INTRUSION DETECTION SYSTEM (IDS)

Made By: Mohd Imran Khan

CONTENT

Introduction

The Intrusion detection system (IDS) are defined as efficient security tools that are used for improving the security of the communicating and the information systems as the primary focus on detecting malicious network traffic. An IDS is seen to very similar process like firewalls, antivirus software can access the control schemes. The IDS is classified depending on detection as the signature detection and anomaly detection systems, the system identified the traffic pattern or the application data as malicious and this requires an updated database for storing all the new attack signatures, whereas the anomaly detection system compare all activates against the normal defined behavior.

The main objective of the IDS system is detecting and them raising an alarm if the network is attacked. The best IDS system detects the new or more malicious attacks within a short period of time and take the necessary action to prevent the devices and data base. The current Intrusion detection system does not providing the 100% security from the cyber attacks and cyber-thefts. Hence this study has been carried out for improving and increasing the Intrusion detection system accuracy. Many machine learning techniques have been used for detecting the network attacks and improving the firewalls accuracy. Machine learning techniques are used to develop effective classification and clustering models. Which can distinguish the normal and abnormal behavior packets. The procedure of detecting the intrusion accurately from the complete traffic, network is a classification problem.

The IDS System is classified as per the detection methods used for identifying all the malicious attacks. A misuse or signature detection techniques identifies signatures or patterns present in the existing attacks within the network traffic. This misuse detection system requires and updated database for storing the new attacks signatures. But new attacks are not detected until the system get trained. The anomaly detection system use and approach based on the detection of traffic anomalies by identifying behavior. Hence this show that can IDS handle new attacks. It is unable to detect or identify or particular attacks.

Several types of researcher have used data mining for improving the IDS by offering an external intrusion detection that identifies the presence of any existing boundaries within a normal network activity. This helps distinguishing between normal and abnormal activity. The machine learning algorithm is also applied in the IDS for identifying if any attacks are present in the system, improving the detection accuracy rate and developing effective classification and clustering models for distinguishing between normal and abnormal behavior packet. The procedure of detecting the intrusion accurately from the complete network traffic is classified as classification problem.

The classification models helps in identifying any malicious attacks, improving the accuracy detection rate, and decreasing false alarms. Many machine learning algorithms have been proposed earlier for generating an efficient IDS like the Random forest classifier, logistic regression, support vector classifier, Decision tree classifier, K-Nearest Neighbors Classifier and XGBoost Classifier etc. All these algorithm are integrated with the different models for distinguishing between the malicious attacks and determing a normal behavior for detecting unknown malicious attacks.

Objective

This project is about Intrusion Detection System. Which is related to cyber security. Network traffic makes it difficult to find many through-system attacks. So we can easily find known or unknown attacks through the network using IDS software which works to easily find the attacks occurring during the every day and every month in the system.

Problem Statement

The goal of this project to is to detect the network intrusion attacks using several classifier. What did I use google colab to create this project. My motive for creating this project is to prevent the traffic happing through the network traffic happening today and the network or system for deteriorating.

Detection method of Intrusion detection system

There are two types of Method of Intrusion detection system

1) Signature - Based method

2)Anomaly - Based method

Signature Based Intrusion Detection system-: Signature based Intrusion detection system detects the basis of the specific patterns such as number of bytes or numbers of 1's or number of 0's in the network traffic. It also detects on the basis of the already known malicious instructions that is used by the malware. The detected patterns in the IDS known as signatures.

Signature based-IDS can easily detect the attack whose pattern (signature) already exist in the system but it is quite difficult to detect the new malware attacks as their pattern (signature) is not known.

Anomaly Based Intrusion detection system-: Anomaly based IDS to detect the unknown malware attacks as new malware are developed rapidly. In anomaly based intrusion detection system there is use of machine learning to create a trustful activity model and anything coming is compared with that model and it is declared suspicious if it is not found in model. Machine learning based method has a better generalized property in comparison to signature based IDS as these models can be trained according to the applications and hardware configuration.

