ISSN 2250-3153

An Efficient Intrusion Detection Based on Decision Tree Classifier Using Feature Reduction

Yogendra Kumar Jain* and Upendra**

*Head of Department (CSE), Samrat Ashok Technological Institute, Vidisha, M.P., India
**Research Scholar M. Tech, (CSE), Samrat Ashok Technological Institute, Vidisha, M.P., India

Abstract— Large computational value has always been a restraint in processing huge network intrusion data. This problem can be extenuated through feature selection to abbreviate the size of the network data involved. In this paper, we first deal existing feature selection methods that are computationally executable for processing vast network intrusion datasets. In this paper, we study and analysis of four machine learning algorithms (J48, BayesNet, OneR, NB) of data mining for the task of detecting intrusions and compare their relative performances. Based on this study, it can be concluded that J48 decision tree is the most suitable associated algorithm than the other three algorithms with high true positive rate (TPR) and low false positive rate (FTR) and low computation time with high accuracy.

Index Terms- Intrusion Detection; Machine Learning; Decision Tree; Bayes Net; NB; KDD 99

I. INTRODUCTION

Recently research on machine learning for intrusion detection has standard much attention in the computational intelligence community. In intrusion detection algorithm, immense strengths of audit data must be analyzed in order to conception new detection rules for increasing number of novel attacks in high speed network. Intrusion detection algorithm should consider the composite properties of attack behaviors to improve the detection speed and detection accuracy. Analyze the large volume of network dataset and the better performances of detection accuracy, intrusion detection become an important research field for machine learning. In this work we have presented J48 decision tree algorithm for intrusion detection based on machine learning. The Intrusion Detection System (IDS) is Process of monitoring the events occurring in a computer system or network and analyzing them for signs of possible incidents. IDS was first introduced in 1980 by James. P. Anderson [3] and then improved by D. Denning [4] in 1987.

They are two basic approaches for Intrusion Detection techniques, i.e. Anomaly Detection and Misuse Detection (signature-based ID) [17]. Anomaly Detection is basically based on assumption that attacker behavior is different from normal user's behavior [1]. In this paper, we present the application of machine learning to intrusion detection. We analyse four learning algorithms (J48, BayesNet, OneR and NB) for the task of detecting intrusions and compare their relative performances. There is only available data set is KDD data set for the purpose of experiment for intrusion detection.KDD data set [2] contain 42 attributes. The classes in KDD99 [18] dataset can be categorized into five main classes (one normal class and four main intrusion classes: probe, Dos, U2R and R2L)

II. RELATED WORK

Intrusion detection started in 1980's and since then a number of techniques have been introduced to build intrusion detection systems [12], [13], [14]. In 2007, Panda and Patra [10] determined a method using naive Bayes to detect signatures of specific attacks. They used KDD99 dataset for experiment, in the early 1980's, Stanford Research Institute (SRI) developed an Intrusion Detection Expert System (IDES) that monitors user behavior and detects suspicious events. Meng Jianliang [6] used the K Mean algorithm to cluster and analyze the data. He used the unsupervised learning technique for the intrusion detection. Mohammadreza Ektefa et al., [8] in 2010, compared C4.5 with SVM and the results revealed that C4.5 algorithms better than SVM in detecting network intrusions and false alarm rate. Zubair A.Baig et al. (2011) proposed An AODE-based Intrusion Detection System for Computer Networks. They suggested that the Naive Bayes (NB) does not accurately detect network intrusions [7]. In 2010, Hai Nguyen et al. [5] applied C4.5 and BayesNet for intrusion detection on KDD CUP'99 Dataset. Jiong Zhang and Mohammad Zulkernine [9] done the intrusion detection using the random forest algorithms in anomaly based NIDS. Cuixio Zhang, Guobing Zhang, Shanshan Sun [15] used the missed approach for the intrusion detection. He designed the mixed combining the anomaly detection and misuse detection in this model the anomaly detection module is built using unsupervised clustering method and the algorithm is an improved algorithm of K means clustering algorithm. The new algorithm learns the strong points from the k-means and improved relations trilateral triangle theorem. Gary Stein [11] applied the genetic algorithm and the decision tree algorithm for the intrusion detection. He used the genetic algorithm technique for the feature reduction.

