Volume 7, Issue 1, January 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Intrusion Detection System and Its Attacks Detection: Comparative

**Apoorv Singh Tomar, Dr. Brijesh Kumar Chaurasia**
Computer Science Department, ITM University, Gwalior,
Madhya Pradesh, India

DOI: 10.23956/ijarcsse/V7I1/0147

***Abstract: Interruption Detection is an issue of distinguishing unapproved clients in a PC framework. It is likewise characterized as the issue of shielding PC organizes frameworks from being bargained. The initially distributed famous writing on PC organize security is where Denning examined different security concerns, exhibited a meaning of Intrusion Detection and talked about various sorts of Intrusion Detection. An interruption identification framework is programming as well as equipment intended to identify unapproved endeavors at getting to, controlling, or potentially crippling of PC framework, principally through a system, for example, the web. One of the fundamental difficulties in the security administration of expansive scale rapid systems is the location of irregularities in system activity.***

***Keywords: Intrusion detection, IDS attacks, Data Mining, Functionality, Techniques***

## I. INTRODUCTION

Intrusion Detection is a problem of identifying unauthorized users in a computer system. It is also defined as the problem of protecting computer network systems from being compromised. The first published renowned literature on computer network security is [2] where Denning discussed various security concerns, presented a definition of Intrusion Detection and discussed different types of Intrusion Detection.

An intrusion detection system is software and/or hardware designed to detect unauthorized attempts at accessing, manipulating, and/or disabling of computer system, mainly through a network, such as the internet. One of the main challenges in the security management of large-scale high-speed networks is the detection of anomalies in network traffic.

***A secure network must provide the following:***

- Confidentiality: Data that are being transferred through the network should be accessible only to those that have been properly authorized.
- Integrity: Data should maintain their integrity from the moment they are transmitted to the moment they are actually received. No corruption or data loss is accepted either from random events or malicious activity.
- Availability: The network should be resilient to Denial of Service attacks.

An *intrusion detection system* is software that automates the intrusion detection process. It can be defined as security systems that can identify attempted or ongoing attacks on a computer system or network. Developing reliable and efficient intrusion detection system that will timely and accurately detect intrusions is challenging. However, it is becoming a necessary security tool in industry. Every year, businesses lose a huge amount of revenue due to improper data manipulation caused by computer network intruders.

Ideally, intrusion detection system should have an attack detection rate (DR) of 100% along with false positive (FP) of 0%. Nevertheless, in practice this is really hard to achieve. The most important parameters involved in the performance estimation of intrusion detection system are shown in Table 1.

Table1

| Parameters | Definition |
|---|---|
| True Positive (TP) or Detection Rate (DR) | Attack occur and alarm raised |
| False Positive (FP) | No attack but alarm raised |
| True Negative (TN) | No attack and no alarm |
| False Negative (FN) | Attack occur but no alarm |

 

Detection rate (DR) and false positive (FP) are used to estimate the performance of intrusion detection system [17], which is given as bellow:

$$DC = \frac{Total\ Detected\ Attacks}{Total\ Attacks} \times 100$$

$$FP = \frac{Total\ misclassified\ process}{Total\ Normal\ Process} \times 100$$

## II. INTRUSION DETECTION SYSTEM

Intrusion detection is a process which is used to identify the intrusion, and is based on the belief that the intruder behavior will be significantly different from the legitimate user. Intrusion Detection System (IDS) are usually deployed along with other preventive security mechanisms, such as access control and authentication, as a second line of defense that protects information systems. There are several reasons that make intrusion detection a necessary part of the entire defense system. First, many traditional system and applications were developed without security in mind. In other cases, system and applications were developed to work in a different environment and may become vulnerable when developed in the current environment.

