Contribution ID: **29**

Type: **Poster**

Evaluating unsupervised data mining methods to assess the utility of such approaches for SIEM event analysis.

Thursday 24 April 2025 10:45 (30 minutes)

Modern organizations generate vast amounts of data, a significant portion of which consists of system, network, and application logs. The sheer volume and scalability of these logs make manual analysis inefficient and highly time-consuming. Automated anomaly detection techniques have been developed and refined to accelerate this process. Currently it is common for open SIEM systems to only support basic anomaly detection features. In our paper, we focus on unsupervised data mining methods for anomaly detection in time-series JSON data. This study compares and explores the utility of several anomaly detection algorithms applied to preprocessed data. The preprocessing stage includes time-bucket segmentation and feature extraction. We have also examined different approaches to preprocessing that are mentioned in the literature. Among the various methods, Isolation Forest and probabilistic algorithms like ECOD, have been proven to perform well on multidimensional semi-structured text datasets. Furthermore, we assessed the performance of selected methods using relevant Key Performance Indicators and compared them in different scenarios. Our findings suggest that some of these methods could be adapted to effectively support security analysts working with SIEM systems. Certain concepts developed within this proof-of-concept study could be further refined and integrated into a dedicated tool for log analysis.

Author: DYDO, Marcin (Warsaw University of Technology)

Presenter: DYDO, Marcin (Warsaw University of Technology)

Session Classification: Session B (Poster)