ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Intrusion Detection System using Adaboost based approach and Fuzzy genetic algorithm

Tanmayee S. Sawant, Prof. Dr. S.A. Itkar

Student, Department of Computer Engineering, PES's Modern College of Engineering, Pune, India

Professor, Department of Computer Engineering, PES's Modern College of Engineering, Pune, India

ABSTRACT*:* Network intrusion detection systems (NIDS) have full-fledged to be a standard feature in security infrastructures. Insufficiently, current systems are underprivileged at detecting the unknown attacks. Data mining is a process to extract information and knowledge from a large number of incomplete, noisy, fuzzy and random data. As a key to this problem we put forward the application of classification techniques of data mining which can be applied to network connection data to detect the various types of attacks, In this paper, we developed an Intrusion Detection System to extend the detection rate of U2R and R2L attacks. We then evaluate existing systems, from commercial systems, to research based intrusion detection systems. Standard datasets and feature extraction turned out to be more important than we had initially expected, so each can be found under its own heading. Next, we review the actual data mining methods that have been implemented. We finish by summarizing the open problems in this area and proposing a new research project to respond some of these open problems.

KEYWORDS*:* Intrusion detection system, Adaboost algorithm and FGA.

I. INTRODUCTION

As network-based computer systems take part in progressively more elemental roles in modern society, they have grown to be the target of intrusions by our enemies and criminals. In addition to intrusion prevention techniques, such as user authentication and authorization, encryption, and defensive programming, intrusion detection is often used as another wall to shield computer systems. The two main intrusion detection techniques are misuse detection and anomaly detection. Misuse detection systems, for example, IDIOT and STAT, use patterns of well known attacks or weak spots of the system to match and identify known intrusions. Misuse detection techniques in broad-spectrum are not effective against new attacks that have no corresponding rules or patterns yet. Anomaly detection systems, for example, the anomaly detector of IDES, flag observed activities that diverge appreciably from the established normal usage profiles as anomalies, that is, possible intrusions. For example, the normal profile of a user may contain the averaged frequencies of some system commands used in his or her login sessions. If for a session that is being monitored, the frequencies are significantly lower or higher, then an anomaly alarm will be raised. Anomaly detection techniques can be effective against unknown or novel attacks since no a priori knowledge about specific intrusions is required. However, anomaly detection systems tend to generate more false alarms than misuse detection systems because an anomaly can just be a new normal behavior. Some IDSs, for example, IDES and NIDES use both anomaly and misuse detection techniques. Till time Intrusion detection systems have been implemented using various methods like data mining methods, neural networks, clustering algorithm, genetic algorithm and hybrid approach etc. Data mining classification algorithms such as decision tree, Bayes classifiers, k- nearest neighbour classifier, case based reasoning, fuzzy logic technique etc. Some researchers also implement it using layered approach furthermore.

II. RELATED WORK

In the paper [1] IDS for distributed environment is proposed and implemented using online Adaboost based approach combined with weak classifiers. This paper overcomes the difficulty of handling multi attribute network connection data with maintaining highest detection rate and accuracy.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Intrusion Detection with feature selection was accomplished to do better than the decision tree algorithm without feature selection. This establishment filtering is able to advance the classification capabilities of the decision tree in a shorter time [2]. Out of selected three features filter algorithm, it was found that Chi square and Information Gain was giving a better performance than ReliefF when KDD data set was taken. The work may be further extended by considering the four major attacks in the KDD data set.

S.Vijayarani, M.Divya analysed the performance of the three classification rule algorithms, namely C4.5, RIPPER and PART algorithms [3]. From the experimental results it is accomplished that in the case of time factor & number of rules generation, Part algorithm seems better than the other two algorithms for Breast Cancer Dataset and Heart Disease dataset.

Mrutyunjaya Panda, Manas Ranjan Patra [4] compares the performance of three well known data mining classifier algorithms namely, ID3, J48 and Naïve Bayes were evaluated based on the 10-fold cross validation test. Experimental results by means of the KDDCup'99 IDS data set conveys that Naïve Bayes is one of the most valuable inductive learning algorithms; decision trees are more remarkable as far as the detection of new attacks is concerned.

