IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 16, May 2023

Employing Multi-Class Classification Techniques to Identify Harmful URLs

Mr. Vedant Rahane[1], Mr. Rushikesh Pokharkar[2], Mr. Aditya Wagh[3], Prof. K. U. Rahane[4]

[1,2,3]Student, Amrutvahini College of Engineering, Sangamner, Maharashtra, India

[4]Assistant Professor, Amrutvahini College of Engineering, Sangamner, Maharashtra, India

Abstract: *The main method for hosting unsolicited content, such as spam, malicious ads, phishing, and drive-by exploits, to mention a few uses of a malicious Uniform Resource Locator (URL), sometimes known as a malicious website. It is crucial to quickly identify the rogue URLs. The contemporary community of Internet users, which is becoming more integrated, is seriously threatened by malicious events. A popular method to find harmful events is detection based on network traffic features and machine learning techniques. Cybersecurity threats including ransomware, phishing, malware injection, etc. have significantly increased recently on many websites throughout the world. Numerous financial institutions, e-commerce businesses, and individuals suffered significant financial losses as a result. Since new attack types are being developed daily, controlling a cybersecurity attack in such a situation is a significant problem for cybersecurity specialists. Identifying dangerous URLs using lexical characteristics and a boosted tree-based machine learning strategy will be used in this project. Three well-known machine learning ensemble classifiers—Random Forest, Light GBM, and XGBoost—will be applied. We will be using a Malicious URLs dataset from Kaggle of 6,51,191 URLs, out of which 4,28,103 benign or safe URLs, 96,457 defacement URLs, 94,111 phishing URLs, and 32,520 malware URLs.*

Keywords: *Malicious URL, Machine Learning, Cybersecurity, Detection of Fraudulent Traffic, Lexical Features, Multi-class Classification.*

REFERENCES

[1] Xess, L. S., Khera, M., Prasad, T., Singh, R., & Aiden, M. K. (2022). Malicious Website Detection using Machine Learning. International Journal of Engineering Research & Technology, Published On 2022.

[2] Manyumwa, T., Chapita, P. F., Wu, H., & Ji, S. (2021). Towards Fighting Cybercrime: Malicious URL Attack Type Detection using Multiclass Classification. IEEE, Published On 2021.

[3] Rong, C., Gou, G., Cui, M., Xiong, G., Li, Z., & Guo, L. (2020). MalFinder: An Ensemble Learning-based Framework for Malicious Traffic Detection. IEEE, Published On 2020.

[4] Joshi, A., Lloyd, L., Westin, P., &Seethapathy, S. (2019). Using Lexical Features for Malicious URL Detection - A Machine Learning Approach. ARXIV, Published On 2019.

[5] Desai, A., Jatakia, J., Naik, R., & Raul, N. (2017). Malicious Web Content Detection Using Machine Learning. IEEE, Published On 2017.

[6] Murch, R., Milne, J. (2012). System Development and LifeCycle Management (SDLCM) Methodology. United States Nuclear Regulatory Commission, Washington, DC, Vol-3, pages 665.

[7] Mohammed, M., Khan, M. B., Bashier, E. B. M. (2016). Machine Learning: Algorithms and Applications. CRC Press, ISBN: 9781498705394.

DOI: 10.48175/IJARSCT-11130