Volume 4, Issue 1, January 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Exploration of Novel Layered Models for Improving Minority Attack Detection in IDS

Dr. Neelam Sharma
Department of Computer Science,
Rajasthan College of Engineering for Women
Jaipur, Rajasthan, India

Yatendra Mohan Sharma
Research Scholar, Department of Computer Science,
Banasthali University, Jaipur
(Rajasthan), India

Abstract—The security on the network is a critical issue due to the wideness of the network. The goal of intrusion detection system (IDS) is to detect illegitimate use that deviates significantly from normal behaviour, through constantly monitoring unusual user activity. While variety of security techniques are being developed and a lot of research is going on intrusion detection, but the field lacks an integrated approach with high detection rate for minority attacks namely R2L and U2R. Minority attacks are more dangerous than majority attacks like DoS and Probe. Hence, it is essential to improve the detection performance for the minority intrusions while maintaining a reasonable overall detection rate. The present day standalone IDS are not effective in detecting the minority attacks. In this paper we proposed layered model integrated with naïve bayes classifier. The result shows that this model drastically increase the prediction of minority attacks without hurting the prediction performance of the majority class.

Keywords— Network Security, Intrusion detection system, Feature selection, Naive Bayes classifier

I. INTRODUCTION

Today computers are part of networked; distributed systems that may span multiple buildings sometimes located thousands of miles apart. The network of such a system is a pathway for communication between the computers in the distributed system. The network is also a pathway for intrusion. An intrusion can be defined as: Any set of actions that attempt to compromise the integrity, confidentiality or availability of a resource [1], and they can be categorized into two main classes: Misuse intrusions: They are well defined attacks on known weak points of a system. They can be spotted by watching for certain actions being performed on certain objects. Anomaly intrusions: These are based on observations of deviations from normal system usage patterns. They are caught by examining log messages resulting from system calls. This can be done using a pattern matching approach such as in [2].

The network traffic is made up of attack and the normal traffic. The number of attacks on the network is typically a very small fraction of the total traffic. Even with the attack traffic, some attacks are rare or minor. On the basis of this the attacks can also be categorized into two classes, minority and majority attack class. The DoS and probe attacks belong to majority class whereas U2R and R2L belongs to minority class also called as rare class of attacks. The present day standalone IDS are not effective in detecting minority attacks. The crucial step in successfully detecting intrusions is to develop a model that describes most known as well as novel unseen attacks, while keeping a low false alarm rate. In real world environment, the minority attacks namely R2L and U2R/Data attacks are more dangerous than the majority attacks like probe and DoS. The U2R attacks are very difficult to detect since they involve the semantic details that are very difficult to capture at an early stage. Most of the present IDS fail to detect such attacks with acceptable reliability Thus the existing approaches of intrusion detection have focused to improve the detection rate of the minority class type.

In this paper, the data mining algorithm naïve bayes classifier (one of the most versatile machine learning algorithms) has been evaluated on the NSL KDD dataset to detect attacks on the four attack categories: Probe (information gathering), DoS (deny of service), U2R (user to root)and R2L (remote to local) with layered approach for improving the minority attack detection rate. The naive bayes classifiers results are computed to show that our proposed model is more efficient for network intrusion detection. Rest of the paper is organized as follows: Section 2 gives overview of intrusion detection system. Section 3 is discussing the related work, Section 4 gives overview of naive bayes classifier. Section 5 describes the approach used for the layered model in intrusion detection and proposed methods described in section 6, section 7 discussed experimental setup. The section 8 presents the result and finally the paper is concluded with their future work in section 9.

II. INTRUSION DETECTION SYSTEM

Intrusion detection is needed in today's computing environment because it is impossible to keep pace with the current and potential threats and vulnerabilities in our computing systems. The environment is constantly evolving and changing field by new technology and the Internet. An intrusion detection system (IDS) inspects all inbound and outbound network activity and identifies suspicious patterns that may indicate a network or system attack from someone attempting to break into or compromise a system. An IDS gathers and analyzes information from various areas within a computer or a

network to identify possible security breaches, which include both intrusions and misuse. Intrusion detection products are tools to assist in managing threats and vulnerabilities in this changing environment. Threats are people or groups who have the potential to compromise your computer system. These may be a curious teenager, a disgruntled employee, or espionage from a rival company or a foreign government [3]. Attacks on network computer system could be devastating and affect networks and corporate establishments. We need to curb these attacks and Intrusion Detection System helps to identify the intrusions. Without an NIDS, to monitor any network activity, possibly resulting in irreparable damage to an organization's network. A network intrusion falls into one of four categories.

