Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 29, 2021
A Regional Conference of the Society for Industrial and Systems Engineering

Command and Control Anomalies: A Framework for Identifying Potential Malware

Katherine Baumeister[1], Anastasiya Joyner[1], Charles Harrington[2], Brenden Shutt[2], and Steven Henderson[1]

[1] United States Military Academy
Department of Systems Engineering
West Point, NY

[2] United States Military Academy
Department of Mathematical Sciences
West Point, NY

Corresponding author's Email: charles.harrington@westpoint.edu

Author Note: Cadets Baumeister and Joyner are 4th year students in the Department of Systems Engineering at the United States Military Academy. Cadets Harrington and Shutt are 4th year students in the Department of Mathematical Sciences at the United States Military Academy. They are participating in a year-long capstone design course under the supervision of Dr. Steven Henderson, Associate Professor in the Department of Systems Engineering, the Capstone group's advisor. The client for this project is Army Cyber Command, with the main point of contact being LTC Gregory Bew. The Capstone team would like to thank LTC Bew, his team, and Dr. Henderson for their support and guidance throughout the project.

Abstract: An enduring threat to large-scale computer networks is malware. Cybersecurity analysts are tasked with identifying malicious actors on a network; however, high-volume networks pose a steep computational challenge in accomplishing this task. Our research answers the question: how can the search for potentially infected machines on a network be expedited? Our work proposes an analytic framework, titled the Command and Control Anomaly Algorithm (C2A2), that leverages unsupervised machine learning techniques to identify potentially infected machines on a network by analyzing the characteristics of machine-specific dataflows. C2A2 incorporates a backend functionality that assigns standardized outlier scores to machines on a network and classifies a machine's dataflow as anomalous with some degree of confidence. C2A2 then communicates the findings in a comprehensive anomaly report to a user-interface wherein cybersecurity analysts can conduct further investigations on suspicious network machines. Within this framework, analytic-specific data storage and transfers are facilitated by an application programming interface (API). Our analytic intends to reduce the time required for a cybersecurity analyst to find potentially infected network machines by highlighting machines whose dataflows fit the profile of an infected machine. Preliminary tests determine that C2A2 is effective in identifying network machines with anomalous dataflows and can meet the scale of operations expected in its use case.

Keywords: Big Data, Cybersecurity, Anomaly, Unsupervised Learning