Decision tree based Intrusion Detection System

Apoorvi Nagar[1], Ritu Chauhan[2]
[1, 2] Dept of Computer Science and Engineering
[1, 2] ITM University

Abstract- *An intrusion detection method is software that automates the intrusion detection process. It can be defined as security systems that can identify attempt or ongoing attacks on a computer system or network. Rising consistent and efficient IDS that will correct and accurately detect intrusion in testing. However, it becomes a necessary security tool in industry. Each and every year, business loses a huge amount of proceeds due to improper data management caused by computer network intruders. If possible, IDS should have an attack detection rate (DR) of 100% along with false positive (FP) of 0%. Even so, in practice this is in reality hard to achieve. The mainly essential parameters involved in the performance estimation of intrusion detection*

Keywords- J48, fuzzy logic, neural network, attacks, decision tree, IDS-intrusion detection system.

I. INTRODUCTION

Intrusion detection is a method of examine the events stirring in a computer system or network and analyzes them for symbols of possible incident, which are violations or standard security practices. Occurrence have many cause, such as malware, unauthorized user attack and access to system from the Internet, and authorized users who exploitation their privileges or attempt to develop additional privileges for which they are not authorized. Although many incidents are malicious in nature, several others are not; for example, a person might mistype the address of a computer and accidentally attempt to tie to a different system without authorization. It is the method to identify those who are using computer network assets lacking authorization or attempting to prevent authorized users from accessing network resources. In an organization, intrusion can take place from the internet and inside the organization’s computer network system.

These things to see the two different types of IDS; Host Based and Network Based Intrusion Detection System.

A Host Based IDS can be defined as a security system that is able of detecting inside exploitation in a computer network.

A Network Based IDS is competent of identify unpleasant uses or attempts of unauthorized procedure of the computer network from outside of the system.

There are quite few forms of network intrusions:

A. Denial-of-service Attack

It a serious form of attack that is developed in compensation worth millions of dollars above the earlier period. While a remarkable problem, it is normally quite simple. They typically engage an attacker disable or rendering difficult to get to a network-based information resource.

B. Denial-of-service Attack

It a serious form of attack that is developed in compensation worth millions of dollars above the earlier period. While a remarkable problem, it is normally quite simple. They typically engage an attacker disable or rendering difficult to get to a network-based information resource.

C. Guessing rlogin Attack

The intruders try to assumption the password that protects the computer network in sort to expand access to it.

D. Scanning Attacks

The intruders try to scan different ports of the victim’s system to find some vulnerable points from where they can launch other attacks.

II. RELATED WORK

Hidden Markov Model: In [1], [2], [3] to detect irregular traces of system call in honored processes Hidden Markov Model are applied. However, modeling system un-accompanied does not provide accurate classification always in such cases various connection level features are ignored. Further, it is generative systems and fails to model long-range dependency between the observations.

Decision Tree: In [3], [4] duration the construction of the tree based some well defined criteria are construct in this process the decision tree choose the best features for each decision node. One such norm is to use the information gain ratio. Generally decision tree have very high speed process and also high accuracy to detect attack even if dealing with a bulky amount of data.

Genetic Algorithms (GAs): In [3], [4] Genetic algorithms imitate the natural reproduction system. In nature where only the fittest entity in a generation will be reproduce in subsequent generations, after undergoing recombination and random change.

Support Vector machine (SVMs): [5] while the neural networks can work resourcefully with noisy data, it necessitate huge amount of data for training and also frequently firm to pick the best achievable architecture for a neural network. It is used to intrusion detection and also map real valued input aspect vector to a higher dimensional. Characteristic gap through non-linear mapping it can provide real time detection capability and deal with large dimensionality of data.

Fuzzy Logic: A [8] set of laws can be formed to describe a connection between the input and output variables, which may indicate whether an intrusion occurred.

III. RESEARCH METHODOLOGY

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

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

A. Probability of Detection

In a given environment during a particular time frame, probability of detection determines proper the rate of attacks by IDS [7]. The difficulty in evaluate is detection rate that is the achievement of an intrusion detection system is largely dependent upon the place of attacks used during the test. Also, the probabilities of recognition vary with the false positive rate, and IDS can be configure or tuned to favor either the ability to identify attacks or to minimize false positives. One has to be cautious to use the same configuration during testing for false positives and hit rates.

Further, a network IDS can be evaded by surreptitious version of attacks. A network IDS may detect an attack when it is launched in a simple straight forward manner, but not when even simple approaches to stealthiest are used. Techniques used to make attacks surreptitious include fragmenting packets, using various types of data encoding, using unusual TCP flags, encrypting attack packets, spreading attacks over multiple network sessions, and launching attacks from various sources.

