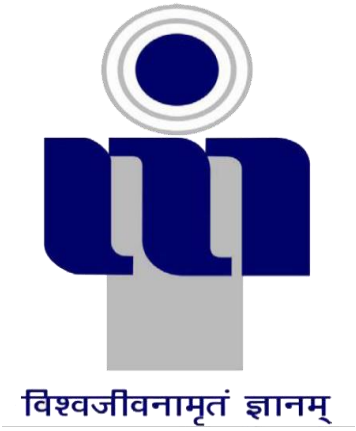
ATAL BIHARI VAJPAYEE - INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT GWALIOR - 474015



Machine Learning Techniques to improve Intrusion detection

A Project report,
Submitted in partial fulfillment of the requirements for B. Tech Project

by

Himanshu Pandey(2018IMT-038)

Under the supervision of:
Dr. Saumya Bhadauriya

CONTENTS

Background Information	2
Motivation	2
Objectives	3
Techniques	3
Datasets	3
Programming Languages	3
Timeline/Gantt Chart	4

Background

During the 1970s, the growth of computer networks created new problems related to the monitoring of user activities and access. Intrusion Detection Systems have two categories: signature-based detection (or “misuse detection”) and anomaly detection. In signature-based detection, the data monitored by the IDS is compared to known patterns of attacks and it can only detect known attacks while anomaly detection builds a model of normal behavior of the system and then looks for deviations in the monitored data. Anomaly detection has a wider aspect but it can also detect unknown attacks which can generate irrelevant alarms.

A number of machine learning techniques have come into picture to detect anomaly. Some of them can rely on algorithms with the ability to learn directly from the data, without being explicitly programmed. Despite of so many advantages, these algorithms are rarely deployed just because of higher false positivity rate. Indeed, even a false positive rate of 1% can create so many false alarms on a high traffic network that they become impossible for an administrator to process.

Motivation

With the passage of the time, a number of various computer attacks are stretching their arms with an intention to harm the targeted system. For a company, antiviruses or firewalls are no longer sufficient to ensure the security of the systems of the company. Now we need multiple layers of security to ensure truly secured systems among which one of the most important layers is provided by Intrusion Detection System (IDS) which is designed to protect its target against any potential attack through a continuous monitoring of the system. In our thesis, we are going to deploy machine learning models which can detect known attacks through supervised intrusion detection and unknown attacks through unsupervised intrusion detection. Further, in the thesis model will be deployed to predict bandwidth utilization and at last all the machine learning techniques will be compared in detecting intrusions so that we can com up with better and efficient algorithms and others can have a knowledge to which algorithm is to be used in case of intrusion detection.

Objectives

1. Techniques for Supervised Intrusion Detection
2. Techniques for Unsupervised Intrusion Detection
3. Predicting Bandwidth Utilization
4. To check the performance of Machine Learning techniques in detecting intrusions by comparing their sensitivity, specificity and accuracy.

Techniques

- Cascade-structured neural networks
- Naive ensemble learning
- Transfer learning
- Protocol-based intrusion detection
- The Autoregressive Integrated Moving Average (ARIMA)
- The Multilayer Perceptron (MLP)
- The Long Short-Term Memory network (LSTM)

Datasets

- KDD Cup 9
- NSL-KDD
- CICIDS2017
- CSE-CIC-IDS2018

Programming Language

- Octave
- Python
- Matlab

Timeline

GNATT CHART