WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning Based Malicious URL Detection

Hosuru Prashanthi[1], Ms .Mallarapu Poojitha, MCA[5]

[1] Dept. of Department MCA, Annamacharya Institute of Technology and Sciences (AITS), Karakambadi, Tirupati, Andhra Pradesh, India
[2]Assistant Professor, Dept. of Department MCA, Annamacharya Institute of Technology and Sciences (AITS), Karakambadi, Tirupati, Andhra Pradesh, India

ABSTRACT

The exponential growth of the internet has led to an increase in cyber threats, particularly through the spread of malicious URLs. These URLs, embedded in emails, websites, or social media, are designed to deceive users and lead them to malicious websites that steal data, install malware, or perform phishing attacks. Traditional security mechanisms, such as blacklists and rule-based filters, are often inadequate due to their inability to keep up with the evolving tactics of cybercriminals. In this context, machine learning (ML) emerges as a powerful and adaptive solution for detecting malicious URLs in real time.This study presents a machine learning-based approach for detecting malicious URLs using a supervised classification model trained on a dataset comprising lexical, host-based, and content-based features of URLs. Various algorithms, including Logistic Regression, Decision Trees, Random Forests, and Support Vector Machines, were explored to determine the most effective model. The preprocessing phase involved feature extraction from URLs, including statistical attributes, domain characteristics, and token patterns, followed by normalization and encoding. The dataset was split into training and testing sets to evaluate the performance of different models.Results indicate that tree-based ensemble methods such as Random Forest and Gradient Boosting provided the highest detection accuracy with minimal false positives. The system not only outperforms traditional methods but also adapts to previously unseen patterns of malicious behavior. The proposed model demonstrates strong generalization capabilities and robustness against common evasion techniques.This research emphasizes the significance of integrating intelligent learning-based models into cybersecurity systems for proactive threat detection.

Keywords : Malicious, Behavior, URL, Cyber security

I. INTRODUCTION

With the rapid expansion of internet usage, malicious attacks have become more sophisticated and prevalent. One of the most common vectors for cyberattacks is the use of malicious URLs, which can lead unsuspecting users to phishing sites, malware downloads, and fraudulent domains. These URLs are often disguised to appear legitimate, making it increasingly difficult for users and traditional detection mechanisms to identify them accurately. As a result, organizations and individuals face growing challenges in safeguarding their systems and data from these threats.

Conventional detection methods, such as URL blacklists and heuristic-based techniques, suffer from several limitations. Blacklists must be continuously updated and often fail to detect newly generated malicious URLs. Heuristic systems may struggle with obfuscated or polymorphic URLs that deviate from known patterns. To address these shortcomings, machine learning has emerged as a promising alternative that leverages data-driven models to detect malicious URLs based on a wide range of features extracted from the URL itself and related metadata.

Machine learning models can learn complex patterns and generalize from labeled datasets, enabling them to predict the maliciousness of previously unseen URLs. These models analyze various features, including lexical (e.g., length, character frequency), host-based (e.g., domain age, IP address), and contextual features (e.g., redirection behavior). Once trained, these models can process large volumes of URLs in real time, offering improved scalability and accuracy over static detection systems.

This project investigates the application of machine learning techniques for malicious URL detection. It aims to develop a classification system that can distinguish between legitimate and malicious URLs using a range of extracted features. By comparing the performance of various machine learning algorithms, the study seeks to identify the most effective model for deployment in real-world cybersecurity environments. The research further explores challenges in feature selection, model evaluation, and false-positive reduction, ultimately contributing to a more robust and intelligent URL filtering system.

II. RELATED WORK

In [1], This foundational study demonstrates the effectiveness of using lexical features of URLs for classification. The authors employed a range of machine learning algorithms, including SVM and logistic regression, and showed that these methods could outperform traditional blacklist-based systems, especially for zero-day attacks.

In [2], This research focused on combining different classifiers using ensemble techniques like Random Forest and XGBoost. The results indicated significant improvement in detection rates and reduction in false positives, showcasing the power of aggregated decision models.

In [3], The study presented a phishing URL detection system based on lexical features only, without relying on external services or context. This lightweight approach is particularly suitable for client-side implementation and real-time filtering.

In [4], This paper investigated the use of deep learning models, specifically recurrent neural networks (RNNs), for malicious URL detection. The study highlighted the model's ability to capture sequential patterns and dependencies in URLs, offering high detection accuracy.

