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 3, December 2023

Evaluating Robustness of Learning-Based Malware Detection System using A Obfuscation Dataset

Ketan M. Deore[1], Pooja J. Kale[2], Vedant B. Wagh[3], Sumeet K. Patil[4], Prof. R. S. Khule[5]

Students, Department of Information Technology[1,2,3,4]

Professor, Department of Information Technology[5]

Matoshri College of Engineering and Research Centre, Nashik, India

Abstract: *Malware poses a significant threat to cybersecurity, continually evolving to evade traditional detection methods. Among the sophisticated evasion techniques employed by malware authors, code obfuscation stands out as a formidable challenge for security analysts. This research presents an innovative approach to combating obfuscated malware through the development of an Obfuscated Learning-Based Malware Detection System.*

The proposed system leverages advanced machine learning techniques to recognize and classify obfuscated code patterns commonly employed by malware. Traditional static and dynamic analysis methods struggle to cope with the rapidly changing landscape of obfuscation, prompting the need for a more adaptive and resilient detection system.

The system's foundation lies in a comprehensive dataset curated with a diverse set of obfuscated and non-obfuscated code samples. Through feature extraction, the model identifies subtle yet characteristic patterns indicative of obfuscation. The learning algorithm is trained on this dataset, utilizing a combination of supervised and unsupervised learning to enhance its capability to generalize across various obfuscation methods.

To address the dynamic nature of obfuscation, the system incorporates continuous learning mechanisms, allowing it to adapt and evolve alongside emerging obfuscation techniques. The learning model is regularly updated with new malware samples, ensuring its proficiency in identifying novel and polymorphic threats.

Furthermore, the system employs ensemble learning, combining the strengths of multiple models to achieve a higher detection accuracy rate. By integrating both static and dynamic analysis features, it establishes a more holistic approach to malware detection..

Keywords: Malware Detection, Deep Learning, Machine Learning, Obfuscation.

I. INTRODUCTION

The Obfuscated Learning-Based Malware Detection System represents a pioneering advancement in cybersecurity, specifically tailored to counter the escalating threat posed by obfuscated malware. This innovative system relies on state-of-the-art machine learning techniques to discern and categorize patterns within obfuscated code, a formidable challenge for traditional detection methods. By constructing a diverse dataset encompassing obfuscated and non-obfuscated code samples, the system undergoes comprehensive training, incorporating both supervised and unsupervised learning methodologies. Key to its efficacy is the system's adaptability. Recognizing the dynamic nature of obfuscation, it embraces continuous learning mechanisms, facilitating ongoing adjustments to evolving obfuscation techniques. This adaptability ensures the system's relevance and effectiveness against emerging polymorphic threats that mutate to evade detection. The driving force behind the creation of the "Obfuscated Learning-Based Malware Detection System" is the pressing need to address the evolving and sophisticated nature of malware threats. The conventional methods of cybersecurity often fall short in effectively identifying and countering malware that employs obfuscation techniques. Recognizing the limitations of traditional approaches, there is a compelling necessity for an innovative solution that leverages machine learning. The system is conceived as a response to the imperative of staying

DOI: 10.48175/568

ahead in the cybersecurity landscape, offering a dynamic and adaptive approach to tackle the challenges posed by obfuscated code. Its development is fuelled by the commitment to enhance the robustness of cybersecurity measures in the face of ever- changing