Some researchers predictable apriori algorithm [5] which scans the dataset only twice and builds FP-tree once while it still requests to generate candidate item sets.

Christian Borgelt described an implementation of the FPGrowth algorithm [6], which contains two methods for competently foretelling an FP-tree-the core operation of the FP-growth algorithm. The experimental results showed that, this implementation clearly outperforms Apriori and Éclat, even in highly optimized versions.

Implementation of Intrusion Detection System for disseminated environment using online Adaboost based approach combined with weak classifiers [7] defeat the complexity of handling multi attribute network connection data with maintaining highest detection rate and accuracy of different types of attacks.

Salem Benferhat el at projected a method for alert correlation and intrusion detection for revising decision tree classifiers outputs in order to fit the available expert knowledge. This work focused on an application of our revision procedures on a real world problem from the computer security field but the contributions of the paper can be valuable to any classification problem where a posterior probability distribution is output by the used classifier. It showed how to correct a probability distribution estimated from a decision tree classifier by a new probability distribution given by some expert knowledge and proposed a polynomial algorithm [8] that adjusts the classifier's predictions in order to take into account the expert knowledge.

Multi attributed frame algorithm and five types of core vectors are estimated in [9]. Experimental results concluded that this algorithm works network intrusion dataset because it does not restrain any excessive value for any attribute. It can work well even if dataset does not contain any outlier in it. Reema Patel, Amit Thakkar, Amit Ganatra proves that [10] by combining more than one data mining algorithms may be used to diminish the disadvantages of one another. Combining a number of trained classifiers pilot to a better performance than any single classifier. Luigi Coppolino, Salvatore D'Antonio, Alessia Garofalo, Luigi Romano [11] implemented IDS for wireless sensor network using local agent and central agent, measure the performance of system using CART,CHAID,C5.0 and Bayesian networks and attain highest accuracy of detection rate.

A machine learning approach [12] known as Genetic Algorithm, to be acquainted with such attack type of connections. Intrusion detection system used information in the form of audit trails or packet of the network. In layered model three layers used for detecting DOS, probe, R2L and U2R attacks. Each layer is individually trained with a small set of pertinent features and then deployed successively. Layered model is used to reduce computation and the overall time required to notice abnormal events. Principal Component Analysis is used for feature reduction. It is a prevailing tool for analyzing data and found analogous patterns in the data.

The paper [13] compares the performance of three well known data mining classifier algorithms namely, ID3, J48 and Naïve Bayes were evaluated based on the 10-fold cross validation test. Experimental results using the KDDCup'99 IDS data set articulate that Naïve Bayes is one of the most effective inductive learning algorithms, decision trees are more interesting as far as the detection of new attacks is concerned. The methods discussed in this paper are being implemented as a network mail filter and implementing a network-level email [14] filter that uses our algorithms to grab malicious executables before users take delivery of them through their mail. They can envelop the potential malicious executable or can block it. This has the potential to bring to a halt some malicious executables in the network and prevent Denial of Service (DoS) attacks by malicious executables and proved that both the Naive Bayes and Multi-Naive Bayes methods are probabilistic.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

III. SYSTEM OVERVIEW

The following figures shows overview of proposed Intrusion Detection System which consists of two modules namely local module design and Classification module.

