Back Propagation Neural Network Based Framework for Unknown Malware Detection

Chen, Yi-Jhe (School: The Affiliated Senior High School of National Taiwan Normal University)

This project explores the application of language models in creating backpropagation-based neural networks for the detection of known and unknown malware. A representation learning model (RPLM) is proposed in this research. The detection framework involves three key steps: (1) disassembly of executable samples into assembly code, (2) extraction of assembly text to form representation vector using the proposed RPLM language models, and (3) classification of the representation vectors by a fully-connected neural network to detect presence of malware. Experiments on malware collected from 2012 through 2023 showed a 98.28% accuracy. Furthermore, experiments on unknown family of malware achieved 97.25% average detection rate. Finally, experiments on training model with known malware up to year N for detection of unknown malware in future years, showed that this framework can maintain 95.63% and 93.01% detection rates in year N+1 and N+2, respectively. An ensembled model was also proposed to enhance the robustness of the detector. Combining features extracted from different language models using fusion techniques could enhance the robustness of the malware detection framework. Experiments shown that fusion can improve known and unknown malware detection rate to 99.39 % and 96.16%, respectively. This research demonstrates feasibility of leveraging natural language models for the detection of malware from low-level assembly code. In comparison to signature-based anti-virus software, the proposed natural language based neural network, by having high and reliable detection rate for known and unknown malwares, is a robust new way to safeguard system security.