Data Mining Techniques for Network Intrusion Detection Systems

Rashmi Shree Veeraiah, Reddysaketh Reddy Chappidi, Saran Parasa, Siddharth Solanki

Department of Applied Data Science, San Jose State University

DATA 240: Data Mining

Dr. Shayan Shams

February 24, 2024

Businesses and organizations confront increased difficulties protecting digital assets and sensitive data as cyberattacks continue to develop and get more complex. There is always a need for sophisticated Network Intrusion Detection Systems that can proactively detect and mitigate potential security breaches in this evolving threat landscape. By continuously monitoring network traffic, spotting suspicious patterns/ abnormalities, and facilitating quick responses to reduce risks and stop unauthorized access or data breaches, it is essential for strengthening an organization's defense measures.

The project aims to employ data-mining techniques to develop a system for identifying real-world network intrusions, prioritizing improvements in feature selection and the application of machine learning and deep learning models for precise threat detection, the aim is to improve network security and efficiency by evaluating effective methodologies. Through experiments with advanced approaches, emphasis on practical applications, and a review of existing literature, the objective is to identify the most effective means of detecting, categorizing, and alerting about network attacks.

Javaid et al. (2016b) suggested a deep learning-based Network Intrusion Detection System (NIDS) on the NSL-KDD dataset that employs Self-taught Learning (STL). STL, which combines Unsupervised Feature Learning with classification, overcame issues with feature selection and labeled dataset scarcity. In comparison to previous approaches, the NIDS outperformed them in terms of accuracy, precision, recall, etc. Javaid et al. (2016b) suggest researching more upgrades with approaches such as Stacked Autoencoder and real-time implementation for on-the-fly feature learning from raw network traffic headers. Overall, their research improves NIDS capabilities by demonstrating the effectiveness of deep learning in handling emerging network security threats. Jose et al. (2022) focused on improving network security by developing a predictive model that can differentiate between intrusions, attacks and good connections. They also discussed how data preprocessing, data cleaning, and visualization helped to achieve good efficiency and accuracy of models developed. Various metrics like precision, recall, accuracy, false positive rate, and specificity are used to determine their effectiveness in predicting network attacks. Future scope will be exploring advanced machine learning techniques, optimizing model performance, and integrating real-time monitoring capabilities for threat detection. Doriguzzi-Corin et al. (2023), introduced a P4DDLe framework designed to enhance NIDS using P4-based programmable data planes. Unlike traditional methods that compress network data, that leads to information loss and restricting access to essential details like packet payloads and categorical features, P4DDLe focuses on efficient packet-level feature extraction. P4DDLe extracts raw packet features, including categorical ones, and arranging them to preserve the semantic flow of traffic. P4DDLe effectively captures distinct and high-quality network flow patterns, resulting in reliable detection of DDoS attacks with a noticeably reduced system False Negative Rate. The evaluation was conducted using a CNN and a dataset of DDoS attacks. The limits of the research highlight the need for additional investigation into more efficient resource use, control channel optimization, data plane memory management, and security enhancements to prevent potential problems with harmful traffic.

References

Doriguzzi-Corin, R., Knob, L. a. D., Mendozzi, L., Siracusa, D., & Savi, M. (2023). Introducing Packet-Level analysis in programmable data planes to advance network intrusion detection. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2307.05936

Javaid, A. Y., Niyaz, Q., Sun, W., & Alam, M. S. (2016b). A Deep Learning Approach for Network Intrusion Detection System. *BICT 2015*. https://doi.org/10.4108/eai.3-12-2015.2262516

Jose, A. V., Selvan, M. P., Mary, V. A., Grace, J., Jancy, S., Helen, S., & P, A. (2022). Prediction of network attacks using supervised machine learning algorithm. *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*. https://doi.org/10.1109/ic3iot53935.2022.9767948