Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection

Kazi Abu Taher
Department of ICT, Bangladesh University of Professionals (BUP)
Dhaka, Bangladesh
kataher@yahoo.com

Billal Mohammed Yasin Jisan
Department of ICT, Bangladesh University of Professionals (BUP)
Dhaka, Bangladesh
gesagn@yahoo.com

Md. Mahbubur Rahman *Department of CSE, Military Institute of Science & Technology*, Dhaka, Bangladesh
mahbubcse@yahoo.com

***Abstract* – A novel supervised machine learning system is developed to classify network traffic whether it is malicious or benign. To find the best model considering detection success rate, combination of supervised learning algorithm and feature selection method have been used. Through this study, it is found that Artificial Neural Network (ANN) based machine learning with wrapper feature selection outperform support vector machine (SVM) technique while classifying network traffic. To evaluate the performance, NSL-KDD dataset is used to classify network traffic using SVM and ANN supervised machine learning techniques. Comparative study shows that the proposed model is efficient than other existing models with respect to intrusion detection success rate.**

***Index Terms*—intrusion, machine learning, deep learning, neural network, support vector machine, feature selection.**

I. INTRODUCTION

With the wide spreading usages of internet and increases in access to online contents, cybercrime is also happening at an increasing rate [1-2]. Intrusion detection is the first step to prevent security attack. Hence the security solutions such as Firewall, Intrusion Detection System (IDS), Unified Threat Modeling (UTM) and Intrusion Prevention System (IPS) are getting much attention in studies. IDS detects attacks from a variety of systems and network sources by collecting information and then analyzes the information for possible security breaches [3]. The network based IDS analyzes the data packets that travel over a network and this analysis are carried out in two ways. Till today anomaly based detection is far behind than the detection that works based on signature and hence anomaly based detection still remains a major area for research [4-5]. The challenges with anomaly based intrusion detection are that it needs to deal with novel attack for which there is no prior knowledge to identify the anomaly. Hence the system somehow needs to have the intelligence to segregate which traffic is harmless and which one is malicious or anomalous and for that machine learning techniques are being explored by the researchers over the last few years [6]. IDS however is not an answer to all security related problems. For example, IDS cannot compensate weak identification and authentication mechanisms or if there is a weakness in the network protocols.

Studying the field of intrusion detection first started in 1980 and the first such model was published in 1987 [7]. For the last few decades, though huge commercial investments and substantial research were done, intrusion detection technology is still immature and hence not effective [7]. While network IDS that works based on signature have seen commercial success and widespread adoption by the technology based organization throughout the globe, anomaly based network IDS have not gained success in the same scale. Due to that reason in the field of IDS, currently anomaly based detection is a major focus area of research and development [8]. And before going to any wide scale deployment of anomaly based intrusion detection system, key issues remain to be solved [8]. But the literature today is limited when it comes to compare on how intrusion detection performs when using supervised machine learning techniques [9]. To protect target systems and networks against malicious activities anomaly-based network IDS is a valuable technology. Despite the variety of anomaly-based network intrusion detection techniques described in the literature in recent years [8], anomaly detection functionalities enabled security tools are just beginning to appear, and some important problems remain to be solved. Several anomaly based techniques have been proposed including Linear Regression, Support Vector Machines (SVM), Genetic Algorithm, Gaussian mixture model, k-nearest neighbor algorithm, Naive Bayes classifier, Decision Tree [3,5]. Among them the most widely used learning algorithm is SVM as it has already established itself on different types of problem [10]. One major issue on anomaly based detection is though all these proposed techniques can detect novel attacks but they all suffer a high false alarm rate in general. The cause behind is the complexity of generating profiles of practical normal behavior by learning from the training data sets [11]. Today Artificial Neural Network (ANN) are often trained by the back propagation algorithm, which had been around since 1970 as the reverse mode of automatic differentiation [12].

The major challenges in evaluating performance of network IDS is the unavailability of a comprehensive network based data set [13]. Most of the proposed anomaly based techniques found in the literature were evaluated using KDD CUP 99 dataset [14]. In this paper we used SVM and ANN –two machine learning techniques, on NSL-KDD [15] which is a popular benchmark dataset for network intrusion.

The promise and the contribution machine learning did till today are fascinating. There are many real life applications we are using today offered by machine learning. It seems that machine learning will rule the world in coming days. Hence we came out into a hypothesis that the challenge of identifying new attacks or zero day attacks facing by the technology enabled organizations today can be overcome using machine learning techniques. Here we developed a supervised machine learning model that can classify unseen network traffic based on what is learnt from the seen traffic. We used both SVM and ANN learning algorithm to find the best classifier with higher accuracy and success rate. This paper is organized as follows: overview of the system model and design is explained in section II, the system experimental analysis is given in section III. Section

978-1-5386-8014-8/19/$31.00 ©2019 IEEE

IV contains the implementation discussion. Finally, section V concludes the paper.

