Chapter 3
Network Intrusion Detection Using Linear and Ensemble ML Modeling

Shilpi Hiteshkumar Parikh
U and P U. Patel Department of Computer Engineering, CSPIT, Charotar University of Science and Technology (CHARUSAT), Changa, India

Anushka Gaurang Sandesara
U and P U. Patel Department of Computer Engineering, CSPIT, Charotar University of Science and Technology (CHARUSAT), Changa, India

Chintan Bhatt
U and P U. Patel Department of Computer Engineering, CSPIT, Charotar University of Science and Technology (CHARUSAT), Changa, India

ABSTRACT

Network attacks are continuously surging, and attackers keep on changing their ways in penetrating a system. A network intrusion detection system is created to monitor traffic in the network and to warn regarding the breach in security by invading foreign entities in the network. Specific experiments have been performed on the NSL-KDD dataset instead of the KDD dataset because it does not have redundant data so the output produced from classifiers will not be biased. The main types of attacks are divided into four categories: denial of service (DoS), probe attack, user to root attack (U2R), remote to local attack (R2L). Overall, this chapter proposes an intense study on linear and ensemble models such as logistic regression, stochastic gradient descent (SGD), naïve bayes, light GBM (LGBM), and XGBoost. Lastly, a stacked model is developed that is trained on the above-mentioned classifiers, and it is applied to detect intrusion in networks. From the plethora of approaches taken into consideration, the authors have found maximum accuracy (98.6%) from stacked model and XGBoost.

DOI: 10.4018/978-1-7998-6988-7.ch003

INTRODUCTION

Currently, we are thriving in a world that is limitless and with no boundaries. With the augment in advances technologically and scientifically there are high chances of attacks, breaches, and other vulnerabilities in the network. Besides this, the surge of internet facilities and online utilities available in a fraction of seconds have resulted in high cases of cyber-crime. Before, two decades the detection of breaches and attacks were carried independently by users without any intervention from the machine. But nowadays due to the high-amount of cyber crimes and intrusions in networks, it is not possible to solve the crime manually and hence it is more efficient with the machine learning and deep learning methods available. Still, there is a huge demand for a novel technique that predicts the intrusions as well as guides the users of the network on how to resolve them.

When we talk about data in wireless networks, different types of data in structure, dimension, size come into picture. According to the authors (Yuanwei et al., 2019), big data resources are utilized by analytical and statistical machine learning tools to support new intelligent applications which are proposed in wireless networks. Accordingly, the prevailing variant types of data can be categorized into major three forms: Wireless Data, Social Data and Cloud Data. The most notable challenge to perform data analytics in wireless networks is to accurately predict user preference distribution. Performing data analytics on wireless networks can also help to look into the odd behaviour of some data and help to figure out the malicious activity taking place inside the wireless networks.

Basically, an intrusion detection system can be of two types-software or hardware. It is the choice of the manufacturer to select a software or hardware system and the system can be attached to the different network domains such as Ethernet, FDDI, or any other. The IDS system continuously inspects the traffic from the original point of installation and performs an observation of passing traffic on the entire subnet and matches the traffic that is passed on the subnets to the collection of known attacks. Such advanced systems are not possible to attack by invaders because any malevolent activity is directly reported to the administrator. The IDS system developed by the researchers here supervises both inbound and outbound traffic on the network, as well as data traversing between systems within the network.

Most present-day businesses require top-tier safety to protect their credentials for work. Even though there are conventional techniques such as authentication and authorization (Xu et al., 2014) but they are not able to ensure complete security in the systems, Intrusion Detection System on the other hand provides a great level of safeguard for protecting the system from attacks and other threats. One important advantage of the IDS system is that it provides an immediate alert to the administrator about the prevailing attacks on the network so that the administrator is at least aware that the network has been infected. Being aware of future possible attacks and breaches, an IT person can take appropriate steps to stop the attacker or prevent it from happening. So, the basic step of any IDS system is to detect the type of attack that would be taking place. Although the system is not able to resolve the attack, perceiving the intrusion will benefit the security officials and hence Intrusion Detection (ID) is the first and foremost step.

This chapter takes into consideration four basic attacks. Amongst them, the DOS attack (Alharbi et al., 2018) is the most hazardous because it generates a lot of traffic making it so full of memory and extraneous resources that the system fails to recognize legal user requests. The main purpose of the R2L attack is to get an illicit permit to the system's resources and the privacy of the whole network is disrupted. The U2R attack gives access to the attacker as a root user so confidentiality of data is again lost. The Probe attack is where the attacker investigates the network for weaknesses that can prove useful to recognize services that can be executed (Chao-yang, 2011).