Project overview

Team: 4

Project Title: Smart Intrusion detection and prevention system

Project Summary:

In Today's time, there is a rapid increase in network traffic due to which data breaches and leaking sensitive information is very common. The goal of this project was to Distinguishing benign network traffic from malicious network-based attacks and leverage machine learning techniques to solve this problem. In this project, we used the knowledge gained from a Known dataset like KDD 99 and applied it to a modern dataset like CICDS 2018 to detect attacks like Probing, Denial of service, U2R, R2L, Dos, etc. We developed ML models using algorithms like KNN, Decision Tree, Random forest, SVM, Neural networks, etc to detect these attacks. **The novelty** of this project is the **features extracted from the random forest and decision tree can be used as a custom rule in sniffing tools like snort to detect intrusion.**

Role of each member:

We have divided the team into two subgroups one group will work on the KDD99 dataset and the others on CICIDS2018 and the final submission will be made by combining the work of both.

Group 1: [CICIDS2018]

- **Akash Kumar:** Exploring related research papers, Exploring dataset, Exploring different types of attack possible in that dataset, worked on KNN
- **Saptarshi Manna:** Data preprocessing, feature selection, outlier removal, data visualization, worked on Random forest, explored different tools.
- **Anupam Misra:** statistical techniques, anomaly detection, model building, hyperparameter tuning, worked on decision tree

Group 2: [KDD99]

- **Deeksha Sahu:** research related to datasets, working on threats taxonomy like Network threats, host threats, software threats.
- **Kritika Singh:** Exploratory data analysis, Defining metrics for IDS evaluation, Exploring limitations of given datasets.
- **Aditya Mohan Gupta:** Feature selection, model building, hyperparameter tuning, setup DNS server for detection
- **Aditya Gupta:** data visualization, comparison with existing IDS, combining learnings from both data sets, worked on SVM.