- ❖ Denial-of-Service (DoS): Attackers tries to prevent legitimate users from accessing the service in the target machine.
- ❖ Probe: Attackers scanning the target machine to gain information about potential vulnerabilities.
- ❖ User-to-Root(U2R):Attackers has local access to the victim machine andtries to gain super user privileges,
- ❖ Remote-to-Local (R2L): Attackers does not have an account on the victim machine, hence tries to gain access.

III. RELATED WORK

IDSs are still experiencing difficulties in detecting intrusive activity on their networks since novel new attacks are consistently being encountered, and analysts can miss legitimate alarms when reviewing large alarm logs that contain a high number of false positives. There has been research investigating the use of data mining techniques to effectively detect malicious activity in an enterprise network. In [4] author shows that the accuracy and performance of an IDS can be improved through obtaining good training parameters and selecting right feature to design any Artificial Neural Network (ANN). In [5], author used PCA to project features space to principal feature space and select features corresponding to the highest eigen values using Genetic Algorithm. In [6] author propose "Enhanced Support Vector Decision Function " for feature selection , which is based on two important factors . First , the feature's rank, and second the correlation between the features. In [7], author propose an automatic feature selection procedure based on Correlation –based Feature Selection (CFS).

In [8] author investigate the performance of two feature selection algorithm involving Bayesian network(BN) and Classification & Regression Tee (CART) and ensemble of BN and CART and finally propose an hybrid architecture for combining different feature selection algorithms for intrusion detection. In [9], author proposes two phases approach in intrusion detection design. In the first phase, develop a correlation-based feature selection algorithm to remove the worthless information from the original high dimensional database. Next phase designs an intrusion detection method to solve the problems of uncertainty caused by limited and ambiguous information. In [10], Axellson wrote a well-known paper that uses the Bayesian rule of conditional probability to point out that implication of the base-rate fallacy for intrusion detection. In [11], a behavior model is introduced that uses Bayesian techniques to obtain model parameters with maximal a-posteriori probabilities. In [12] author presents an Intelligent Intrusion Detection and Prevention System (IIDPS), which monitors a single host system from three different layers; files analyzer, system resource and connection layers. The approach introduced, a multi – layered approach, in which each layer harnesses both aspects of existing approach, signature and anomaly approaches, to achieve a better detection and prevention capabilities. In [13] authors presented "A frame work using a layered approach for intrusion detection". They have addressed two main issues of ID i.e. accuracy and efficiency by using conditional random fields and layered approach. They have shown that layered CRFs have very high attack detection rate 98.6% for probe and 97.40% for DOS. However, they were outperformed by a significant percent for the R2L and U2R attacks. [14], Author presents a novel approach for learning from imbalanced data sets, based on a combination of the SMOTE algorithm and the boosting procedure. Unlike standard boosting where all misclassified examples are given equal weights, SMOTEBoost creates synthetic examples from the rare or minority class, thus indirectly changing the updating weights and compensating for skewed distributions.

In [15], Author presented the data-dependent sensor fusion architecture to reduces the false positive rate and improves the overall detection rate and also the detection rate of minority class types in particular. In [16] Author consider three levels of attack granularities depending on whether dealing with whole attack or grouping them in four main categories or just focusing on normal and abnormal behaviours. In the whole experimentation, compare the performance of naïve bayes networks with decision trees and found the performance of naïve byes better than the decision trees.

IV. NAIVE BAYES CLASSIFIER

The Naïve Bayes classifier technique is based on the Bayesian theorem and is particular suited when the dimensionality of the input is high. Despite its simplicity Naïve Bayes can often outperforms more sophisticated classification method. It works on strong independence relation assumption [17], that is, features are independent in the context of a session class and the probability of one attribute does not affect the probability of the other. It is defined as follows:

$$P(c|X) = \frac{P(X|c)P(c)}{P(X)}$$

i.e.

$$P(c|X) = P(x1|c) \times P(x2|c) \times \ldots \times P(xn|c) \times P(c)$$

Where,

—P(c|x) is the posterior probability of class (target) given

predictor (attribute).
—P(c) is the prior probability of class.
—P(x|c) is the likelihood which is the probability of predictor given class.
—P(x) is the prior probability of predictor.

