C18 H.Y.D.R.A. (Hyper Yield Data-Driven Real-Time Analysis)

Anthony Skjellum, PhD, Auburn University, Dept of Computer Science and Software Eng, 345 W Magnolia, 3101 Shelby Center, Auburn, AL 36849-5347; Austin Hancock, BS*, Auburn University, 3101 Shelby Center, Auburn, AL 36849-5347; Janice Canedo*, Auburn University, 3101 Shelby Center, Auburn, AL 36849-5347; and Erby Fischer, Auburn University, 3101 Shelby Center, Auburn, AL 36849-5347*

After attending this presentation, attendees will better understand a dynamic malware detection architecture for forensic science on Android™ devices that is also applicable to Linux®-based systems.

This presentation will impact the forensic science community by verifying and illustrating in detail the steps necessary to derive a data set composed of process control block variables that, when monitored in real time, are capable of identifying malicious behavior within milliseconds of its occurrence. This provides a dynamic architecture for forensic science on Android™ devices and supports live response as well as network defense and continuous monitoring.

Android™ and Linux® malware is an important class of threats to digital systems including mobile phones, tablets, Internet of Things (IoT), and portable devices. In this presentation, the focus is on advancements based on previous work that combines machine learning and process behavior to detect malware dynamically without resort to signature-based or other static methods. This research enables forensics studies of running systems.

Current static detection, including signature-based detection, fail to adapt quickly to the changing nature of malware in mobile devices. This inability to adapt quickly is inefficient at providing malicious behavior analysis, particularly in light of digital forensics. By utilizing a dynamic detection technique, attendees will overcome many of these aforementioned shortcomings. For example, the dynamic technique does not require familiarity with a given sample, knowledge of its signature, nor is it impeded by code obfuscation.

Research conducted recently by Dr. Farrukh Shahzad has shown that the execution behavior of a malicious Android™ application on Android™ 2.3 is markedly different from that of a benign Android™ application. To define execution behavior, a data set was determined by using a custom kernel module to monitor all variables in the Process Control Block (PCB) for an android process. Once a model of malicious behavior was obtained, it could be utilized in real time to classify an unknown application's behavior as malicious or benign.

This study intends to investigate the methodology demonstrated on the legacy system by Shahzad on current Android™ versions (e.g., Android™ 4.3, 4.4, 5.0) further. Dataset investigations from the legacy system will be demonstrated to determine the extent of changes vis-à-vis current Android™ versions. A comparison of the mathematics and methodologies utilized by Shahzad to both Bayesian Classification and Exploratory Data Analysis techniques will be analyzed to determine if they perform equivalently for malware detection.

For mining the data, system processes including core device processes and network performance are considered. Each system process is examined in an effort to find non-trivial correlations. For each correlation discovered, investigations are pursued further in an effort to prove a causation, since not all correlations lead to a causation. Both Exploratory Data Analysis (EDA) techniques and Bayesian classification techniques are explored in order to find causation, if present.

Both the validity and capability of alternative classification techniques on process behavior are covered. The work demonstrates the dynamic nature of the dataset as the Android™ operating system version progresses. The results from this investigation validate and refine a dynamic architecture for forensic science on Android™ devices.

Beyond Android™ and Linux®, the methodology created and advanced in this presentation can be applied in the future to other embedded and mobile operating systems.

Dynamic, Android, Malware