The Analysis and Classification of Android Malware

Kimberly Tam

"With the integration of mobile devices into our daily lives, smartphones are privy to increasing amounts of sensitive information. As of 2016, Android is the leading smartphone in popularity with sophisticated mobile malware targeting its data and services. Thus this thesis attempts to determine how accurate and scalable Android malware analysis and classification methods can be developed to robustly withstand frequent, and substantial, changes within the Android device and in the Android malware ecosystem.

First, the author presents a comprehensive survey on leading Android malware analysis and detection techniques, and their effectiveness against evolving malware. Through the systematized survey, the author identifies under-developed areas of research which lead to the development of the novel Android malware analysis and classification solutions within in this thesis.

This thesis considers the usefulness and feasibility of reconstructing high-level behaviours via system calls intercepted while running Android apps. Previously, this method had only been rudimentarily implemented. However, the author was able to remedy this and developed a robust, novel, framework, to automatically and completely reconstructs all Android malware behaviours by thoroughly analysing dynamically captured system calls.

Next, the author investigates the efficacy of using our reconstructed behavioural profiles, at different levels of abstractions, to classify Android malware into families. Experiments in this thesis show our reconstructed behaviours to be more effective, and efficient, than raw system call traces. To classify malware, we utilized support vector machines to achieve high accuracy, precision and recall. Deviating from previous methods, we further apply statistical classification to achieve near-perfect accuracies.

Finally, the author explores an alternative Android malware analysis method using memory forensics. By extrapolating from these experiments, the author theorizes how to use this method to assist in capturing behaviours our previous methods could not, and how they could assist classification."