International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Intelligent Detection of Malicious Websites Using Machine Learning

[1]Dr. P. P. Sadhu Naik,[2]G. V. S. K. Prasad,[3]Ch. Naveen,[4]K. Venkatasubbaiah,[5]M. Nagalakshman, [6]A. Venkateswarulu

[1]Professor & HoD, Dept of CSE, Dr.Samuel Gerorge Institute of Engineering & Technology, Darimadugu Village, Markapur, Prakasam, Idupur, Andhra Pradesh 523320.

[2,3,4,5,6]U. G Student, Dept of CSE, Dr.Samuel Gerorge Institute of Engineering & Technology, Darimadugu Village, Markapur, Prakasam, Idupur, Andhra Pradesh 523320.

ABSTRACT-*Social media platforms have become an essential medium for communication, business, and information sharing. However, the increasing growth of these platforms has also given rise to fake accounts that pose threats such as misinformation, cyber fraud, spam, and identity impersonation. Detecting and preventing fake accounts is therefore a critical challenge to ensure the security, reliability, and trustworthiness of online communities. This project focuses on the detection of fake accounts on social media using machine learning techniques. By analyzing user profile features, activity patterns, content characteristics, and network behavior, the system can differentiate between genuine and fake users. The study explores both traditional rule-based methods and modern algorithms such as Decision Trees, Random Forests, and XGBoost, which improve detection accuracy. The proposed approach emphasizes building an intelligent model that learns from real-world datasets and adapts to evolving patterns of fake account creation. The outcome of this project will help in reducing the spread of spam, safeguarding users from fraudulent activities, and enhancing the overall trust of social media platforms. This work not only contributes to the technical understanding of fake account detection but also highlights the importance of ethical considerations, user privacy, and scalability in real-world applications.*

Keywords*—Malicious websites, Machine learning, Cyber security, URL analysis, Network traffic.*

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

INTRODUCTION

The internet has become an essential platform for communication, business, and information sharing. However, this growth has also led to an increase in cyber attacks, particularly those involving malicious URLs, Attackers use deceptive links to steal sensitive information, spread malware, or redirect users to fake websites. Existing security mechanisms are often ineffective against newly created malicious URLs. Machine learning provides a dynamic solution by learning patterns from historical data and identifying unknown threats. This project focuses on detecting malicious URLs using supervised machine learning techniques. By converting URLs into numerical features and training effecient models, the system aims to provide accurate and proactive threat detection. The proposed solution enhances online safety by preventing threat detection. users from accessing harmful websites.

LITERATURE SURVEY

Several studies have explored machine learning approaches for malicious website detection. Researchers have shown that lexical URL features and network traffic characteristics are effective indicators of malicious behavior. Supervised learning models such as Support Vector Machines, Random Forest, and Naïve Bayes have demonstrated high accuracy in classifying malicious websites. Despite promising results, many existing approaches suffer from limitations such as high false-positive rates, limited scalability, and inability to detect evolving attack patterns. These challenges highlight the need for an improved detection mechanism that combines feature analysis with intelligent learning models.

RELATED WORK

Traditional blacklist-based methods are ineffective against newly generated malicious websites. Machine learning techniques have been widely adopted to classify websites using URL and behavioral features. Studies show that combining network traffic analysis with supervised learning improves detection accuracy. Algorithms such as Random Forest and Support Vector Machine are commonly used due to their robustness. These approaches form the basis for intelligent malicious website detection systems.

EXISTING SYSTEM

Existing systems for malicious URL detection primarily rey on blacklists and

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

signature-based techniques. These methods can only detect known malicious URLs and fail to identify newly generated threats. Manual rule-based approaches exist, but they often use limited features and simple classifiers resulting in lower accuracy. Additionally, most existing systems do not provide reasons for their predictions, reducing user confidence. These limitations highlight the need for a more robust and explainable solution.

PROPOSED SYSTEM

The proposed system uses machine learning techniques to identify malicious URLs proactively. URLs provided by users are preprocessed and converted into numerical features such as length, special characters, protocol type, and keyword presence. These features are fed into advanced machine learning models, particularly XGBoost, to perform classification. The system not only determines whether a URL is safe or malicious but also generates an evidence summary explaining the decision. This improves accuracy, adaptability, and transparency. The proposed approach effectively mitigates cyber threats in real time.