III. METHODOLOGICAL APPROACH

Decision tree technology is a common, intuitionist and fast classification method [21]. Its construction process is top-down, divide-and-rule. Essentially it is a greedy algorithm. Starting from root node, for each non-leaf node, firstly choose an attribute to test the sample set; Secondly divide training sample set into several sub-sample sets according to testing results, each sub-sample set constitutes a new leaf node; Thirdly repeat the above division process, until having reached specific end conditions. In the process of constructing decision tree, selecting testing attribute and how to divide sample set are very crucial. Different decision tree algorithm uses different technology. In practice, because the size of training sample set is usually large, the branches and layers of generated tree are also more. In addition, abnormity and noise existed in training sample set will also cause some abnormal branches, so we need to prune decision tree. One of the greatest advantages of decision tree classification algorithm is that: It does not require users to know a lot of background knowledge in the learning process. As long as training samples can be expressed as the form of attribute-conclusion, you can use this algorithm to study. But decision tree technology also has a lot of deficiency, such as: When there are too many categories, classification accuracy is significantly reduced; It is difficult to find rules based on the combination of several variables. At present, there are a lot of decision algorithms, such as: ID3, SLIQ, CART, CHAID and so on. But J48 algorithm is the most representative and widely used. It is proposed by Quinlan in 1993.

A Naive Bayes classifier [19] is a simple probabilistic classifier based on applying Bayes' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple Depending on the precise nature of the probability model; naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods.

A. Information Gain by an Example Data Set

The proposed feature reduction technique can be easily understood by the following example. To demonstrate efficiency of the proposed technique, we have used customer database [20] to calculate information gain.

TABLE I. CUSTOMER EXAMPLE DATASET

RID	Age	Income	Student	Credit Rating	Class:buys
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	middle_aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	middle_aged	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	middle_aged	Medium	No	Excellent	Yes
13	middle_aged	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

Table I. presents a training set, D, of class-labelled tuples randomly selected from the All Electronics customer database. In this example, each attribute is discrete-valued. The class label attribute, buys computer, has two distinct values (namely, yes, no); therefore, there are two distinct classes (that is, m = 2).Let class C1 correspond to yes and class C2 correspond to no. There are nine tuples of class yes and five tuples of class no. A (root) node N is created for the tuples in D. We compute the information gain of each attribute. We first compute the expected information needed to classify a tuple in D:

Info (D)= - 9/14 log2 (9/14) –5/14 log2 (5/14)=0.940 bits....(1)

Next, we need to compute the expected information requirement for each attribute. Let's start with the attribute age. We need to look at the distribution of yes and no tuples for each category of age. For the age category youth, there are two yes tuples and three no tuples. For the category middle aged, there are four yes tuples and zero no tuples. For the category senior, there are three yes tuples and two no tuples. Now we calculate the Info for an attribute Age. The expected information needed to classify a tuple in *D* if the tuples are partitioned according to age is:

Info age (D) = 5/14 × (- 2/5 log2 2/5 – 3/5 log2 3/5) + 4/14 × (-4/4 log2 4/4 -0/4 log2 0/4) + 5/14 × (-3/5 log2 3/5 – 2/5 log2 2/5)

= 0.694 bits..... (2)

Hence, the gain in information from such a partitioning would be equation (1) – (2)

Gain (age) = Info (D) – Info age (D)

= 0.940 – 0.694

= 0.246 bits

Similarly, we can compute Gain (income) = 0.029 bits, Gain (student) = 0.151bits, and Gain (credit rating) = 0.048 bits

Using the method above for calculation of information gain, we calculate the info gain of the all the attribute of the KDD99 data set. The info gain of the all the attribute is given below in table I. In our proposed technique we are using the KDD99 dataset with these selected features and train and test the algorithm. For the testing we are using the 10 fold cross validation. Features selection techniques have been employed by Researchers. In other domain to extract important

features.Skurichina and Duin [16] suggested that predictive accuracy can be improved by combining feature sets.

TABLE II. A SAMPLE CONFUSION MATRIX

	Predicted Class Positive	Predicted Class Negative
Actual Class Positive	a	b
Actual Class Negative	c	d

In this confusion matrix, the value a is called a true positive and the value d is called a true negative. The value b is referred to as a false negative and c is known as false positive.