Assume that, the effect of the value of a predictor(X) on a given class (C) is independent of the values of other predictor. The work reported in[18] examines the circumstances under which the naïve bayes classifier performs well and why. It states that the error is a result of three factors: training data noise, bias, and variance. Training data noise can only be minimized by choosing good training data. The training data must be divided into various groups by the machine learning algorithm. Bias is the error due to groupings in the training data being very large. Variance is the error due to those groupings being too small.

Discretization for Naïve Bayes Classifier

Research study shows that Naïve Bayes classification works best for discretized attributes and discretization effectively approximates a continuous variable [19]. We used the entropy-based supervised discretization (EBD) method proposed by Fayyad and Irani [20]. It discretizes numeric attributes first using Minimum Description Length(MDL) method.

Given a set of samples *I*, the basic method for EBD of an attribute A is as follows:

1. Each value *v* of *A* can be considered as a potential interval boundary Band thereby can create a binary discretization (e.g. $A < v$ and $A \geq v$).
2. Given I, the boundary value selected is the one that maximizes the information gain resulting from subsequent partitioning. The information gain is:

$$\text{InfoGain (I,B)} = E(I) - CIE(I,B)$$

where CIE(I,*B*) is the *class information entropy* determined by the formula:

$$CIE(I,B) = \frac{|I_1|}{|I|}E(I_1) + \frac{|I_2|}{|I|}E(I_2),$$

where $|I_1|$ and $|I_2|$ correspond to the examples of *I* satisfying the conditions $A < B$ and $A \geq B$ respectively. The entropy function E for a given set I_i is calculated based on the class distribution of the samples in the set, i.e.:

$$E(I_i) = -\sum_{j=1}^{m} \frac{c_j}{c} \log_2 \left(\frac{c_j}{c}\right),$$

Where $\frac{c_j}{c}$ is the probability of class c_j in I_i , determined by the proportion of samples of class $c_j$in the set I_i and m is the number of classes in I_i.
3. The process of determining a new interval boundary is recursively applied to each interval produced in previous steps, until the following stopping criterion Δ based on MDL principle is satisfied:

$$\text{InfoGain(I,B)} < \Delta$$

$$\Delta = \frac{log_2(n-1) + log_2(3^m - 2) - [mE(I) - m_1E(I_1 - mE(I_2)]}{n},$$

Where m_i is the number of classes represented in the set I_i and n is the number of samples in I.

Since the described above procedure is applied independently for each interval, it is possible to achieve the final set of discretization intervals with different size that is, some areas in the continuous spaces will be partitioned very finely whereas others (with relatively low entropy) will be partitioned roughly

V. LAYERED MODEL FOR INTRUSION DETECTION

In layered model we define four layers that correspond to the four attack groups i.e. DoS layer for detecting DoS attacks, Probe layer for detecting probe attacks, R2L layer for detecting R2L attacks and U2R for U2R attacks. Each layer is separately trained with a small set of relevant features. This is because all the 41 features are not required for detecting attacks belonging to a particular attack group.

The goal behind the layered approach is to improve the minority attack detection rate, while maintaining a reasonable overall detection rate. During the analysis of intrusion detection we observe two main challenging issues in this system. First, the number of intrusions on the network is typically a very small fraction of the total traffic. Therefore the essential

step in successfully detecting intrusions is to develop a model that describes most known as well as novel unseen attacks. Second, the attack groups are different in their impact and hence, it becomes necessary to treat them differently. We select features for each layer, based upon the type of attacks that the layer is trained to detect feasibility of each feature before selecting it for a particular layer. The framework for the Layered Model shown in Fig.1

