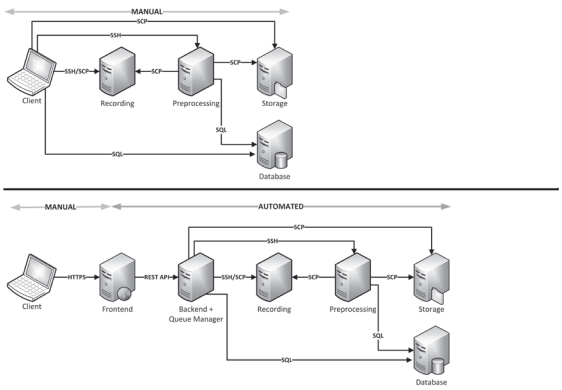
Extension of the KAN system for semi-automated memory tracing

Degree programme: Master of Science in Engineering | Specialisation: Information and Communication Technologies
Thesis advisor: Dr. Endre Bangerter
Expert: Reto Inversini (MELANI)

During the last years, the Security Engineering Lab (SEL) has been developing the KAN toolset for malware analysis. In this work, several additions to KAN were developed. These range from data acquisition through system call recording, over extended data extraction using system call arguments, call stacks and handle information from the acquired data, to adding automation and queueing mechanisms. Lastly, the KAN components have been cloud-enabled to increase the scalability.

The KAN system is a dynamic malware analysis system framework based on information found in memory. Instead of working on a single memory snapshot, KAN uses a technology called «memory tracing», which has been developed at the SEL. Contrary to classic memory snapshots, memory tracing creates hundreds to thousands of automated snapshots at a high frequency. This results in a much bigger data basis and allows to follow the complete lifecycle of a malware execution.

Prior to this work, the KAN workflow started with the selection of the sample one would like to record. After transferring the sample to the recording host, the recording had to be launched manually. Once the recording was done, the data needed to be copied to the preprocessing host and the preprocessing would again be launched manually. After the preprocessing step, the used files had then to be moved to the storage location. Although technically reliable, its implementation suffered from a few drawbacks. The whole process was divided in several small steps that had to be executed in the right order. Also it needed constant supervision to execute the next command at the right time. Furthermore, the information available from system calls, stack back traces and network communication was not recorded and lost to the analyst.



KAN Automated Workflow

During my thesis, I worked on several aspects of the KAN system to improve these shortcomings. In a first phase, I developed the system call argument recording and resolution logic, which is implemented in a way allowing for simple extension for other operating systems.
Another part of my work consisted of the conception and integration of a method to record network activity initiated during an analysis. As malware samples can generate dangerous traffic, it was necessary to set up an isolated network to protect our lab network. This isolated network was completed with mock-up services providing correct responses to a majority of network queries. This was accomplished through the integration of an established third-party product, in this case the Inetsim network simulation suite.

The last part of the thesis was the development of the automation and the subsequent unattended KAN workflow. The first step towards higher automation were scripts to control the recording and preprocessing phases. These scripts were then integrated into the KAN Dispatcher which works with Celery-based queues and a RabbitMQ backend to schedule and fulfill tasks asynchronously. The management user interface was integrated in the existing frontend. The latest iteration then lead to the integration of cloud technologies developed by Canonical, specifically MAAS (Metal As A Service) and Juju (Service Deployment).

With these new additions to the KAN system, the previously missing information is now also recorded and easily accessible in the analysis database. The KAN Dispatcher now handles the entire workflow and therefore allows the analyst to easily submit a sample with a corresponding configuration. In combination with the cloud technologies this allows to increase the throughput from some tens of samples per day to potentially several hundreds or thousands of samples per day.

Thomas Marcel Ender