Prof. Dr. Christoph Meinel
Internet-Technologies & -Systems

Master Project 2019

Cyber Threat Hunting/Detection via Data Science and Engineering

Background

Nowadays, the majority of organizations collect and store valuable event logs and telemetry generated by different components in the organization's premises, e.g., proxy servers, DNS servers, firewalls, workstations, etc. These event logs are then shipped to a centralized system known as Security Information and Event Management (SIEM). Traditionally, this collection and storage have been mostly done for compliance reasons. However, today, the organizations monitor these events within their Security Operations Center (SOC), constantly seeking indicators of compromise. In this regard, more and more organization are starting to realize the value and the potential of analyzing this data. This is due to the fact that SIEMs are expected to be the centralized repository for all events and information. Therefore, even if there is a threat that has managed to successfully bypass the perimeters of defense such as firewall, intrusion detection system, anti-virus, etc., it is quite likely that there are traces of its activities somewhere in the log-data shipped to the SIEM system. However, analyzing this big and dark data is an extremely challenging task that requires not only cybersecurity expertise but also advanced data science/engineering skills.

Tasks

In this project, the students are expected to start with a hypothesis for a threat hunting/detection approach based on advanced data science techniques (e.g., machine learning, graph analytics, etc.) and examine their hypothesis by implementing and testing it with real-world data. The outcome is expected to be a set of advanced approaches that intend to operate on a set of specific security-related data sources, hunting a potentially advanced threat.

Deliverables

- Advanced threat detection/hunting approach implemented, tested and documented
- Intermediate and final presentations and demos
- Scientific paper

Requirements

- Knowledge of
    - IT security
    - Data science and engineering
- Experiences with data science technologies (e.g., MapReduce, Apache Spark, TensorFlow, etc.)
- Programming Languages: Scala, Python, R

Contacts

- Prof. Dr. Christoph Meinel, Dr. Feng Cheng (HPI)
    - Email: security-analytics@hpi.de
    - Room: H-1.13 / Tel: +331-5509-519