II. SYSTEM MODEL

The system proposed is composed of feature selection and learning algorithm show in Fig.1. Feature selection component are responsible to extract most relevant features or attributes to identify the instance to a particular group or class. The learning algorithm component builds the necessary intelligence or knowledge using the result found from the feature selection component. Using the training dataset, the model gets trained and builds its intelligence. Then the learned intelligences are applied to the testing dataset to measure the accuracy of home much the model correctly classified on unseen data.

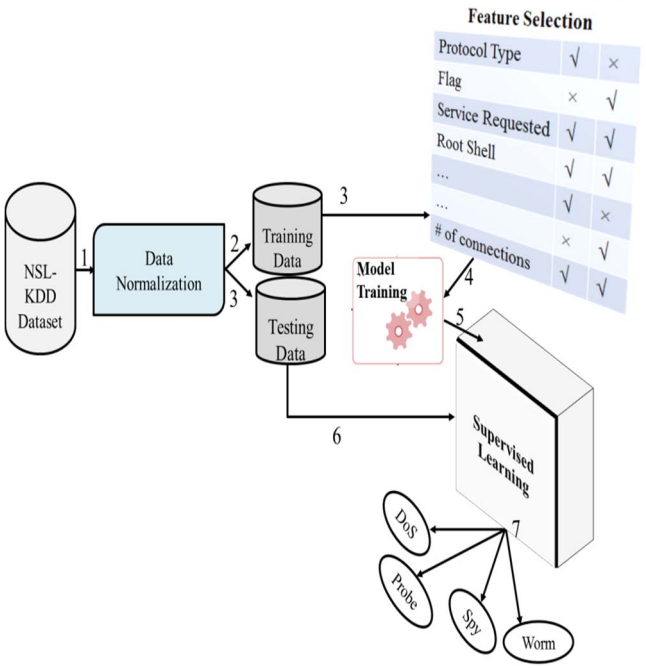

Fig 1: Proposed supervised machine learning classifier system

A. Feature Selection

Feature selection is an important part in machine learning to reduce data dimensionality and extensive research carried out for a reliable feature selection method. For feature selection filter method and wrapper method have been used. In filter method, features are selected on the basis of their scores in various statistical tests that measure the relevance of features by their correlation with dependent variable or outcome variable. Wrapper method finds a subset of features by measuring the usefulness of a subset of feature with the dependent variable. Hence filter methods are independent of any machine learning algorithm whereas in wrapper method the best feature subset selected depends on the machine learning algorithm used to train the model. In wrapper method a subset evaluator uses all possible subsets and then uses a classification algorithm to convince classifiers from the features in each subset. The classifier consider the subset of feature with which the classification algorithm performs the best. To find the subset, the evaluator uses different search techniques like depth first search, random search, breadth first search or hybrid search. The filter method uses an attribute evaluator along with a ranker to rank all the features in the dataset. Here one feature is omitted at a time that has lower ranks and then sees the predictive accuracy of the classification algorithm. Weights or rank put by the ranker algorithms are different than those by the classification algorithm. Wrapper method is useful for machine learning test whereas filter method is suitable for data mining test because data mining has thousands of millions of features.

B. Building Machine Intelligence

Based on the best features found in the feature selection process, learning models are developed. To develop the learning model, machine learning algorithm is used. Training dataset is used to train the algorithm with the selected features. In supervised machine learning, each instance in the training dataset has the class it belongs to. The algorithm build the learning model based on which machine learning algorithm is being used.

C. Support Vector Machine (SVM)

In SVM a separating hyper plane defines the classifier depending on the type of problem and available datasets. In case where dataset is one dimensional, the hyper plane is a point, for two dimensional data it is a separating line as shown in Fig 2, for three dimensional dataset, it is a plane and if the data dimension is higher it is a hyper plane. For a linearly separable dataset, the classifier or the decision function will have the form -