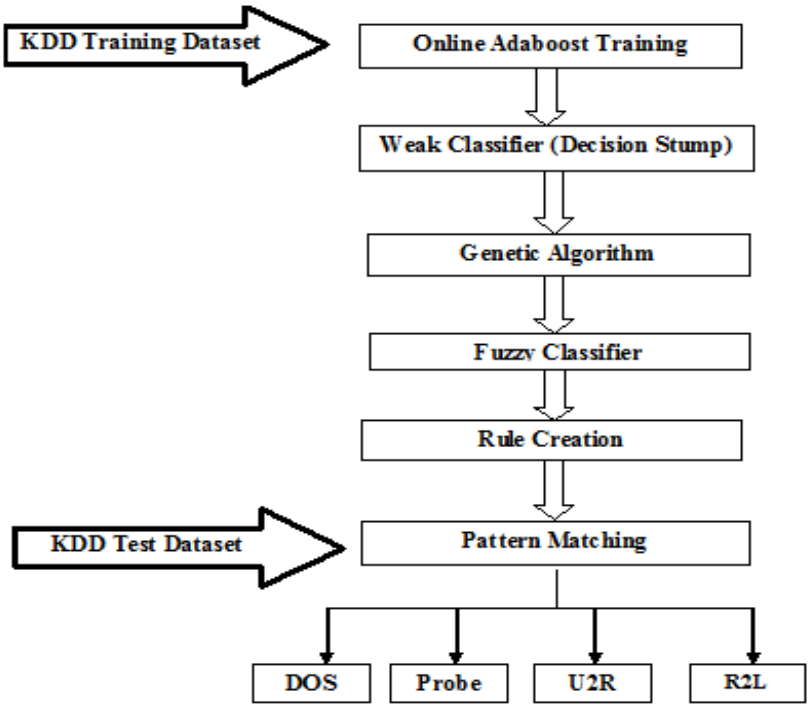


Fig 1: Proposed system Architecture

1. Local module design: In this module we assemble a local detection model at each node which includes the design of weak classifiers and Adaboost-based training. Each individual feature component corresponds to a weak classifier. So that the mixed attribute data for the network connections can be handled without human intercession, and full use can be made of the information in each feature. The Adaboost training will be implementing using only the local training samples at each node. To become accustomed to online training, in which each training sample is used only once for learning the strong classifier. Online Adaboost algorithm selects a number of weak classifiers according to a certain rule and updates them simultaneously for each input training sample.

1. Weak classifier: Decision stump

A decision stump is a decision tree with a root node and two leaf nodes. A decision stump is constructed for each feature component of the network connection data. Intact set of attributes are alienated into two subsets ^{Cf}i (intrusion connections set) and ^{Cf}n (Normal connections set). If the attack samples are more than (greater than) number of normal samples then connection is assigned into Cfi i.e. to intrusion connection set else it is assigned to Normal connection set.

2. Traditional online Adaboost algorithm:

Step 1: Initialize two weighted counters

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Counter1 = 0 and Counter2 = 0

Step 2: For every new sample i.e. for each network connection

a) Initialise weight (Cnt) = 1.
Randomly sample it using Poisson distribution.

b) Renew the weak classifier as,

i. If it correctly classifies the sample,

Counter1 = counter1+ cnt;

Approximate weighted classification rate (CR) is updated by,

CR = counter2 / counter1 + counter2;

Weight of connection (Cnt) is updated by,

Cnt = cnt (1 / 2(1 – CR))

ii. If it misclassifies,

Counter2 = counter2 + cnt;

Approximate weighted classification rate (CR) is updated by,

CR = counter2 / counter1 + counter2;

Weight of connection (cnt) is updated by,

Cnt = cnt (1/2*CR).

Step 3: The final weight for weight (α) for weak classifier is,

α = log (1-CR / CR).

This weight reflects the implication of sample or feature for detecting intrusion.

After training, each node will restrain a parametric model that consists of the parameters of the weak classifiers and the ensemble weights.

2. Classification module: In this module, we applied Fuzzy genetic algorithm for generating rule set and further we apply rule set to the test dataset to discover detection rate of different types of attacks.

Fuzzy Genetic Algorithm:

Fuzzy logic is the structure of many – valued logic. A fuzzy genetic algorithm is deliberate as a GA that uses a fuzzy logic based techniques. The term fuzzy logic was introduced with 1965 proposal of fuzzy set theory by Lotfi A. Zadeh. Genetic algorithm is the search technique used in computing to find true or approximate solution to optimization and search problems. Genetic algorithms are exacting class of evolutionary algorithm that use techniques

