|  |  |
| --- | --- |
| Technische Hochschule Wildau  FB Wirtschaft, Informatik, Recht  Prof. Dr. – Ing. Stephan Rein |  |

Bachelor oder Master Arbeit

**KI-Modellbildung mit Web-Crawler für die Analyse von Security-Problemen**

**Beschreibung**

Bei der Administration von IT-Systemen ist die Fehlerbehebung eine der schwierigsten Aufgaben. Ein typischer Ansatz zur Behebung von schwierigen Fehlern ist es, zunächst die betreffende Log-Meldung zu analysieren. Da die Systeme immer spezieller und komplexer werden, kann die Meldung allerdings schwer interpretierbar sein. Viele Administratoren bemühen deswegen eine Internet-Recherche, um weitere Hinweise zu erhalten. Dabei ist es häufig notwendig, in entsprechenden Foren zu recherchieren, was aufgrund der Vielzahl und teilweise auch nicht weiterhelfender Beiträge zeitaufwendig und mühsam ist. (Ein Beispiel für eine nicht zielführende Antwort auf Foren, die mehrere unnötige Beiträge verursachen kann, ist die die folgende: „Warum verwenden Sie überhaupt dieses System, benötigen Sie das wirklich …?“ ☺ ).

In dieser Arbeit soll mittels eines Web-Crawlers ein KI-Modell gebildet werden, welches nützliche Hinweise zu ausgewählten IT-Problemen, insbesondere mögliche Security-Probleme, anhand der Log-Meldungen liefern kann. Als Crawler soll die Software *Scrapy* ( <https://scrapy.org/> ) zum Einsatz kommen. Als Log-Meldungen sollen beispielhaft ausgesuchte Meldungen von dem Logging-Tool *Falco* ( <https://falco.org> ) verwendet werden. Hierzu wurde am Fachgebiet bereits eine Vorarbeit durchgeführt (siehe Literaturteil). Die Logging-Meldungen sind System-Call-Meldungen, die von Containern geliefert werden.

Für die Modellbildung und die anschließende Verwendung zur Lösung von Problemen sollen Open-Source-Tools auf Basis von Python verwendet werden, u.a. bietet sich hierfür PyTorch (<https://pytorch.org> ) an. Das Tool zur Modellbildung erhält die vom Web-Crawler gefundenen Daten über Files, die bei der Web-Suche erstellt werden. Das Testen des Modells kann dann später über Kommandozeilen-Eingaben und Skripten erfolgen.

Teil der Arbeit ist die Evaluation des breitzustellenden Systems. Hierbei sind mehrere Forschungsfragen zum Thema „Modellbildung“ zu beantworten:

* Untersuchung der benötigten Zeit für die Bildung der verschieden großen Modelle
* Bewertung der Nützlichkeit der verschiedenen Modelle
* Aussagen dazu, ob sich die Ergebnisse sich auch für größere Modelle verwenden lassen (Skalierbarkeit)

**Aufgaben**

* Literatur-Recherche
* Lösung oder Ansatz entwickeln
* Lösung realisieren und evaluieren
* Arbeit schreiben / dokumentieren

**Voraussetzungen**

* Gute Linux-Kenntnisse
* Python- oder C-Kenntnisse
* Idealerweise Besuch der Veranstaltungen „IT-Administration I-III“ , „Cloud-Technologien“ oder „Cloud-Administration“ an der TH.

**Literatur**

* **Constantin Moye**, Praktikabilität von Standard-Security-Regelsätzen für Container-Systemcall-Logging, Bachelorarbeit TH Wildau, Oktober 2023
* **Luca Guerry and Nicolas Lacasse**, Strengthen K8s & Container Security without Loosing Visibility, Cloud Native Computing Foundation Webinar, Sept 2022
* **Federico Lago**, Building a SIEM: centralized logging of all Linux commands with ELK + auditd, Infosec Writeups, August 2020, ([link](https://infosecwriteups.com/building-a-siem-centralized-logging-of-all-linux-commands-with-elk-auditd-3f2e70503933))
* **Thanh Nguyen** et al., Live system call trace reconstruction on Linux, Elsevier Journal on digital/multimedia forensic science and evidence-based incident response, Vol. 42, Juli 2022
* **Maya Kaczorowski & Ann Wallace** , Container Forensics: What to Do When Your Cluster is a Cluster, Vortrag Cloud Native Computing Foundation (CNCF) 2019

<https://www.youtube.com/watch?v=MyXROAqO7YI>

**Kontakt:** stephan.rein@th-wildau.de

<https://th-wildau.de/stephan-rein>