Centralized Log Analysis for Information Security using Machine Learning

- Project Charter

Introduction

- In our current security architecture, various third-party appliances are independently responsible for securing different domains, such as network, email, and endpoint security.
- These appliances collect logs and perform analytics in isolation, and only the analysis is forwarded after.
- However, this approach limits the effectiveness of our security posture, as correlations between logs from different domains remain unexamined.
- This report proposes the development of a centralized machine learning-based solution that collects logs from all domains, allowing for a holistic analysis that provides a more comprehensive overview of the organization's security architecture.

Current Prototype

- A prototype model currently exists, that employs Autoencoders to identify anomalies in email logs.
- By analyzing patterns in email traffic, the model is able to detect unusual behaviors that may indicate security threats.
- For a more detailed explanation of this approach, refer to the Email_Logs_Anomaly_Detection-Report.pdf documentation.

Key Takeaways for Future Development

1. **Expand the Centralized Model**: The future model should also work with, and incorporate, network and endpoint (antivirus) logs (and associated analysis). Given that logs collected from different domains will have varying formats, there is also a need to scale and standardize/normalize all logs to a single type before feeding them into the model. This is essential for enabling cross-domain analysis and correlation.

2. **Real-Time Analysis**: The current model solution analyzes logs in batches, iteratively and manually. To enhance practical implementation, it is necessary to translate the solution to operate and analyze logs in real-time.

3. **Action Items**: Implementing the capability to set up action items in the solution, similar to the ‘models’ feature in DarkTrace, will allow for specific pre-defined triggers, each with its associated response. This automation of certain taskflows will facilitate easier usage and operation, also providing immediate reactions to particular critical trigger events (of choosing).

4. **NLP-Enabled Chatbot**: Another possible enhancement to the solution is the integration of a chatbot with Natural Language Processing (NLP) capabilities, thereby enabling easier and more intuitive interaction with the security system.