Evaluation of Intrusion Detection Algorithms for Interoperability Gateways in Ad Hoc Networks

S. P. Manikandan
Department of Computer Science and Engineering
P. B. College of Engineering
Chennai, India.
mkandan_2000@yahoo.com

Dr. R. Manimegalai
Department of Computer Science and Engineering
Vellamal Engineering College
Chennai, India.

Abstract— **Security issues in MANET are of primary importance due to its higher vulnerability to attacks compared to wired networks. Though attack prevention mechanisms using authentication and encryption become the first line of defense, newer types of attacks cannot be detected by such systems. A second line of defense is necessary to detect and respond to intrusion. In MANETs the attack can occur within the network from one of the nodes or could be perpetuated from outside the network through the gateway. Attacks from external node not only affect the target node but also the participating nodes in the network. In this paper we evaluate Intrusion Detection Classification Techniques on a MANET gateway connecting to a wired network. Bayesian and Decision tree induction techniques are evaluated and results presented.**

Keywords- Ad Hoc Network, MANET, Intrusion Detection, Attack Classification, Gateways.

I. INTRODUCTION

Wireless network can be broadly classified into Infrastructure based network and Ad hoc network. In an infrastructure based network all nodes access external world using fixed access points. The wireless communication between the node and the access point enables the mobility of the node if required. Infrastructure based wireless networks using the IEEE standard 802.11 have been successfully deployed throughout the world and is used in laptops and mobile phones [1]. Ad hoc networks are formed by a group of nodes communicating with each other without any infrastructure and are often called as Mobile Ad hoc Network (MANET). In MANETs routing activity is performed by participating nodes for communication between source and destination which are out of radio range. The source can reach the destination by multi hops using the intermediate nodes.

Routing protocols used in traditional wired networks fail in wireless networks due to inherent problems such as noisy physical layer, larger latency and dynamic nature of wireless networks. In Ad hoc networks, routing protocols can be broadly classified into proactive routing protocols and **r**eactive routing protocols. In proactive routing**,** which is also called as table driven routing, the routes are discovered and updated periodically irrespective of data communication between nodes. Though routes are available immediately, the network overheads tend to be huge in high mobility and large networks. Popular pro-active routing protocols are Optimized Link State Routing (OLSR) routing [2] and Destination Sequence Distance Vector routing (DSDV) [3]. Reactive routing on the other hand discovers routes only when a node requires a communication channel to send data for a particular destination. Hence reactive routing is also called as on demand routing protocols. Popular reactive routing protocols include Ad hoc-On-demand Distance Vector (AODV) routing protocol [4] and Dynamic Source Routing (DSR) [5]. Routing and security play a very important role for successful implementation of MANETs. MANETs can communicate with the external world using Internet Gateways where available.

Though MANETs are rapidly deployable, they have inherent vulnerabilities and hence susceptible to all types of malicious attacks [6]. The increased security threats in MANET is due to the co-operative nature of Ad hoc networks which happens to be a primary requirement for successful implementation of Ad hoc networks. Typical Ad hoc networks do not have a centralized authority for authentication and hence it is more susceptible for attacks and compromise. Attacks can be either active or passive on all layers of the OSI model. At the link layer the adversary can jam a radio channel creating a condition similar to Denial of Service (DOS). At the network layer an adversary can add, modify or delete communication packets or spoof as another node**.** MANETs are more susceptible to attacks with the increasing number of hops.

A typical Network Intrusion Detection Systems (NIDS) consist of tools, processes and resources to identify and report abnormal and unauthorized network activity [7]. They are broadly classified into signature based intrusion detection systems and anomaly based intrusion detection systems. Signature based Intrusion Detection Systems (IDS) use known attack scenarios to monitor incoming packets and cannot identify new types of attacks. Anomaly based IDS attempts to find activities that are different from the normal network behavior. Learning algorithms are extensively used in anomaly based intrusion detection system. Both classification and clustering are extensively used to detect anomalies. Various data mining algorithms based on Bayesian theory, Decision tree and solutions based on Support Vector Machine (SVM) systems have been proposed in [8, 9, 10].