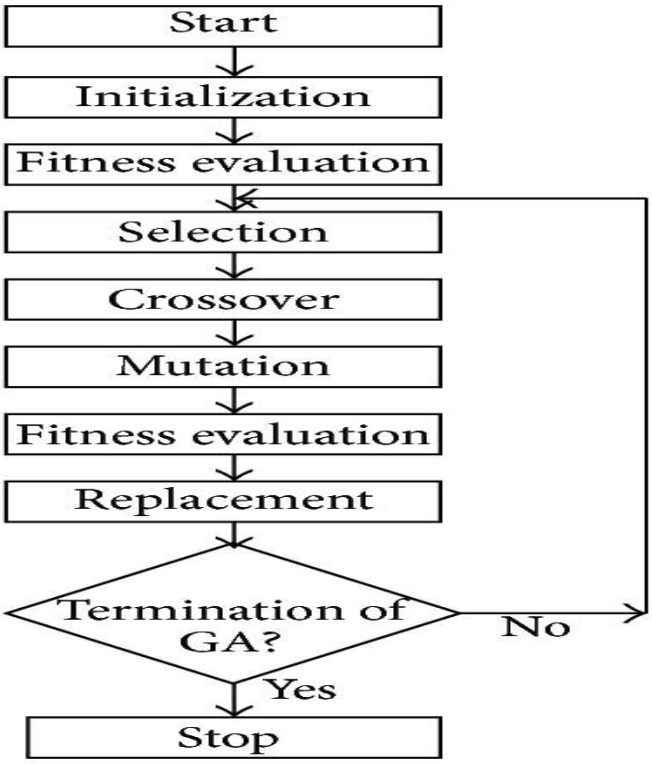
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

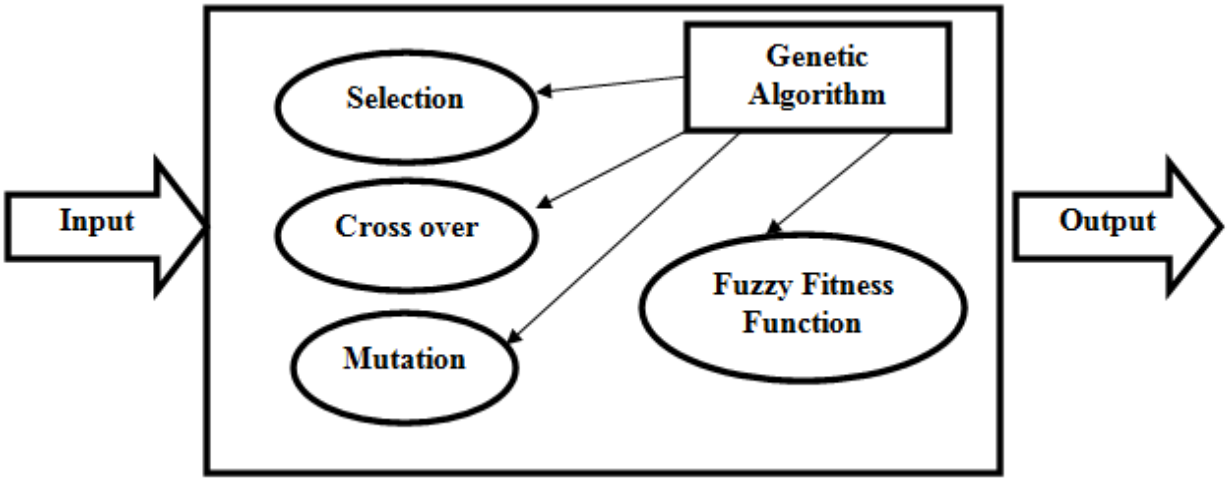
Vol. 4, Issue 2, February 2016

optimistic by evolutionary biology such as inheritance, mutation, selection and crossover (also called recombination). Genetic algorithm flow chart is as follows,



- The evolution in general starts with population of randomly generated individuals.
- Individual solutions are favored through a fitness based process.
- This process has continual till termination condition has been reached.
- Progress the solution through cyclical application of mutation, crossover, inversion and selection operators.