B. True Positive Rate,Fale Positive Rate

In the context of intrusion detection, a true positive is an instance which is normal and is also classified as normal by the intrusion detector. For a good IDS TP rate should be high. False positive means no attack but IDS detect the attack. For a good IDS FP should be low.

C. Accuracy

This is the most basic measure of the performance of a learning method. This measure determines the percentage of correctly classified instances. From the confusion matrix, we can say that:

$$\text{Accuracy} = \frac{a + d}{a + b + c + d}$$

This metric gives the number of instances from the dataset which are classified correctly i.e. the ratio of true positives and true negatives to the total number of instances.

D. Algorithm

J48_Tree, generate a decision tree from the given training data
Input: training sample set T, the collection of candidate attribute attribute_list
Output: a decision tree.

- Create a root node N;
- IF T belongs to the same category C, then return N as a leaf node, and mark it as class C;
- IF attribute_list is empty or the remainder samples of T is less than a given value, then return N as a leaf node, and mark it as the category which appears most frequently in attribute_list, for each attribute, calculate its information gain ratio.
- Suppose test_attribute is the testing attribute of N, then test_attribute = the attribute which has the highest information gain ratio in attribute list:
- If testing attribute is continuous, then find its division threshold;
- For each new leaf node grown by node N {
  Suppose T' is the sample subset corresponding to the leaf node. If T' has only a decision category, then mark the leaf node as this category; Else continue to implement J48_Tree (T ', T'_attributelist) .
  }
- Calculate the classification error rate of each node, and then prune the tree.

IV. RESULT ANALYSIS

The Tables III, IV and V Shows the performance of four classification methods based on correctly classified Instances, incorrectly classified Instances ,Kappa statistic, Mean absolute error, Root Mean Squared Error and Relative Absolute Error and Root Relative Squared error and Time taken to build the models respectively. The comparison is performed for 41 and 11 attributes. The four classifier models on the dataset are built and tested by means of 10-fold cross-validation. The Java Heap size was set to 1024 MB for WEKA 3.6.2, the simulation platform is an Intel™ Core i3-2100 processor system with 3 GB RAM under Microsoft Windows XP™ Service Pack-2 operating system, 3.10 GHz with 500 GB memory.

TABLE III. COMPARISION OF THE RESULTS FOR J48, BAYESNET, ONER AND NB WITH ALL ATTRIBUTE

Parameter	Classifier			
	J48	***BayesNet***	***OneR***	***NB***
Correctly Classified Instances	99.5594%	96.5624%	96.1893%	89.5919 %
Incorrectly Classified Instances	0.4406 %	3.4376 %	3.8107 %	10.4081 %
Kappa statistic	0.9911	0.9307 %	0.9237 %	0.7906 %
Mean absolute error	0.0064	0.0378 %	0.0381 %	0.1034 %
Root mean squared error	0.0651	0.175 %	0.1952 %	0.3152 %
Relative absolute error	1.2854 %	7.6037 %	7.6566 %	20.7817 %
Root relative squared error	13.059 %	35.0792%	39.132 %	63.1897 %

TABLE IV. COMPARISION OF THE RESULTS FOR J48, BAYESNET, ONER AND NB WITH 11 ATTRIBUTE

Parameter	Classifier			
	J48	***BayesNet***	***OneR***	***NB***
Correctly Classified Instances	99.9039%	99.1073%	97.6761%	92.697 %
Incorrectly Classified Instances	0.0961 %	0.8927 %	2.3239 %	7.303 %
Kappa statistic	0.998	0.982	0.9529 %	0.8551 %
Mean absolute error	0.0006	0.004	0.0093 %	0.0297 %
Root mean squared error	0.019	0.056	0.0964 %	0.1641 %
Relative absolute error	0.2997 %	2.0366 %	4.7346 %	15.144 %

Root relative squared error	6.0564 %	17.8797%	30.7728%	52.3814 %

From table III and IV.It is clear that The J48 gave the best performance.

Now we compare the result of the J48, BayesNet, OneR and NB algorithms. Firstly we compare the result after run the algorithm with all attribute. Secondly we compare the result after run the algorithm with reduced 11 attribute than only we conclude that which one algorithm is good best for the intrusion detection.

