IJFEAT

INTERNATIONAL JOURNAL FOR ENGINEERING APPLICATIONS AND TECHNOLOGY

DECISION TREE BASED INTRUSION DETECTION SYSTEM

Miss.Yogeshwari D.Hardas[1] Prof.Ankush Pawar[2]

[1] ME Scholar Computer Science, ARMIET, Sapgaon,9158770776,yohardas143@gmail.com

[2] ME Guide (H.O.D) Dept. Of Computer Science, DRIEMS, Neral,9960333280,ankushpawar1981@gmail.com

Abstract:

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

Keywords: j48, fuzzy logic, neural network, attacks, decision tree, intrusion detection system.

Introduction:

Intrusion detection is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible *incidents*, which are violations or imminent threats of violation of computer security policies, acceptable use policies, or standard security practices. Incidents have many causes, such as malware, attackers gaining unauthorized access to systems from the Internet, and authorized users of systems who misuse their privileges or attempt to gain additional privileges for which they are not authorized. Although many incidents are malicious in nature, many others are not; for example, a person might mistype the address of a computer and accidentally attempt to connect to a different system without authorization. It is the process of identifying individuals who are using computer network resources without authorization or attempting to prevent authorized users from accessing network resources. In an organization, intrusions can take place from the Internet or from inside the organization's computer network system. This highlights the two different types of Intrusion Detection Systems; Host Based Intrusion Detection System and Network Based Intrusion Detection System. A Host Based Intrusion Detection System can be defined as a security system that is capable of detecting inside abuses in a computer network. A Network Based Intrusion Detection System is capable of identifying abusive uses or attempts of unauthorized usage of the computer network from outside the system.

There are several forms of network intrusions:

• *Denial-of-service Attack* - This is particularly a serious form of attack that has resulted in damages worth millions of dollars over the past few years. While a significant problem, *Denial-of-service* attacks are usually quite simple. They typically involve an attacker disabling or rendering inaccessible a network-based information resource.

• *Guessing rlogin Attack* - Here the intruder tries to guess the password that protects the computer network in order to gain access to it.

• *Scanning Attacks* - The intruder goes about scanning different ports of the victim's system to find some vulnerable points from where they can launch other attacks

Literature review:

Hidden Markov Model: In [1], [2], [3] To detect anomalous traces of system calls in privileged processes Hidden Markov Model are applied. However, modeling the system alone may not always provide accurate classification as in such cases various connection level features are ignored. Further, HMMs are generative systems and fail to model long-range dependencies between the observations.

Decision Tree: In [3], [4] 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.

Support Vector machine (SVMs): In [41], [36] Though the neural networks can work efficiently with noisy data, they require large amount of data for training and it is often hard to select the best possible architecture for a neural network. Support vector machines have also been used for detecting intrusions. Support vector machine map real valued input feature vector to a higher dimensional feature space through nonlinear mapping and can provide real time detection capability, deal with large dimensionality of data, and can be used binary class as well as multiclass classification.

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

Fuzzy Logic: In [8] 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.

Research methodology:

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

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

Probability of Detection

This measurement determines the rate of attacks detected correctly by an intrusion detection system in a given environment during a particular time frame[34]. The difficulty in measuring the detection rate is that the success of an intrusion detection system is largely dependent upon the set of attacks used during the test. Also, the probability of detection varies with the false

positive rate, and an intrusion detection system can be configured or tuned to favor either the ability to detect attacks or to minimize false positives. One must be careful to use the same configuration during testing for false positives and hit rates.

Further, a network intrusion detection system can be evaded by stealthy versions of attacks. A network intrusion detection system may detect an attack when it is launched in a simple straightforward manner, but not when even simple approaches to stealthiness are used. Techniques used to make attacks stealthy 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 multiple sources.

Resistance to Attacks Directed at the intrusion detection system

This measurement demonstrates how resistant an intrusion detection system is to an attacker's attempt to disrupt the correct operation of the intrusion detection system [34]. Attacks against an intrusion detection system may take the form of:

1. Sending a large amount of non-attack traffic with volume exceeding the intrusion detection system's processing capability. With too much traffic to process, an intrusion detection system may drop packets and be unable to detect attacks.

2. Sending to the intrusion detection system non-attack packets that are specially crafted to trigger many signatures within the intrusion detection system, thereby overwhelming 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 large 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 intrusion detection system packets containing data that exploit vulnerability within the intrusion detection system 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.

Ability to Handle High Bandwidth Traffic

This measurement demonstrates how well an intrusion detection system will function when presented with a large volume of traffic [34]. Most network-based Intrusion detection system will begin to drop packets as the traffic volume increases, thereby causing the intrusion detection system to miss a percentage of the attacks. At a certain threshold, most Intrusion detection system will stop detecting any attacks. This measurement is almost identical to the "resistance to denial of service measurement" when the attacker sends a large amount of non-attack traffic to the intrusion detection system. The only difference is that this measurement calculates the ability of the intrusion detection system to handle particular volumes of normal background traffic.

Ability to Correlate Events

This measurement demonstrates how well an intrusion detection system correlates attack events [19] [34]. These events may be gathered from intrusion detection system, routers, firewalls, application logs, or a wide variety of other devices. One of the primary goals of this correlation is to identify staged penetration attacks. Currently, Intrusion detection system has only limited capabilities in this area.

Ability to Detect Never Before Seen Attacks

This measurement demonstrates how well an intrusion detection system can detect attacks that have not occurred before [19][34]. For commercial systems, it is generally not useful to take this measurement since their signature-based technology can only detect attacks that had occurred previously (with a few exceptions). However, research systems based on anomaly detection or specification-based approaches 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.

