# Network Intrusion Detection using Machine Learning Algorithms

Team 11: Rutuja Manohar Kute, Ruthwik Bommenahalli Gowda

## Introduction:

In today’s world, the increasing demand for network usage has led to the development of advanced technologies such as the Internet of Things (IoT), cloud computing, and Software Defined Networks (SDN). While these technologies have greatly enhanced our ability to access services and data, they have also made us vulnerable to various online threats. Intrusion Detection Systems and Prevention Systems are crucial tools for network users to defend against these threats. In this project, we aim to use machine learning algorithms to detect network intrusion in an SDN.

## Motivation and Goals:

The primary motivation behind this project is to develop an accurate and efficient network intrusion detection system that can protect SDNs against various types of online threats. The goal of this project is to explore and evaluate the performance of various machine learning algorithms for detecting network intrusion in an SDN. We aim to achieve high accuracy, precision, and recall while minimizing false positives and false negatives.

## Methodology:

To achieve our goals, we will perform the following steps:

* Data Preprocessing: The raw data we will use for our project contains 79 quantitative and qualitative features. We will clean and preprocess the data to ensure that it is suitable for machine learning algorithms. This includes handling missing values, removing duplicates, and converting qualitative features to quantitative features using one-hot encoding.
* Feature Selection: We will select the most important features that will help to detect network intrusion. We will use various feature selection techniques such as correlation analysis, principal component analysis (PCA), and recursive feature elimination (RFE).
* Model Training: We will train various machine learning algorithms such as Naive Bayes, K-Nearest Neighbors (KNN), Decision Trees, Random Forest, Support Vector Machines (SVM), Linear and Logistic Regression on the preprocessed and selected features to detect network intrusion.
* Model Evaluation: We will evaluate the performance of the trained models based on various metrics such as accuracy, precision, recall, and F1-score. We will also use techniques such as cross-validation and ROC analysis to assess the robustness of our models.

## Description of the Dataset:

The dataset we will use for our project is the Network Intrusion Detection System (NIDS) 2017 dataset, which is publicly available and widely used in research. The dataset was collected in a laboratory environment using the Kali Linux operating system and various attack tools such as Metasploit, Nmap, and Hping3. The dataset contains the record of real-time traffic captured daily in a tabular format. The dataset originally belongs to a Packet Capture file (PCAP) and contains 79 quantitative and qualitative features. Out of these, 1 feature represents qualitative attributes, and 78 features represent quantitative attributes.

We selected a dataset that contains the records of DDoS, XSS Intrusion, Brute Force Intrusion, SQL Injection, and benign traffic. This dataset contains 1,188,333 rows of observation of the network intrusion and whitelisted traffic along with 79 features. The count of all types of traffic is as follows:

BENIGN Traffic with 798,322 Observations

DDoS Traffic with 383,439 Observations

Web Attack Brute Force Traffic with 4,550 Observations

Web Attack XSS Traffic with 1,962 Observations

Web Attack Sql Injection Traffic with 60 Observations

**Data Sources**: <https://www.kaggle.com/datasets/subhajournal/sdn-intrusion-detection>