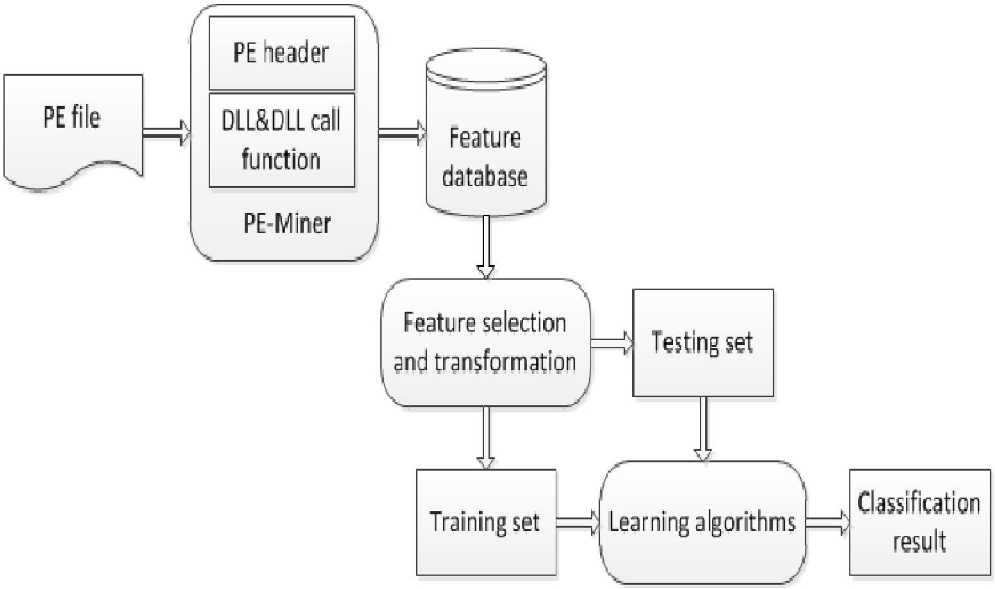
II. PURPOSE

To develop malware Detection System by using Machine Learning and Deep Learning which monitor Various malware fetures using a very unique dataset of thousands of features of malware detection which guides the User to detect malwares. Malwares are crucial system manipulators to cause harmful fatal viruses to cause loss towards software, These malwares can vary everyday as they evolved and caught by various detection systems. It has been adopted internationally for its Software maintenance benefit. Among several techniques, Obfuscation based malware detection system is very efficient as it uses the dataset with huge number of features which can counteract the malwares in day-to-day lifes

III. OBJECTIVE OF SYSTEM

The objectives of the "Obfuscated Learning-Based Malware Detection System" are to enhance detection accuracy and adapt to emerging threats. By leveraging advanced machine learning, the system aims to continuously learn and identify obfuscated code patterns, contributing to more resilient and effective cybersecurity defenses against evolving malware threats.

IV. SYSTEM ARCHIOTECTURE



It shows the general system architecture of the application wherein the the user and the admin interact with the web server which fetches the data from the database. We proposed an image processing-based method to detect malware. The architecture includes data collection, preprocessing, machine learning, dynamic detection, threat intelligence integration, user interface, and real-time protection for effective malware detection.

V. CONCLUSION

the Obfuscated Learning-Based Malware Detection System offers a sophisticated solution to the escalating threat of obfuscated malware. By integrating advanced machine learning, dynamic detection, and real-time protection, the system adapts rapidly to evolving obfuscation techniques. Its architecture ensures seamless compatibility with existing cybersecurity infrastructures, while the user-friendly interface empowers security analysts in training the model and interpreting results.

VI. ACKNOWLEDGMENT

First and the foremost I, express my deep sense of gratitude, sincere thanks and deep sense of Appreciation to project Guide **Dr. Rupali S Khule** Department of Information Technology, Matoshri College of Engineering and Research Center, Nashik. Your Availability at any time throughout the Semester, valuable guidance, view, comments, critics, encouragement and support tremendously boosted this research work.Again, Lot and lots of thank to head Information Technology Department **Prof. N. L. Bhale** for providing me the support I ever had. Your opinion, view, Comments and thoughts have been really helped me to improve my writing.

REFERENCES

[1] H.S. Anderson, P. Roth, EMBER: An open dataset for training static PE malware machine learning models, 2018, ArXiv e-prints arXiv:1804.04637.

[2] E. Raff, J. Barker, J. Sylvester, R. Brandon, B. Catanzaro, C.K. Nicholas, Malware detection by eating a whole EXE, 2017, ArXiv arXiv:1710.09435.

[3] G.E. Dahl, J.W. Stokes, L. Deng, D. Yu, Large-scale malware classification using random projections and neural networks, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 3422–3426, http://dx.doi.org/10.1109/ICASSP.2013.6638293.

[4] K. Rieck, P. Trinius, C. Willems, T. Holz_aff2n3, Automatic Analysis of Malware Behavior Using Machine Learning, 19 (4) (2011) 639–668.

[5] J. Saxe, K. Berlin, Deep neural network based malware detection using two dimensional binary program features, in: 2015 10th International Conference on Malicious and Unwanted Software, 2015, pp. 11–20, http://dx.doi.org/10. 1109/MALWARE.2015.7413680.