POLYMORPHIC MALWARE DETECTION USING MACHINE LEARNING

Lakshika Sammani Chandradeva

A dissertation submitted in partial fulfilment of the requirement for
Master of Science degree in Cyber Security and Forensics

Department of Computing

Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK

2021

Abstract

In recent years, a great number of malware has spread indiscriminately, resulting in a variety of serious cyberspace security crises across the world. As a result, malware detection has emerged as a critical study area in cyberspace security. However, at present, practical training for malware detection relies mostly on theory and skills, with little emphasis on actual combat training. Most malware detection techniques rely on malware signatures. While detecting known dangerous programmes in a system is straightforward, the difficulty emerges when dealing with unknown malware. Since unknown malware cannot be identified using established malware signatures, approaches relying on signatures are incapable of identifying unknown or zero-day attacks.

Therefore, having analysed the methodologies used in existing malware detection solutions, it was determined that there is a requirement for malware detection solutions to detect polymorphic malware. Polymorphic malware is a subtype of malware that is continually changing its identifying traits to evade detection. Numerous common varieties of malware, such as viruses, worms, bots, trojans, and keyloggers, are polymorphic in nature. Polymorphic approaches require continuously modifying recognizable attributes such as file names and types or encryption keys to render malware undetectable by various detection techniques. Polymorphism is used to circumvent pattern-matching detection, a technique employed by security systems such as the current endpoint threat detection solutions. While many characteristics of polymorphic malware alter, its functional objective remains constant.

The proposed malware detection framework has addressed the inability of existing solutions to detect malware that changes its distinguishing characteristics in order to avoid detection. This research was performed using a novel behavioural malware detection method based on Deep Graph Convolutional Neural Networks (DGCNNs) to learn directly from API call sequences and their related behavioural graphs.

Keywords

Deep Graph Convolutional Neural Networks, Long Short Term Memory (LSTM), Polymorphic Malware Detection