***Misuse/Signature detection:*** misuse detection catches intrusions in terms of the characteristics of known attacks. Any action that conforms to the pattern of a known attack or vulnerability is considered as intrusive. The main issues in misuse detection system are how to write a signature that encompasses all possible variations of the pertinent attack. And how to write signatures that do not also match non-intrusive activity. In This paper, we transform flow records into datasets with a small number of features for predefined time intervals and service-specific port numbers. Our goal is to identify time intervals showing anomalous traffic behavior, as it may be caused by network malfunctions or malicious attack traffic. The processing steps of our approach can be summarized as follows:

1) Training data containing flow records of both normal and anomalous traffic are transformed into feature datasets.
2) The datasets are divided into different clusters for normal and anomalous traffic using the K- means clustering algorithm.
3) The resulting cluster centroids are deployed for fast detection of anomalies in new monitoring data based on simple distance calculations.

While clustering monitoring data and identifying anomalies based on outlier detection has already been tried before, we are not aware of previous attempts generating additional clusters for anomalous traffic as we do.

***Types of IDS****:* Figure 1 shows the different types of Intrusion detection

Host based IDS

Network based IDS

Application based IDS

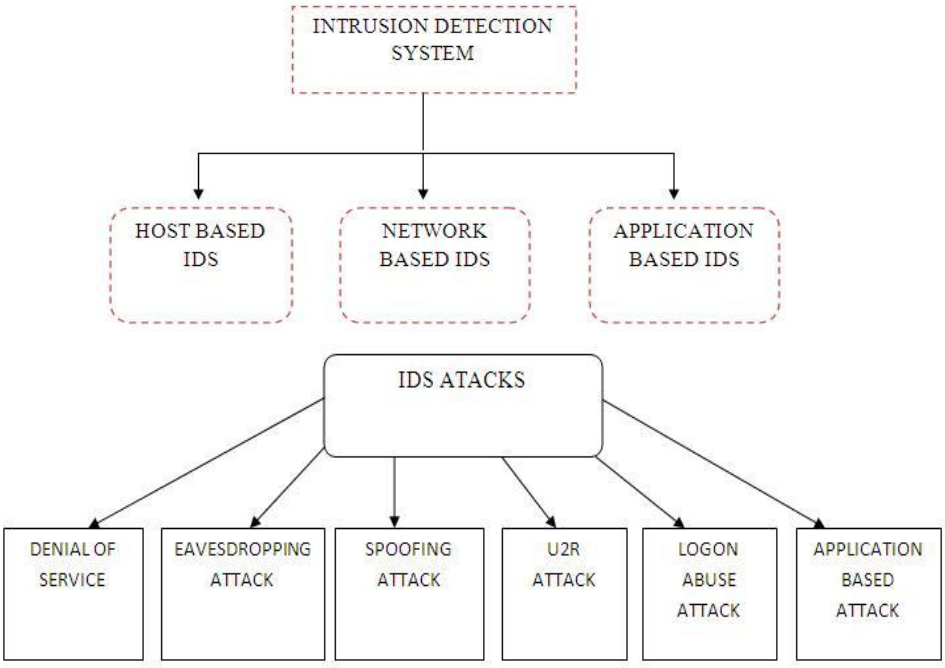


Fig 1: Intrusion Detection Attacks

***Functions of IDS:*** The IDS consist of four key functions namely, data collection, feature selection, analysis and action, which is given in Figure 2.

 

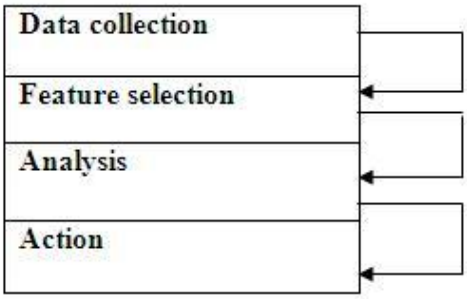


Fig:2

## III. LITERATURE SURVEY

The idea of intrusion detection system has been introduced by James Anderson in 1980. And since then so much work has been done by various researches. There are a numerous artificial intelligence techniques incorporated in the development of intrusion detection system. Data Mining techniques have proved as a very significant approach for the enhancement of performance of intrusion detection system.

In 1998, Wenke Lee and Salvatore J. Stolfo integrated Data mining technique in intrusion Detection. Two data mining technique i.e. association rule mining and frequent episode mining were implemented for the development of framework which are essential in describing program or user behavior, sendmail system call data and network tcpdump data was used for the construction of the detection model and performance evaluation. Perliminary experiments showed the promising results. The main problem faced by this model was the high false alarm rate.