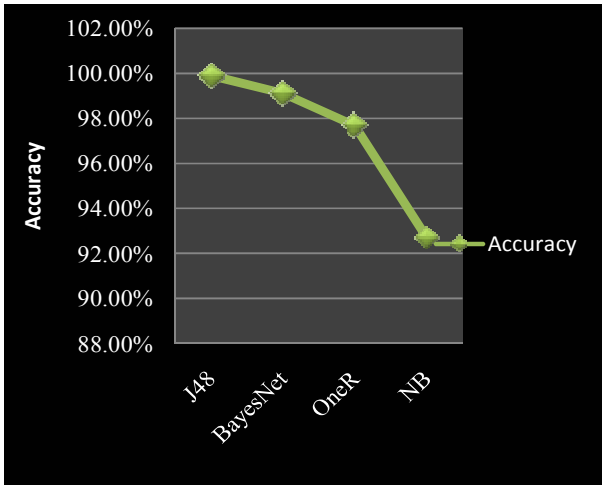

Figure 1. Comparison of accuracy for J48, BayesNet, OneR and NB.

From above figure 1. It is clear that information gain feature reduction method gives the better accuracy which is desirable for good Intrusion Detection System. Especially in the case of J48 accuracy is 99.9%.

Now we compare the TPR of the J48, BayesNet, OneR and NB algorithm with all attribute and with selected 11 attributes.

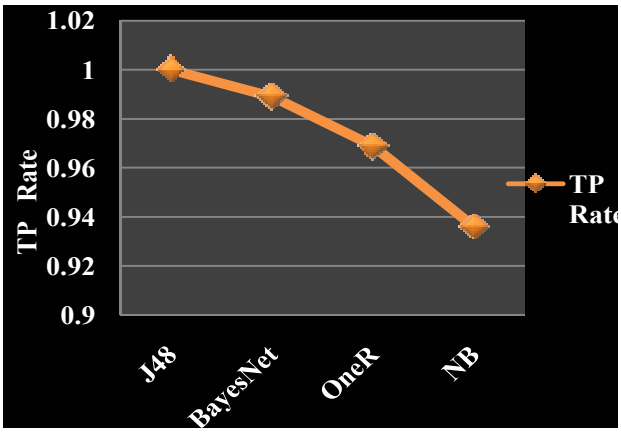

Figure 2. TPR comparison of J48, BayesNet, OneR and NB.

For a good IDS TP Rate should be high. Above figure 2. Shows that TP Rate of the J48 algorithm is higher when we reduce the feature of the data set using information gain. Especially in the case of J48 TPR is 1

Figure 2 and Figure 3 above shows the TPR (True Positive Rate) and FPR (False Positive Rate) of the J48, BayesNet, OneR and NB algorithm when run with the all attributes of the data set. Figure 2. Shows that TPR of the J48 is higher than the remaining three algorithms which is desirable. Figure 3. Also shows that FPR of the J48 is almost zero which is desirable for a good intrusion detection algorithm.

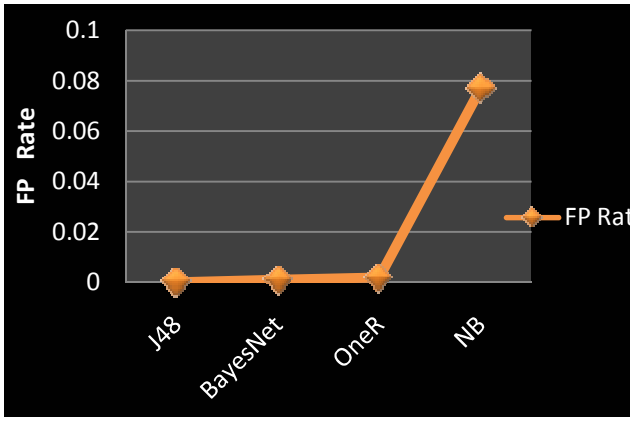

Figure 3. FPR comparison of J48, BayesNet, OneR and NB.

For a good IDS FPR should be low. Above figure 3 shows that FPR of the J48 algorithm is lower when we reduce the feature of the data set using information gain. Especially in the case of J48 FPR is 0.In the case of BayesNet, OneR and NB algorithm FPR of the greater than 0. From above figures 1, 2 and 3 it is clear that J48 algorithm Accuracy, TPR and FPR is better than other three algorithms. So we can say that reduction of the feature using information gain is better technique.

