Doctoral Dissertation

Doctoral Program in Control and Computer Engineering

Machine Learning Methods for Hardware-Based Malware Detection

Candidate: Cristiano Pegoraro Chenet

Supervisors: Prof. Stefano Di Carlo and Prof. Alessandro Savino

Cyber insecurity is among the most significant global risks, alongside major threats such as climate change and involuntary migration. Malware, short for malicious software, is the primary vector for cybercrimes. It includes any code modification within a software system aimed at causing harm or disrupting the system's intended function. Malware attacks cover spying, intrusive ads, email abuse, system damage, ransom demands, data release, slowdown, browser manipulation, and unauthorized access to sensitive information. The latest attempt to boost malware detection is a hardware-based approach, also called Hardware-Based Malware Detection (HMD). It involves dynamically analyzing architecture and microarchitecture events in a Central Processing Unit (CPU) using Machine Learning (ML) algorithms to distinguish between benign applications and malware. This thesis investigated this approach in depth, focusing on how the core component - the ML methods – should be employed to optimize performance and efficiency. A survey on the field deepens knowledge about malware, its properties and classification; discusses and compares the main available approaches for malware detection; identifies and discusses key features of HMD approaches, such as the CPUs' Performance Monitoring Units (PMUs), starting point that enables the detection by hardware; and elaborates a landscape of recent papers in HMD, assessing proposed solutions in terms of performance and efficiency. Two case studies employ a simulation environment and an anomaly HMD framework to investigate the approach performance capabilities in detecting the challenging zero-day malware, a threat that exploits previously unknown vulnerabilities. These studies implement the detection for the emerging Reduced Instruction Set Computer V (RISC-V) to disseminate the hardware-based approaches to the community surrounding this platform. The findings presented in this thesis help the malware detection field. Vulnerability exploitation, propagation method, and concealment strategy, together Common Vulnerabilities Exposures (CVE) and Common Weakness Enumeration (CWE) lists, are straightforward resources to deal with malware. The advantages of HMD are the ability for runtime, zero-day, and stealthy malware detection, resilience against subverting the protection, low-performance overhead, and reduced detection cost. ML methods to improve

performance in the field were discussed, such as ensemble learning, specialization, adaptive detection, and time series. Moreover, the case studies showed the ability of anomaly HMD to detect zero-day threats, identifying some points to overcome the poor detection: classifiers should be tailored to specific applications, the randomness/unpredictability of protected applications is an obstacle, detection algorithms should match the application, and an optimal number of ML features (Hardware Performance Counters (HPCs)) helps tune performance. Finally, HMD is a promising solution, especially if combined with software detection, generating effective and lightweight detectors.

Keywords: Cybersecurity, zero-day malware, malware detection, hardware-based detection, RISC-V