

In this Thesis, we propose a machine-learning based classification algorithm of applications for a popular mobile phone operating system Android that can distinguish malicious samples from benign ones. Feature extraction for the machine learning is based on static analysis of the application's bytecode with focus on API and method calls. We show various ways to transform the most frequent API and method calls into numeric (histogram-based) features. We further examine the specifics of the extracted features and discuss their importance. The dataset used for experiments in this Thesis contains more than 200,000 samples with approximately half of them malicious and half of them benign. Further, multiple machine learning algorithms are examined and their performance is evaluated. The size of our dataset prevents overfitting and hence provides a reliable basis for training of classification models. The results of the experiments show that the proposed algorithm achieves very low false positive rate under 2.9% while preserving specificity over 93.6%.