Ability to Identify an Attack

This measurement demonstrates how well an intrusion detection system can identify the attack that it has detected by labeling each attack with a common name or vulnerability name or by assigning the attack to a category[10][34].

Ability to Determine Attack Success

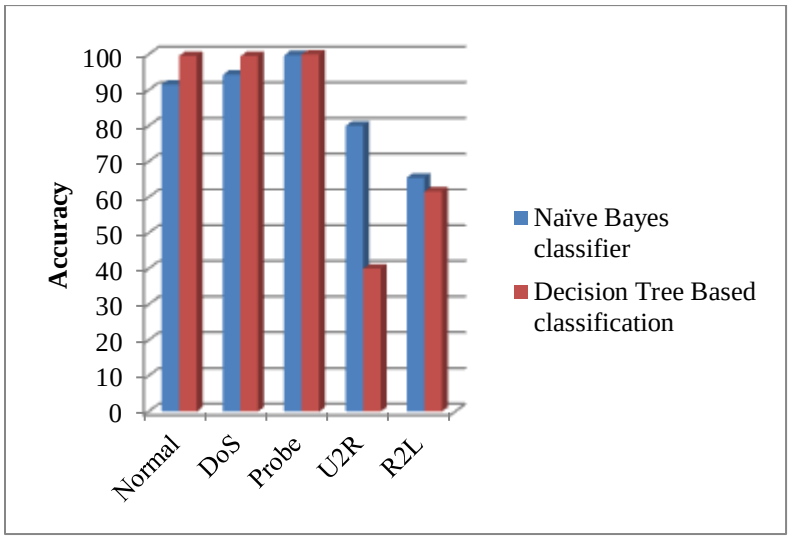
This measurement demonstrates if the intrusion detection system can determine the success of attacks from remote sites that give the attacker higher- level privileges on the attacked system[18][34]. In current network environments, many remote privilege-gaining attacks (or probes) fail and do not damage the system attacked. Many Intrusion detection system, however, do not distinguish the failed from the successful attacks. For the same attack, some Intrusion detection system can detect the evidence of damages (whether the attack has succeeded) and some Intrusion detection system detect only the signature of attack actions (with no indication whether the attack succeeded or not). The ability to determine attack success 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. Measuring this capability requires the information about failed attacks as well as successful attacks.

Capacity Verification for Network Intrusion Detection System

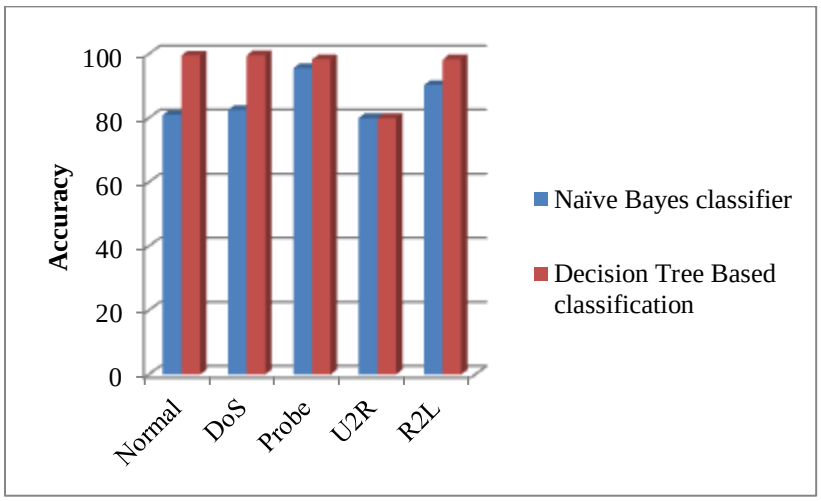
The network intrusion detection system demands higher- level protocol awareness than other network devices such as switches and routers [18][34]; it has the ability of inspection into the deeper level of network packets. Therefore, it is important to measure the ability of a network intrusion detection system to capture, process and perform at the same level of accuracy under a given network load as it does on a quiescent network. The network intrusion detection system customers can then use the standardized 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.

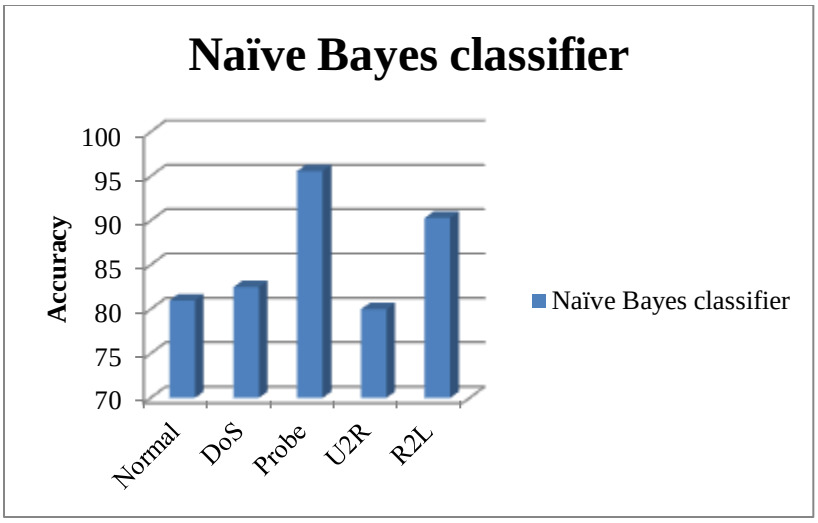
Simulation and results:

Summary of overall measurement using training data set



Summary of overall measurement using testing data set





Conclusion and Future Work

Conclusion

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.

Future Work

Many recommendations can be proposed for the future work like:

• 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.

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