International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

Detection of Cyber Attack in Network Using Machine Learning Techniques

Bhuvaneswari N[1], Jaya varshni N[1], Prabu M[1], Tarshana A[1], Dr. S. Jothi lakshmi[2]

[1]UG Scholar, Department of Computer Science and Engineering, Akshaya College of Engineering and Technology, Coimbatore, Tamil Nadu, India

[2]Assosiate Professor, Department of Computer Science and Engineering, Akshaya College of Engineering and Technology, Coimbatore, Tamil Nadu, India

Article Info
Volume 9, Issue 3

Page Number : 357-360

Publication Issue :
May-June-2022

Article History
Accepted : 15 May 2022
Published: 30 May 2022

ABSTRACT

Intrusion Detection is one of network security area of technology main research directions. Data mining technology will be applied to Network Intrusion Detection System (NIDS), may automatically discover the new pattern from the massive network data, to reduce the workload of the manual compilation intrusion behavior patterns and normal behavior patterns. This article reviewed the current intrusion detection technology and the data mining technology briefly. Focus on data mining algorithm in anomaly detection and misuse detection of specific applications. For misuse detection, the main study the classification algorithm; For anomaly detection, the main study the pattern comparison and the cluster algorithm. In pattern comparison to analysis deeply the association rules and sequence rules. Finally, has analyzed the difficulties which the current data mining algorithm in intrusion detection applications faced at present, and has indicated the next research direction.

Keywords: IDS, Network, Cyber Security, Cyber Attacks

I. INTRODUCTION

Intrusion detection plays a vital role in the network defense process by aiming security administrators in forewarning them about malicious behaviors such as intrusions, attacks, and malware. IDS are a mandatory line of defense for protecting critical networks against ever evolving issues of intrusive activities. Research in IDS has, hence, flourished over the years. An Intrusion detection system or IDS is a system developed to monitor for suspicious activity and issues alerts when such activity is discovered. The primary aim of IDS is to detect anomalous activities, but some systems are also able to take action against these intrusions like blocking traffic from the suspicious IP address. An IDS can also be used to help analyze the quantity and types of attacks. Organizations can use this data to change their security systems or implement more effective systems. An IDS can also help organizations to identify bugs or problems with their network device configurations. These metrics can then be used to predict future risks . On the down side, IDS is also prone to giving false alarms, this is in part due to insufficient data in data sets. Intrusion detection systems detect these intrusions in different ways.

- A network intrusion detection system (NIDS): It is deployed at strategic points in the network. It monitors inbound and outbound traffic from all devices on the network.
- Host intrusion detection systems (HIDS): It runs on all computers or devices in the network with direct access to both the internet and the enterprise internal network. HIDS can detect anomalous network packets that originate from inside the organization or malicious traffic that a NIDS fails to detect. It may also be able to identify malicious traffic that originates from the host itself, as when the host has been infected with malware and is attempting to spread to other systems.
- Signature-based intrusion detection systems: It monitors all the packets traversing the network and compares them against a database of signatures or attributes of known threats, similar to antivirus software.
- Anomaly-based intrusion detection systems: It monitors network traffic and compare it to an already established baseline, to determine what is considered 6normal for the network with respect to bandwidth, protocols, ports and other devices. This type of IDS alerts administrators to malicious activity.

II. PROBLEM STATEMENT

Many IDS suffer from high false positives and a lot of research goes into trying to reduce this high false positive rate. . We believe that intrusion detection is a data analysis process and can be studied as a problem of classifying data correctly. From this standpoint, it can also be observed that any classification scheme is as good as the data presented to it as input. Hence, it can be said that the cleaner the data is the more accurate the results will be. From this point of view, extracting specific features that contribute more to demarcating the normal data from the abnormal data will increase processing speed and efficiency as well as memory usage.

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on models and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers.

III. PROPOSED SYSTEM

With the enormous growth of computer networks usage and the huge increase in the number of applications running on top of it, network secrity is becoming increasingly more important. All the computer systems suffer from security vulnerabilities which are both technically difficult and economically costly to be solved by the manufacturers. Therefore, the role of Intrusion Detection Systems (IDSs), as special purpose devices to detect anomalies and attacks in the network, is becoming more important.

Traditionally, data analysis was always being characterized by trial and error, an approach that becomes impossible when data sets are large and heterogeneous. Machine learning comes as the solution to all this chaos by proposing clever alternatives to analyzing huge volumes of data. By developing fast and efficient algorithms and data-driven models for real-time processing of data, machine learning is able to produce accurate results and analysis.

In unsupervised learning, the algorithm builds a mathematical model of a set of data which contains only inputs and no desired outputs. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories, as in feature learning. Dimensionality reduction is the process of reducing the number of "features", or inputs, in a set of data.

An unsupervised learning method is a method in which we draw references from datasets consisting of input data without labeled responses. Generally, it is used as a process to find meaningful structure, explanatory underlying processes, generative features, and groupings inherent in a set of examples.

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups. It is basically a collection of objects on the basis of similarity and dissimilarity between them.

