World Academy of Science, Engineering and Technology
International Journal of Electrical and Information Engineering
Vol:9, No:05, 2015

A Comparative Study of Malware Detection Techniques Using Machine Learning Methods

Authors : Cristina Vatamanu, Doina Cosovan, Dragos Gavrilut, Henri Luchian

Abstract : In the past few years, the amount of malicious software increased exponentially and, therefore, machine learning algorithms became instrumental in identifying clean and malware files through semi-automated classification. When working with very large datasets, the major challenge is to reach both a very high malware detection rate and a very low false positive rate. Another challenge is to minimize the time needed for the machine learning algorithm to do so. This paper presents a comparative study between different machine learning techniques such as linear classifiers, ensembles, decision trees or various hybrids thereof. The training dataset consists of approximately 2 million clean files and 200.000 infected files, which is a realistic quantitative mixture. The paper investigates the above mentioned methods with respect to both their performance (detection rate and false positive rate) and their practicability.

Keywords : ensembles, false positives, feature selection, one side class algorithm
Conference Title : ICISS 2015 : International Conference on Information Systems Security
Conference Location : Paris, France
Conference Dates : May 18-19, 2015