Equation 1 gives the Naïve Bayes Classification that is used in NIDS. Given the class label, we can derive the Naïve Bayes Classification for conditionally independent attributes from Equation 1 [17].

$$p(y\mid x)=\frac{p(x,y)}{p(x)}=\frac{p(x\mid y)p(y)}{\sum_{y'=1}^{C}p(x\mid y')p(y')} \tag{1}$$

$$p(x\mid y=c)=\prod_{i=1}^{D}p(x_i\mid y=c) \tag{2}$$

The attributes used in this work for intrusion prediction are independent and hence equation 2 will be justified for the prediction algorithm.

Decision tree induction [9] is a popular and powerful classification consisting of a tree that consists of branch nodes where a decision can be made and terminated with leaves which represents the predicted class. A simple decision tree is depicted in Figure. I

.

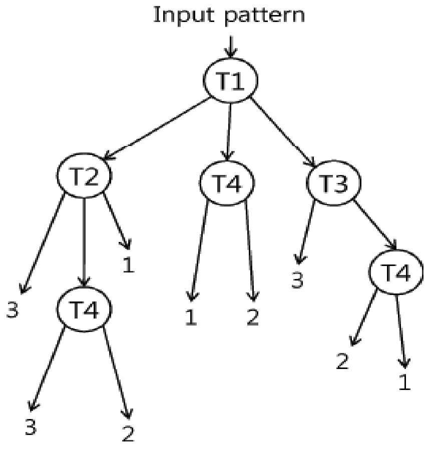

Figure I : Decision tree for pattern classification

In Figure I, the input pattern attribute is compared and adjudged to the second level node based on how best they fit in the next level. The best fit is determined using information gain or gini index. The tree process is continued till the end of a leaf node is reached which becomes the assigned class for the pattern. In Figure I, T1, T2, T3 and T4 represents the attribute and 1,2,3,4 represents the class label. In our dataset the numbers 1, 2, 3 and 4 represent the type of attacks.Decision forests are an ensemble of decision trees and the mode of the outputs obtained from the different trees becomes the predicted class.

In this paper, we investigate the classification efficiency of various classification algorithms for Network intrusion detection at the network gateway of the proposed MANET. The network architecture of the proposed system is shown in Figure II. The proposed MANET consists of nodes which dynamically enter the network or leave the network and move randomly within the network. The network has a single gateway connecting to the internet cloud. Nodes in the network can access the internet by establishing a communication link with the gateway. The connection between the source and the gateway can either be single hop or multihop. This paper is organized into the following sections, Section II gives the details of related work done in the area of data mining and IDS. Our proposed methodology is explained in section III. Section IV discusses the experimental results and section Vconcludes the paper.