The research in the intrusion detection field has been mostly focused on a nomaly-based and misuse-based detection techniques for a long time. While misuse-based detction is generally favoured in commercial products due to its predictability and high accuracy, in academic research anomally detection is typically conceived as a more powerful method due to its theoretical potential for addressing novel attacks.

Conducting a through analysis of the recent research trend is anomaly detection, one will encounter several machine learning methods reported to have a very high detection rate of 98% while keeping the false alarm rate at 1%. However, when we look at the state of art IDS solutions and comercial tools, there is no evidence of using anomaly detection approaches, and practitioners still think that it is an immature technology. To find the reason of this contrast, lots of research was done in anomaly detection and considered various aspects such as learning, and detection approaches, training data sets, testing data sets, and evaluation methods.

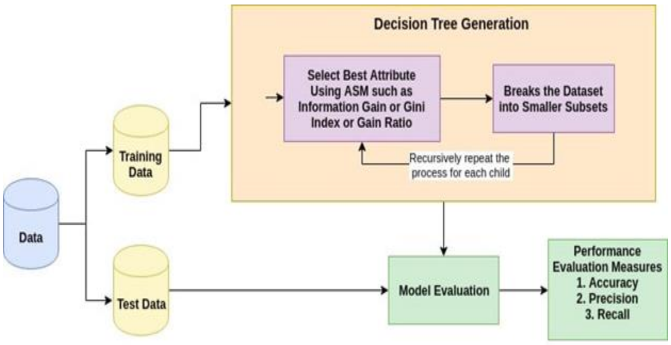
IV. DECISION TREE CLASSIFIER

- A decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.

Decision Tree algorithm work:

- Select the best attribute using Attribute Selection Measures(ASM) to split the records.
- Make that attribute a decision node and breaks the dataset into smaller subsets.
- Starts tree building by repeating this process recursively for each child until one of the condition will match:
  - All the tuples belong to the same attribute value.
  - There are no more remaining attributes.
  - There are no more instances.

Decision tree block diagram:


Decision Tree Generation
Select Best Attribute Using ASM such as Information Gain or Gini Index or Gain Ratio
Breaks the Dataset into Smaller Subsets
Recursively repeat the process for each child
Training Data
Data
Test Data
Model Evaluation
Performance Evaluation Measures
1. Accuracy
2. Precision
3. Recall

Logistic Regression

Classification techniques are an essential part of machine learning and data mining applications. Approximately 70% of problems in Data Science are classification problems. There are lots of classification problems that are available, but the logistics regression is common and is a useful regression method for solving the binary classification problem. Another category of classification is Multinomial classification, which handles the issues where multiple classes are present in the target variable. Logistic Regression can be used for various classification problems such as spam detection. Diabetes prediction, if a given customer will purchase a particular product or will they churn another competitor, whether the user will click on a given advertisement link or not, and many more examples are in the bucket.

Block diagram:

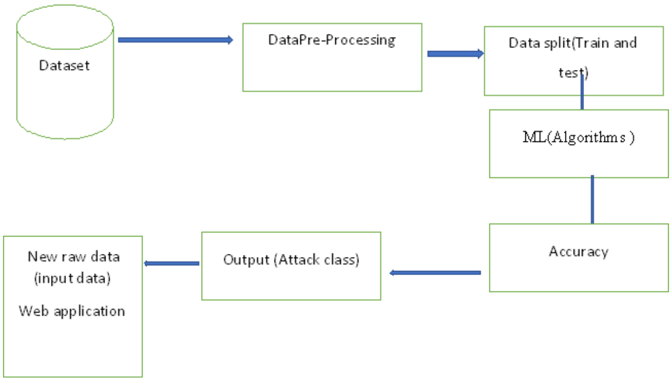


Figure 1. Block Diagram

V. CONCLUSION

Data mining algorithms is available, intrusion detection based on data mining has developed rapidly. It advances in the ability to handle massive data, but it also has problems like, for instance, searching for more effective data mining algorithms, how to improve the correct rate of intrusion detection, how to control the rate of false alarm in anomaly detection and etc. These can be the topics for future research, meanwhile they also need lots of work and experiments to develop a system that is more effective and more appropriate. There are many types of techniques in intrusion detection, in which that based on data mining becomes the hot spot in the current intrusion detection technology

VI. REFERENCES

[1]. Zhen-Ya Zhang, Hong-Mei Cheng, et al. From data mining to Opportunity / symptoms found. Computer Science. 2007.

[2]. Dai Yingxia, lian Yi-feng, Wang hang. Security and intrusion detection systems [M]. Beijing: Tsinghua University Press, 2002.

[3]. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4]. XuZhuoqun. Construction of data [M]. Beijing: China Broadcast television University Publishing house, 2001: 260- 272.

[5]. Liu xin, Zhang Yongping, Wanyanli. Decision Tree Algorithm in Intrusion Detection System Analysis and Improvement. Computer Engineering and Design 2006 .

Cite this Article

Bhuvaneswari N, Jaya varshni N, Prabu M, Tarshana A, Dr. S. Jothi lakshmi, "Detection of Cyber Attack in Network Using Machine Learning Techniques", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 3, pp. 357-360, May-June 2022.

Journal URL : https://ijsrset.com/IJSRSET2293144 |