

Exploring the Use of Artificial Intelligence to improve Security Operations Centre Activities

Purpose

This document outlines a research program to be carried out as part of the Career Fit Plus Marie Curie Cofund. The research is carried out by Athlone Institute of Technology (AIT) in Ireland in conjunction with Hewlett Packard Enterprise (HPE). Career Fit Plus background and application details can be found at <https://www.horizon2020.ie/career-fit-plus/>

Contact Brian Lee at AIT (blee@ait.ie) for more information

Aim

This research program will investigate a number of targeted topics whose aim is to increase the efficiency of the SOC team through the use of artificial intelligence techniques. This research will be conducted in collaboration with the Security Operation Centres (SOC) of Hewlett Packard Enterprise (HPE) in Galway, Ireland. HPE is one of the world's leading information technology companies with a wide suite of IT solutions. The SOC team in Galway operates at the heart of HPE's security defense system

It is a widely known fact that teams in Security Operation Centres (SOC) and SecOps teams struggle to deal with the vast volumes of information that are produced by all the various security sensors across an enterprise. For a large company this can run to billions of events per week. Sifting through even greatly aggregated data to find true threats is made even more difficult by the large number of false positives that are produced even by the best of security management suite of tools. Dealing with these false readings not only delays finding threats but also reduces the morale of the security team. Nor is it just immediate interaction with the data that poses challenges- data must be retained for weeks or even months in order to provide a back-view on possible latent or well concealed attacks. This poses challenges for the retention, storage and subsequent processing of that data for activities such as threat hunting.

Machine learning/AI techniques have been widely used for a long time in cyber security. The vast majority of such research has been applied to intrusion detection. More recently AI research is being directed at improving the efficiency of SOC team operations e.g. [1], as part of Security Orchestration Automation and Response (SOAR) activities. Other suggested area where AI could be included in the SOC process include, [2]

- Abstracting lessons from individual incidents, generalizing them across systems and networks, and applying those lessons to increase attack and defense effectiveness elsewhere.
- Identifying strategic and tactical trends from large datasets and using those trends to adapt attack and defense tactics.
- Using natural language sentiment analysis to automate security processes,[3]

SOC activities span a wide range of roles and responsibilities with many possibilities for AI.

Research Objectives

This proposal has the following objectives

TO1 – To investigate the use of Interactive Machine Learning (IML) to improve analyst workloads and efficiency.

Interactive Machine Learning (IML) is an iterative learning process that tightly couples a human with a machine learning model. The technology is widely used by researchers and practitioners to effectively solve a wide variety of real-world application problems [4]. IML enables machine learning models to be interactively steered by humans and is more advantageous for the tasks where human knowledge is needed in the analysis process. Researchers have recently begun evaluating the use of IML techniques to improve SOC teams performance. Sopan [5] reports on the use of IML to help analysts classify oncoming alerts more efficiently and gain insight into how a machine learning model generates predictions. Arnaldo, [6], describes an IML system to improve threat hunting based on interactive anomaly detection and notes threat hunting “remains vastly unexplored in the research community” and entails open challenges in combining the fields of outlier analysis, explainable machine learning, and recommendation systems.

This objective will develop an IML system to improve the threat hunting process within HPE. Scenarios will be developed through consultation with the threat-hunting analysts and a functional systems will be researched, designed and prototyped. The prototype will be evaluated in-situ with the team.

TO2- To explore existing datasets to uncover new data relationships to improve threat detection and response.

HPE stores and aggregates data at different granularities over hourly, daily, weekly and monthly periods. These datasets contain a rich vein of data whose potential for threat detection and evasion is not yet fully explored. This objective will work with analysts to identify potential detection scenarios to mine from this data and will then develop the algorithms and tools sets to evaluate the scenario. While the specific scenario will be identified as part of the work one promising novel scenario may be to correlate network event data with host event data to identify multi-step attacks. This will require combining data from diverse data sets and developing algorithms and code to identify new correlations.

TO3- To explore the use of sentiment analysis to improve the security automation.

Interaction between different SOC teams drives the cyber defence. Each team has different responsibilities and deal with information of different forms, granularities and time scales. Hence much of the information received and analysed by the threat intelligence analysts may be in the form of unstructured text and natural language. Very often this concerns the tactics, techniques and procedures of different adversaries and may point to new indicators of compromise (IoC) or other information that can be used by the security event analysts. The useful lifetime of such information may change quickly e.g. botnet C&C servers however and it may not always be update such changes in an optimal time resulting in spurious alerts. This objective will apply NLP deep learning to analyse incoming textual information to see if any new information exists that can be presented to the event triage team and conversely can also note the absence of certain information that may point to a need to update firewall or IDS rules.

TO4- To investigate the use of AI to optimise event storage

This objective will look for ways to reduce the amount of information that needs to be stored thus reducing the amount of storage needed but also the amount of information that needs to be manipulated to derive meaningful results. Security event information is very heterogeneous and a large number of event/data relationships may exist. It will use AI learning to correlate which types of information and which parts i.e. attributes or fields of events and alerts is most relevant for different tasks. The task may be carried out as an extension of TO1 to have analysts provide input to the learning system.

Outline Work Plan

WP1 – Requirements and State of Art

[M1-M3] Athlone IT (AIT)- Induction; Literature review;

[M4-M6] HPE Galway – Induction; Familiarisation with SOC; Scenario definition

WP2 – IML Development

[M7-M15] AIT - Development of an IML systems (TO1 and TO4)

[M15-20] HPE - Deployment and evaluation of design; Scenario development

WP3 – Data Analysis

[M21-M31] AIT – Development of Data Analysis algorithms (TO2, TO3)

[M32-M34] HPE - Deployment and evaluation of design

WP4- Dissemination

[M12-36] AIT – Publication of results.

Works Cited

- [1] S. Barends, "Security Automation Soars to the Top of the Agenda," [Online]. Available: <https://www.infosecurity-magazine.com/opinions/security-automation-soars/>.
- [2] B. Schneier, "Artificial Intelligence and the Attack/Defense Balance," March 2018. [Online]. Available: https://www.schneier.com/essays/archives/2018/03/artificial_intelligence.html.
- [3] R. V. e. al., "TwitterOSINT: Automated Cybersecurity Threat Intelligence Collection and Analysis using Twitter Data Feeds," 2018. [Online]. Available: <https://csce.ucmss.com/cr/books/2018/LFS/CSREA2018/SAM9750.pdf>.
- [4] L. J. e. al., "Recent Research Advances on Interactive Machine Learning," *Journal of Visualisation*, vol. 22, no. 2, 2019.
- [5] A. S. e. al., "Building a Machine Learning Model for the SOC, by the Input from the SOC, and Analyzing it for the SOC," in *IEEE SYMPOSIUM ON VISUALIZATION FOR CYBER SECURITY*, 2018.
- [6] e. a. I. Arnaldo, "eX2: a framework for interactive anomaly detection," in *IUI Workshops*, L.A., 2019.

