Politecnico di Torino

Machine Learning and other Computational-Intelligence Techniques for Security Applications

Supervisor:
Prof. Giovanni Squillero

Candidate:
Andrea Marcelli

Abstract

Machine learning and evolutionary computation are powerful tools that achieved incredible results in the most variegate fields. While the techniques are quite known, their application requires a deep knowledge in the field of usage. This thesis explores the application of computational intelligence methodologies to open problems in computer security, mainly in the field of malware families detection.

Malware is a big business. With hundreds of thousands of malware delivered every day, manual analysis in not an option. Malicious samples are commonly detected using a combination of techniques, ranging from machine learning to hash-based content. However, the industry mostly relies on signatures, which are patterns extracted from the code or behavior of selected samples. Generating effective signatures, with 0-false positives, and low false negatives rates, is a task that requires a considerable amount of time and resources from skilled experts, while automatically generating them is an open problem.

In this thesis, we propose a semi-supervised methodology for the automatic identification of malware families, used to safely extend experts knowledge on new malicious samples, and to reduce the amount of applications to manually analyze. Then, newly discovered samples are submitted to an automatic signature generation procedure, which produces a formal rule which has a limited risk of detecting false positives in the future, yet it is general enough to catch future threats.

The effectiveness of the approach is assessed running experiments on 1.5 million Android applications, the largest dataset ever used in a public research on Android malware. The procedure has been implemented in two frameworks which have been publicly released: *YaYaGen* for Android applications, and *YaYaGenPE* for Windows. Furthermore, since January 2018, part of the proposed approach is in use in Koodous, a collaborative analysis platform for Android, developed by Hispasec.