***Association Rules:*** Data mining approaches for intrusion detection include association rules and frequent episodes, which are based on building classifiers by discovering relevant patterns of program and user behavior. Association rules and frequent episodes are used and the features can be defined in the form of packet and connection details. However, mining of features is limited to entry level of the packet and requires the number of records to be large and sparsely populated; otherwise, they tend to produce a large number of rules that increase the complexity of the system

***Decision Tree:*** The decision trees select the best features for each decision node during the construction of the tree based on some well defined criteria. One such criterion is to use the information gain ratio. Decision trees generally have very high speed of operation and high attack detection accuracy even if dealing with a large amount of data.

***Fuzzy Logic:*** A set of rules can be created to describe a relationship between the input variables and the output variables, which may indicate whether an intrusion occurred. In 2012, Li Hanguang, and Ni Yu discussed an intrusion detection technology research based on Apriori algorithm. The author used Apriori algorithm which is the classic of association rules in Web-based Intrusion Detection System and applied the rule base generated by the Apriori Agorithm to identify a variety of attacks, improving the overall performance of the detection System. The author proposed an improved Apriori algorithm for intrusion detection. Improved algorithm without traversing the database computes support but the algorithm complexity increases while the data is huge taking up considerable memory and processor resources, computation time is not significantly improved.

***Proposed Framework for Intrusion Detection***

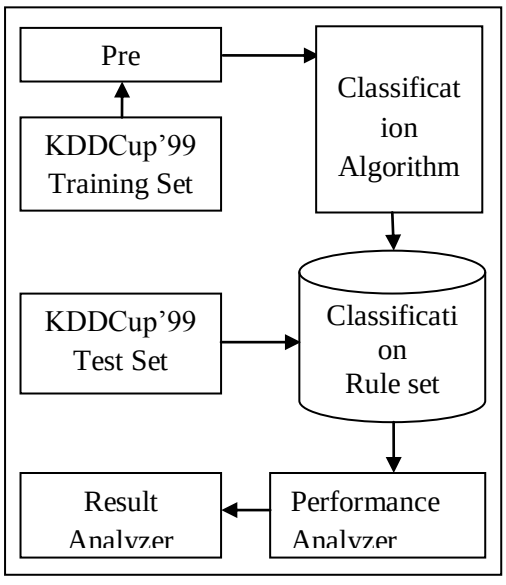


Fig:3

Intrusion detection system by using decision tree classification

 

***Data Preprocessing and Normalization***

$$f_i = normalized_{f_i} = I(X, Y_{f_i}) \frac{f_i}{\max(F_i)}$$

## IV. PROPOSED WORK

***J48 Algorithm:*** The attribute with the highest information gain (or greatest entropy reduction) is chosen as the test attribute for the current node. Such an information-theoretic approach minimizes the expected number of tests needed to classify an object and guarantees that a simple (but not necessarily the simplest) tree is found.

Imagine selecting one case at random from a set S of cases and announcing that it belongs to some class $C_j$. The probability that an arbitrary sample belongs to class $C_j$ is estimated by

$$p_i = \frac{freq(C_j, S)}{|S|}$$

***Performance analysis:*** Ideally, intrusion detection system should have an attack detection rate (DR) of 100% along with false positive (FP) of 0% Nevertheless, in practice this is really hard to achieve. The most important parameters involved in the performance estimation of intrusion detection system are shown in Table 2

Table 2

| Parameters | Definition |
|---|---|
| True Positive (TP) or Detection Rate (DR) | Attack occur and alarm raised |
| False Positive (FP) | No attack but alarm raised |
| True Negative (TN) | No attack and no alarm |
| False Negative (FN) | Attack occur but no alarm |

## V. EXPERIMENTATION AND PERFORMANCE ANALYSIS

***KDDCup'99 Dataset:*** This is the data set used for The Third International Knowledge Discovery and Data Mining Tools Competition, which was held in conjunction with KDD-99 The Fifth International Conference on Knowledge Discovery and Data Mining [9]. The competition task was to build a network intrusion detector, a predictive model capable of distinguishing between ``bad'' connections, called intrusions or attacks, and ``good'' normal connections [9]. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment.

