Enhancing Network Security Through Machine Learning-Based Intrusion Detection

Mohan Babu Kunchala
A20524765
mkunchala@hawk.iit.edu

Sanjitha Reddy Pathuri
A20524383
spathuri2@hawk.iit.edu

Sirisha Gandham
A20544789
sgandham1@hawk.iit.edu

Project Proposal

1 Problem Description

Network intrusion detection plays a critical role in protecting against malicious attacks and cybersecurity is still a top priority for organizations worldwide. Conventional rule-based methods frequently don't catch evolving attack tactics. By learning from large datasets, machine learning provides a promising way to improve detection capabilities and efficiently identify anomalous network behaviors.

This project uses the UNSW-NB15 dataset to precisely classify both normal and attack traffic with the goal of developing a strong machine learning-based approach for network intrusion detection.

2 Brief survey of the literature and proposed work

Machine learning's potential for network intrusion detection has been the subject of numerous studies. Moustafa et al. (2015) proposed the use of Principal Component Analysis (PCA) for feature selection and presented the UNSW-NB15 dataset. Alazab et al. (2016) carried out an analysis of anomaly-based detection methods and proposed a framework utilizing several machine learning algorithms. Surveys were conducted by Mukherjee et al. (2020) and Islam and Ahmed (2019) in which machine learning techniques were divided into three categories: supervised, unsupervised and hybrid approaches. Deep learning techniques were found to be promising.

Building on these findings, our project focuses on using neural networks (NNs) for network intrusion detection, namely CNNs, RCNNs and MLP Classifiers. We also want to investigate feature selection techniques to improve classifier performance. In order to evaluate NNs efficacy, the study will compare them to more conventional classifiers such as XGBoost, SVM, Random Forest, Decision Tree, and Logistic Regression.

3 Preliminary Plan (Milestones):

1. Dataset Preparation:
    - Extract and clean the UNSW-NB15 dataset obtained from kaggle, paying attention to categorical variables and missing values.

- Encode categorical variables and standardize numerical variables.

2. Exploratory Data Analysis (EDA)
    - Use visualizations like histograms, heatmaps and boxplots to comprehend feature distributions, correlations, and outlier identification through EDA.
3. Feature Selection:
    - The top k features that have the biggest impact on classification accuracy are chosen using Mutual Information Score.
4. Model Selection:
    - Examine several neural network architectures (CNNs RCNNs and MLP Classifier) in addition to more conventional classifiers such as XGBoost, SVM, Random Forest, Decision Tree, and Logistic Regression.
5. Model Evaluation:
    - Evaluate the performance of the classifier using metrics such as F1-Score, AUC-ROC, Accuracy, Precision and Recall.
    - Examine confusion matrices to find misclassified samples and gain knowledge that will help in improving the model.

Conclusion

This project aims to improve network intrusion detection capabilities by utilizing machine learning techniques and the UNSW-NB15 dataset. Our objective is to create detection models that are both efficient and successful in identifying known as well as unknown attack patterns through thorough testing and assessment. The project's ultimate goal is to support continued efforts to improve cybersecurity safeguards for businesses around the globe.

References

[1] Mamoun Alazab, Michael Hobbs, and Jemal Abawajy. Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 60:84–102, 2016

[2] Md Rafiul Islam and Kazi Mohammed Ahmed. Machine learning approaches for network intrusion detection: A comprehensive survey. IEEE Access, 7:27459–27484, 2019.

[3] Ahmed Moustafa, Jill Slay, and Gregory Creech. Unsw-nb15: A comprehensive data set for net- work intrusion detection systems (unsw-nb15 network data set). In Military Communications and Information Systems Conference (MilCIS), pages 1–6. IEEE, 2015.

[4] Nour Moustafa and Jill Slay. The evaluation of network anomaly detection systems: Statistical analysis of the unsw-nb15 dataset and the comparison with the kdd99 dataset. Information Security Journal: A Global Perspective, 25(1-3):1–14, 2016.

[5] Sourav Mukherjee, Ananda Roy Chowdhury, Shukla Das, Sayan Chakraborty, and Mita Nasipuri. A comparative study of deep learning approaches for network intrusion detection. Future Generation Computer Systems, 107:1063–1077, 2020