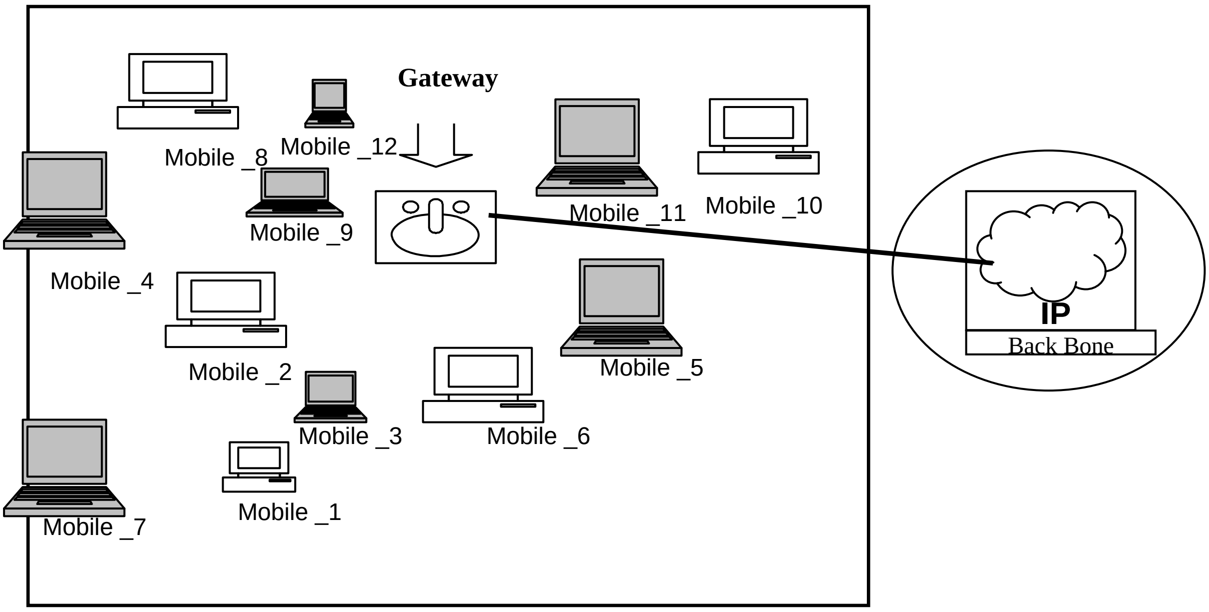

Figure II : Gateway where the NIDS classification algorithm is proposed.

II. RELATED WORK

Mitrokotsa et al., [11] have evaluated the performance of intrusion detection systems for MANET using the linear classifier, Gaussian mixture model and Support vector machine. Classification error is in the range of 0.35 to 0.6. The performance of the classification algorithms was evaluated with changing traffic conditions and mobility patterns. The various attacks included in this work are Black Hole, Forging, Packet Dropping, and Flooding attacks.

A modification of the Dynamic Source Routing (DSR) protocol to improve the security and accommodate intrusion detection is proposed in [12]. The proposed architecture consists of fixed wired-infrastructure based network which supports the multihop wireless nodes which could be PDA, Laptop or other mobile devices.

A new anomaly detection system containing detection subsystems in MAC layer, IP layer and application layer has been proposed in [13]. Markov chain and Bayesian classifiers are used for anomaly detection in [13]. Test data obtained from the network traffic was fed into the detection framework in each layer. If there is any deviation from normal behavior, it is considered as abnormal or anomaly based on predefined thresholds. Intrusion results from detection subsystems of all the three layers are integrated at local integration module and the final result is sent to the global integration module.

Shrestha et al., 2010 proposed a novel cross layer intrusion detection architecture to discover the malicious nodes and different types of denial of service attacks by exploiting the information available across different layers of protocol stack in order to improve the accuracy of detection. Cooperative anomaly intrusion detection with data mining technique is used in [14].

III. Classification accuracy measurement

The TCP dump containing HTTP data from the KDD 99 dataset [15] is used to evaluate our methodology. The dump consists of normal data along with attacks including portsweep, ipsweep, backdoor and Neptune. 0.875 % of the available dump contained abnormal packets. The attributes captured for the classification includes status of the connection, number of bytes from source to destination, number of bytes from destination to source, number of compromised conditions, number of connections on access control files and traffic features.

Information gains are used to preprocess the data, specifically for data reduction. Information gain provides the effectiveness of an attribute in classifying the training data based on the attribute which has to be predicted also called as the class label, the information gain is computed. Let *Attr* be the set of all attributes

Ex the set of all training examples, *value*(*x*,*a*)

$x \in Ex$ defines the value of a specific example *x* for attribute $a \in A$,

H specifies the entropy (In information theory, entropy is a measure of the uncertainty associated with a random variable)

$|x|$ is the number of elements in the set *x*.