In 1998, DARPA intrusion detection evaluation program, a simulated environment was set up to acquire raw TCP/IP dump data for a local-area network (LAN) by the MIT Lincoln Lab to compare the performance of various intrusion detection methods [9]. It was operated like a real environment, but being blasted with multiple intrusion attacks and received much attention in the research community of adaptive intrusion detection. In KDD99 dataset [9], each example represents attribute values of a class in the network data flow, and each class is labeled either normal or attack.

The classes in KDD99 dataset [9] can be categorized into 5 main classes (one normal class and four main intrusion classes: probe, DOS, U2R, and R2L) [4][6].

1) *Normal* connections are generated by simulated daily user behaviour such as downloading files, visiting web pages [4].
2) *Denial of Service (DoS)* attack [6] causes the computing power or memory of a victim machine too busy or too full to handle legitimate requests. DoS attacks are classified based on the services that an attacker renders unavailable to legitimate users like apache2, land, mail bomb, back, etc.
3) *Remote to Local (R2L)* [6] is an attack that a remote user gains access of a local user/account by sending packets to a machine over a network communication, which include send-mail, and Xlock.
4) *User to Root (U2R)* [6] is an attack that an intruder begins with the access of a normal user account and then becomes a root-user by exploiting various vulnerabilities of the system. Most common exploits of U2R attacks are regular buffer-overflows, load module, Fd-format, and Ffb-config.
5) *Probing (Probe)* [6] [8] is an attack that scans a network to gather information or find known vulnerabilities. An intruder with a map of machines and services that are available on a network can use the information to look for exploits.

In 1999, the original TCP dump files were preprocessed for utilization in the Intrusion Detection System benchmark of the International Knowledge Discovery and Data Mining Tools Competition [2][9]. To do so, packet information in the TCP dump file is summarized into connections. Specifically, "a connection is a sequence of TCP packets starting and ending at some well defined times, between which data flows from a source IP address to a target IP address under some well defined protocol" [14]. This process is completed using the Bro intrusion detection system [13], resulting in 41 features for each connection.

 

Features are grouped into four categories:

- *Basic Features:* Basic features can be derived from packet headers without inspecting the payload. Basic features are the first six features listed in Table 5.2[9].
- *Content Features*: Domain knowledge is used to assess the payload of the original TCP packets. This includes features such as the number of failed login attempts[9];
- *Time-based Traffic Features*: These features are designed to capture properties that mature over a 2 second temporal window. One example of such a feature would be the number of connections to the same host over the 2 second interval[9];
- *Host-based Traffic Features*: Utilize a historical window estimated over the number of connections – in this case 100 – instead of time. Host based features are therefore designed to assess attacks, which span intervals longer than 2 seconds [9].