B. Resistance to attacks directed at the intusion detection

These measurements demonstrate how resistance IDS is to an attacker's attempt to interrupt the accurate operation of the intrusion detection system [7]. Attacks against IDS may take the form of:

1) Sending a outsized amount of non-attack traffic with volume above the IDS processing capability. With too much traffic to process, IDS may drop packets and be not capable to detect attacks.

2) Sending to the intrusion detection system non-attack packets that are specially craft to generate many signatures within the intrusion detection system, thereby crushing the intrusion detection system's human operator with false positives or crashing alert processing or display tools.

3) Sending to the intrusion detection system a huge number of attack packets intended to distract the intrusion detection system's human operator while the attacker instigates a real attack hidden under the "smokescreen" created by the multitude of other attacks.

4) Sending to the IDS packets containing data that exploit liability within the IDS processing algorithms. Such attacks will only be successful if the intrusion detection system contains a known coding error that can be exploited by a clever attacker. Fortunately, very few Intrusion detection system have had known exploitable buffer overflows or other vulnerabilities.

C. Ability to handle high Bandwidtg Traffic

These dimensions demonstrate how intrusion detection system will gathering when accessible with large dimensions of traffic [7]. This measurement is almost identical to the "resistance to denial of service measurement" when the attacker sends a bulky amount of non-attack traffic to the IDS. The only difference is that this measurement calculates the ability of the intrusion detection system to handle particular volumes of normal background traffic.

D. Ability to corelate events

In this dimension demonstrate how well an intrusion detection system correlates attack events [6] [7]. These events may be gathering from intrusion detection system, routers, firewalls, application logs, or an extensive variety of other devices. For the most component of this association is to recognize theatrical diffusion attacks. Currently, Intrusion detection system has limited capability in this area.

E. Ability to Detect Never Before Seen Attacks

For commercial systems, [6] [7] generally it is not useful to get this measurement since their signature-based technology can only identify attacks that had occurred previously (with a few exceptions). However, research systems based on irregularity detection or requirement based approach may be suitable for this type of measurement. Usually systems detecting attacks that had never been detected before produce more false positives than those that do not have this feature.

F. Ability to Identify an Attack

Demonstrate how well IDS can recognize the attack that is detected by tagging each attack with a frequent name or vulnerability name or by assigning the attack to a category [7] [10].

G. Ability to Determine Attack Success

If IDS can resolve the achievement of attacks from distant sites that gives the attacker higher- level privileges on the attacked system[7][9]. In current network environment, several remote privilege-gaining attacks (or probes) fail and do not damage the system attack. Many Intrusion detection systems, however, don't distinguish the failed from the successful attacks. For the same attack, some IDS can detect the evidence of damages (whether the attack has succeed) and some Intrusion detection system detect only the signature of attack actions (with no warning whether the attack succeeded or not). The capability to resolve attack achievement is essential for the analysis of the attack correlation and the attack scenario; it also greatly simplifies an analyst's work by distinguishing between more important successful attacks and the usually less damaging failed attacks. Measure this ability requires the information about failed attacks as well as successful attacks.

H. Capacity Verification for Network Intrusion Detection System

The network IDS demands higher- level protocol awareness than other network devices such as switches and routers [7] [9]; it has the skill of inspection into the deeper level of network packets. Therefore, it is significant to measure the capability of a network IDS to imprison, process and perform at the equivalent level of accuracy under a given network load as it does on a quiescent network. In network IDS clients can use the same capacity test results for each metric and a profile of their networks to determine if the network intrusion detection system is even capable of sustaining inspection of the traffic.

IV. SIMULATION AND RESULTS

Summary of overall measurement using training data set.

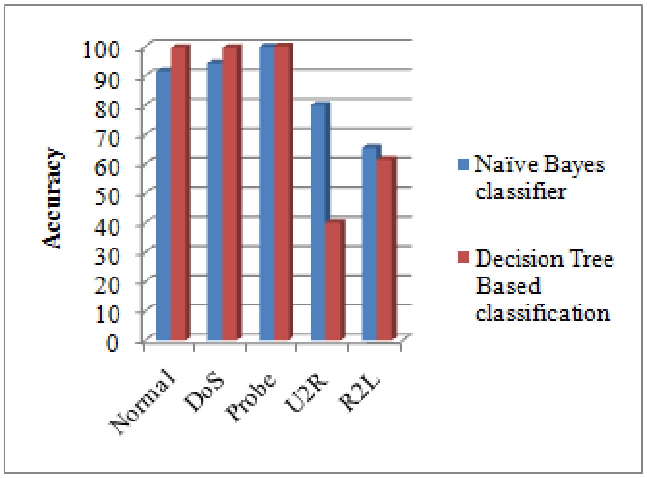

Fig 1: Summary of overall measurement using testing data set.

