INTRUSION DETECTION USING NAÏVE BAYES FOR REAL TIME DATA

Shubhangi S. Gujar and B. M. Patil
P. G. Dept., MBES College of Engineering, Ambajogai, India

ABSTRACT

Recently, network security has become a key issue in information technology as there is increasing in security threats. A variety of intrusion detection systems (IDS) have been proposed for protecting computers and networks from malicious network-based or host-based attacks. A network intrusion detection system monitors traffic on a network looking for suspicious activity, which could be an attack or unauthorized activity. Most of the researchers were used KDD99 dataset but in this paper we are creating our own dataset by real time packet capturing. We use Naïve bayes classifier for intrusion detection which classifies whether the attack is present or not.

KEYWORDS: *Intrusion detection, Naïve Bayes classifier, Real time data.*

I. INTRODUCTION

An intrusion detection system refers to the instruments that are developed to detect infringement of system security policy. Intrusive activities are observed different from normal activities and thus it is noticeable. It is not introduced to replace the prevention based methods such as authentication and access control. It is proposed to complement existing security measures and detect actions that bypass the security monitoring and control component of the system. Basically an intrusion detection would cause of loss of integrity confidentiality, denial of resources, or unauthorized use of resources which include the some of these are illegal modifications of system files so as to facilitate illegal access to either system or user information. Not permitted modifications of tables or other system information in network components unauthorized use of computing resources [1]. Intrusion detection system should able to detect different types of attacks and must not recognize any legitimate activity as an attack. At the same time, the IDS must not fail to recognize any real attacks.

The early intrusion was implemented for host-based that located in servers to examine the internal interfaces [2]-[4], latter the focus was shifted toward network-based. Network based intrusion detection system performs packet logging, real-time traffic analysis of IP network, and tries to discover if an intruder is attempting to break into the system [5]-[7]. Basically there are two types of detection models misuse and anomaly are commonly using by IDS. The Misuse detection model performs simple pattern matching techniques to match an attack pattern corresponding to known attack patterns in the database and produces very low false positives. Anomaly detection model identifies new attacks by evaluate the anomalous behaviors from normal behaviors [8], and achieves high detection rates for new attacks, but produces many false positives (FP). Anomaly based IDS generate rules by observing collected audit data that is the records of activities generated by the operating system. Currently adaptive intrusion detection aims to solve the problems of analyzing the huge volumes of audit data and realizing performance optimization of detection rules [9]-[13]. Intrusion Detection System plays vital role of detecting various kinds of attacks and secures the networks. Naïve bayes classifier is a simple probabilistic classifier based on applying bayes theorem with strong independence assumptions. Naive bayes classifiers have worked quite well in many complex real-world situations.

The rest of the paper is organized in the following manner; Section 2 describes related work on intrusion detection system, Section 3 describes our proposed method and section 4 presents the experimental results. Finally, section 5 provides the concluding remarks and future scope of the work.

II. RELATED WORK

Panda and Patra [14] proposed a framework of NIDS based on Naïve bayes algorithm. It performs better in terms of false positive rate, cost, and computational time when applied to KDD99 dataset compared to a back propagation neural network based approach. Yoshimasa Tsuruokaand Junichi Tsujii [15] have combined the Naive bayes classifier with the well-established EM algorithm to exploit the unlabeled data. A class distribution constraint is introduced into the iteration process of the EM algorithm. This constraint keeps the class distribution of unlabeled data consistent with the true class distribution estimated from labeled data, preventing the EM algorithm from converging into an undesirable state. A new learning algorithm for adaptive network intrusion detection using Naive Bayesian classifier and decision tree is presented, which performs high detection rates (DR) and significant reduce false positives (FP) for different types of network intrusions using limited computational resources[16].In [18], authors developed Decision Support in Heart Disease Prediction System using Naive Bayesian Classification technique. The system extracts hidden knowledge from a historical heart disease database. This model could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy.
For a wide range of benchmark datasets, Naive bayes models learned using EM has accuracy and learning time comparable to Bayesian networks with context-specific independence [21].
Naive bayes inference is orders of magnitude faster than Bayesian network inference using Gibbs sampling and belief propagation. Puttini et al. [22] have presented a new anomaly IDS design using a parametric mixture model for behavior modeling and Bayesian based detection. Continuous model update is accomplished by model parameter re-estimation. Their preliminary experiments show that proposed algorithms present real-time feasibility with no special hardware requirement.
Intrusion detection is the process of monitoring and analyzing the events in computer systems or networks to discover the signals of possible incidents, which attempt to compromise the confidentiality, integrity, and availability of computer resources [16]. There are two types of intrusion detection systems that employ one or both of the intrusion detection methods. Host-based systems base their decisions on information obtained from a single host, while network-based intrusion detection systems obtain data by monitoring the traffic in the network to which the hosts are connected.