The Proposed approach of multi stage system

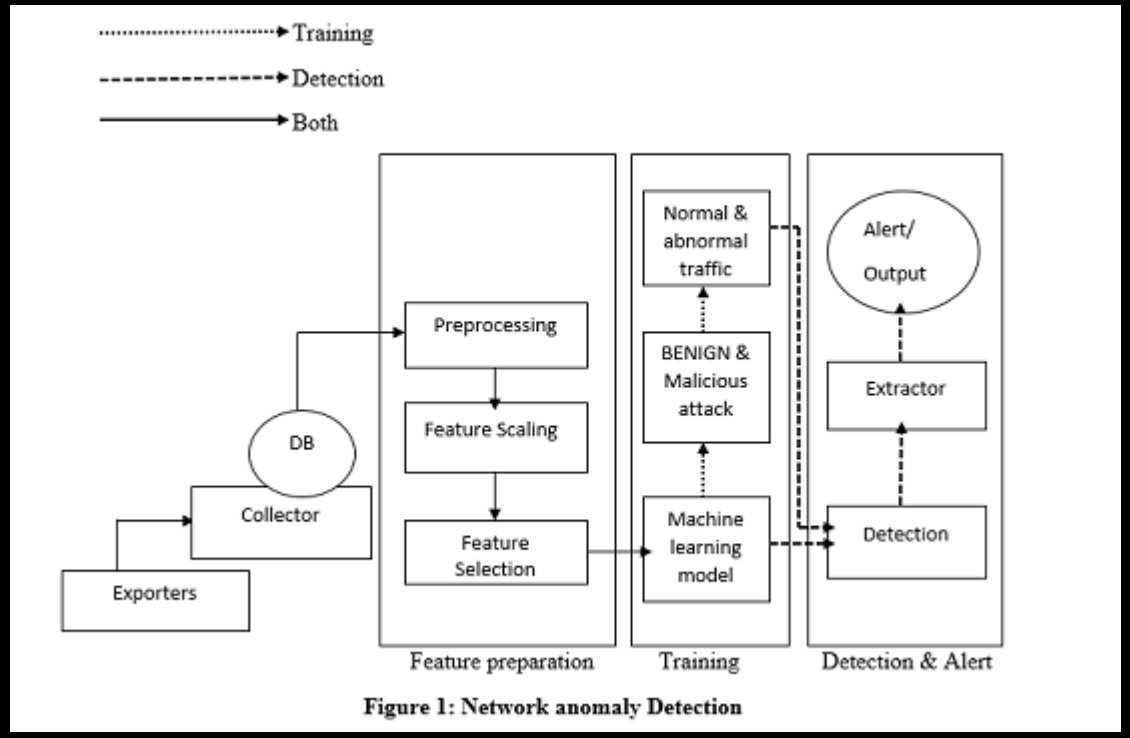

Figure 1: Network anomaly Detection

Exporters

BRO enabled devices which permanently monitor network traffic, account statistics, and export connection data to our system.

Collector

The function of this module is to collect connection data exported from one or several exporters.

Feature Preparation Module

Traffic distributions can be different on weekdays and weekends, so we segregated the flow data accordingly. It is very unlikely that a single connection log can give much information about the activities happening in the network. Therefore, collected flows are analyzed for constant-length time intervals (say every 1 minute). A probability mass function (PMF) is created for every feature for a time interval. Next, depending on the version, Tsallis or Renyi entropy of positive and negative α values are calculated for traffic feature distributions and a time series is created out of the same.

Training

Over the years, several anomaly detection techniques have been proposed in the literature. The problem at hand was completely unsupervised and needed a sound approach for anomaly detection. Trained Machine learning model to learn an approximation of the original data and filtered out the top X% connection logs whose reconstruction error was high and thus capped the reconstruction error threshold. It is very likely that the network traffic at previous time intervals can show an effect on the traffic under study. Also, during the training phase, it builds a network profile of long term behavior over the data points which have less reconstruction error and are assumed to contain few or no attacks/anomalies. The network profile captures the hourly min and max concentration and dispersion Entropy Statistic for all the connection features to reflect the changes on weekdays and weekends.

Detection

The final stage of the proposed system involves identifying the events that occurred, and gathering other related information as support,-from other related logs like Bro dns logs, http logs, file logs, smtp logs - to present status of the network to the administrator and create alarms with all the details as output.

