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

IJARSCT

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.

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.

DOI: 10.48175/568