III. THE PROPOSED MODEL

Naive bayes Rule is the basis for many machine-learning and data mining methods. This algorithm is used to create models with predictive capabilities. It provides new ways of exploring and understanding data. Generally, Naive bayes classifier technique is used when the data is high and when the attributes are independent of each other [18]. Naive bayes classifier algorithm is used to model normal and suspicious network activity. The Naive bayes classifier is a supervised learning algorithm based largely off of bayes Theorem,

$$P(B / A) = \frac{P(A / B)\, P(B)}{P(A)}$$

We can calculate the probability that an attack is occurring based on some data by first calculating the probability that some previous data was part of that type of attack and then multiplying by the probability of that type of attack occurring [20].
In this paper we are considering 3 attacks and 1 normal behavior of the packet as follows:

- A SYN flood is a form of denial-of-service attack in which an attacker sends a succession of SYN requests to a target's system in an attempt to consume enough server resources to make the system unresponsive to legitimate traffic.
- A UDP flood attack is a denial-of-service attack that can be initiated by sending a large number of UDP packets to random ports on a remote host.

- A TCP Data flood is the denial of service attack in which an attacker sends TCP data as fast as the air interface will allow.
- Normal connections are generated by simulated daily user behavior.

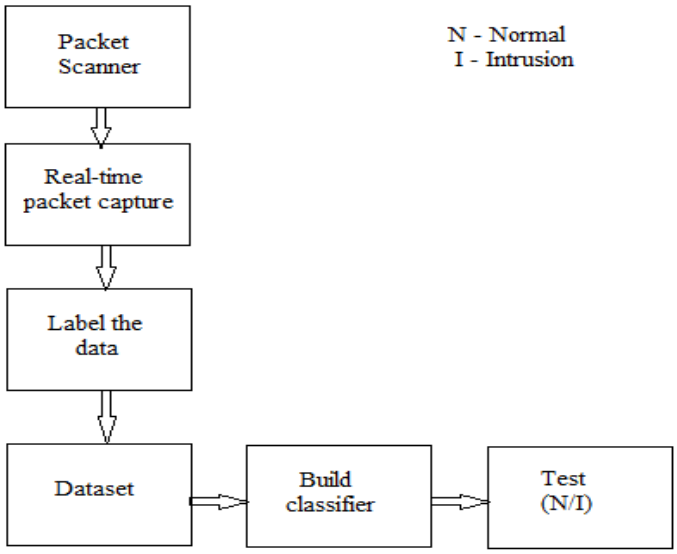

Figure 1: Proposed framework of NIDS

The proposed framework of network intrusion detection system is shown in the Figure 1. The first step is scanning the packets from the network traffic. Capture the real time packets and assign the labels in order to create training dataset. Build the classifier model on training dataset and finally use the Naïve bayes classifier on real time data to detect the given packet is NORMAL or INTRUSION.

IV. EXPERIMENTAL RESULTS

In this paper we are creating our own dataset instead of using KDD99 dataset. The first step for creating dataset is packet capturing. Figure 2 shows the raw packet capturing.

