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 Supervison of

Dr. Saumya Bhadauriya

ABV INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT
GWALIOR-474 015
2021

Contents

List of Figures

1 Abstract

Hacking has grown in popularity in recent years, resulting in an increase in the quantity and variety of cyber-attacks. Malware, denial of service attacks, phishing, and social engineering are all examples of computer network threats. Antiviruses and firewalls are no longer sufficient for effective cyber security. To prevent these dangers, you can't only rely on antivirus and firewalls: you need multiple levels of defence. With the capacity to monitor packets from OSI layer 2 (Data link) through layer 7 (Application), network-based Intrusion Detection Systems (IDSs) offer a supplementary technique of increasing security. IDSs that use anomaly detection can detect unknown assaults, but they are less accurate, resulting in a high amount of false alarms. Machine learning methods are investigated in this thesis in order to develop IDSs that can be deployed in real-world computer networks. To begin, a three-step optimization strategy is given to increase detection quality: 1) data augmentation to rebalance the dataset, 2) model performance optimization, and 3) ensemble learning to integrate the findings of the best models. This method has a disadvantage of requiring labelled datasets, which are rarely available in real-life situations. As a result, transfer learning is explored in order to train machine learning models on huge labelled datasets and subsequently finetune them on innocuous network traffic. This method also has problems because the models are trained on previously known assaults and so do not do anomaly detection. As a result, an unsupervised learning-based method is provided. It makes advantage of network prototyping. When anomalies are discovered, they are grouped into attacks or disregarded if they are isolated. Finally, network congestion detection is investigated. The bandwidth usage of various links is forecasted in order to prevent problems from arising.

2 Introduction

Computers have become an inextricable aspect of our lives. They are widely employed in all government agencies, businesses, private organisations, hospitals, and private residences. Because of their importance in our lives, protecting them from invasions is a huge task for us. Intrusions are still the most persistent hazards in the cyber world, despite significant advancements in protection technologies. Many ways for analysing intrusions have been developed to date.

3 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 behaviour 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.

4 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 come up with better and efficient algorithms and others can have a knowledge to which algorithm is to be used in case of intrusion detection.

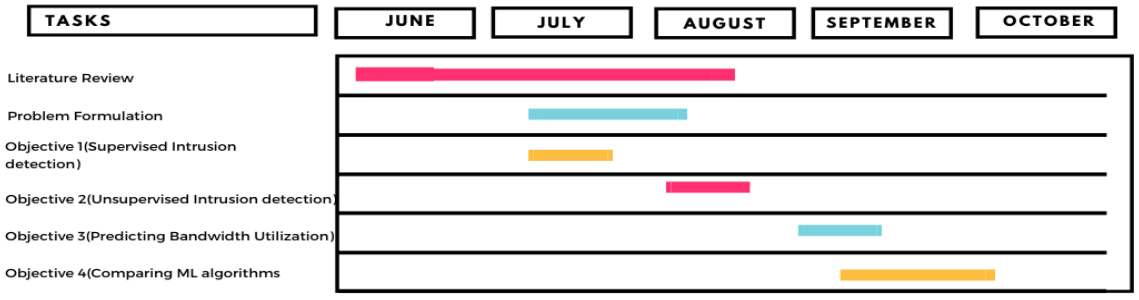
5 Problem Statement

Techniques to detect intrusions are to be found. It may include supervised learning or unsupervised learning algorithms. We need to check the accuracy, scalability and specificity of the algorithms.

5.1 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.

6 Gantt Chart



7 References

References

[1] M. Roesch, “Snort - lightweight intrusion detection for networks,” in Proceedings of the 13th USENIX Conference on System Administration, LISA ’99, (USA), p. 229–238, USENIX Association, 1999.
[2] “Suricata | open source ids / ips / nsm engine,” Dec. 2009. [Online; accessed 2019-12-16].
[3] S. Axelsson, “The base-rate fallacy and the difficulty of intrusion detection,” ACM Trans. Inf. Syst. Secur., vol. 3, p. 186–205, Aug. 2000.
[4] M. Labonne, A. Olivereau, and D. Zeghlache, “A survey of neural network classifiers for intrusion detection (submitted),” 2020.
[5] M. Labonne, A. Olivereau, and D. Zeghlache, “Automatisation du processus d’entraînement d’un ensemble d’algorithmes de machine learning optimisés pour la détection d’intrusion,” in Proceedings of Journées C&ESAR 2018, pp. 1–10, Journées C&ESAR, 11 2018.
[6] M. Labonne, A. Olivereau, B. Polvé, and D. Zeghlache, “A cascade-structured meta-specialists approach for neural network-based intrusion detection,” in 16th IEEE Annual Consumer Communications & Networking Conference, CCNC 2019, Las Vegas, NV, USA, January 11-14, 2019, pp. 1–6, 2019.
[7] M. Labonne, B. Polvé, and A. Olivereau, “Procédé et système de détection d’anomalie dans un réseau de télécommunications,” 2019.
[8] M. Labonne, A. Olivereau, B. Polve, and D. Zeghlache, “Unsupervised protocolbased intrusion detection for real-world networks,” in ICNC 2020: International Conference on Computing, Networking and Communications, 2020 International Conference on Computing, Networking and Communications (ICNC), (Big Island, United States), pp. 299–303, IEEE, Feb. 2020.
[9] M. Labonne, C. Chatzinakis, and A. Olivereau 2020.
[10] Maxime Labonne. Anomaly-based network intrusion detection using machine learning. Cryptography and Security [cs.CR]. Institut Polytechnique de Paris, 2020. English. ffNNT : 2020IPPAS011ff. fftel02988296
[11] “Isms family of standards,” standard, International Organization for Standardization, Geneva, CH, 2018.