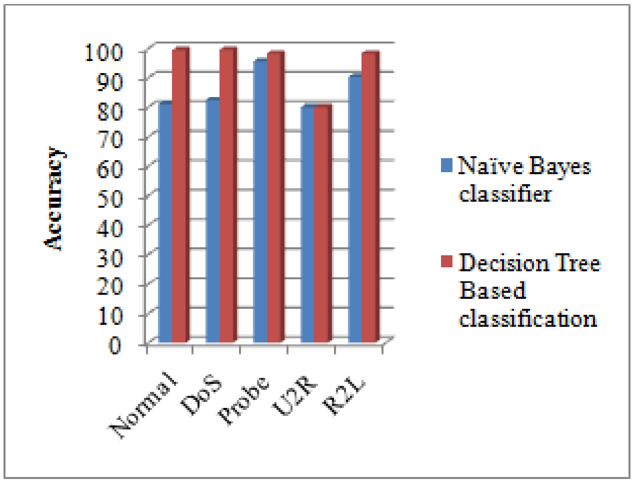

Fig 2: Comparison between Naïve Bayes classifier and Decision Tree Based classification.

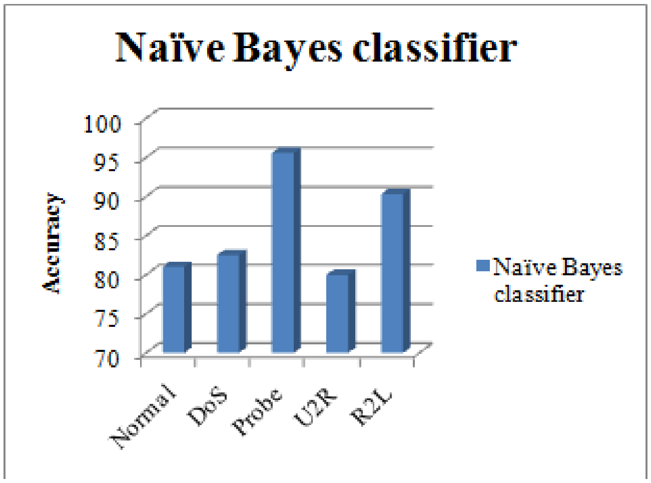

Fig 3: Graph of Naïve Bayes classifier.

V. CONCLUSION

The proposed approach Decision Tree Based classification is evaluated and compared with the single Naïve Bayes classifier using KDD Cup '99 data set. The new outcome show that the k Decision Tree Based classification approach achieves improved precision and detection rates while reducing the false alarm by detecting novel intrusions accurately. The show of Naïve Bayes classifier has been improved by applying Decision Tree Based classification. However, Decision Tree Based classification has limitation to identify intrusions that are very similar with each other such as U2R and R2L.

VI. FUTURE WORK

Many recommendations can be proposed for the upcoming work like:

- Put and test all previous models in the real world.
- To construct 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 upcoming work.

REFERENCES

[1] Richard Power, "1999 CSI/FBI Computer Crime and Security Survey," Computer Security Issues & Trends, Computer Security Institute, winter 1999

[2] Denning D E, "An Intrusion-Detection Model," In IEEE Transaction on Software Engineering, Vol. Se-13, No. 2, pp. 222-232, February 1987.

[3] 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.

[4] Satinder Singh, Guljeet Kaur, "Unsupervised Anomaly Detection In Network Intrusion Detection Using Clusters," Proceedings of National Conference on Challenges & Opportunities in Information Technology RIMT-IET, Mandi Gobindgarh. March 23, 2007.

[5] Eric Bloedorn , Alan D. Christiansen , William Hill , Clement Skorupka , Lisa M. Talbot , Jonathan Tivel, "Data Mining for Network Intrusion Detection: How to Get Started," CiteSeer, 2001

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

[7] Theodoros Lappas and Konstantinos Pelechrinis, "Data Mining Techniques for (Network) Intrusion Detection Systems," http://citeseerx.ist.psu.edu/viewdoc/ download?doi=10.1.1.120.2533&rep=rep1&type=pdf.

[8] Dewan Md. Farid, Nouria Harbi, Suman Ahmmed, Md. Zahidur Rahman, and Chowdhury Mofizur Rahman, "Mining Network Data for Intrusion Detection through Naïve Bayesian with Clustering", World Academy of Science, Engineering and Technology, 2010.

[9] The KDD Archive. KDD99 cup dataset, 1999. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[10] X. Li and N. Ye., "A supervised clustering algorithm for computer intrusion detection," Knowledge and Information Systems, 8, pp498-509, ISSN 0219-1377, 2005