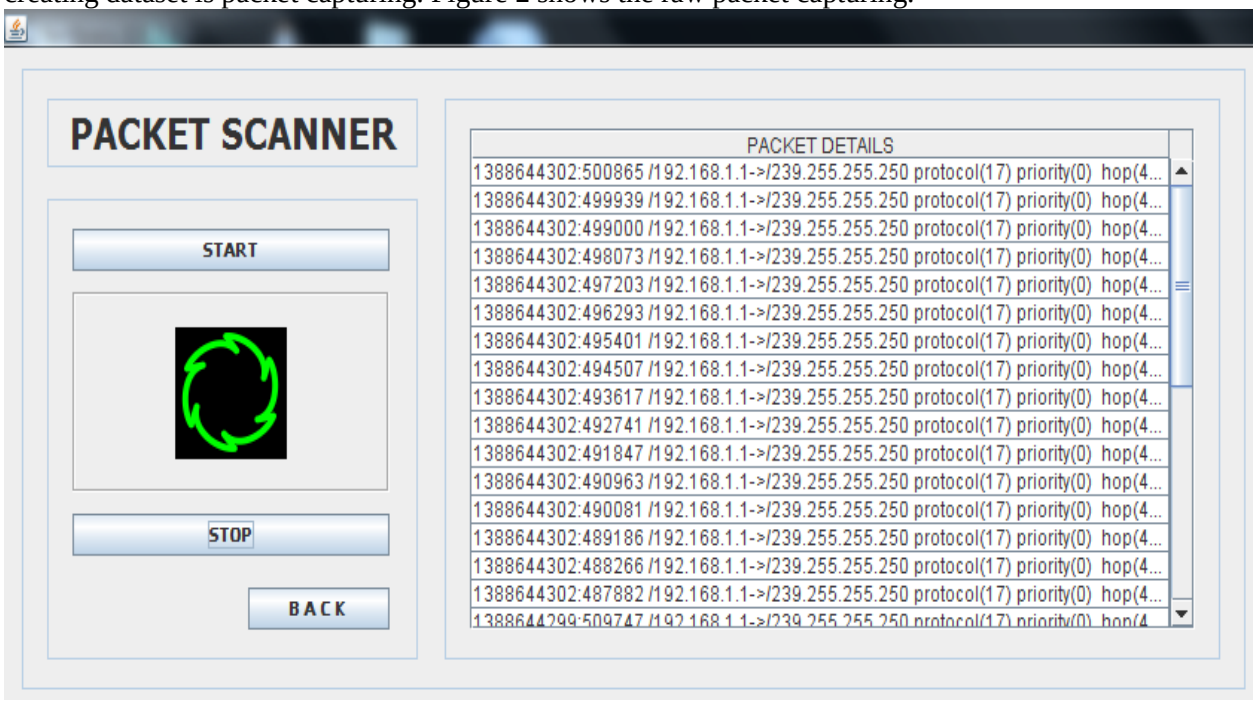

Figure 2: Raw packet capture

For training dataset, capture the real time data as shown in Figure 3, after packet capturing, we can add those captured packets into the dataset and manage the dataset by labeling the packets as normal or intrusion as shown in Figure 4.

