Using Spam Filters to Detect Malware: A Machine Learning Approach to Malware Detection

Kim, Chan Woo (School: Shanghai American School - Puxi Campus)
Liu, Austin (School: Shanghai American School - Puxi Campus)

As computing systems become increasingly advanced and as users increasingly engage themselves in technology, security has never been a greater concern. In malware detection, static analysis has been the prominent approach. This approach, however, quickly falls short as malicious programs become more advanced and adopt the capabilities to transform its code to execute the same malicious functions, making static analysis virtually inapplicable to newer variants. The approach we propose uses dynamic analysis of malware, analyzing behavior of programs instead. We used widely used document classification techniques to detect malware by doing such analysis on system call traces, a form of dynamic analysis. Features considered are extracted from system call traces of benign and malicious programs, and the task to classify these traces is treated like a binary document classification task using sparse features. The system call traces were processed to remove the parameters to only leave the system call function names. The features were processed with the Chi squared test, grouped into various n-grams, and weighted with Term Frequency-Inverse Document Frequency weighting. We applied Support Vector Machines optimized using a Stochastic Gradient Descent algorithm that uses L1, L2, and Elastic-Net regularization terms, the best of which achieved a highest of 98% accuracy with 98% recall score. Additional contributions include the identification of significant system call sequences that could be avenues for further research.

Awards Won:
China Association for Science and Technology (CAST): Award of $1,200
King Abdul-Aziz &amp
his Companions Foundation for Giftedness and Creativity: Award of $1000 for research in Cyber Security
Fourth Award of $500