**H-Tree based IDS**

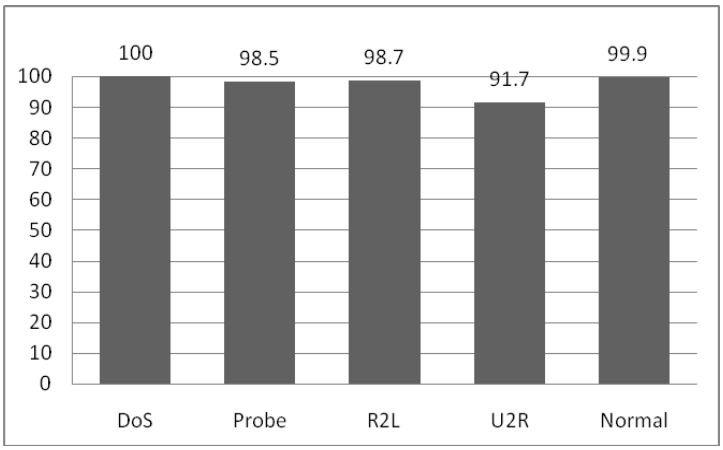


Fig 4: Accuracy comparison graph

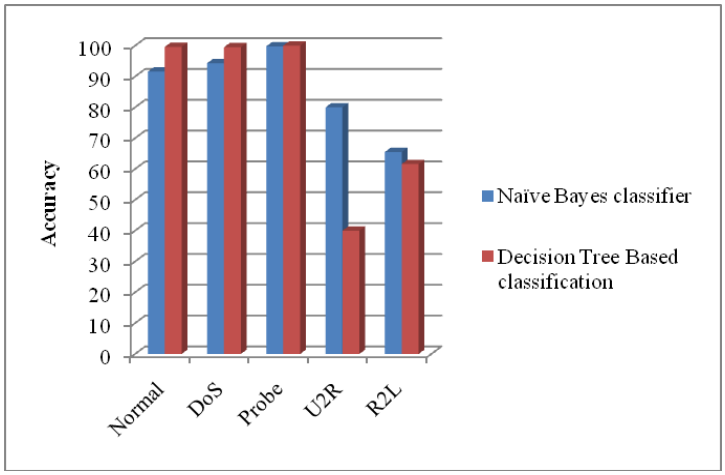


Fig:5 Accuracy comparison graph by using training data set

## VI. IDS IN DATA MINING

Data mining is the process of extracting the hidden knowledge from the databases. IDS are very important in data mining. Intrusion detection includes identifying a set of malicious actions that compromise the integrity and availability of information resources [22].

Intrusion detection in data mining has two divisions, they are, misuse detection and anomaly detection. In misuse detection the labeled data are built using anticipating model [23]. In anomaly detection there is a deviation between models. To use the data first it should be converted featured data and the data mining models are applied to it and they are summarized to produce the result

***TECHNICAL CHALLENGES***

Large data size

Higher dimensionality Data preprocessing

## VII. CONCLUSION AND FUTURE WORK

The suggested approach called Decision Tree Based classification is evaluated and compared with the single Naïve Bayes classifier using KDD Cup ’99 data set. The experimental results show that the k Decision Tree Based classification approach achieves better accuracy and detection rates while reducing the false alarm by detecting novel

 

intrusions accurately. The performance of Naïve Bayes classifier has been improved by applying Decision Tree Based classification. However, Decision Tree Based classification has limitation to detect intrusions that are very similar with each other such as U2R and R2L.

Many recommendations can be proposed for the future directions of works in this area can be summarized as follows:

- Need to effectively detect the U2R and R2L types of attacks.
- Need to suggest preparation for handling intrusion and fraud detection at application level.
- Theses system need to automatic responding to any detected intrusion.
- The KDDCup'99 dataset requires a revision in order to learn about the existing unknown types of attack.
- Put and test all previous models in the real world.
- To make the previous models as general as possible, the training data set must be as variant as much as possible.

Since U2R and R2L attacks are primary attack strategies used by attackers, honey net like techniques can be considered for the future work

## ACKNOWLEDGMENT

The work was carried out M. Tech at Institute of Technology and Management University, Gwalior under guidance of Dr. Brijesh Kumar Chaurasia. I am very thankful to Institute of Technology and Management University, Gwalior for giving me opportunity for doing this work.

## REFERENCES

[1] Lee, W, Stolfo S and Mok K , "Adaptive Intrusion Detection: A Data Mining Approach," In Artificial Intelligence Review, Kluwer Academic Publishers, 14(6), pp. 533 - 567, December 2000.

[2] L. Portnoy, "Intrusion Detection with Unlabeled Data Using Clustering," Undergraduate Thesis, Columbia University, 2000

[3] William Stallings, Cryptography and Network Security: Principles and Practices, Pearson Education, 4th Edition, 2011

[4] Debar H., Becker M., Siboni D., "A Neural Network Component for an Intrusion Detection System". Proceedings of the 1992 IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, CA May 1992

[5] Mé L. et V. Alanou. Intrusion detection: A bibliography. Technical Report SSIR-2001-01, Supélec, Rennes, France, September 2001.

[6] Chundong Wang, Quancai Deng, Qing Chang,Hua Zhang and Huaibin Wang " A New Intrusion Detection System Based on Protocol Acknowledgement" IEEE 2010

[7] Lawrence R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, 1989.

 