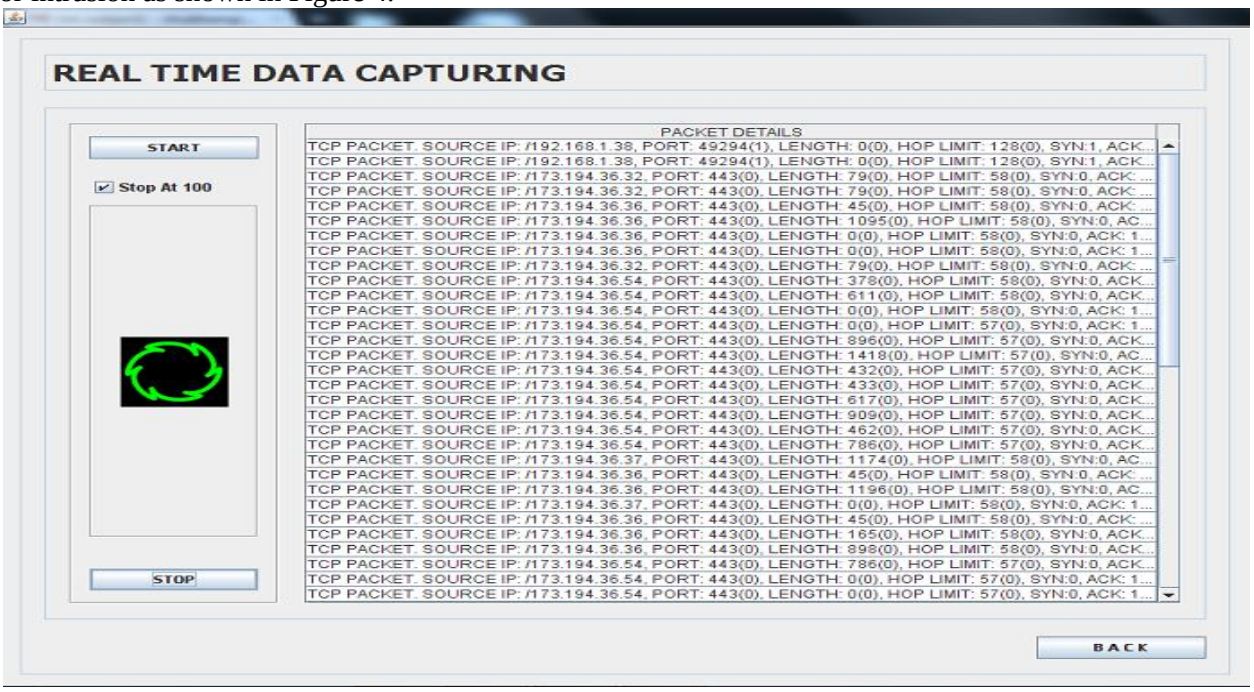

Figure 3: Real-time packet capture

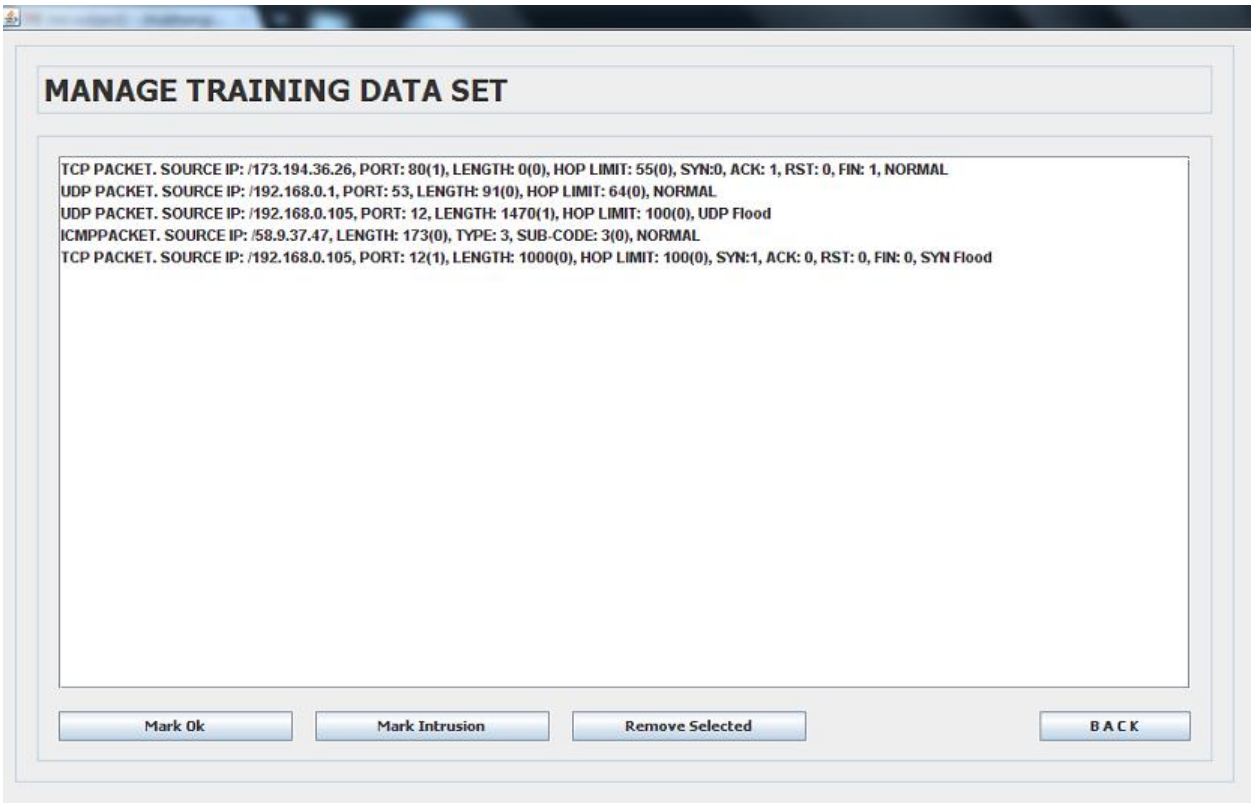

Figure4: Training dataset

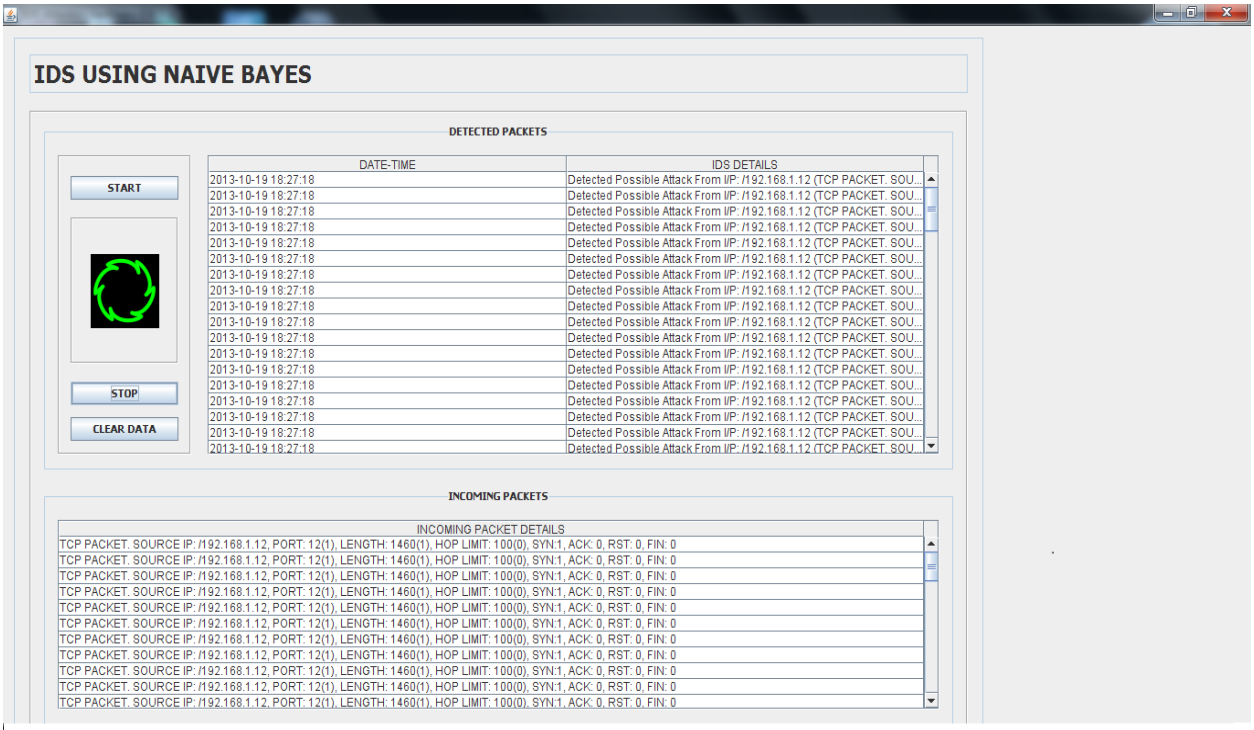

Figure5: IDS using Naïve Bayes

Finally, Figure 5 shows the IDS using Naïve bayes consists of incoming packets and detected packets. Incoming packets are the real time packets and detected packets are the intrusions detected from real time packets (data). Table 1 shows the experimental results of detecting attacks and the normal behavior of packets.

Table 1. Experimental results

Type	**Result (in %)**
SYN flood	90.46
TCP data flood	92
UDP flood	87.58
Normal	96

V. CONCLUSIONS AND FUTURE WORK

Intrusion detection is very important part of network security. Many machine learning algorithms for intrusion detection require a training data set and most of the researchers use the KDD99 dataset. In this paper, we have proposed a framework of NIDS based on Naïve bayes algorithm. We are creating our own dataset by real time data capturing. An experimental result shows that the proposed method is best for real time data. This work cane extended by using different classifiers in real time environment and also they can extend this method for detecting different types of networking attacks in real time environment.

REFERENCES

[1] Bishop Matt. Computer security e art and science: Addison Wesley; 2003.
[2] Jackson, T., Levine, J., Grizzard, J., Owen, and H., "An investigation of a compromised host ona honeynet being used to increase the security of a large enterprise network," IEEE workshop on Information Assurance and Security, IEEE, 2004.
[3] D. Y. Yeung, and Y. X. Ding, "Host-based intrusion detection using dynamic and static behavioral models," *Pattern Recognition*, 36, 2003, pp. 229-243.