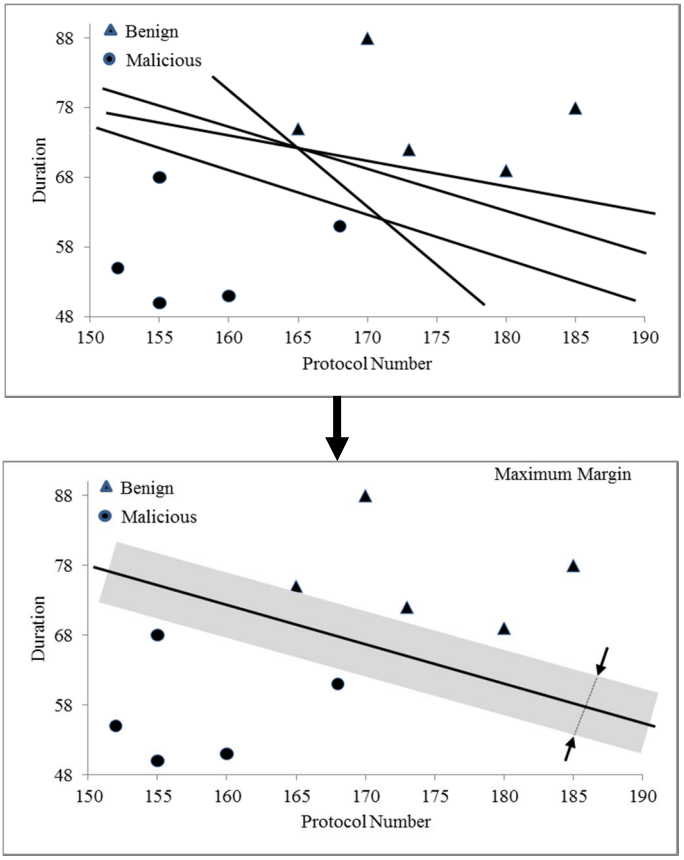

Fig 2: SVM classifier in two dimensional problem spaces

$$ax + by + c = 0 \quad (1)$$

For a given data points (x,y), the above decision function will classify the point in one class if ax + by ≥ c or it will categorize if ax + by < c. The equation of a line y=ax+b can be rewritten as y−ax−b=0 that can be represent using two vectors as below-

$$\mathbf{w}\begin{pmatrix}-b\\-a\\1\end{pmatrix} \text{ and } \mathbf{x}\begin{pmatrix}1\\x\\y\end{pmatrix} \quad (2)$$

which says we can write the linear equation of a line using two vectors as below-

$$\mathbf{w}^T\mathbf{x} = (-b)\times(1)+(-a)\times x+1\times y, \text{ or}$$
$$\mathbf{w}^T\mathbf{x} = y - ax - b \quad (3)$$

The reason of using the hyper plane equation $\mathbf{w}^T\mathbf{x}$ instead of y=ax+b is because it is easier to work in more than two dimensions with this notation and the vector **w** will always be normal to the hyper plane. Once the hyper plan with maximum margin has been found, this hyper plane can be used to make predictions [11]. The hypothesis function h will be-

$$h(x_i) = \begin{cases} +1; \ if \ \ \mathbf{w.x}+b \geq 0 \\ -1; \ if \ \ \mathbf{w.x}+b < 0 \end{cases} \quad (4)$$

D. Artificial Neural Network (ANN)

Artificial Neural Network is another tool used in machine learning. As it name suggests, ANN is a system inspired by human brain system and replicate the learning system of human brain. It consists of input and output layers with one or more hidden layers in most cases as shown in Fig 3. The ANN uses a technique called back propagation to adjust the outcome with the expected result or class.

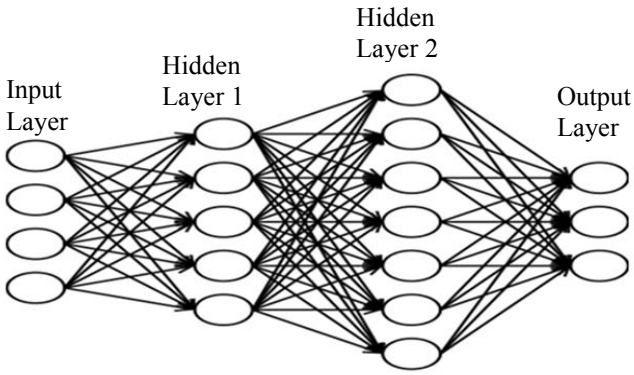

Fig 3: Artificial neural network showing the input, output and hidden layers

III. EXPERIMENTAL ANALYSIS OF THE SYSTEM

A. Feature Selection

The experiment carried out using Weka open source software suite popular for data mining and machine learning and consists of two parts. In the first part, we extracted most relevant features using different feature selection (FS) methods. In the wrapper method we used SVM classification algorithm with cross-validation to avoid over fitting and under fitting problem. In the filter method a ranker algorithm is used to find the best result suitable for our proposed classifier. The training data we used from NSL-KDD dataset contains 25,191 labeled instances. Results of the feature selection experiment are shown in Table I.

TABLE I
RESULT OF FEATURE SELECTION

FS Technique	FS Type	Input Features	Output Features
Correlation Based	Wrapper	41	17
Chi-Square Based	Filter	41	35

Correlation based feature selection found total 17 features most relevant from 41 features present in the training dataset whereas Chi-Square algorithm retained 35 features to be more relevant to the resultant class. These 17 and 35 retained features were used to train the model using training or seen dataset as well as to test the model using unseen or testing dataset.