The experimental results shows that Performance Evaluation of four classification models, J48 have much better performance than other three methods and it is also observed that the overall performance of J48 classification has increased their performance using feature reduction method a notable improvement in their classification, means the classification accuracy increases better after feature selection.

In this paper, the performance of four well known data mining classifier algorithms namely J48, BayesNet, OneR and Naïve Bayes are evaluated based on the 10-fold cross validation test, Experimental results using the KDD CUP99 IDS data set demonstrate that while J48 is one of the most effective inductive learning algorithms, decision trees are more interesting as far as the detection of new attacks is concerned.

TABLE V. COMPARISION OF THE RESULTS FOR J48, BAYESNET, ONER AND NB

Feature Used	Classifier	Accuracy	normal		dos		probe		r 2 l		u 2 r	
			TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate	TP Rate	FP Rate
11	J48	99.9039%	1	0.002	1	0	0.971	0	0.75	0	0.2	0
11	Bayes Net	99.1073%	0.989	0.001	0.996	0.001	0.951	0.006	0.824	0.001	0.6	0.001
11	OneR	92.697 %	0.969	0.002	0.999	0.035	0.682	0.001	0.794	0	0	0
11	NB	92.697 %	0.936	0.077	0.917	0.015	0.811	0.007	0.824	0.019	0.6	0.007

V. CONCLUSIONS

In this paper, we reduced the features of the data set using information gain of the attributes. This study is approached to discover the best classification algorithm for the applications of machine learning to intrusion detection. Our simulation results show that, in general, the J48 has the highest classification accuracy performance with the lowest error rate. On the other hand, we also found that drastically decreased in learning time of the algorithm and increase in accuracy and TPR. Comparison shows that reduction of the feature using information gain technique is suitable for the feature reduction. Using Weka, we analysed four algorithms towards their suitability for detecting intrusions from KDD99 dataset. We showed that machine learning can be effectively applied to detect novel intrusions and focused on anomaly detection. The four learning algorithms J48, BayesNet, OneR and NB were compared at the task of detecting intrusions. J48 with an accuracy rate of approximately 99% was found to perform much better at detecting intrusions than BayesNet, OneR and NB Based on the experiments done in the paper and their Corresponding results, we can state the following: Machine learning is an effective methodology which can be used in the field of intrusion detection.

REFERENCES

[1] Lida Rashidi,Sattar Hashem and Ali Hamzeh, "Anomaly detection in categorical datasets using bayesian networks," AICI'11 Proceedings of the Third International Conference on Artificial Intelligence and Computational Intelligence, Volume Part II, Springer-Verlag, Berlin ,Heidelberg, 2011, pp.610–619.

[2] Knowledge Discovery in Databases DARPA archive. Task Description ,KDDCUP 1999 DataSet, http://www.kdd.ics.uci.edu/databases/kddcup99/task.htm

[3] James P. Anderson, "Computer Security Threat Monitoring and Surveillance," Technical Report, James P.Anderson Co.,Fort Washington, Pennsylvania, USA , pp.98–17, April 1980.

[4] Dorothy E. Denning,"An Intrusion Detection Model," IEEE Transaction on Software Engineering (TSE), volume–13, No.2, pp.222–232,February 1987.

[5] Hai Nguyen, Katrin Franke and Slobodan Petrovi'c, "Improving Effectiveness of Intrusion Detection by Correlation Feature Selection," International Conference on Availability, Reliability and Security, pp. 17–24, IEEE 2010.

[6] Meng Jianliang, Shang Haikun, "The application on intrusion detection based on K-Means cluster algorithm," International Forum on Information Technology and Application, 2009.

[7] Zubair A. Baig, Abdulrhman S. Shaheen, and Radwan AbdelAal, "An AODE-based Intrusion Detection System for Computer Networks," pp. 28–35, IEEE 2011.

[8] Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lilly Suriani Affendey,"Intrusion Detection Using Data Mining Techniques," Proceedings Of IEEE International Conference on Information Retrieval & Knowledge Management,Exploring Invisible World, CAMP'10,2010, pp.200-203.

