Image-based Malware Family Detection: An Assessment between Feature Extraction and Classification Techniques

Giacomo Iadarola[1], Fabio Martinelli[1], Francesco Mercaldo[1,2], and Antonella Santone[2]
[1]*Institute for Informatics and Telematics, National Research Council of Italy (CNR), Pisa, Italy*
[2]*Department of Biosciences and Territory, University of Molise, Pesche (IS), Italy*
{*giacomo.iadarola, fabio.martinelli, francesco.mercaldo*}*@iit.cnr.it,* {*francesco.mercaldo, antonella.santone*}*@unimol.it*

Keywords: Machine Learning, Mobile Security, Android, Malware Classification, Image Texture Analysis

Abstract: The increasing number of malware in mobile environment follows the continuous growth of the app stores, which required constant research in new malware detection approaches, considering also the weaknesses of signature-based anti-malware software. Fortunately, most of the malware are composed of well-known pieces of code, thus can be grouped into families sharing the same malicious behaviour. One interesting approach, which makes use of Image Classification techniques, proposes to convert the malware binaries to images, extract feature vectors and classifying them with supervised machine learning models. Realizing that researchers usually evaluate their solutions on private datasets, it is difficult to establish whether a model can be generalized on another dataset, making it difficult to compare the performance of the various models. This paper presents a comparison between different combination of feature vector extraction methods and machine learning models. The methodology aimed to evaluate feature extractors and supervised machine learning algorithms, and it was tested on more than 20 thousand images of malware, grouped into 10 different malware families. The best classifier, a combination of GIST descriptors and Random Forest classifiers, achieved an accuracy of 0.97 on average.

ACKNOWLEDGEMENTS

This work has been partially supported by MIUR - SecureOpenNets and EU SPARTA contract 830892, CyberSANE projects, and the EU project CyberSure 734815.