SYSTEM ARCHITECTURE

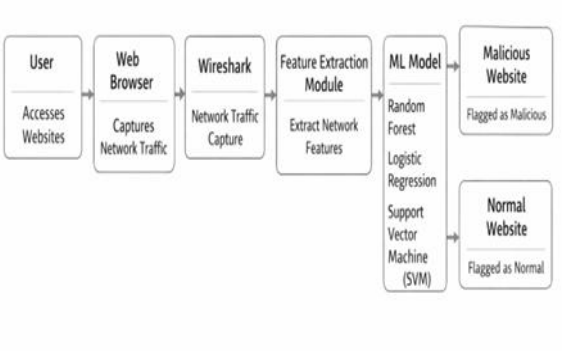

Fig 1: Architecture diagram

METHODOLOGY DESCRIPTION

The system detects malicious websites using machine learning based on network traffic behavior. When a user opens a website, network traffic is captured using the Wireshark tool. Important traffic features such as duration, packet flow, and response time are extracted from the captured data. These features are sent to a Flask-based web application for processing. The extracted values are analyzed using trained machine learning models, including Random Forest, Logistic Regression, and SVM. The model classifies the website as malicious or normal based on traffic patterns. Finally, the prediction result is displayed to the user through the web interface.

RESULTS AND DISCUSSION

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

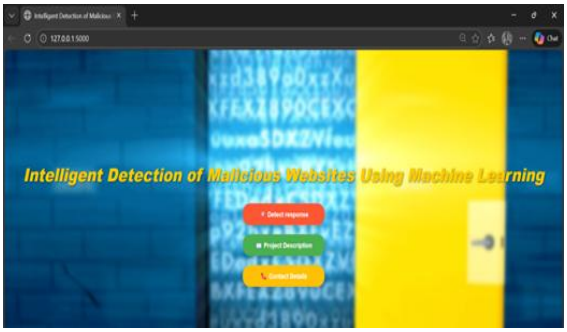

Fig 2: Home Page

This output shows the application's main interface, which enables users to access the contact, project description, and detection modules. It attests to the web-based system's successful deployment.

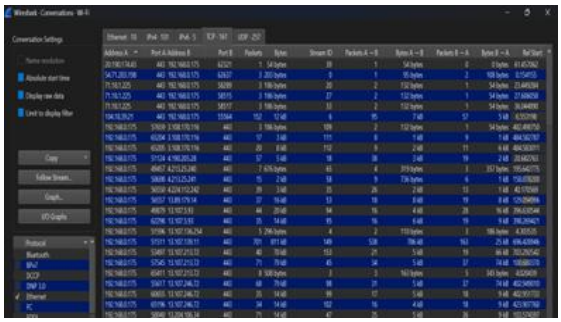

Fig 3: Wire Shark Traffic Screenshot

The network traffic that is recorded when a website is accessed using the Wireshark tool. The screenshot shows packet-level information that is used as input features for the detection of malicious websites, such as data flow, packet length, and response timing. This collected traffic is the main source of data for the machine learning model and aids in the analysis of website behavior.

Fig 4: Network Traffic Values

This screen shows the input form where users enter traffic parameters such as duration, flow bytes, packet length, and response time captured from Wireshark. It validates the interaction between the user and the backend system.

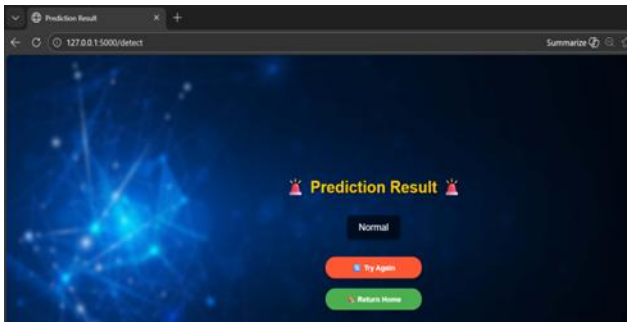

Fig 5: Result-Normal Website Detection

This screen shows the classification of a website as **"Normal"** when the traffic behavior matches legitimate browsing patterns. It demonstrates the model's ability to distinguish normal traffic from malicious activity.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