Fuzzy genetic algorithm model is as follows,



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

Genetic Operators:
1. Population: This is the set of randomly generated individuals.
2. Crossover: It is the genetic operator used to contrast the chromosomes from one generation to the next.
3. Mutation: It is the genetic operator used to vary one or more gene values in chromosomes.
4. Fitness function: It is the genetic operator used to discover the fittest chromosome for next generation.

Fuzzy classifier:
This classifier used to appraise the probability of being an attack identified by each attribute. The fuzzy logic is encoded into four parameters which are a, b, c and d. The probability is calculated as follows,

If (data value is between “b” to “c”)
Then prob =0.0

else if (data value between “a” to “b”)
then prob = attribute value - a / b-a

else if (data value between “c” to “d”)
then prob = d - attribute value / d - c
else prob = 1.0

After calculating probability next step is,

```
for each record
{
for each rule
                    {
for each attribute
{
Prob = fuzzy ();
Totalprob = totalprob + prob;
}
If (totalprob > threshold)
{
Class is attack;
True negative ++;
}
Else
 {
Class is normal;
True positive++;
}
```

IV. EXPERIMENTAL RESULTS

We use the knowledge discovery and data mining (KDD) CUP 1999 dataset to educated guess the performance of our algorithms as well as for training purpose. This dataset is still the most trustful and believable public benchmark dataset for evaluating network intrusion detection algorithms. In the dataset, 41 features including nine categorical features and 32 continuous features are extracted for each network connection. Attacks in the dataset fall into the following four main categories,

1.Denial of service (DoS) attacks: Attackers put out of articulation a host or network service to make legal users can not access to a machine, E.g. Back, Smurf, Land, SYN Flood, Ping of death, Teardrop, neptune.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

2. Probe: It is a category of attacks where an attacker examines a network to discover well-known vulnerabilities. E.g. IPsweep, Satan, Nmap, Portsweep.
3. User to Root (U2R) attacks: Local users get admittance to local machine without permission and then make the most of the machine's vulnerabilities, E.g. Loadmodule, Perl, buffer overflow, root kit.
4. Remote to Local (R2L) attacks: unsanctioned attackers gain local access from a remote machine and then exploit the machine's vulnerabilities, E.g. phf, ftp_write, Imap, Spy, multihop, warezmaster, warezclient, guesspsw.
The detailed description of KDD dataset is shown in following table. It describes the number of sample it contains in entire dataset.

Sr. No.	Category	Training Data	Test Data
1	**Normal**	97278	60593
2	**DoS**	391458	223298
3	**Probe**	4107	2377
4	**U2R**	52	39
5	**R2L**	1126	5993
6	**Other**	0	18729
7	**Total**	494021	311029

Table I: KDD Description

RESULTS
Our proposed Intrusion Detection system is skilled to discover, identify and categorize attacks. It gives the classification of attack so that we are able to take remedial action by alerting the user or network administrator of the system and secure the network from different types of attacks. Following graphs shows the execution time and detection rate of system.

Sr. No.	Name of Test Dataset	No of Records Present				No. Of Records Detected			
		DoS	Probe	U2R	R2L	DoS	Probe	U2R	R2L
1	KDD test21	5960	939	38	2182	5955	930	35	2170
2	NSLKDDtest	5957	948	43	2208	5956	939	35	2205
3	NSLKDDTest21	5963	958	45	2209	5958	936	37	2155

Table II: Identified attack count in Test Dataset

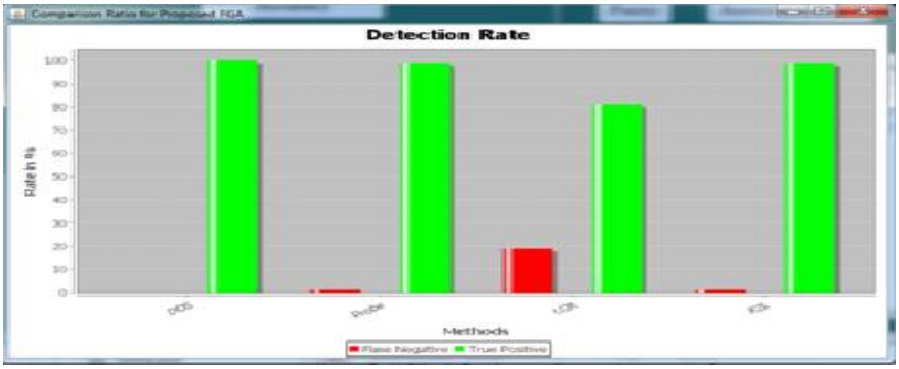


Fig. 2 Graph of Detection Rate

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

V. CONCLUSION

In this paper, we implement the Intrusion detection system which will be able to correctly detect, identify and classify the incoming attacks using online Adaboost based approach with minimum time consumption.