B. Classification

With the features found in feature selection part, total four models are built in Weka software suite using the training dataset. Classification using supervised machine learning first requires training the model using training dataset. We used 20% of NSL-KDD dataset as training data that have 25,191 labeled data instances. To training the model we used SVM and ANN learning algorithm for each type of feature selection method. Hence we build four learning models, two model using SVM and another 2 using ANN. Among the 2 model built for each learning algorithm, one is built using 17 features and another one is built using 35 features found in the feature selection part. Next these four trained models were evaluated using 22,542 instances of testing data picked from the NSL-KDD testing dataset. The findings are summarized in Table II as below -

TABLE II
RESULT OF CLASSIFICATION

Learning Type	Number of Features	Detection Accuracy
SVM	17	81.78%
SVM	35	82.34%
ANN	17	94.02%
ANN	35	83.68%

In Table III, we listed our results with recently published results in the literature. While comparing the performance of the proposed model with the others works, we picked works having hypothesis of comparable aspects related to learning algorithm and benchmarking datasets. But there are other aspects like attribute reduction, number of instances, the number layers and learning rates used. The detection success rate of the proposed model is also compared with other existing models in Table III as below-

TABLE III
PERFORMANCE COMPARISON WITH EXISTING MODELS

Learning Type	Our Model Accuracy	Existing Model	Existing Model
SVM	82.34%	92.84% [16]	69.52% [17]
ANN	94.02%	81.2% [18]	77.23% [19]

IV. DISCUSSION ON SYSTEM IMPLEMENTATION

To implement and evaluation the system we have used widely used open source machine learning software suite called Weka. Along with machine learning algorithm implemented, Weka also has several algorithm and search technique implemented to perform feature selection. In the ANN model, we experimented with different number of hidden layer and found that the detection success rate varies with the number of hidden layer. After several trial and error methods, we found best detection rate with 3 hidden layers and 0.1 learning rate. In the wrapper feature selection method, we also used SVM algorithm as classifier. The model implemented in Weka has been run on a computing platform having 64 bit 2.6 GHz Intel core i5 CPU with 8 GB RAM on Windows 7 environment with limited network traffic instances. Implementing the solution on large scale network will require additional infrastructure with some higher capacity server platform.

V. CONCLUSION

In this paper, we have presented different machine learning models using different machine learning algorithms and different feature selection methods to find a best model. The analysis of the result shows that the model built using ANN and wrapper feature selection outperformed all other models in classifying network traffic correctly with detection rate of 94.02%. We believe that these findings will contribute to research further in the domain of building a detection system that can detect known attacks as well as novel attacks. The intrusion detection system exist today can only detect known attacks. Detecting new attacks or zero day attack still remains a research topic due to the high false positive rate of the existing systems.

REFERENCES

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," *American Journal of Criminal Justice*, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labeled data," in *Web Research (ICWR), 2017 3th International Conference on*, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in *International Conference on Networked Systems*, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," *International Journal of Scientific and Engineering Research*, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," *arXiv preprint arXiv:1312.2177*, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," *International Journal of Computing and Business Research (IJCBR) ISSN (Online)*, pp. 2229–6166, 2013.

[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *computers & security*, vol. 28, no. 1–2, pp. 18–28, 2009.

[9] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," *Procedia Computer Science*, vol. 89, pp. 117–123, 2016.

[10] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," *Neural Computing and Applications*, vol. 22, no. 5, pp. 1023–1035, 2013.

[11] F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning techniques for intrusion detection," in *Communication Networks and Services Research, 2007. CNSR'07. Fifth Annual Conference on*, 2007, pp. 350–358.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[13] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Military Communications and Information Systems Conference (MilCIS), 2015*, 2015, pp. 1–6.

[14] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and KDDCUP'99 datasets," in *Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on*, 2017, pp. 1881–1886.

[15] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 6, pp. 446–452, 2015.

[16] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, 2016, pp. 21–26.

[17] M. Panda, A. Abraham, and M. R. Patra, "Discriminative multinomial naive bayes for network intrusion detection," in *Information Assurance and Security (IAS), 2010 Sixth International Conference on*, 2010, pp. 5–10.

[18] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in *Signal Processing And Communication Engineering Systems (SPACES), 2015 International Conference on*, 2015, pp. 92–96.

[19] L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, "A comparison study for intrusion database (Kdd99, Nsl-Kdd) based on self organization map (SOM) artificial neural network," *Journal of Engineering Science and Technology*, vol. 8, no. 1, pp. 107–119, 2013.