Group 2

Network Anomaly Detection

Chandana Chevuturi 11664381

Naveen Ajay Karasu 11646981

Surya simha reddy chintha 11702127

PAVAN KUMAR REDDY BHUMIREDDY 11656786

Introduction

In today’s age of cybersecurity the detection of network anomalies is very important to prevent cyberattacks and to ensure the integrity of network. As part of this course, in this project we are trying to focus on the development of an anomaly detection model using machine learning. By using AI, we aim to automate the process of detecting the anomalies within network traffic. Which can help in providing early warning for the signs of cyber intrusion. The main goal is to enhance the capabilities of analyst by creating a robust detection system that can be used to classify normal and anomalous network activities.

Data Used

For this project we are using the dataset from Kaggle's "Network Anomaly Detection" dataset, which has various different attributes which are used for describing network traffic. This data will includes a mix of normal and anomalous records. The data is labeled to facilitate supervised learning for classification. Each row will represent a network event with feature’s like packet counts, protocol types, and connection durations etc..,

Features

Intrusion Detection: The main feature of the system we are developing will be its ability to detect potential intrusions by analyzing network traffic.

Real-time Analysis: We are planning to optimize the model to allow near real-time classification to provide the cybersecurity analysts with quick insights into potential threats in real time.

User-friendly Interface: We are planning to design it for cybersecurity analyst with output which will be easy to interpret by showing alerts when anomalies are detected, along with details of the suspicious activity.

Requirements

Clustering Algorithm: We are planning to use clustering machine learning techniques for identifying patterns in network traffic data that may correspond to potential intrusions. We are using Python programming language and its relevant libraries Scikit-learn will be used for developing it.

Data Preprocessing: The network traffic data, we are using for developing the model will be preprocessed by using normalization, feature selection and transformation which can help us to ensure that the clustering algorithm can perform optimally to have more accuracy in the obtained result.

Performance Evaluation: The level of performance of the system can be determined by accuracy, recall and F1-score metrics.

Scalability: The model will be scalable so that it can handle large volumes of network traffic data.

Evaluation Criteria

The effectiveness of the system can be evaluated by using the following technique's:

Accuracy: It can be defined as the proportion of correctly classified instances (both normal and anomalous) to the total instances.

Recall: It can be defined has ability of the model to identify actual anomalies among all the anomalies present in the dataset.

F1-Score: It can be defined has the harmonic mean of precision and recall. Which can provide a measure of the model's performance for when handling imbalanced datasets.

Links

https://www.kaggle.com/datasets/anushonkar/network-anamoly-detection