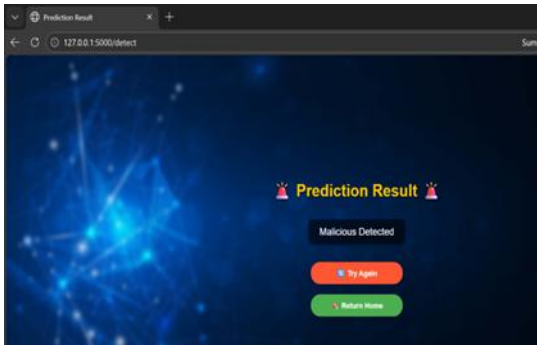

Fig 6: Result-Malicious Website Detection

CONCLUSION

This Project presents an efficient machine learning-based system for detecting malicious URLs. By extracting important URL features and applying classification models, the system accurately identifies unsafe links. The proposed approach helps in preventing cyber-attacks such as phishing and malware distribution. Overall, the system improves user security and trust while browsing the internet

FUTURE SCOPE

In the future, the proposed system can be enhanced by integrating deep learning models to further improve detection accuracy. The system can be developed as a real-time browser extension or mobile application for instant URL verification. Safe URLs can be highlighted in blue color, while malicious URLs can be displayed in red color along with alert messages to warn users. Continues model updates using live internet data can help in detecting newly emerging threats. Additionally, the system can be extended to identify malicious link in emails, text messages, and QR codes. These enhancements will improve usability,security, and real-time threat prevention.

REFERENCES

[1] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[2] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[3] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, “Beyond blacklists: Learning to detect malicious web sites from suspicious URLs,” in *Proc. ACM SIGKDD*, pp. 1245–1254, 2009.

[4] M. Aburrous, T. Hossain, K. Dahal, and F. Thabtah, “Intelligent phishing website detection using random forest classifier,” *Int. J. Advanced Computer Science and Applications*, vol. 8, no. 1, pp. 180–186, 2017.

[5] D. Canali, M. Cova, G. Vigna, and C. Kruegel, “Prophiler: A fast filter for the large-scale detection of malicious web pages,” in *Proc. World Wide Web Conf.*, pp. 197–206, 2014.

[6] S. Singh and R. Kaur, "Network traffic based malicious website detection using machine learning," *Journal of Information Security*, vol. 11, no. 3, pp. 150–162, 2020.

[7] H. Zhang and G. Liu, "A survey of malicious websites detection techniques," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2094–2121, 2018.

[8] M. Al-Fawa'reh, S. A. Rashaideh, and A. Aljawarneh, "Detection of malicious websites using machine learning techniques," *Security and Communication Networks*, vol. 2019, pp. 1–10, 2019.

[9] R. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," *Neural Computing and Applications*, vol. 25, no. 2, pp. 443–458, 2018.

[10] D. Sahoo, C. Liu, and S. C. Hoi, "Malicious URL detection using machine learning: A survey," *ACM Computing Surveys*, vol. 52, no. 3, pp. 1–36, 2019.

[11] M. Shafiq, X. Yu, A. A. Laghari, and L. Yao, "Network traffic classification using machine learning techniques," *Journal of Network and Computer Applications*, vol. 90, pp. 1–16, 2016.

[12] Y. Zhao, Y. Chen, and Z. Wang, "Detection of malicious web traffic using machine learning," *IEEE Access*, vol. 8, pp. 104199–104210, 2020.

[13] J. Kim and S. Lee, "Cyber threat detection based on network traffic behavior," *Computers & Security*, vol. 92, pp. 101–112, 2020.

[14] K. Srinivasan and S. Ramachandran, "Behavior-based malicious website detection using traffic features," *International Journal of Cyber Security*, vol. 3, no. 2, pp. 45–53, 2021.

[15] A. Buczak and E. Guven, "A survey of data mining and machine learning methods for cybersecurity intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[16] R. Verma and A. Das, "What works and what does not: A study of phishing detection," in *Proc. IEEE Int. Conf. Data Mining*, pp. 1–10, 2017.

[17] T. Nguyen and G. Armitage, "A survey of techniques for internet traffic classification," *IEEE Communications Surveys & Tutorials*, vol. 10, no. 4, pp. 56–76, 2008.

[18] J. Saxe and K. Berlin, "Deep neural network based malware detection using two dimensional binary program features," in *Proc. IEEE Malware Conf.*, pp. 1–8, 2015.

[19] Y. Zhang, X. Luo, and S. Yu, "Detection of malicious websites based on traffic analysis," *Future Generation Computer Systems*, vol. 102, pp. 481–489, 2019.

[20] A. Kumar and T. Lim, "Early detection of malicious websites using machine learning," *International Journal of Computer Networks & Communications*, vol. 10, no. 5, pp. 25–36, 2018.