ACKNOWLEDGMENT

A large measure of any credit for the "Implementation of Intrusion Detection System using Adaboost based approach and Fuzzy Genetic Algorithm" must go to my project guide Prof. Dr. Mrs. Suhasini A. Itkar, Assistant Professor of PES's Modern College of Engineering, Pune and our ME Coordinator Ms. Deipali V. Gore who with the author has assisted in the preparation of this paper. I admire their infinite patience and understanding that they guided us in field we had no previous experience. We are grateful to them for having faith in us.

REFERENCES

1. Weiming Hu, Jun Gao, Yanguo Wang, Ou Wu, and Stephen Maybank," Online Adaboost-Based Parameterized Methods for Dynamic Distributed Network Intrusion Detection", IEEE Transactions on Cybernetics 2013.
2. Luigi Coppolino, Salvatore D'Antonio, Alessia Garofalo, Luigi Romano," Applying Data Mining Techniques to Intrusion Detection in Wireless Sensor Networks", 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing.
3. Vikas Sharma, Aditi Nema," Innovative Genetic approach For Intrusion Detection by Using Decision Tree", 2013 International Conference on Communication Systems and Network Technologies.
4. Dr. T. Subbulakshmi, Ms. A. Farah Afroze," Multiple Learning based Classifiers using Layered Approach and Feature Selection for Attack Detection", 2013 IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology (ICECCN 2013).
5. P. Jongsuebsuk+, N. Wattanapongsakorn+, C. Charnsripinyo*," Real-Time Intrusion Detection with Fuzzy Genetic Algorithm", 978-1-4799-0545-4/13/$31.00 ©2013 IEEE
6. Manish Kumar, Dr. M. Hanumanthappa, Dr. T. V. Suresh Kumar," Intrusion Detection System Using Decision Tree Algorithm", proceeding for IEEE, 2012.
7. Jinhua Huang and Jiqing Liu," Intrusion Detection System Based on Improved BP Neural Network and Decision Tree", 2012 IEEE fifth International Conference on Advanced Computational Intelligence (ICACI).
8. Shina Sheen, R Rajesh, Member IEEE ,"Network Intrusion Detection using Feature Selection and Decision tree classifier", 2012 IEEE International Conference on Tools with Artificial Intelligence.
9. M Suman , T Anuradha ,K Gowtham, A Ramakrishna," A Frequent Pattern Mining Algorithm Based On FP-Tree Structure And Apriori Algorithm", IJERA, Vol. 2, Issue 1, Jan-Feb 2012, pp.11'4-116.
10. Mr. V. K. Pachghare, Parag Kulkarni," Pattern Based Network Security using Decision Trees and Support Vector Machine", System engineering and Electronics, Vol.27,No.7,July 2011.
11. Krung pinasinoparan, Narudam Techaval Network Intrusion Detection using multi attributed decision tree 2011 IEEE International Conference on Tools with Artificial Intelligence.
12. Vijayarani, M.Divya," An Efficient Algorithm for Generating Classification Rules", IJCST Vol. 2, Issue 4, Oct. - Dec. 2011.
13. Mrutyunjaya Panda, Manas Ranjan Patra," A Comparative Study of Data Mining Algorithms for Network Intrusion Detection", First International Conference on Emerging Trends in Engineering and Technology, 2008.
14. Matthew G. Schultz and Eleazar Eskin, Erez Zadok," Data Mining Methods for Detection of New Malicious Executables", Proceedings of the 2008 International Virus Bulletin Conference, 2008.