Sean Kilgallon

Magic: Malware Analysis to Generate Important Capabilities

Manually constructed malware analysis platforms that identify important capabilities in malicious software cannot keep up with the massive amounts of malware being released on a daily basis. Traditional approaches that detect the functional capabilities of malware usually contain brittle handcrafted heuristics that quickly become outdated, and can be exploited by nefarious actors. As a result, it is necessary to change the way software security is approached by using advanced analytics, i.e., machine learning, and significantly more automation, to develop more adaptable malware analysis engines that correctly deduce the important capabilities of malware. In this presentation, I will discuss our novel work using machine learning (i.e., decision tree) models to automatically identify a malware's important capabilities. Our training data consists of features extracted from extremely fast static analysis of malicious code, as well as information derived from slower dynamic analysis of binaries in a malware analysis sandbox. Our experimental results demonstrate that by learning from relatively large amounts of malware, we are able to accurately predict important capabilities of malicious executables with an accuracy of up to 98% without having to run malware through expensive dynamic analysis.

Biography

Sean Kilgallon is a postdoctoral researcher at the University of Delaware and Lead Data Scientist at Cyber 20/20, Inc. His research focusses on large scale machine learning for the detection and classification of malware. Using scalable cloud based static and dynamic analysis, he is able to analyze millions of malware files to be used as features for machine learning. His research interests include deep learning, malware analysis, and high performance computing (HPC).