CSE 847: Project Proposal

Project Title: Network Anomaly Detection Using Various Machine Learning Algorithms - A Comparative Study

Project Team: Yashashvini Rachamallu

Problem Description: In the increasingly interconnected digital landscape, network security plays a critical role in safeguarding sensitive data and ensuring the uninterrupted operation of organizations. The sophistication and scope of cyberattacks and intrusions are constantly increasing, posing a serious threat to both individuals and companies. Network anomalies, often indicative of cyberattacks, can take various forms, including denial-of-service (DoS) attacks, probes, and unauthorized access attempts. Detecting and mitigating these anomalies in real-time is crucial for maintaining the integrity and availability of network resources. **Network Anomaly Detection** aims to create a strong and proactive solution to strengthen cybersecurity defenses. The creation, deployment, and comparison of machine learning models for detecting network anomalies are the main objectives of this project. Our goal is to quickly detect malicious activity and unexpected network behavior by utilizing data-driven insights and predictive analytics.

The primary objective of this project is to develop and evaluate machine learning models for network anomaly detection using the **KDD Cup 1999 Dataset.**
Specifically, we aim to:

1. Identify and categorize diverse types of network anomalies present in the dataset.
2. Implement various machine learning algorithms to detect these anomalies.
3. Conduct a comparative analysis to determine which algorithms are most effective at detecting specific anomaly types.
4. Provide recommendations for selecting the most suitable algorithm(s) for network anomaly detection based on the dataset's characteristics and the nature of the anomalies.

Preliminary Plan:

1. Data Collection and Exploration: Acquire the KDD Cup dataset, which includes labeled network connection data. Perform exploratory data analysis (EDA) to understand the dataset's structure and characteristics.
2. Feature Engineering: Preprocess the dataset by handling missing values, scaling features, and encoding categorical variables. Select relevant features that are indicative of network anomalies.
3. Model Selection: Choose a diverse set of machine learning or deep learning algorithms, like Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), PCA, Neural Networks (GAN or Encoders)
4. Model Training: Train each selected machine learning algorithm on the preprocessed dataset. Optimize hyperparameters and model configurations.
5. Model Evaluation: Evaluate the performance of each model using appropriate metrics (e.g., accuracy, precision, recall, F1-score). Compare the algorithm's ability to detect distinct types of attacks in the dataset.
6. Comparative Analysis: Conduct a comprehensive comparative analysis to determine which algorithms excel in identifying specific attack types and overall anomaly detection. Identify strengths and weaknesses of each algorithm in this context.
7. Recommendations: Provide recommendations for selecting the most suitable machine learning algorithms for network anomaly detection based on comparative analysis.

References:

1. https://avinetworks.com/glossary/anomaly-detection/
2. Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, Fatima Mohamad Dakalbab, “Machine Learning for Anomaly Detection: A Systematic Review,” IEEE Access, Volume 9, 2021.
3. D. Kwon, K. Natarajan, S. C. Suh, H. Kim and J. Kim, "An Empirical Study on Network Anomaly Detection Using Convolutional Neural Networks," 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), Vienna, Austria, 2018, pp. 1595-1598, doi: 10.1109/ICDCS.2018.00178.
4. KDD Dataset
5. https://www.spiceworks.com/tech/networking/articles/network-behavior-anomaly-detection/

Timeline:

1. Project Proposal: October 4, 2023
2. Data Preprocessing, Feature Engineering – October 10, 2023
3. Model Selection: October 13, 2023
4. Model Training: October 20, 2023
5. Intermediate Report: November 1, 2023
6. Model Training: [Rest Algorithms] November 7, 2023
7. Model Evaluation: November 14, 2023
8. Comparative Analysis: November 19, 2023
9. Recommendations and Documentation: December 10, 2023

By the end of this project, we aim to provide valuable insights into the effectiveness of various machine learning algorithms for network anomaly detection, assisting in enhancing network security measures.