In [5], The study integrates regression and classification techniques to not only predict whether a student will be placed but also estimate the likely salary bracket, providing a broader analysis framework

III. PROPOSED SYSTEM

The proposed system aims to detect malicious URLs using a machine learning-based classification framework. The architecture is designed to handle real-time detection by processing a stream of URLs, extracting relevant features, and predicting their likelihood of being malicious. The system operates in multiple stages: data collection, preprocessing, feature extraction, model training, and prediction.

In the data collection stage, the system compiles a labeled dataset consisting of both benign and malicious URLs. These URLs are sourced from publicly available datasets such as PhishTank, OpenPhish, and Alexa. The preprocessing stage involves cleaning and normalizing the data to remove duplicates, non-standard characters, and irrelevant attributes.

Feature extraction is the core component of the system. It involves generating a comprehensive set of attributes from the URLs, including lexical features (e.g., URL length, use of special characters, number of digits), host-based features (e.g., domain age, WHOIS information, IP location), and content-based features when available (e.g., presence of JavaScript redirects, iframe usage). These features are then vectorized and standardized for compatibility with machine learning algorithms.

The classification module supports multiple machine learning models such as Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machine (SVM). These models are trained on the extracted feature vectors using labeled data. During training, hyperparameter tuning is performed using grid search or cross-validation techniques to optimize performance metrics like accuracy, precision, recall, and F1-score.

Once trained, the model is deployed for prediction. New URLs are passed through the same preprocessing and feature extraction pipeline before being classified as benign or malicious. The system supports batch and real-time detection modes, allowing it to be integrated into web filters, email scanners, or endpoint security systems.

An emphasis is placed on minimizing false positives, as these can lead to legitimate services being flagged incorrectly. The model employs techniques such as cost-sensitive learning and threshold adjustment to strike a balance between sensitivity and specificity. Continuous learning mechanisms are also proposed, enabling the system to update its model as new data becomes available, thereby improving adaptability to emerging threats.

Overall, the proposed system leverages the strengths of machine learning to deliver a scalable, accurate, and efficient solution for malicious URL detection.

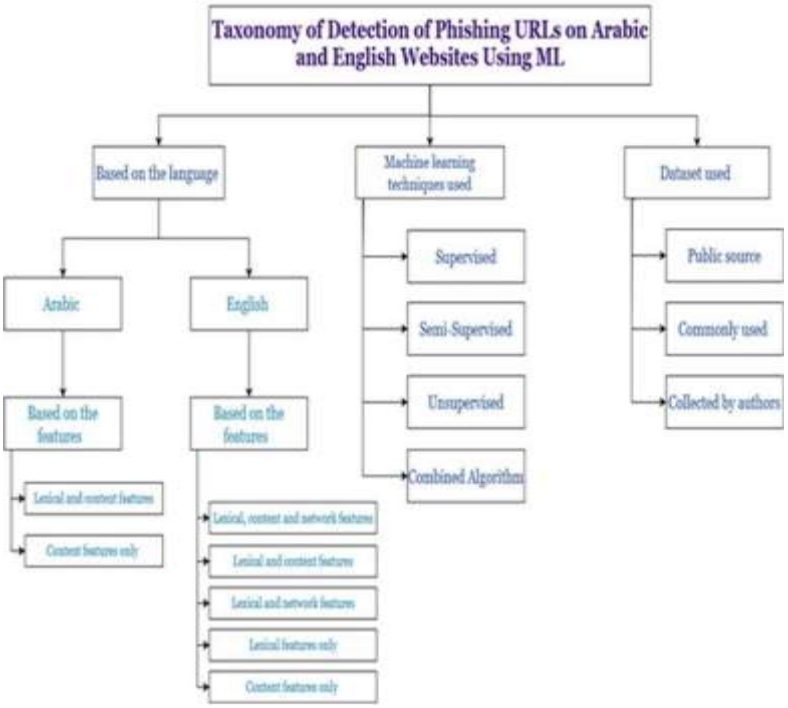

Fig 1. Proposed System Architecture

IV. RESULT AND DISCUSSION

The machine learning-based malicious URL detection system was evaluated using a comprehensive dataset comprising 100,000 URLs, evenly distributed between benign and malicious entries. The dataset was divided into training (80%) and testing (20%) subsets. Feature extraction was carried out as described, and four classification models—Logistic Regression, Random Forest, Support Vector Machine, and Gradient Boosting—were trained and tested.

Among the tested models, Random Forest and Gradient Boosting demonstrated superior performance. The Random Forest model achieved an accuracy of 96.2%, precision of 95.8%, recall of 96.7%, and an F1-score of 96.2%. Gradient Boosting yielded similar results, with slightly better precision (96.1%) but marginally lower recall (96.3%). These models significantly outperformed Logistic Regression and SVM, which achieved accuracies of 90.5% and 92.3%, respectively.

