techniques such as DNS analysis, IP address verification, traffic monitoring, and server response evaluation are widely used to identify suspicious behavior associated with phishing activities. These approaches analyze features such as URL structure, domain characteristics, packet flow, and redirection patterns to distinguish between legitimate and malicious websites. Compared to traditional methods, network-based detection systems provide better adaptability and real-time monitoring capabilities. Overall, these techniques enhance detection performance, improve reliability, and play a crucial role in strengthening cybersecurity against modern phishing attacks.

6. Output

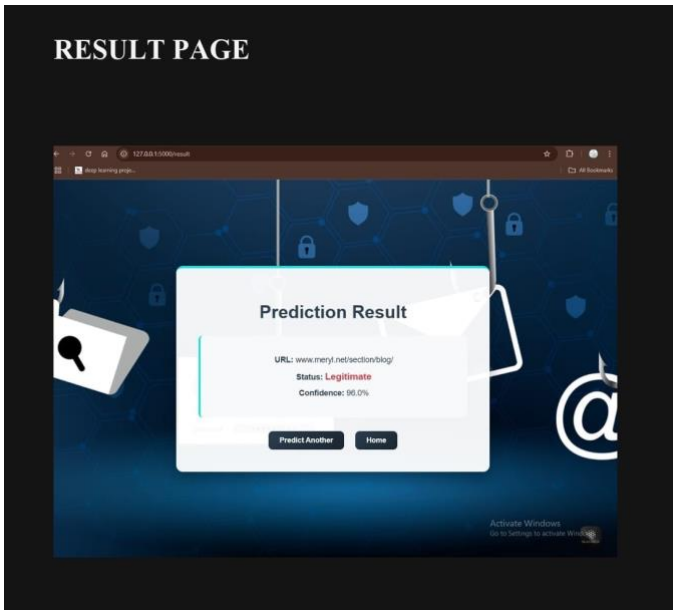

Fig 2.Result Page

7. Results

The results obtained from the proposed system are presented in a structured format to evaluate its effectiveness in detecting phishing websites. The performance is analyzed based on key metrics such as accuracy, precision, recall, and F1 score. The results are organized to highlight the efficiency of different network-based techniques used in the system. This arrangement enables easy comparison and helps in identifying the most effective techniques for phishing detection. It also provides a clear understanding of the system's strengths and its ability to handle real-time cyber threats.

Recent studies have emphasized the importance of network-based approaches in improving phishing detection systems. Researchers have proposed various techniques that analyze domain information, IP address behavior, and traffic patterns to identify malicious websites. These approaches focus on detecting abnormal network activities and inconsistencies in domain records. By integrating multiple network-level features, the detection system becomes more reliable and efficient.

8.Conclusion

The proposed system for phishing website detection using network-based techniques provides effective and reliable results in identifying malicious websites. By analyzing URL characteristics along with network-related features such as DNS records, IP address behavior, and traffic patterns, the system is able to accurately distinguish between legitimate and phishing websites. Among the techniques used, network analysis proved to be highly effective in detecting suspicious activities and improving overall detection performance. The system demonstrated strong accuracy and consistency when evaluated using different datasets.

By incorporating factors such as domain information, URL structure, and real-time network behavior, the system can identify complex phishing patterns and prevent potential cyber threats. This enables better decision-making in securing online platforms and protecting user data. Although challenges such as dynamic phishing techniques and evolving attack patterns exist, continuous improvements in network monitoring and analysis methods can further enhance system performance. Overall, the proposed approach provides a scalable and efficient solution for real-time phishing detection and contributes to strengthening modern cybersecurity systems.

9. References

[1] Abu-Nimeh, S., Nappa, D., Wang, X., and Nair, S., "A comparison of phishing detection techniques," *Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit*, 2007, pp. 6–69.

[2] Zhang, Y., Hong, J.I., and Cranor, L.F., "CANTINA: A content-based approach to detecting phishing websites," *Proceedings of the 16th International World Wide Web Conference*, 2007, pp. 639–648.

[3] Garera, S., Provos, N., Chew, M., and Rubin, A.D., "A framework for detection and measurement of phishing attacks," *Proceedings of the 2007 ACM Workshop on Recurring Malcode*, pp. 1–8.

[4] Ma, J., Saul, L.K., Savage, S., and Voelker, G.M., "Beyond blacklists: Learning to detect malicious web sites from suspicious URLs," *Proceedings of the 15th ACM SIGKDD International Conference*, 2009, pp. 1245–1254.

[5] Xiang, G., Hong, J., Rose, C.P., and Cranor, L., "CANTINA+: A feature-rich machine learning framework for detecting phishing web sites," *ACM Transactions on Information and System Security*, 2011.

[6] Le, A., Markopoulou, A., and Faloutsos, M., "PhishDef: URL names say it all," *IEEE INFOCOM*, 2011, pp. 191–195.

[7] Verma, R. and Das, A., "What's in a URL: Fast feature extraction and malicious URL detection," *Proceedings of the 3rd ACM Workshop on Security and Artificial Intelligence*, 2010.

[8] Thomas, K., Grier, C., Ma, J., Paxson, V., and Song, D., "Design and evaluation of a real-time URL spam filtering service," *IEEE Symposium on Security and Privacy*, 2011.

[9] Sahoo, D., Liu, C., and Hoi, S.C.H., "Malicious URL detection using machine learning: A survey," *ACM Computing Surveys*, 2017.

[10] Almomani, A., Gupta, B.B., Atawneh, S., Meulenberg, A., and Almomani, E., "A survey of phishing email filtering techniques," *IEEE Communications Surveys & Tutorials*, 2013.

[11] Mohammad, R.M., Thabtah, F., and McCluskey, L., "Phishing detection: A recent intelligent machine learning comparison based on models content and features," 2014 IEEE International Conference on Intelligence and Security Informatics, pp. 72–77.

[12] Basnet, R.B., Mukkamala, S., and Sung, A.H., "Detection of phishing attacks: A machine learning approach," Studies in Fuzziness and Soft Computing, Springer, 2008, pp. 373–383.

[13] Bergholz, A., De Beer, J., Glahn, S., Moens, M.F., Paaß, G., and Strobel, S., "New filtering approaches for phishing email," Journal of Computer Security, vol. 18, no. 1, 2010, pp. 7–35.

[14] Khonji, M., Iraqi, Y., and Jones, A., "Phishing detection: A literature survey," IEEE Communications Surveys & Tutorials, vol. 15, no. 4, 2013, pp. 2091–2121.

[15] Jain, A.K. and Gupta, B.B., "Phishing detection: Analysis of visual similarity-based approaches," Security and Communication Networks, vol. 7, no. 5, 2014, pp. 863–875.

10. Acknowledgment

This article is the outcome of the research work carried out in the **Department of Computer Science.** The authors would like to express their sincere gratitude to the Department for providing the necessary support and resources to successfully complete this work. We also extend our thanks to the faculty members and mentors for their valuable guidance and encouragement throughout the research. Their continuous support has greatly contributed to the successful development of this project.