The information gain for an attribute $a \in A$ is given in Equation 3 [17].

$$IG(Ex,a) = H(Ex) - \sum_{v \in values(a)} \frac{|\{x \in Ex \mid value(x,a) = v\}|}{|Ex|} . H(\{x \in Ex \mid value(x,a) = v\}) \quad (3)$$

The top 65% ranked values were selected for validating the various classification algorithm. 65% of the data were used as the training set and the balance as test set. Weka was used for measuring the classification accuracy. Weka is a open source tool for data mining and written in Java. It contains tools for visualization, preprocessing and predictive modeling.

IV. Experiment results and discussion

The classification results obtained are shown in Figure III and Table I. It is seen that tree based algorithms perform marginally better than probability based algorithms.

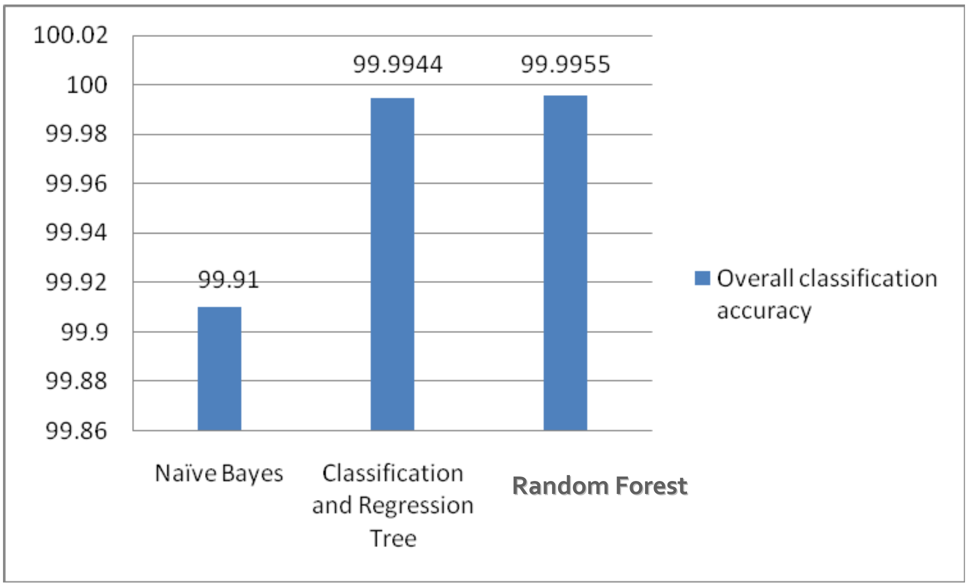

Figure III. Overall Classification accuracy

Table I : Miscalculation percentage between Normal and Anomaly data.

Misclassification percentage	Normal	Anomaly
Naïve Bayes	100	0
Classification and Regression Tree	0	100
Random Forest	0	100

From Table I, it can be seen that the misclassification of tcpdump occurs only in the anomalous data for decision tree based classifiers whereas in Naïve Bayes, misclassification occurs only in the normal data. Typically the misclassification rate should be low for normal data as this will avoid true negatives and high for anomalous data, however a balanced approach need to be taken.

In the second setup, the NSL-KDD dataset was used. This dataset consists of two class namely normal and anomalous attributes. The attributes were selected based on the information gain as in the previous setup. The classification accuracy are shown in Figure IV and Table II.