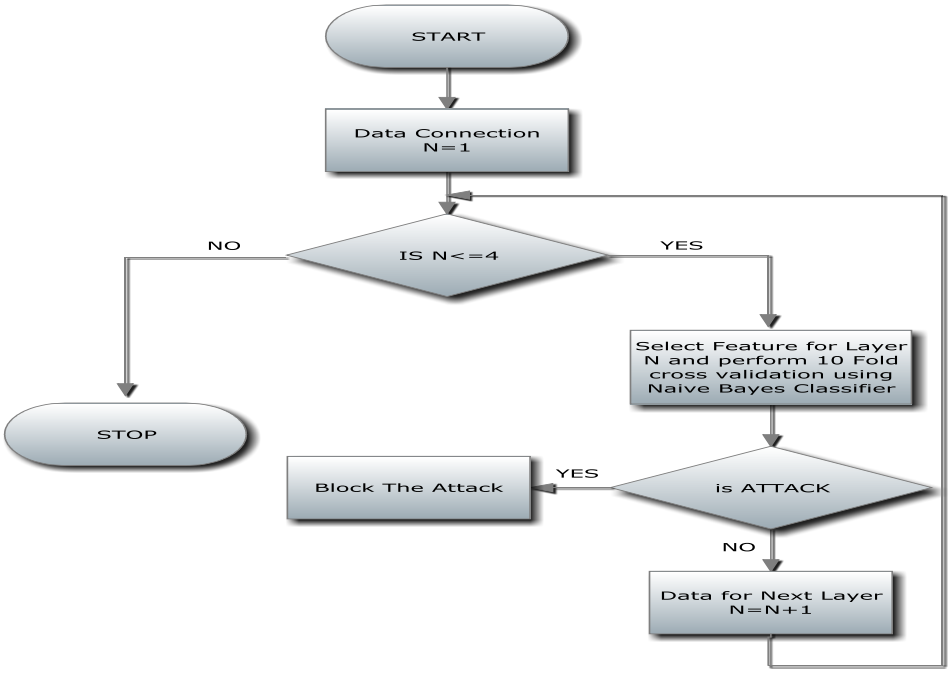

Fig.1 Framework of Layered Model for IDS

Feature Selection

Feature selection is an effective and an essential step in successful high dimensionality data mining applications [21]. It is often an essential data processing step prior to applying a learning algorithm. Reduction of the attribute space leads to a better understandable model and simplifies the usage of different visualization technique. In proposed model every layer is independent of every other layer.

We used domain knowledge and the sequential search to identify the important set of features, starting with the set of all features; one feature was removed at a time until the accuracy of the classifier was below a certain threshold. In other words, the feature selection of is "leave-one-out" remove one feature from the original dataset , redo the experiment , then compare the new results with the original results. Thus, from the total 41 features, we selected 16 features for U2R layer, 10 for R2L, 5 for probe and 3 for DoS Layer. The selected feature set of proposed model for all the four layers are:

Table 1
Feature Selected for DoS Layer

Feature Number	Name of the Feature
5	src_bytes
6	dst_bytes
24	srv_count

Table 2: Feature Selected for Probe Layer

Feature Number	Name of the Feature
1	duration
5	src_bytes
6	dst_bytes
30	diff_srv_rate
33	dst_host_srv_count

Table 3
Feature Selected for U2R Layer

Feature Number	Name of the Feature
1	duration
3	service
5	src_bytes
6	dst_bytes
10	hot
11	num_failed_login
13	num_compromised
14	root_shell
16	num_root
17	num_file_creations
31	srv_diff_host_rate
32	dst_host_count
33	dst_host_srv_count
34	dst_host_same_srv_rate
36	dst_host_same_src_port_rate
37	dst_host_srv_diff_host_rate

Table 4
Feature Selected for R2L Layer

Feature Number	Name of the Feature
1	duration
3	service
5	src_bytes
6	dst_bytes
23	count
24	srv_count
30	diff_srv_rate
31	srv_diff_host_rate
32	dst_host_count
36	dst_host_same_src_port_rate

In order to make the layers independent, some features may be present in more than one layer i.e. the feature set for the layer is not disjoint. We represent the working of layered approach in fig. 2.

