Aufbau eines ML-basierten Intrusion Detection Systems (IDS)

Erkennen von Angriffen in HTTP-Anfragen mittels maschinellem Lernen

Diplomanden

Christoph Landolt

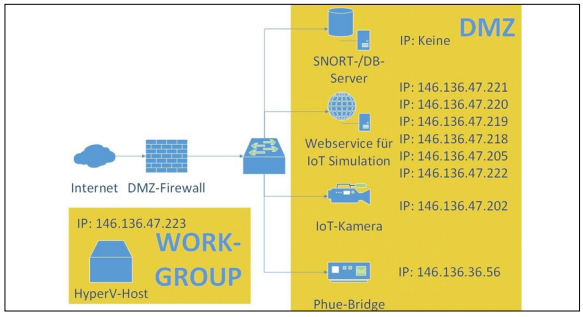
Moritz Bättig

Ausgangslage: Das Erkennen von Attacken auf HTTP-fähige Systeme stellt für IT-Sicherheitsverantwortliche eine erhebliche Schwierigkeit dar, da in der Praxis jeder Zugriff mit einer Liste von statischen Regeln abgeglichen werden muss, um einen Angriff zu erkennen. Damit diese regelbasierten Intrusion Detection Systeme zuverlässig funktionieren, muss die Liste mit statischen Regeln in immer kürzeren Zeitintervallen auf neue bekannte Angriffe und Angriffsmuster aktualisiert werden.
Es gibt bereits auf dem Markt erhältliche Systeme, welche maschinelles Lernen zur Erkennung von Angriffen einsetzen. Diese Systeme benötigen aber eine grosse Menge an Trainingsdaten, sehr viel Rechenleistung und eine Anbindung an das Security Operations Center des Herstellers, um die eingegangenen Daten in Angriff und Nicht-Angriff einzuteilen.

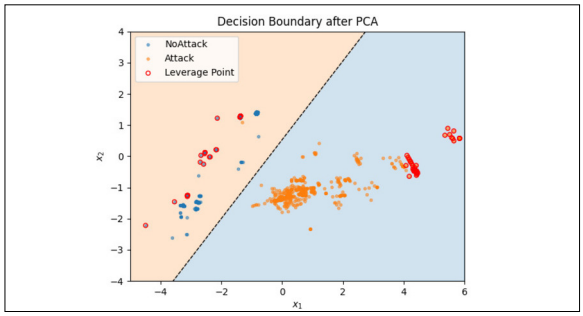
Vorgehen: In der vorliegenden Arbeit wird untersucht, wie HTTP-Zugriffe quantifiziert werden können, um mit begrenzter Rechenleistung ein Machine Learning Modell zu trainieren, um Angriffe über das HTTP-Protokoll effizient erkennen zu können. Dazu wurde eine Software entwickelt, mit welcher HTTP-Zugriffe aufgezeichnet, analysiert und an das Zielsystem weitergeleitet werden können. Damit die Algorithmen neben akademischen Datensätzen ebenfalls mit realen Daten getestet werden können, wurde ein Honeypot aufgebaut, um möglichst viele Angriffe aufzuzeichnen.

Ergebnis: Das in der Arbeit vorgestellte Modell konnte in den realen Daten eine Korrektklassifikationsrate von über 96.6% und in den öffentlichen Datensätzen, je nach Datensatz, über 99.7% erreichen. Die Arbeit zeigt somit, dass eine zuverlässige Intrusion Detection mittels maschinellen Lernens möglich ist, und bildet dank der modularen Softwarestruktur eine gute Grundlage für weitere Optimierungen und akademische Untersuchungen im Bereich der Merkmalsextraktion in HTTP-Anfragen und des maschinellen Lernens zur Unterstützung der Intrusion Detection.

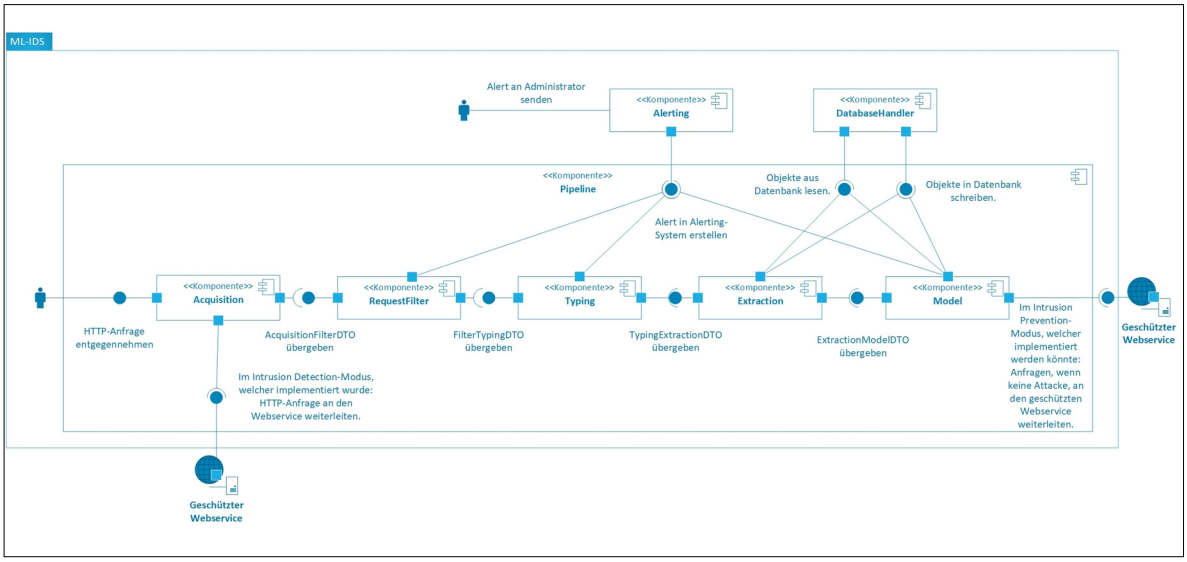
Aufbau des Netzwerks mit Honeypot
Eigene Darstellung



Logistische Regression und Decision Boundary
Eigene Darstellung



Komponentendiagramm des aufgebauten Systems
Eigene Darstellung



Referent
Prof. René Pawlitzek

Korreferent
Prof. Dr. Klaus Frick

Themengebiet
Ingenieurinformatik, Informations- und Kommunikationssysteme, Computational Engineering

Projektpartner
Noser Engineering AG