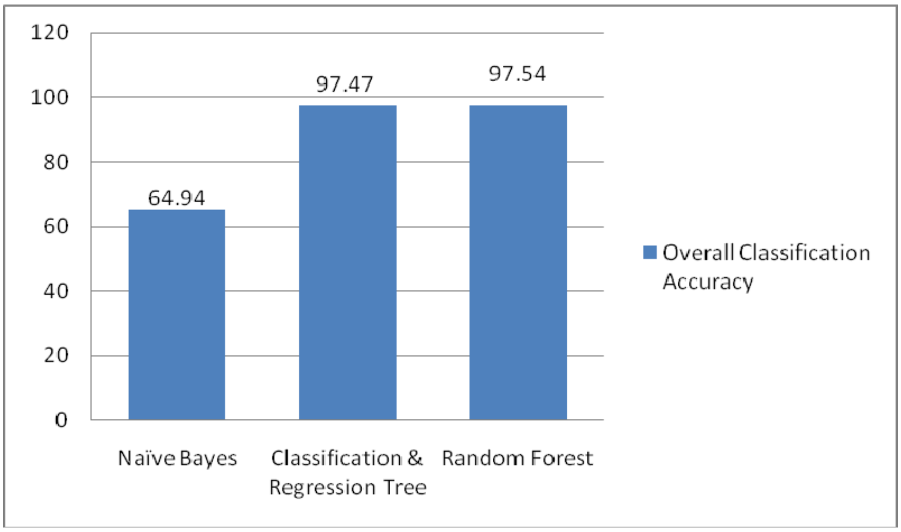

Figure IV : Overall classification accuracy of NSL-KDD dataset sample

Table II : Misclassification percentage of NSL-KDD dataset.

Misclassification percentage	Normal	Anomaly
Naïve Bayes	8.4	91.58
Classification and Regression Tree	51.67	48.33
Random Forest	57.88	42.12

NSl-KDD is generally considered a challenging dataset for predictive modeling. From Figure IV and table II it is seen that Naïve Bayes does not perform well with most of the anomalous data misclassified. However decision tree based algorithms show equal misclassification for normal and anomalous data.

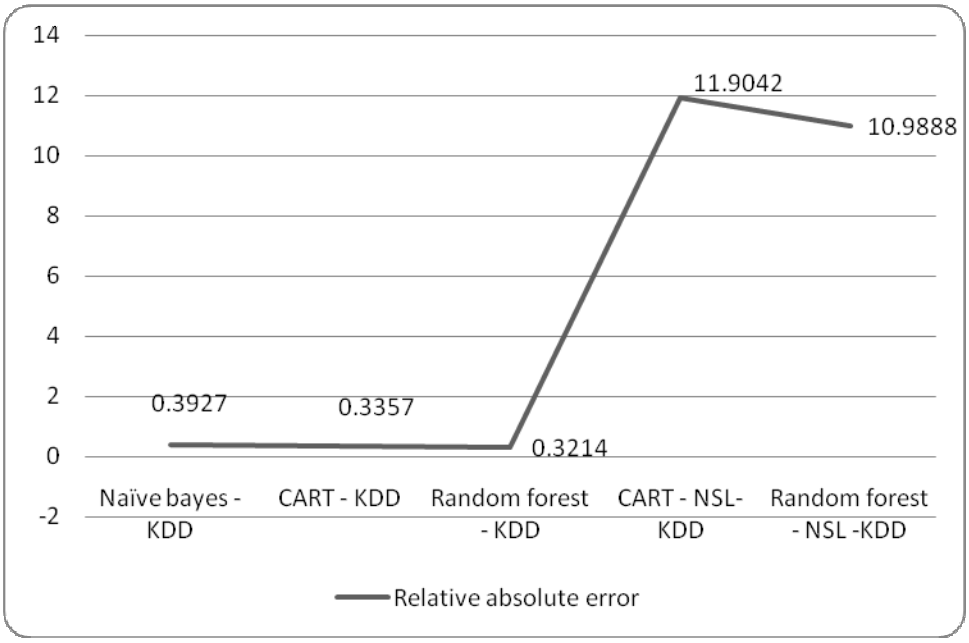

Figure V : Relative Absolute Error from both the data set used in this work.

We also investigate the classification efficiency when the percentage of anomaly and normal data is varied in the dataset under test. The classification accuracy obtained is shown in Figure VI. Figure VI illustrates that Naïve Bayesian algorithm is able to classify as the percentage of anomaly data decreases, whereas CART and Random Forest perform consistently irrespective of the percentage of anomalous data.

