Abstract No: SO-06 **Software Intensive Systems**

Machine learning based model for Android malware analysis and detection

G. A. Sathindu Kavneth* and Shantha Jayalal

Department of Industrial Management, Faculty of Science,
University of Kelaniya, Sri Lanka
**Email: sathindukavneth@gmail.com*

Rapid advancement of technology has enabled smartphones to become extremely powerful. They are capable of attracting a considerable amount of users with new features provided by mobile device operating systems such as *Android* and *iOS*. Android extended its lead by capturing 86 percent of the total market in 2017, and became the most popular mobile operating system. The Android operating system, which is found on a wide range of devices is owned by Google and powered by the Linux kernel. It is an open source operating system that enables mobile application developers to access unlocked hardware and develop new apps as they wish. However, this huge demand and freedom has made the hackers and cybercriminals more curious to generate malicious apps towards the Android operating system. They constantly target the security vulnerabilities in the operating system to gain access within the system. Even though, Google provides a primary set of security services, there are possibilities for potentially harmful applications in the Google Play store and other third party application stores. Thus, research on effective and efficient mobile threat analysis becomes an emerging and important topic in cybersecurity research area. Many researchers proposed various security analysis and evaluation strategies such as static analysis and dynamic analysis. In this research, we propose a hybrid approach, which aggregates the static and dynamic analysis for detecting security threats and attacks by Android malware application. This approach has two phases. First phase is the static analysis for applications, which will analyze the mobile application without execution. This focuses on extracting app *APK* file and examining permission requests made by Android apps that have declared in *AndroidManifest.xml*, as a means for detecting malwares. Because, in most of cases extra permissions granted by applications will allow the attacker to exploit the device. As the next phase, we perform dynamic analysis for mobile application. This phase focuses on runtime data obtained from the applications such as CPU, scheduler information from every running application, network calls, sensor data and so forth. For both phases, we have used supervised, machine learning based algorithms to train models and recognize malwares. In the first phase, potentially harmful applications can be identified as well as in the proposed hybrid mechanism, which is a combination of both phases. Data that was collected by several cybersecurity research centers were used for the evaluation of the proposed hybrid approach and both real-life malware and benign app data demonstrated a good detection performance with high scalability. The initial findings have been more accurate in identifying Android malwares rather than separating those two static and dynamic behaviors.

Keywords: Android, Machine learning, Malware detection, Security