Proposal for a master thesis WS2022/23:

Web-server log anomaly detection

using Self-Supervised Learning (SSL) with NLP approach

1 Background

The web-server is the front door that provides services that expose organizations to the outside world, thus making them an attractive target. Attackers could manipulate web requests in order to exploit possible vulnerabilities within web-servers (e.g., XSS, SQL injections, etc.). Nowadays the majority of web-servers' logs contain the details of their interaction with a particular user. Analyzing these logs could allow us to detect possible exploitation. However, Web-server log analysis could be challenging due to the lack of labels, making the classical machine learning algorithms inefficient, particularly for big data. Therefore, enhancing the existing cybersecurity detections using a novel Self-Supervised Learning (SSL) model with the right annotation approach can improve weblog analysis in the absence of labels.

2 Objetives

The goal of this master thesis is to explore NLP-based augmentation along with Self-Supervised Learning (SSL) for the analysis of Web/App-server logs (i.e., HTTP-logs) in the context of Web-server Log Anomaly Detection (WLAD) in order to find malicious attempts. Follow the below guidelines:

- Analyze the current state-of-the-art concerning log analytics for WLAD applications to identify webserver attacks.
- Describe the underlying NLP augmentation technique that fit log-based data for developing (automated) annotation.
- Develop and implement the best NLP-based approach along with SSL for Web-server log analytics and describe its process.
- Perform tests on large Web-server logs and evaluate the results generated by the used method.

3 Deliverables

- Master Thesis
- Running prototype
- Scientific publications on international conferences/journals **(Expected)**

4 Requirements

- M.Sc. Programs: Cybersecurity, IT Systems Eng., or Data Eng.
- **(Expected)** knowledge and experiences/skills on:

  - Network/System/Application security, IT/Security operations

  - Web-server logs, (Big) Data science and engineering, Regex patterns, Log Parsing (Templatization), NLP .

Further details about master thesis available via the slides at: www.sec.hpi.de.

Contact

Internet Technologies and Systems chair
Prof. Dr. Christoph Meinel
Sec-Eng@HPI: Dr.Cheng, Mehryar Majd

Email I	mehryar.majd@hpi.de
Email II	security-analytics@hpi.uni-potsdam.de
Website	www.sec.hpi.de
Phone	+49 (0331) 5509 587
Twitter	twitter.com/HPI_DE

Address

Hasso Plattner Institute
Digital Engineering Department,
University of Potsdam
Security Engineering (Sec-Eng) Team
Campus III (Griebnitzsee) Building H
Prof.-Dr.-Helmert-Straße 2 - 3
14482 Potsdam — Germany