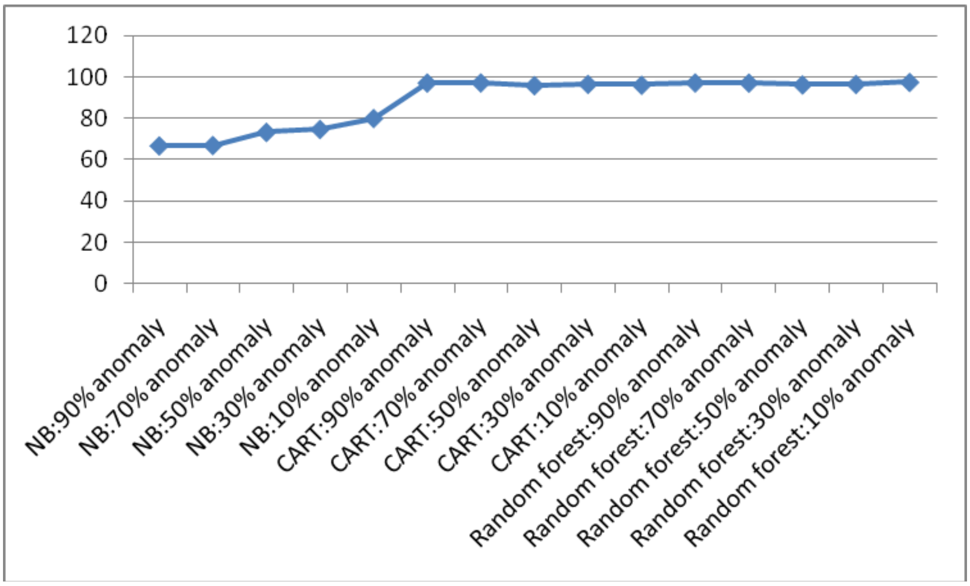

Figure VI : Classification accuracies as the percentage of anomalous data and normal data % is changed

V. CONCLUSIONS

In this paper, we have investigated the classification accuracy of three popular algorithms, namely Naïve Bayes [8], Random forest [9] and CART [10] in a MANET by selecting attributes using information gain. Two different large datasets consisting of 254032 packets were used in this work with one dataset containing very few anomalous data and the other set containing more anomalous data. As MANET's operate in a power constrained environment, reducing the computational time is essential and hence only 65% of the attributes available were used. Information gain was used for feature selection. The classification accuracy obtained by all the three algorithms is greater than 99.9%. Since part of the intrusion detection is handled by the gateway, the overheads of the node’s CPU decreases. Security mechanism need to be implemented to prevent attacks within the network as the gateway handles the security concerns external to the network. The proposed method can improve the QOS of MANET containing CPU constrained nodes, as part of the intrusion detection process is handled by the gateway. It is also found that decision tree based algorithms were able to consistently classify intrusion data as against Naïve Bayesian method. Further investigation needs to be done on effect of malicious node in the performance of an Ad hoc Network.

REFERENCES

[1] J. Liu, F. R. Yu, C. H. Lung and H. Tang, “Optimal Combined Intrusion Detection and Biometric-based Continuous Authentication in High Security Mobile ad hoc networks”. IEEE Transactions on Wireless Communications-2009, Volume : 8 , Issue:2 pp 806.

[2] T. Clausen, P. Jaquet, et.al. “Optimized link state routing protocol”. Internet Draft, draft-ietfmanet-olsr-06.txt, work in progress-2001, pp 186 -190.

[3] C. Perkins and P. Bhagwat, “Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers,” ACM SIGCOMM Computer Communication Review-1994, pp 234 – 244.

[4] C. Perkins and E Royer, “Ad Hoc On-Demand Distance Vector Routing,” Second IEEE Workshop on Mobile Computer Systems and Applications-February 1999, pp 90-100.