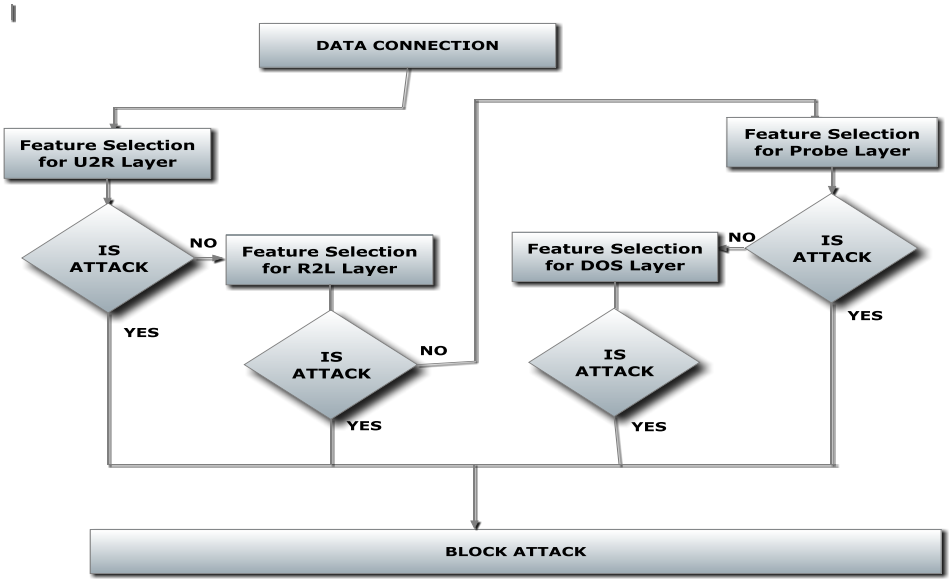

Fig. 2 Working of Layered Approach

VI. PROPOSED METHOD

In this paper we proposed two different approaches for intrusion detection using layered model with naïve bayes classifier on discretized values. Since results using discretized features are usually more compact, shorter and accurate using continuous values. We implement the layered approach by selecting small set of features for every layer rather than using all the 41 features. We integrate layered approach and the naïve bayes classifier to build a single system.

Working of the Layered Model 1

In the first approach, we train and test each layer to detect only a particular type of attack. For example, first layer of our proposed model is trained to detect U2R attacks only. When such a system is deployed online, other attacks such as probe can either be seen as normal or attack, we expect them to be detected as attack at other layers in the system. Hence, for four attack classes, we have four independent layer models, which are trained separately with specific features to detect attacks belonging to that particular group.

Working of the Layered Model 2

We implement the layered approach 2 almost as the similar way of layered approach 1. The difference between the both is that, in approach 1 at layers we are detecting only a particular type of attack. For example at U2R layer we are interested only in detection of U2R attack and after correctly detection of attack we block only those instances which are affected by U2R attack. Then second layer such as R2L, it detects instances which are infected by R2L attacks. The same process repeated for Probe and Dos attacks at layer third and fourth.

In the second approach we are not only interested to block attacks of U2R types but also the other three attacks namely R2L, Probe and DoS are also blocked after the detection as U2R. However, if the probe attacks are detected as U2R, it must be considered as an advantage since the attack is detected at an early stage. Similarly, if some U2R attacks are not detected at the U2R layer, they may be detected at subsequent layers. By this approach we are able to block more attacks at layer first in comparison of layered approach 1.The same process repeated at layer second, third and fourth for attack type R2L, Probe and DoS respectively.

VII. EXPERIMENTAL SETUP

We used WEKA 3.6 a machine learning tool [22], to compute the feature selection subsets for different layers, and to measure the classification performance on each of these feature sets. We choose the Naïve Bayes classifier with full training set and 10-fold cross validation for the testing purposes. In 10-fold cross-validation, the available data is randomly divided into 10 disjoint subsets of approximately equal size. One of the subsets is then used as the test set and the remaining 9 sets are used for building the classifier. The test set is then used to estimate the accuracy. This is done repeatedly 10 times so that each subset isused as a test subset once. The accuracy estimates is then the mean of the estimates for each of the classifiers. Cross-validation has been tested extensively and has been found to generally work well when sufficient data is available.

Dataset Description

The data set to be used in our experiments is NSL-KDD labeled dataset. NSL-KDD dataset suggested to solve some of the inherent problems of the KDD'99 data set[23].The number of records in the NSL-KDD train and test sets are reasonable. This advantage makes it affordable to run the experiments on the complete set without the need to randomly select a small portion. Consequently, evaluation results of different research work will be consistent and comparable. For our experiment dataset is either labeled as normal or as one of the 24 different kinds of attack. These 24 attacks can be grouped into four classes ; Probe, DoS, R2L and U2R We have used 1,25,973 NSL-KDD dataset connections for training and testing. Table 5 shows the distribution of classes in the actual training data for classifiers evaluation and the percentage of attacks is displayed using bar chart in Fig 3.

Table 5 Exemplify distribution of classes and the percentage of attacks

Category of Class (Class)	Number of instances/records	Percentage of Class Occurrences (Approximate)
Normal	67,343	53.46
DoS	45,927	36.46
Probe	11,656	9.25
U2R	52	0.04
R2L	995	0.79
Total	1,25,973	100%