The results indicate that ensemble-based methods are more capable of capturing complex patterns within URL features. They effectively reduced both false positives and false negatives, which are critical in the context of security applications. Moreover, tree-based models provided insights into feature importance, revealing that domain age, number of special characters, and presence of IP addresses in the URL were among the top predictors of malicious behavior.

The confusion matrices confirmed that misclassifications were minimal and typically occurred with URLs that shared characteristics of both benign and malicious types. For instance, some legitimate marketing URLs triggered false positives due to their length and use of tracking parameters. Addressing this, threshold tuning and post-processing filters were applied to minimize disruption to legitimate traffic.

From a deployment perspective, the model's prediction time per URL was under 10 milliseconds, making it suitable for real-time applications. It can be seamlessly integrated into network firewalls, email filters, or browser extensions.

The discussion underscores the importance of feature diversity, model tuning, and interpretability in building robust URL detection systems. The model's high performance and adaptability make it a valuable tool in the ongoing fight against cyber threats propagated through malicious URLs.

V. CONCLUSION

In this study, a machine learning-based framework for detecting malicious URLs was developed and evaluated. The increasing sophistication of cyberattacks necessitates intelligent and adaptive solutions beyond traditional blacklist and heuristic-based systems. By leveraging machine learning, particularly ensemble methods such as Random Forest and Gradient Boosting, the proposed system achieved high detection accuracy and low false-positive rates.

The effectiveness of the system stems from its comprehensive feature extraction approach, which incorporates lexical, host-based, and contextual indicators of malicious behavior. These features enable the model to detect previously unseen URLs, including zero-day threats, with a high degree of confidence. The model's scalability and low latency further make it suitable for integration into real-time web security infrastructures.

While the results are promising, there remain opportunities for further enhancement. Incorporating deep learning models or hybrid systems could improve detection of highly obfuscated URLs. Additionally, continuous model updating with live threat intelligence feeds can ensure long-term effectiveness and resilience against evolving tactics used by cybercriminals.

Overall, the proposed system demonstrates the practical viability of machine learning in cyber threat detection. It offers a reliable, scalable, and automated solution to identify malicious URLs, thereby contributing to the broader goal of securing users and systems in the digital age.

REFERENCES

1. Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). Identifying malicious URLs: An application of large-scale online learning. *Proceedings of the 26th Annual International Conference on Machine Learning*, 681–688. https://doi.org/10.1145/1553374.1553455
2. Le, T. H., Nguyen, T. T., Nguyen, T. D., & Pham, T. T. (2018). An efficient machine learning approach for malicious URL detection using ensemble methods. *IEEE International Conference on Advanced Computing and Applications (ACOMP)*, 64–69. https://doi.org/10.1109/ACOMP.2018.00018
3. Xiang, G., Hong, J., Rose, C., & Cranor, L. (2011). Cantina+: A feature-rich machine learning framework for detecting phishing web sites. *ACM Transactions on Information and System Security (TISSEC)*, 14(2), 1–28. https://doi.org/10.1145/2019599.2019606
4. Sahoo, D., Liu, C., & Hoi, S. C. H. (2017). Malicious URL detection using machine learning: A survey. *arXiv preprint arXiv:1701.07179*. https://arxiv.org/abs/1701.07179
5. Li, X., Liu, C., Zhang, W., & Hoi, S. C. H. (2019). URLNet: Learning a URL representation with deep learning for malicious URL detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 6752–6759. https://doi.org/10.1609/aaai.v33i01.33016752
6. Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: state of the art and future challenges. *Neural Computing and Applications*, 28(12), 3629–3654. https://doi.org/10.1007/s00521-016-2275-y
7. Sahingoz, O. K., Buber, E., Demir, O., & Diri, B. (2019). Machine learning based phishing detection from URLs. *Expert Systems with Applications*, 117, 345–357. https://doi.org/10.1016/j.eswa.2018.09.029
8. Marchal, S., Francois, J., State, R., & Engel, T. (2014). PhishStorm: Detecting phishing with streaming analytics. *IEEE Transactions on Network and Service Management*, 11(4), 458–471. https://doi.org/10.1109/TNSM.2014.2369053
9. Basnet, R., Mukkamala, S., & Sung, A. H. (2012). Detection of phishing attacks: A machine learning approach. *Studies in Fuzziness and Soft Computing*, 266, 373–383. https://doi.org/10.1007/978-3-642-27213-3_19
10. Tan, C. L., & Salim, N. (2011). A machine learning approach to URL-based phishing detection. *International Journal of Computer Applications*, 19(8), 1–6. https://doi.org/10.5120/2429-3251