World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:18, No:05, 2024

Using Autoencoder as Feature Extractor for Malware Detection

Authors : Umm-E-Hani, Faiza Babar, Hanif Durad

Abstract : Malware-detecting approaches suffer many limitations, due to which all anti-malware solutions have failed to be reliable enough for detecting zero-day malware. Signature-based solutions depend upon the signatures that can be generated only when malware surfaces at least once in the cyber world. Another approach that works by detecting the anomalies caused in the environment can easily be defeated by diligently and intelligently written malware. Solutions that have been trained to observe the behavior for detecting malicious files have failed to cater to the malware capable of detecting the sandboxed or protected environment. Machine learning and deep learning-based approaches greatly suffer in training their models with either an imbalanced dataset or an inadequate number of samples. AI-based anti-malware solutions that have been trained with enough samples targeted a selected feature vector, thus ignoring the input of leftover features in the maliciousness of malware just to cope with the lack of underlying hardware processing power. Our research focuses on producing an anti-malware solution for detecting malicious PE files by circumventing the earlier-mentioned shortcomings. Our proposed framework, which is based on automated feature engineering through autoencoders, trains the model over a fairly large dataset. It focuses on the visual patterns of malware samples to automatically extract the meaningful part of the visual pattern. Our experiment has successfully produced a state-of-the-art accuracy of 99.54 % over test data.

Keywords : malware, auto encoders, automated feature engineering, classification
Conference Title : ICCSIS 2024 : International Conference on Computer Science and Intelligent Systems
Conference Location : Montreal, Canada
Conference Dates : May 23-24, 2024