Working of Intrusion detection system

An intrusion detection system is a monitor-only application designed to identify and report on anomalies before hackers can damage your network infrastructure. IDS is either installed on your network or a client systems (host-based IDS). Typical intrusion detection systems look for known attaks signature or abnormal deviations from set norms. These anomalous patterns in the network traffic are then sent up in the stack for further investigation at the protocol and application layers of the OSI (open system interconnection) model.

An IDS is placed out of the real time communication band (a path between the information sender and reciver) within your network infrastructure to work as a detection system. It instead leaverage as a span or tap port for network monitoring and analyzes a copy of inline network packets (fetched through port mirroring) to make sure the streaming traffic is not malicious or spoofed in any way. The IDS efficiently detects infected elements with the potential to impact your overall network performance, such as malformed information packets. DNS poisonings Xmas scans, and more.

Result and Discussion

The CSE-CIC-IDS2018 on AWS dataset I have implemented in Two ways

1)Multiple classification problem

2)Binary classification problem

Multiple classification problem

A classification predictive modeling problem where all example belong to one or three classes, that means those data set who contain more than two target or labels. In CSE-CIC-IDS2018 on AWS dataset target or label had Benign and DDoS, Bot, etc. so these data set was classified as classification problem and these data set are purely imbalanced. So for the balancing data set I have used Random over sampling.

According to these result after analysis. I have shown on some result of multi classification problem in Table 1 given on below.

Multiclass classification problem

Table 1

	scores	Precision	Recall	f1_score	Accuracy (%)
BernoulliNB	0.545333	0.545333	0.545333	0.545333	54.533333
DecisionTreeClassifier	0.948667	0.948667	0.948667	0.948667	94.866667
GaussianNB	0.940444	0.940444	0.940444	0.940444	94.044444
KNeighborsClassifier	0.972222	0.972222	0.972222	0.972222	97.222222
LogisticRegression	0.922667	0.922667	0.922667	0.922667	92.266667
MultinomialNB	0.645778	0.645778	0.645778	0.645778	64.577778
RandomForestClassifier	0.981778	0.981778	0.981778	0.981778	98.177778
SGDClassifier	0.927111	0.927111	0.927111	0.927111	92.711111
SVC	0.948667	0.948667	0.948667	0.948667	94.866667
XGBClassifier	0.995111	0.995111	0.995111	0.995111	99.511111

Binary Class Classification Problem

A classification predictive modeling problem where all example belongs to one of two classes, ie. 0 and 1 or True or False.

According to these result after analysis. I have shown some result of Binary class classification problem. In table 2.

Table 2

	scores	Precision	Recall	f1_score	Accuracy (%)
BernoulliNB	0.803794	0.803794	0.803794	0.803794	80.379353
DecisionTreeClassifier	0.987562	0.987562	0.987562	0.987562	98.756219
GaussianNB	0.838464	0.838464	0.838464	0.838464	83.846393
KNeighborsClassifier	0.995025	0.995025	0.995025	0.995025	99.502488
LogisticRegression	0.980100	0.980100	0.980100	0.980100	98.009950
MultinomialNB	0.803794	0.803794	0.803794	0.803794	80.379353
RandomForestClassifier	0.990672	0.990672	0.990672	0.990672	99.067164
SGDClassifier	0.974502	0.974502	0.974502	0.974502	97.450249
SVC	0.995802	0.995802	0.995802	0.995802	99.580224
XGBClassifier	0.998601	0.998601	0.998601	0.998601	99.860075

Advantage and Disadvantage of Intrusion detection system

Advantage

- Analyzes on going traffic, activity, transection and behavior for anomalies.
- Potential to detect perversely unknown types of attacks.
- CatLog the difference between baseline behavior and on going activity.

Disadvantage

- Prone to false positive
- Heavy processing overhead
- Vulnerable to attack while creating time consuming statically significant base line.

Conclusion

Intrusion detection system look for attacks signatures which are specific patterns that usually indicate malicious or suspicious intent. IDS schemes for detecting various novel attacks rather than individual instantiations.