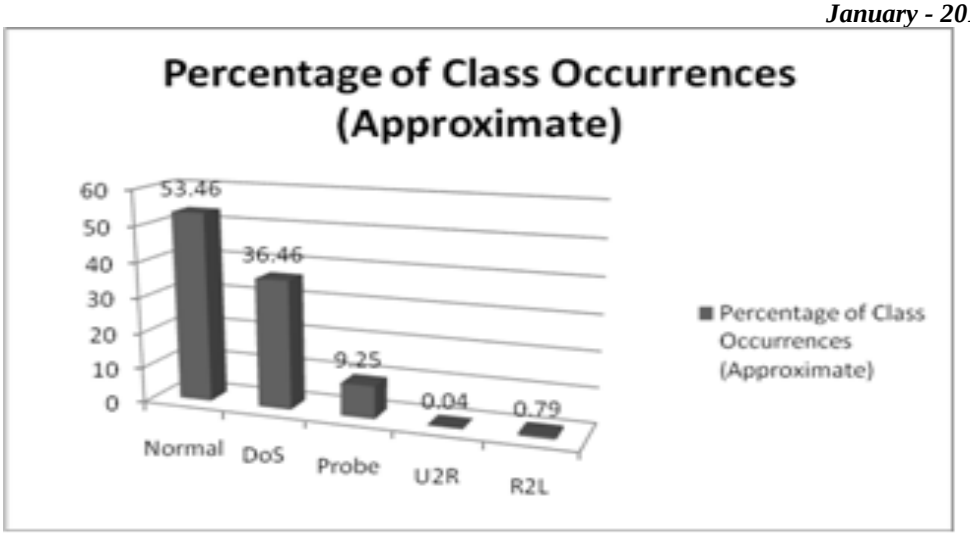

Fig 3 Percentage of Class Occurrences in Dataset

VIII. RESULTS

To evaluate the results of classifier, we have used standard metrics such as confusion matrix, true positive rate, false positive rate, and classifier's accuracy.

Confusion Matrix- This may be used to summarize the predictive performance of a classifier on test data. It is commonly encountered in a two-class format, but can be generated for any number of classes. A single prediction by a classifier can have four outcomes which are displayed in the following confusion matrix.

Confusion Matrix		Predicted Class	
		Class=Yes	Class=No
Actual Class	Class=Yes	TN	FP
	Class=No	FP	TN

True Positive (TP), the actual class of the test instance is positive and the classifier correctly predicts the class as positive. False Negative (FN), the actual class of the test instance is positive but the classifier incorrectly predicts the class as negative. False Positive (FP), the actual class of the test instance is negative but the classifier incorrectly predicts the class as positive. True Negative (TN), the actual class of the test instance is negative and the classifier correctly predicts the class as negative.

Positive Rate (TPR) or *Sensitivity* or *Recall (R)* is defined as:

$$TPR=TP/(TP+FN)$$

False Positive Rate (FPR) is:

$$FPR=FP/(TN+FP)$$

We can obtain the accuracy of a classifier by

$$Accuracy =(TP+TN)/(TP+FN+FP+TN) * 100\ \%$$

We perform two sets of experiments. First is for layered approach 1 and the second for layered approach 2. Table 3 and 4 clearly indicates empirical results of high detection rate/ TPR for minority attacks as well as majority attacks.

Table 6 : Result of Layered Approach 1

Attacks	Recall (%) with layered	Classifier's Accuracy at every layer
U2R	80.8	96.82
R2L	97	98.07
PROBE	98.8	97.94
DoS	99.9	96.88
Average Classifier's Accuracy		**97.43**

Table 7 Result of Layered Approach 2

Attacks	Recall (%)				
	U2R Layer	R2L Layer	Probe Layer	DoS Layer	Total Detection (%)
U2R	80.8	0	3.8	0	**84.6**
R2L	1.61	96.28	0	0.4	**98.2**
PROBE	0.14	0	98.74	0.19	**99.07**
DoS	0.06	0.008	0.68	99.17	**99.9**
Classifier's Accuracy at every layer	96.81	98.12	97.92	98.06	
Average Classifier's Accuracy					**97.72**

