Project Milestone 1

By Danny Tandet, Hudson Wesel, Jeffery Andreski

Introduction:

According to Varonis (Sobers), a large-cap software cybersecurity company, cited from IBM, the average time to identify a breach is 194 days. Moreover, according to Forbes (Mariah St. John, “Cybersecurity Stats: Facts and Figures You Should Know”), In 2023, security breaches saw a 72 percent increase from 2021. Given both this delayed response time and the upward trend of cyber attacks, it’s vital to formulate ethical cybersecurity solutions. This project aims to be a preliminary proof of concept to address this problem. The resolution leverages a machine-learning-powered agent that will serve as part of a greater intrusion detection system. This agent will accept parsed network-traffic data in real time, and attempt to identify anomalous and malicious activity on the network. Given this project’s scope, our team is focused on detecting Distributed Denial of Service (DDoS) attacks, reconnaissance activities, and command and control (C2) communications. Traditional threat detection methods, which depend on User behavior analytics (UBA) are used to identify abnormal behavior - possibly indicating an unknown threat - across your network. In contrast, our machine-learning model will identify what is happening at the network layer. Given this more intelligent alert system, taking in real-time data, it will more rapidly detect and alert security operation center analysts to emerging attack patterns. While out of scope for this project, given enough data and tuning, our machine-learning model would be able to identify new tactics, techniques, and procedures, by which attackers can gain access to a corporate network. Doing this at an enterprise level is also beyond the scope of a singular project, especially when factoring in scaled-up noise. This is fundamentally a binary classification problem, at least preliminarily. Simply feeding network traffic logs into the model, the model will identify whether the traffic is malicious or not. From there, it becomes a multi-class classification problem, to identify which, of the three types of malicious in this case plus benign traffic data, any given section of the log. Using this knowledge, we will be creating models to identify unusual patterns and potential threats in real time. Support vector machines (SVM), decision trees, and random forests are the most used methods in classification tasks. This will involve applying supervised learning techniques. Our ultimate goal is to build a system that not only provides detailed reports on potential threats but also sends real-time alerts to help fend off network attacks proactively.

Related work:

Research in the realm of network traffic analysis and the use of AI and machine learning for identifying security threats has gained significant attention in recent years. Several approaches have been developed to improve the detection of anomalies in network traffic. The following is a review of existing studies categorized by their approach and methodology

1. **Machine-Learning Based Approaches:**
According to ("Real_Time_Network_Traffic_Analysis_Using_Artificial_Intelligence_Machine_earning_and_Deep_Learning_A_Review_of_Methods_Tools_and_Applications") Machine Learning (ML) methods are frequently used to automate network traffic analysis. The paper then discusses the problems for which this machine-learning approach solves and its application. While there are various methods to solving these problems including the following: support vector machines, decision trees, and random forests, each method is used for a specific application. For instance, are frequently used for classification tasks, whereas decision trees are frequently utilized for anomaly detection jobs. Moreover, machine learning algorithms have the potential to be more precise than conventional rule-based techniques, and can be scaled for high volumes of traffic data.
2. **AI in Network Traffic and IP Routing**:
This article explores how AI revolutionizes traditional network traffic analysis and IP routing, which allows computer network professionals to address more sophisticated problems. Network routing decisions have traditionally relied on static, rule-based systems via a routing table. While these methods have been adequate to an extent, they need more flexibility and scalability to manage modern, dynamic networks. AI can significantly enhance network performance by predicting traffic patterns and automatically adjusting routing to avoid congestion. In network traffic analysis, AI enables real-time data analysis and anomaly detection.
3. **AI-Driven Intrusion Detection Systems (IDS)**:
According to (Rao et al., "as the complexity and sophistication of cyber threats continue to evolve, traditional methods of network anomaly detection fail to identify novel and subtle attacks. In response to this challenge, authors propose a novel approach to network anomaly detection utilizing a Hybrid Convolutional Neural Network (CNN) and Generative Adversarial Network (GAN)." The CNN component extracts features, and data about any given packet, from network traffic data, which it then uses to conjecture complex patterns and relationships within the data. Simultaneously, the GAN component acts as a generator and discriminator, learning to generate normal network traffic patterns and distinguishing anomalies from them. To train the hybrid model, employing a large dataset of labeled network traffic, encompassing both normal and anomalous behavior.

4. **AI Models for Network Attack Detection**:
The paper (Alqudah and Yaseen) discusses different machine-learning approaches for traffic analysis. Machine learning (ML) shows effective capabilities in solving network problems. A review of the techniques used in the traffic analysis is presented in this paper. The paper categorizes these techniques into the following steps: identifying classes of training data, creating model and training model, testing the model with unknown data, results are evaluated for accuracy, and if these results are not satisfactory, the model must be tuned. It also outlines a subset of common ML approaches, that of supervised and unsupervised learning.

