Capstone: Survey Existing Research and Reproduce Available Solutions

29.01.2025

Bruce Walker, UCSD MLE/AI Bootcamp

Research Papers Reviewed

1) Singh, Kuldeep, et al. “A Near Real-time IP Traffic Classification Using Machine Learning” I.J. Intelligent Systems and Applications, February 2013 (https://www.mecs-press.org/ijisa/ijisa-v5-n3/IJISA-V5-N3-9.pdf)
2) Alqudah, Nour and Yaseen, Qussai “Machine Learning for Traffic Analysis: A Review” *International Workshop on Data-Driven Security (DDS 2020)*, April 6-9, 2020. (https://doi.org/10.1016/j.procs.2020.03.111)
3) Dhakad, Aschin, et al. “Real Time Network Traffic Analysis using Artificial Intelligence, Machine Learning and Deep Learning: A Review of Methods, Tools and Applications” *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, October 2023 (https://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)

What I Learned

Network Traffic Analysis (NTA) is the process of examining data traveling across/through a computer network in order to detect anomalies that may indicate malicious use, performance problems, or other traffic that may be undesired. Using NTA to detect aberrant behavior is a complicated and resource-intensive task.

Singh, et al. Dhakad, et al. point out that traditional/legacy anomaly detection has been signature-based which means only patterns that have been previously identified and cataloged can be detected. This approach is of no use in detecting novel and emerging threats. Dhakad, et. al. further point out that the fast-paced, high-volume nature of computer network traffic data makes a signature-based detection scheme inefficient. An AI-powered, deep-learning approach is well-suited to the fast-paced, high-volume, and evolving threat needs of NTA.

Nour Alqudah, et al. highlight many of the same advantages and challenges of applying machine learning to NTA as Singh, et al. and Dhakad, et al. However, Nour Alqudah, et. al. additionally points out the differences between supervised and unsupervised approaches. The ever-changing nature of modern network traffic potentially makes an unsupervised approach better as it can potentially adapt to changing data without the time-intensive process of additional supervised learning cycles.

Existing Projects for Intrusion Detection using ML Network Traffic Analysis

1. XGBoost IoT Malicious Detection - https://www.kaggle.com/code/rem4000/xgboost-iot-malicious-detection-99-99-accuracy
2. malware dection on IoT - https://www.kaggle.com/code/jaimemoranchel/malware-dection-on-iot

My exploration of the existing projects

1. https://github.com/bdwalker1/UCSD_MLE_Bootcamp_Capstone/blob/master/PreviousWorkReview/XGBoost_IoT_Malicious_Detection_Bruce_Walker_Version.ipynb
2. https://github.com/bdwalker1/UCSD_MLE_Bootcamp_Capstone/blob/master/PreviousWorkReview/malware_detection_on_IoT_Bruce_Walker_Version.ipynb