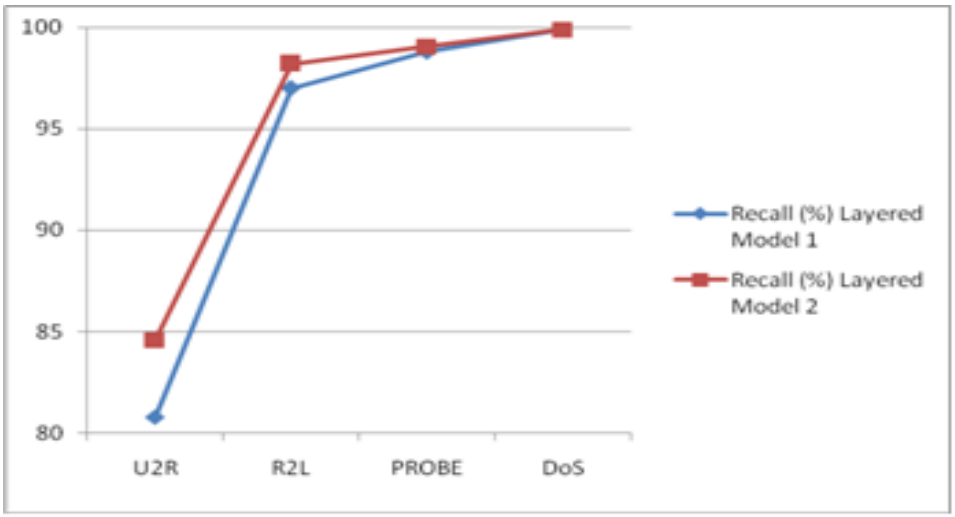

Fig 4 Compare Proposed Layered Result for Each Attack Class

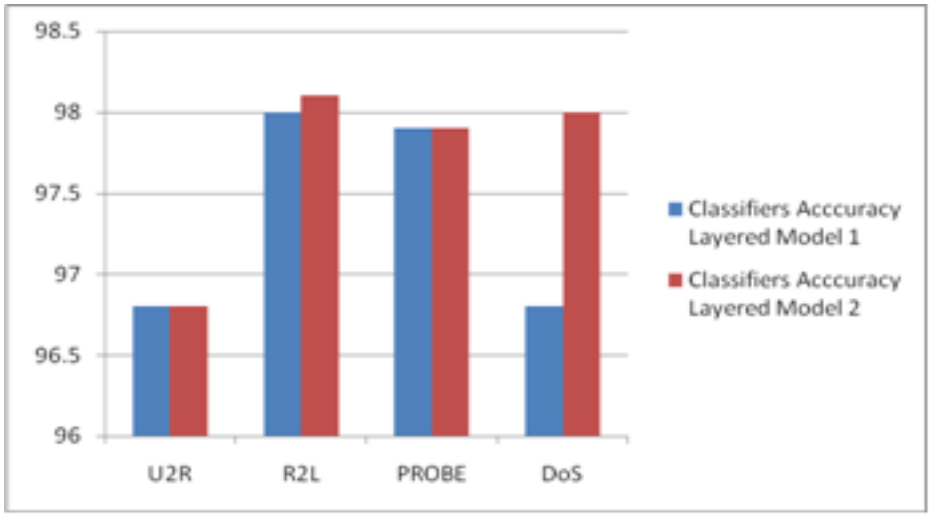

Fig 5 Layer-wise Classifier's Accuracy of Proposed Methods

We observe in Fig. 4 that the layered approach 2 obtained more high detection rate of U2R and R2L attacks in comparison of layered approach 1. Fig 5 displays the accuracy of classifiers at every layer of both proposed models.

IX. CONCLUSION & FUTURE WORK

Experimental results indicate that the proposed layered model with naïve bayes classifier can result in better prediction of minority classes without hurting the prediction performance of the majority class. The area of future research includes improving the recall, without sacrificing the precision. However, the recall and precision goals are often conflicting and attacking them simultaneously may not work well, especially when one class is rare.

REFERENCES

[1] Jian Pei Shambhu J. Upadhyaya Faisal FarooqVenugopalGovindaraju. Proceedings of the 20th International Conference on Data Engineering (ICDE04) 1063-6382/04 $ 20.00 © 2004 IEEE

[2] Debar, H., Dacier, M., and Wespi, A., A Revised taxonomyfor intrusion detection systems, Annales des Telecommunications, Vol. 55, No. 7–8, 361–378, 2000.

[3] Saman M. Abdulla, Najla B. Al-Dabagh, Omar Zakaria, Identify Features and Parameters to Devise an Accurate Intrusion Detection System Using Artificial Neural Network, World Academy of Science, Engineering and Technology 2010.

