AI in der Security – Projektplanung SS2020

In diesem Dokument soll eine grundlegende Planung für das abzuleistende Projekt dargestellt werden. Diese beinhaltet eine grundlegende Projektbeschreibung, das technisches Konzept, eine vorläufige Aufgabenverteilung, den Projektstrukturplan und eine Ablaufplanung.

Projektbeschreibung

Ziel dieses Projekts ist es einen Datensatz zu erstellen, welcher für Supervised-Machine Learning in Form von neuronalen Netzen zur Erkennung von bösartigem Netzwerkverkehr eingesetzt werden kann. Dies soll auf Basis bestimmter Charakteristika im Datensatz möglich sein. Somit ist der Zweck des Projekts einen Beitrag zur automatisierten und verlässlichen Erkennung von schadhaften Angriffen auf Rechnersysteme zu leisten. Um das Ziel dieses Projekts zu erreichen, muss ein entsprechender Netzwerkverkehr simuliert, aufgezeichnet und weiterverarbeitet werden. Letztlich wird der erzeugte Datensatz mit dem neuronalen Netz aus dem Vorprojekt getestet und ein Performance-Vergleich durchgeführt.

Technisches Konzept (Proof of Concept)

Um den Netzwerkverkehr zu erzeugen wird der Netzwerksimulator GNS3 verwendet. Dieser verfügt über eine API, mit welcher es möglich ist Netzwerktopologien (für Angriffe und normalen Netzwerkverkehr) zu erstellen. Darüber hinaus können Container und virtuelle Maschinen in die Simulation integriert werden, wodurch die Weiterleitung des Netzwerkverkehrs an einen Recorder (zum Aufzeichnen des Verkehrs) und einen Extractor (zum Extrahieren relevanter Features) bewerkstelligt werden kann. Um den Netzwerksimulator in einer virtuellen und skalierbaren Umgebung zu betreiben wird Vagrant verwendet. Für das Deployment kommen Ansible Playbooks zum Einsatz, welche zum Definieren des Host-Setups und der verschiedenen Netzwerktopologien dienen. Der Simulations-Workflow beginnt somit bei der Eingabe des gewünschten Netzwerkszenarios. Daraufhin erfolgt das Deployment (von Host-System, Netzwerktopologie, Extractor, Recorder). Die extrahierten Features werden letztlich in einer zentralen Datenbank abgespeichert.

Konkret wird Folgendes durchgeführt:

- Aufbau mehrerer GNS3 Server mittels Ansible Playbooks
- Bereitstellung von GNS3 Netzwerken mittels dem gns3fy Ansible Modul über die GNS3 API. Ein GNS3 Netzwerk beinhaltet eine Netzwerktopologie (Angriff und normaler Netzwerkverkehr), sowie Recorder und Extractor.
- Verschiedene Szenarien können bereitgestellt und getestet werden mittels Ansible Playbooks (samt Parameter zum Extrahieren von Features).
- Features werden erzeugt und in einer Datenbank abgespeichert (genaue Position der Datenbank wird noch definiert).

Abbildung 1 zeigt eine grafische Darstellung des technischen Konzepts:

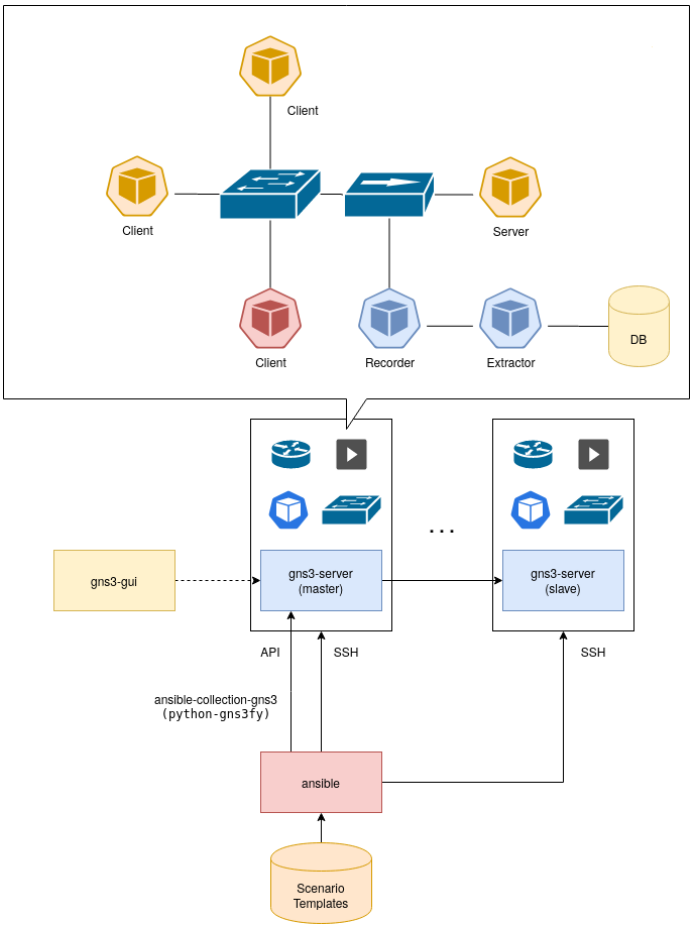

Abbildung 1: Technisches Konzept

Aufgabenverteilung

Aktuell wurden folgende, grundlegende Aufgabenbereiche identifiziert und die jeweilige Verantwortung festgelegt:

- **Netzwerktopologie 1** (Bernhard Gally**)**
  - Angriff: SSH Brute Force + Erzeugung eines normalen Netzwerkverkehrs
  - Aufbereiten der ersten Netzwerktoplogie in GNS3 via API (ansible-collection-gns3) auf dem eigenen Gerät
  - Verwendung von Docker Container für Client und Server Maschinen
- **Recorder** (Damir Marijanovic)
  - Erzeugung eines Docker Images für den Recorder
  - Testen des Containers in GNS3 auf dem eigenen Gerät
- **Extractor** (Boris Stampf)
  - Erzeugung eines Docker Images für den Extractor
  - Erstellen eines Datenbank-Designs
  - Testen des Containers in GNS3 auf dem eigenen Gerät

- **Infrastruktur** (Sebastian Lipp)
    - Unterstützung beim Erstellen von Docker Images
    - Unterstützung bei der Verwendung von Ansible/ansible-collection-gns3
    - Unterstützung bei der Abspeicherung von Daten / Datenbank-Integration
    - (Unterstützung bei der Erzeugung eines normalen Netzwerkverkehrs)
    - Kombinieren von Netzwerktopologie, Recorder und Extractor in einem Ansible Playbook
    - Erstellen eines Ansible Playbooks für das Aufsetzen des GNS3 Servers
- **Projektmanagement** (Maximilian Wech)
    - Erstellung der Projektplanung
    - Projekthandbuch
    - Projektkoordination und Projektcontrolling
- **Unterstützung technische Teile** (Maximilian Wech)
- **Netzwerktopologie 2-5** (noch nicht aktuell)
- **Testen des neuronalen Netzes** (noch nicht aktuell)

Im Laufe des Projektes kann es je nach Bedarf zu Veränderungen bei der Aufgabenverteilung kommen.

Projektstrukturplan

Es folgt eine objektorientierte Darstellung des Projektstrukturplanes. Diese eignet sich für die vollständige Auflistung aller im Projekt relevanten Arbeitspakete und der vergleichsweise einfachen Zuteilung von Verantwortlichkeiten.