5. State-of-the-Art Algorithms for Traffic Analysis:
According to (Nguyen et al.), the study focuses on how Machine Learning methodologies are leveraged to create an advanced network traffic classification system. However, the study is more focused on optimizing the performance of a network, via proper quality of service (QoS) rather than in a cybersecurity context. Their method organizes similar kinds of network traffic into distinct categories based on latency requirements. Furthermore, it decomposes the network traffic stream into multiple, smaller traffic flows, with each flow uniquely carrying a specific service. They trained their model on labeled examples representing different network service types. This approach resulted in more efficient energy consumption, enhances Quality of Service assurance, and optimized= the allocation of network resources. This gives our team insight into how we should construct our own model.

Works Cited

Alqudah, Nour, and Qussai Yaseen. "Machine Learning for Traffic Analysis: A Review." *Procedia Computer Science*, vol. 170, 2020, pp. 911–916, https://doi.org/10.1016/j.procs.2020.03.111.

Kumar, Purnendra, et al. "Network Traffic Analysis and Prediction Using Machine Learning." *International Journal of Research Publication and Reviews Journal Homepage: Www.ijrpr.com*, vol. 4, 2071, ijrpr.com/uploads/V4ISSUE8/IJRPR16324.pdf.

Nguyen, Khuong N., et al. "Towards Intelligent Network Management: Leveraging AI for Network Service Detection." *Google.com*, 2024, www.google.com/url?q=arxiv.org/pdf/2310.09609&sa=D&source=docs&ust=1727123403968012&usg=AOvVaw2zcgcdGsH2iEj48av9GHkO. Accessed 23 Sept. 2024.

OpenAI. “ChatGPT.” *Chat.openai.com*, OpenAI, 2024, chat.openai.com.

Rao, Vuda Sreenivasa, et al. “AI Driven Anomaly Detection in Network Traffic Using Hybrid CNN-GAN.” *Journal of Advances in Information Technology*, vol. 15, no. 7, 2024, pp. 886–895, www.jait.us/articles/2024/JAIT-V15N7-886.pdf, https://doi.org/10.12720/jait.15.7.886-895. Accessed 23 Sept. 2024.

“Real_Time_Network_Traffic_Analysis_Using_Artificial_Intelligence_Machine_Learning_and_Deep_Learning_A_Review_of_Methods_Tools_and_Applications.” *Google.com*, 2024, www.google.com/url?q=www.researchgate.net/publication/376287072_Real_Time_Network_Traffic_Analysis_Using_Artificial_Intelligence_Machine_Learning_and_Deep_Learning_A_Review_of_Methods_Tools_and_Applications&sa=D&source=docs&ust=1727123253628269&usg=AOvVaw0gBFbuSFHFNMZdjX_Ki2oK. Accessed 23 Sept. 2024.

Singh, Shruti , and Aschin Dhakad. “Real Time Network Traffic Analysis Using Artificial Intelligence, Machine Learning and Deep Learning: A Review of Methods, Tools and Applications.” *ResearchGate*, 13 Oct. 2023, www.researchgate.net/publication/376287072_Real_Time_Network_Traffic_Analysis_Using_Artificial_Intelligence_Machine_Learning_and_Deep_Learning_A_Review_of_Methods_Tools_and_Applications.

Sobers, Rob. “166 Cybersecurity Statistics and Trends [Updated 2022].” *Www.varonis.com*, 14 Oct. 2024, www.varonis.com/blog/cybersecurity-statistics#data-breach-hacking.

“The Application of Artificial Intelligence in Network Traffic Analysis and Prediction.” *Advances in Computer, Signals and Systems*, vol. 8, no. 5, 2024, www.clausiuspress.com/assets/default/article/2024/08/17/article_1723910657.pdf, https://doi.org/10.23977/acss.2024.080509. Accessed 23 Sept. 2024.

"Understanding AI's Role in Network Traffic Analysis and IP Routing." *Www.iplocation.net*, www.iplocation.net/understanding-ais-role-in-network-traffic-analysis-and-ip-routing.

Xua , Shengnan, and Qianqian Wangb. "The Application of Artificial Intelligence in Network Traffic Analysis and Prediction." *Advances in Computer, Signals and Systems*, vol. 8, no. 5, 2024, www.clausiuspress.com/assets/default/article/2024/08/17/article_1723910657.pdf, https://doi.org/10.23977/acss.2024.080509. Accessed 23 Sept. 2024.