[4] I Ahmad, A B Abdulah, A S Alghamdi, K Alnfajan,MHussain, Feature Subset Selection for Network Intrusion Detection Mechanism Using Genetic Eigen Vectors, Proc .of CSIT vol.5 (2011)

[5] S Zaman, F Karray Features selection for intrusion detection systems based on support vector machinesCCNC'09 Proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference 2009

[6] H Nguyen, K Franke, S Petrovic Improving Effectiveness of Intrusion Detection by Correlation Feature Selection, 2010 International Conference on Availability, Reliability and Security,IEEE Pages-17-24

[7] S Chebrolu, A Abraham, J P. Thomas Feature deduction and ensemble design of intrusion detection systems, Computers &Security,Volume 24, Issue 4, June 2005, Pages 295-307

[8] T. S. Chou, K. K. Yen, and J. Luo "Network Intrusion Detection Design Using Feature Selection of Soft Computing Paradigms. International Journal of Computational Intelligence 4;3 2008

[9] S. Axelsson, "The base rate fallacy and its implications for the difficulty of Intrusion detection", Proc.Of 6th.ACM conference on computer and communication security 1999.

[10] R.Puttini, Z.marrakchi, and L. Me, "Bayesian classification model for Real time intrusion detection", Proc. of 22nd.International workshop on Bayesian inference and maximum entropy methods in science and engineering, 2002.

[11] OludeleAwodele, Sunday Idowu, OmotolaAnjorin, and Vincent J. Joshua "A Multi-Layered Approach to the Design of Intelligent Intrusion Detection and Prevention System (IIDPS)" Issues in Informing Science and Information Technology Volume 6, 2009.

[12] Kapil Kumar Gupta, BaikunthNath and Ramamohanarookotagiri, "A layered approach using conditional random fields for intrusion detection", IEEE Tranc.on Dependence and secure computing, Vol.7, 2010

[13] Nitesh V. Chawla1, Aleksandar Lazarevic2, Lawrence O. Hall3, Kevin Bowyer4, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting" , 7th European conference on principles and practice of knowledge discovery in databases (PKDD) pp. 107 to 109, dubrovnik, Croatia, 2003.

[14] Ciza Thomas a and N. Balakrishnan "Improvement in minority attack detection with skewness in network traffic" Data Mining, Intrusion Detection, Information Assurance, and Data Networks Security 2008, edited by Belur V. Dasarathy, Proc. of SPIE Vol. 6973, 69730N, (2008) · 0277-786X/08/$18 · doi: 10.1117/12.785623

[15] Nahla Ben Amor, Salem Benferhat, ZiedElouedi "Naive Bayes vs Decision Trees in Intrusion Detection Systems" AC'04, March 14-17, 2004, Nicosia, Cyprus. , March 14-17, 2004, icosia, Cyprus Copyright 2004 ACM 1-58113-812-1/03/2004 ... 5.00.

[16] Ms.Nivedita Naidu, Dr.R.V.Dharaskar "An effective approach to network intrusion detection system using genetic algorithm", International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 2, 2010.

[17] Chris Fleizach, Satoru Fukushima, A naive Bayes classifier on 1998 KDD Cup

[18] Chun-Nan Hsu, Hung-Ju Huang, Tsu-TsungWong,"Why Discretization works for Naïve Bayesian Classifiers", 17th ICML, pp 309-406, 2000.

[19] U.M. Fayyad, K.B Irani, "Multi-interval discretization of continuous-valued attributes for classification learning", In Proceedings of the 13th International Joint Conference on Artificial Intelligence, pp. 1022–1027,1993.

[20] Liu H ,Setiono R, Motoda H, Zhao Z Feature Selection: An Ever Evolving Frontier in Data Mining, JMLR: Workshop and Conference Proceedings 10: 4-13 The Fourth Workshop on Feature Selection in Data Mining.

[21] Weka: http://www.cs.waikato.ac.nz/~ml/weka/

[22] "NSL-KDD dataset for network –based intrusion detection systems" available on http://iscx.info/NSL-KDD/

Appendix A

List of 41 Features of NSL-KDD Dataset

Feature No.	Feature Name	Feature No.	Feature Name
1	duration	22	is guest login
2	protocol type	23	Count
3	service	24	srv count
4	flag.	25	serror rate
5	source bytes	26	srvserror rate bytes
6	destination	27	rerror rate
7	land	28	srvrerror rate
8	wrong fragment	29	samesrv rate
9	urgent	30	diffsrv rate
10	hot	31	srv diff host rate

11	failed logins.	32	dst host count
12	logged in	33	dst host srv count
13	# compromised	34	dst host same srv rate
14	root shell	35	dst host diff srv rate
15	su attempted	36	dst host same src port rate
16	# root	37	dst host srv diff host rate
17	# file creations	38	dst host serror rate
18	# shells	39	dst host srvserror rate
19	# access files	40	dst host rerror rate
20	# outbound cmds	41	dst host srvrerror rate
21	is host login		