[9] Jiong Zhang and Mohhammad Zulkernine, "Anomaly based Network Intrusion detection with unsupervised outlier detection," School of Computing Queen's University, Kingston, Ontario, Canada. IEEE International Conference ICC 2006, Volume-9, pp. 2388-2393, 11-15 June 2006.

[10] M. Panda, and M. R. Patra, "Network intrusion detection using naive Bayes," International Journal of Computer Science and Network Security (IJCSNS), Volume -7, No. 12, December 2007, pp. 258–263.

[11] Gary Stein, Bing Chen," Decision Tree Classifier for network intrusion detection with GA based feature selection," University of Central Florida. ACM-SE 43, Proceedings of 43rd annual Southeast regional Conference. Volume-2,2005.ACM,New York,USA.

[12] Shai Rubin, Somesh Jha, and Barton P. Miller, "Protomatching Network Traffic for High Throughput Network Intrusion Detection," In Proceedings of the Proceedings of the 13th ACM conference on Computer and Communications Security, pages 47–58. ACM, 2006.

[13] Marco Cova, Davide Balzarotti, Viktoria Felmetsger, and Giovanni Vigna. Swaddler, "An Approach for the Anomaly-Based Detection," Symposium on Recent Advances in Intrusion Detection(RAID*)*, pages 63–86. Springer, 2007.

[14] Pavel Kachurka, Vladimir Golovko, "Neural Network Approach to Real-Time Network Intrusion Detection and Recognition," The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Application,15-17 September 2011, pp. 393–397, IEEE 2011.

[15] Cuixiao Zhang, Guobing Zhang, Shanshan Sen., "A mixed unsupervised clustering based Intrusion detection model," Third International Conference on Genetic and Evolutionary Computing, 2009.

[16] M. Skurichina and R.P.W. Duin, "Combining feature subsets in feature selection," Lecture Notes in Computer Science, Vol. 3541,pp-165-175,Springer Verlag, Berlin, 2005.

[17] LI Min and Wang Dongliang, "Anomaly Intrusion Detection Based on SOM," IEEE WASE International Conference on Information Engineering, IEEE Computer Society, 2009, pp. 40-44.

[18] Mahbod Tavallaee,Ebrahim Bagheri,Wei Lu, and Ali A.Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Application(CISDA 2009),IEEE 2009.

[19] R.Dogaru,"A modified Naive Bayes classifier for efficient implementations in embedded systems," Signals Circuits and Systems

(ISSCS), IEEE 10th International Symposium on Lasi, June 30,2011-July 1, 2011, pp.1–4.

[20] Jiawei Han and Micheline kamber, "Data Mining Concepts and Techniques,"Second Edition,University of Illinois at Urbana-Champaign The Morgan Kaufmann Series in Data Management Systems,Elsevier 2007.

[21] Juan Wang, Qiren Yang,Dasen Ren, "An intrusion detection algorithm based on decision tree Technology,"Asia-Pacific Conference on Information Processing,APCIP 2009 ,Shenzhen,IEEE 18-19 July 2009. pp. 333–335.

AUTHORS PROFIL

Dr. Yogendra Kumar Jain presently working as head of the department, Computer Science & Engineering at Samrat Ashok Technological Institute Vidisha M.P India. The degree of B.E. (Hons) secured in E&I from SATI Vidisha in 1991, M.E. (Hons) in Digital Tech. & Instrumentation from SGSITS, DAVV Indore(M.P), India in 1999. The Ph. D. degree has been awarded from Rajiv Gandhi Technical University, Bhopal (M.P.) India in 2010.

Research Interest includes Image Processing, Image compression, Network Security, Watermarking, Data Mining. Published more than 50 Research papers in various Journals/Conferences, which include 30 research papers in International Journals.

Mr. Upendra is a research scholar of the department in Computer Science & Engineering from Samrat Ashok Technological Institute, Vidisha, affiliated to Rajiv Gandhi Technological University, Bhopal (M.P.), India. He secured degree of B.E. in Information Technology from Christian College of Engineering & Technology, Bhilai, affiliated to Pt.Ravi Shankar Shukla University, Raipur (C.G.), India in 2006.

Research Interests includes developing network security applications for the detection of suspicious abnormal behaviors, studying the performance of various network security tools. Designing and implementing various soft computing tools.

Mobile: +91-9907221159. E-mail: upendra.chaurasiya@gmail.com