[4] X. Xu, and T. Xie, "A reinforcement learning approach for host-based intrusion detection using sequences of system calls," In Proc. of International Conference on Intelligent Computing,Lecture Notes in Computer Science, LNCS 3644, 2005, pp. 995-1003.
[5] Krasser, S., Grizzard, J., Owen, H., and Levine. J., "The use of honeynets to increase computer network security and user awareness," Journal of Security Education, vol. 1, 2005, pp. 23-37.
[6] Shon T., Seo J., and Moon J., "SVM approach with a genetic algorithm for network intrusion detection," in Proc. of 20th International Symposium on Computer and Information Sciences (ISCIS 2005), Berlin: Springer-Verlag, 2005, pp. 224-233.
[7] X. Xu, X.N. Wang, "Adaptive network intrusion detection method based on PCA and supportvector machines," Lecture Notes in Artificial Intelligence (ADMA 2005), LNAI 3584, 2005, pp.696-703.
[8] Lazarevic, A., Ertoz, L., Kumar, V., Ozgur,. A., Srivastava, and J., "A comparative study of anomaly detection schemes in network intrusion detection," In Proc. of the SIAM Conference on Data Mining, 2003.
[9] Sebastiaan Tesink, "Improving intrusion detection system through machine learning," Technical Report, Series no. 07-02, ILK Research Group, Tilburg University, March, 2007.
[10] Barbara, Daniel, Couto, Julia, Jajodia, Sushil, Popyack, Leonard, Wu, and Ningning, "ADAM: Detecting intrusion by data mining," IEEE Workshop on Information Assurance and Security, West Point, New York, June 5-6, 2001.
[11] Lee W., "A data mining and CIDF based approach for detecting novel and distributed intrusions," Recent Advances in Intrusion Detection, 3rd International Workshop, RAID 2000,Toulouse, France, October 2-4, 2000, Proc. Lecture Notes in Computer Science 1907 Springer,2000, pp. 49-65.
[12] Lee W., Stolfo S., and Mok K., "Adaptive Intrusion Detection: A Data Mining Approach,"Artificial Intelligence Review, 14(6), December 2000, pp. 533-567.
[13] Stolfo J., Fan W., Lee W., Prodromidis A., and Chan P.K., "Cost-based modeling and evaluation for data mining with application to fraud and intrusion detection," DARPA Information Survivability Conference, 2000.
[14] Mrutyunjaya Panda and ManasRanjanPatra, "Network Intrusion Detection Using Naive Bayes",IJCSNS International Journal of Computer Science and Network Security, vol.7 no.12, December 2007.
[15] YoshimasaTsuruokaand Jun'ichiTsujii, "Training a Naive Bayes Classifier via the EM Algorithm with a ClassDistribution Constraint".*Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*. Association for Computational Linguistics, 2003.
[16] Dewan Md. Farid, NouriaHarbi, and Mohammad Zahidur Rahman, "COMBINING NAIVE BAYES AND DECISION TREEFOR ADAPTIVE INTRUSION DETECTION", International Journal of Network Security & Its Applications (IJNSA), vol 2, no 2, April 2010.
[17] Srilatha Chebrolu, Ajith Abraham and Johnson P. Thomas, "Feature deduction and ensembledesign of intrusion detection systems", Computers& Security (2005) 24, pp.295-307.
[18] Mrs. G. Subbalakshmi, Mr. K. Ramesh and Mr. M. ChinnaRao, "Decision Support in Heart Disease Prediction System using Naive Bayes", Indian Journal of Computer Science and Engineering (IJCSE), ISSN : 0976-5166, vol. 2 no. 2 Apr-May 2011.
[19] Dewan Md. Farid, NouriaHarbi, EmnaBahri, Mohammad Zahidur Rahman, Chowdhury Mofizur Rahman, "Attacks Classification in Adaptive Intrusion Detection using Decision Tree", World Academy of Science, Engineering and Technology 2010.
[20] Jonathan Palmer, "Naive Bayes Classification for Intrusion DetectionUsing Live Packet Capture", Data Mining in Bioinformatics, Spring 2011.
[21] Daniel Lowd& Pedro Domingos, "Naive Bayes Models for Probability Estimation", Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195-2350, USA.
[22] Ricardo S. Puttini, ZakiaMarrakchi, and LudovicMé, "A Bayesian Classification Model for Real-Time Intrusion Detection",Supélec, Campus de Rennes, ÉquipeSécurite des Systèmesd Information et Réseaux, Avenue de la Boulaie, BP 81127, 35511 Cesson Sévigné Cedex, France.

AUTHORS

Shubhangi S. Gujar presently working as Lecture at the Department of Computer Science & Engineering in M. S. Bidve Engineering College,Latur, India. She completed her Bachelor's degree in Computer Science & Engineering Department from M.B.E. society's College of Engineering, Ambajogai under Dr. B.A.M.University, Aurangabad, India. She is pursuing her Master's Degree from the College of Engineering, Ambajogai. Her areas of research interest include Computer Networks &Wireless Network.

B.M. Patil is currently working as a Professor in P.G. Computer Science & Engineering Department in M.B.E.Society's College of Engineering, Ambajogai, India. He received his Bachelor's degree in Computer Engineering from Gulbarga University in 1993, MTech Software Engineering from Mysore University in 1999, and PhD Degree from Indian Institute of Technology, Roorkee, 2011. He has authored several papers in various international journals and conferences of repute. His current research interests include data mining, medical decision support systems, intrusion detection, cloud computing, artificial intelligence, artificial neural network, wireless network and network security.