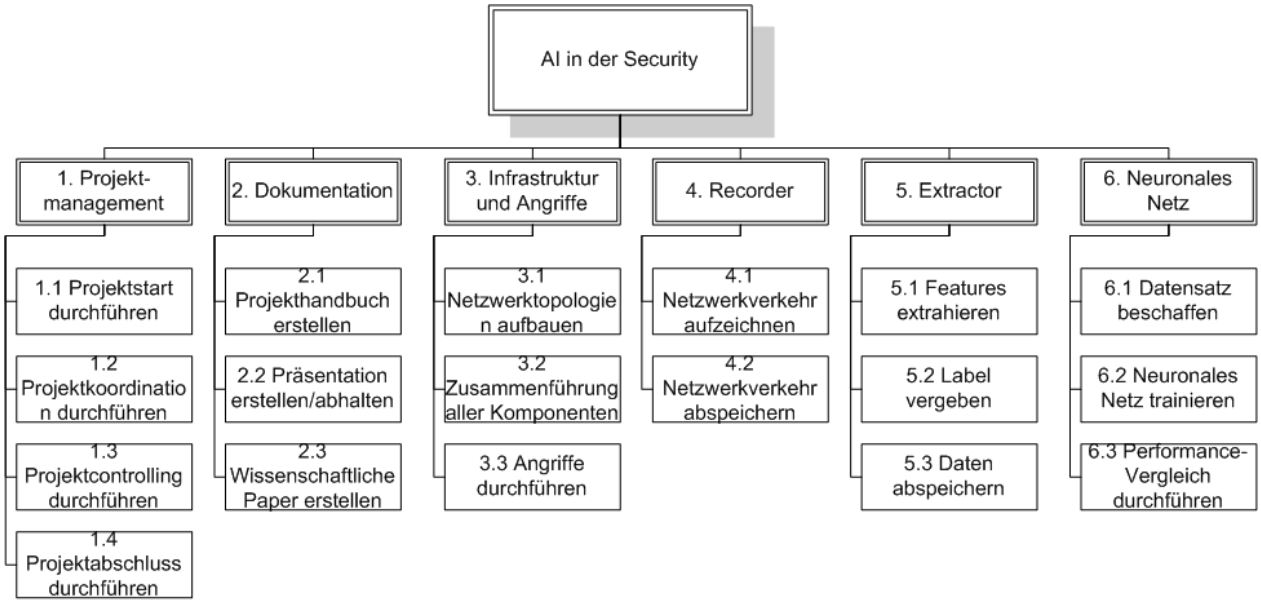

Abbildung 2: Projektstrukturplan (objektorientiert)

Der Projektmanagementblock trägt insgesamt dazu bei, einen zeitgerechten und erfolgreichen Projektabschluss zu bewerkstelligen. Es soll der laufend der Fortschritt gemessen und bei Bedarf steuernd eingegriffen werden.

Die Dokumentation besteht aus mehreren Komponenten. Es soll ein ausführliches und vollständiges Projekthandbuch erstellt werden, welches eine detaillierte Projektplanung enthält. Die Ergebnisse des Projekts werden im Rahmen einer Präsentation vorgeführt und in zwei wissenschaftlichen Papers erläutert.

Der Infrastruktur- und Angriffsteil wird mit einem Netzwerksimulator simuliert und beinhaltet den Aufbau einer entsprechenden Netzwerktopologie, bei welcher mehrere Komponenten (wie Clients, Server, etc.) zusammengeschalten werden. Dabei erfolgt eine Durchführung verschiedener Angriffsarten (z.B. SSH Brute Force).

Die nächsten zwei Bereiche beschäftigen sich mit dem Aufzeichnen des Netzwerkverkehrs (mittels tcpdump), der Extrahierung relevanter Features (mittels CIC-Flowmeter) und der Vergabe von Labels. Die extrahierten Daten werden in einer Datenbank abgespeichert.

Der letzte Teil inkludiert die Thematik des neuronalen Netzes. Dabei soll ein Performance-Vergleich mit dem im Vorprojekt erzeugten Modell durchgeführt und die Eignung unterschiedlicher Datensätze geprüft werden.

Meilenstein-/Ablaufplan

Es folgt eine tabellarische Darstellung bedeutender Termine samt Beschreibung, welche Tätigkeiten bis zu diesen Zeitpunkten fertiggestellt sein müssen. Daran ist auch der Projektablauf zu erkennen, welcher eingehalten werden muss, um einen zeitgerechten Projektabschluss zu gewährleisten.

Daten	Meilenstein (und Beschreibung)
1.5.2020	Proof of Concept erstellt ● Netzwerktopologie 1 umgesetzt ● Eine Angriffsart wurde durchgeführt ● Anfallender Netzwerkverkehr aufgezeichnet ● Entsprechende Features extrahiert ● Labels vergeben
1.6.2020	Netzwerktopologien 2-5 umgesetzt ● Die gleichen Schritte wie oben durchgeführt, jeweils für die unterschiedlichen Szenarien
9.6.2020	Neuronales Netz – Performance Vergleich durchgeführt ● Erzeugter Datensatz wird mit den neuronalen Netz aus dem Vorprojekt getestet ● Vergleich der Performance
16.6.2020	Präsentation durchgeführt ● Präsentation wesentlicher Projektergebnisse und der Vorgangsweise durchgeführt
30.6.2020	Projektabnahme durchgeführt ● Wissenschaftliche Paper erstellt ● Wesentliche Projektmaterialien (Source Code, Datensatz, etc.) und Projekthandbuch an den Projektauftraggeber übergeben

Tabelle 1: Meilensteinplan mit Beschreibungen