[5] Yu, X. Distributed cache updating for the dynamic source routing protocol.IEEE Transactions on Mobile Computing, June-2006, Volume 5, Issue 6, pp 609.

[6] A. K. Rai, R. R. Tewari and S. K. Upadhyay "Different Types of Attacks on Integrated MANET-Internet Communication" International Journal of Computer Science and Security (IJCSS) Volume (4): Issue (3) 265 pp 245-74.

[7] Satria Mandala, Md. Asri Ngadi, A. Hanan Abdullah "A Survey on MANET Intrusion Detection"International Journal of Computer Science and Security, Volume (2) : Issue (1)-1999, pp 1-11.

[8] I. Rish, “An empirical study of the naive Bayes classifier”, Workshop on Empirical Methods in Artificial Intelligence-IJCAI 2001, pp 41-46.

[9] D. P. Bhukya and S. Ramachandram "Decision Tree Induction: An Approach for Data Classification Using AVL-Tree" International Journal of Computer and Electrical Engineering, Vol. 2, No. 4, August-2010, pp 660-665.

[10] Guofei Gu, Alvaro A. Cárdenas, Wenke Lee. Principled Reasoning and Practical Applications of Alert Fusion in Intrusion Detection Systems. ASIACCS-2008, Japan, pp 120-128.

[11] A. Mitrokotsa, M. Tsagkaris, and C. Douligeris, "Intrusion Detection in Mobile Ad Hoc Networks Using Classification Algorithms", presented at Clinical Orthopaedics and Related Research(CoRR)-2008, pp 140-156.

[12] A. T. Nuruzzaman, S. Haque, M. N. Masum, “Modification of DSR and its implementation in Ad Hoc City”, Tenth International Conference on Computer and Information technology-2007, pp 1-6.

[13] S. Bose, S. Bharathimurugan, A. Kannan, “Multi-Layer Integrated Anomaly Intrusion Detection System for Mobile Adhoc Networks”, International Conference on Signal Processing, Communications and Networking-2007, pp 360-365.

[14] R. Shrestha, K. H. Han, D. Y. Choi, S. J. Han. “A Novel Cross Layer Intrusion Detection System in MANET”. Twenty fourth IEEE International Conference on Advanced Information Networking and Applications-2010, pp 647-654.

[15] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[16] K. Iwata, k. Ikeda, H.Sakai, “A new Criterion using Information Gain for Action Selection Strategy in Reinforcement Learning”, IEEE Transactions on Neural Networks-2004, Volume 15, Issue 4, pp 792.

[17] Data mining : Concepts and Techniques, Jiawei Han and Micheline Kamber, 2nd edition, Morgan Kaufmann Publishers

AUTHORS PROFILE

Mr. Manikandan has eleven years of experience in teaching. He has worked as Lecturer in the Department of Information technology in Pavendar College of Enginering and Technology , Trichy. Then worked as Lecturer in the Department of Information technology in M.N.M Jain Engineering College, Chennai. Further worked as Senior Lecturer and Head Incharge in the Department of Computer Science and Engineering in Mohamed Sathak A.J College of Engineering, Chennai. He was Promoted and worked as Assistant Professor in the Department of Computer Science and Engineering in VELTECH, Chennai. Currently he is with P.B. College of Engineering, Chennai as Assistant Professor in the Department of Computer Science and Engineering. His areas of interests include Data Mining and Ad hoc Networks.

Dr. Manimegalai has seventeen years of experience in teaching, research and industry put-together. She has worked as software engineer in DCM Technologies, New Delhi and Xilinx India Technology Services, Hyderabad. Currently she is with Velammal Engineering College, Chennai as Senior Professor in the Department of Computer Science and Engineering. She is life member in Computer Society of India, Institution of Engineers (India) and Indian Society for Technical Education. She is also a member of IEEE and VLSI society of India. Her areas of interests include Algorithms for VLSI/FPGA Design, Reconfigurable Computing, Natural Language Processing and Computer Networks.