Ingestion, parsing and visualization of KAPE output data in Elastic Stack

Degree programme : MAS Digital Forensics & Cyber Investigation

When handling cyber security incidents, Netcloud AG uses the KAPE (Kroll Artefact Parser and Extractor) tool to collect forensic data from an affected host. The output of this tool are multiple CSV files which makes it cumbersome to search, filter and visualize. This master thesis aimed to create a pipeline to parse and ingest this data into Elastic stack. Furthermore, visualizations and dashboards should be created to provide the analyst with valuable tools and information.

Initial situation

Netcloud AG provides various IT services for security, network, collaboration, cloud, datacenter and cyber defence. As part of the cyber defence services, Netclouds cyber defence center (NCCDC) offers incident response and forensic services. To provide the customer with an easy and fast way to collect forensic data from an affected host during cyber security incidents, a toolset with a KAPE script was developed. KAPE is a triage tool for incident response and forensic investigations, it allows an investigator to quickly collect and transform forensic relevant data. It provides an analyst with multiple CSV files containing all sorts of digital evidence. Analysing, searching and filtering CSV files is very cumbersome and takes a lot of time. Ingesting and parsing this data into Elastic stack would allow searching and filtering across all data as well as the ability to easily create visualizations and dashboards. The solution should allow the analyst to easily ingest and parse the data into Elastic stack where pre-defined searches, visualizations and dashboards have been created.

Goals

The goal of the thesis was to create a proof-of-concept (PoC) of the new solution. The new solution should allow the analyst to easily ingest the KAPE output data into Elastic stack without having to manually modify it first. Furthermore, pre-defined searches, visualizations and dashboards should be created. Additionally, possibilities on how to further use, analyse or enhance the ingested data shall be explored. The solution aims to be beneficial for an analyst during investigations and speed up analysis time as well as increasing Netclouds forensic readiness for future incident responses.

Concept

In the concept phase, KAPE modules and their output were analysed and it was evaluated, which modules will be parsed and ingested into Elastic stack. Furthermore, dashboards and their respective modules were defined. Finally, a test concept on how to test the new solution was created.

Implementation

In the implementation phase, the PoC virtual machine was created and Elastic stack was installed. A python script was created, which checks a defined folder for data, unpacks the KAPE output files, modifies some of the files and then starts the Logstash ingestion process. To parse the CSV fields into the correct fields in Elastic, Logstash configuration files were created for each defined module output.
Furthermore, different dashboards with visualizations and saved searches were created according to the concept. After the parsing and ingestion process as well as the visualizations and dashboards were created, different possibilities on how to further use, analyse and enhance the data from Elastic were explored. Finally, the new solution was tested according to the test concept.

Conclusion

The proof-of-concept of the solution showed that the KAPE output data can now be easily parsed and ingested into Elastic stack. The saved searches, visualizations and dashboards provide valuable tools to an analyst. Additionally, the data in Elastic can be further used with Elastics REST API. The solution increases Netclouds forensic readiness and supports the analysts during future forensic investigations. Following the master thesis semester, the solution will be implemented to the lab environment to enable access for all analysts.

Dominique Schlegel