A machine learning based approach towards building an Intrusion Detection System

Problem Description

With the rising amount of network enabled devices connected to the internet such as mobile phones, IOT appliances or vehicles the concern about the security implications of using these devices is growing. The increase in numbers and types of networked devices inevitably leads to a wider surface of attack whereas the impact of successful attacks is becoming increasingly severe as more critical responsibilities are assumed be these devices.

To identify and counter network attacks it is common to employ a combination of multiple systems in order to prevent attacks from happening or to detect and stop ongoing attacks if they can not be prevented initially.
These systems are usually comprised of an intrusion prevention system such as a firewall as the first layer of security with intrusion detection systems representing the second layer.
Should the intrusion prevention system be unable to prevent a network attack it is the task of the detection system to identify malicious network traffic in order to stop the ongoing attack and keep the recorded network traffic data for later analysis. This data can subsequently be used to update the prevention system to allow for the detection of the specific network attack in the future. The need for intrusion detection systems is rising as absolute prevention against attacks is not possible due to the rapid emergence of new attack types.

Even though intrusion detection systems are an essential part of network security many detection systems deployed today have a significant weakness as they facilitate signature-based attack classification patterns which are able to detect the most common known attack patterns but have the drawback of being unable to detect novel attack types.
To overcome this limitation research in intrusion detection systems is focusing on more dynamic approaches based on machine learning and anomaly detection methods. In these systems the normal network behaviour is learned by processing previously recorded benign data packets which allows the system to identify new attack types by analyzing network traffic for anomalous data flows.

This project aims to implement a classifier capable of identifying network traffic as either benign or malicious based on machine learning and deep learning methodologies.

Data

The data used to train the classifier is taken from the CSE-CIC-IDS2018 dataset provided by the Canadian Institute for Cybersecurity. It was created by capturing all network traffic during five

days of operation inside a controlled network environment on AWS where realistic background traffic and different attack scenarios were conducted.
As a result the dataset contains both benign network traffic as well as captures of the most common network attacks.
The dataset is comprised of the raw network captures in pcap format as well as csv files created by using CICFlowMeter-V3 containing 80 statistical features of the individual network flows combined with their corresponding labels.
A network flow is defined as an aggregation of interrelated network packets identified by the following properties:

- Source IP
- Destination IP
- Source port
- Destination port
- Protocol

The dataset contains approximately 16 million individual network flows and covers the following attack scenarios:

- Brute Force
- DoS,
- DDos
- Heartbleed,
- Web Attack,
- Infiltration,
- Botnet

Approach

The goal of this project is to create a classifier capable of categorising network flows as either benign or malicious.
The problem is understood as a supervised learning problem using the labels provided in the dataset which identify the network flows as either benign or malicious. Different approaches of classifying the data will be evaluated to formulate the problem either as a binary classification or a multiclass classification problem differentiating between the individual classes of attacks provided in the dataset in the later case. A relevant subset of the features provided in the dataset will be used as predictors to classify individual network flows.
Machine learning methods like k-nearest neighbours, random forest or SVM will be applied to the problem and evaluated in the first step in order to assess the feasibility of using traditional machine learning approaches.
Subsequently deep learning models like convolutional neural networks, autoencoders or recurrent neural networks will be employed to create a competing classifier as recent research has shown that deep learning methods represent a promising application in the field of anomaly detection.
The results of both approaches will be compared to select the best performing classifier.

Deliverables

The classifier will be deployed and served via a REST API in conjunction with a simple web application providing a user interface to utilize the API.

The REST API will provide the following functionality:

- an endpoint to submit network capture files in pcap format. Individual network flows are extracted from the capture files and analysed for malicious network traffic.
- (optional) an endpoint to stream continuous network traffic captures which are analysed in near real-time combined with
- (optional) an endpoint to register a web-socket in order to get notified upon detection of malicious network traffic.

To further showcase the project, a testbed could be created against which various attack scenarios can be performed. This testbed would be connected to the streaming API for near real-time detection of malicious network traffic.

Computational resources

The requirements regarding the computational resources to train the classifiers are given below:

CPU	Intel Core i7 processor
RAM	32 GB
GPU	1 GPU, 8 GB RAM
HDD	100 GB