International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

NETWORK ENCROACHMENT RECOGNITION USING ANN

[1]Prof. Shilpa M, [2]Prakrithi Singh, [3]Radhika Singhal, [4]Yashi Sikligar

[1]Assistant Professor, Dept of Computer Science, Sapthagiri College of Engineering, Bangalore, Karnataka, India
[2,3,4]Student, Dept of Computer Science, Sapthagiri College of Engineering, Bangalore, Karnataka, India

---

Abstract – *Attacks on computers and data networks have been a regular and sophisticated issue. Intrusion detection has shifted its attention from hosts and operating systems to networks and has become a way to provide a sense of security to these networks. The unseen network traffic is classified as either a benign network or as a harmful network ,The proposed model uses Machine learning techniques (MLT) to classify the network traffic datasets using different algorithms and feature selection technique and thus helps in identifying an intruder.*

ANN learning algorithms was used to find the best classifier with higher accuracy and success rate.

***Key Words*: Intrusion detection system, Machine Learning Technique, ANN, Feature Selection**

1. INTRODUCTION

With the wide spreading usages of internet and increases in access to online contents, cybercrime is also happening at an increasing rate. Intrusion detection is the first step to prevent security attack. Hence the security solutions such as Firewall, Intrusion Detection System (IDS), Unified Threat Modelling (UTM) and Intrusion Prevention System (IPS) are getting much attention in studies. An ID detects attacks from a variety of systems and network sources by collecting information and then analyzes the information for possible security breaches.

An IDS detects attacks from a variety of systems and network sources by collecting information and then analyzes the information for possible security breaches. The network based IDS analyzes the data packets that travel over a network and this analysis are carried out in two ways.

- Anomaly based detection.
- Signature based detection.

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. 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. IDS however is not an answer to all security related problems.

2. MODULE SPECIFICATION

2.1 Data Collection and Data Normalization

Data collection is the process of gathering and measuring information from countless different sources. We are collecting the NSL-KDD training and test network intrusion data from UCI-Machine learning repository. The NSL-KDD data set has 41 attribute and one class attribute. The dataset is subjected to normalization. Data normalization is the process of cleaning data here we remove the repeated data and the empty rows. Data normalization is used to transform features to be on a similar scale. This improves the performance and training stability of the model. Data normalization gets rid of a number of anomalies that can make analysis of the data more complicated.

2.2 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. 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. The training data we used from NSL-KDD dataset contains 25,191 labelled instances.

In Feature Selection the values which system have got are compared with trained dataset and only some features are selected based on the algorithm. 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.

2.3 Classification of Intruders:

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 labelled data instances. To training the model we used Decision tree, Random Forest, Logistic Regression, ANN which is fabricated using the feature selected by the algorithm. Based on the training models different results will be generated.

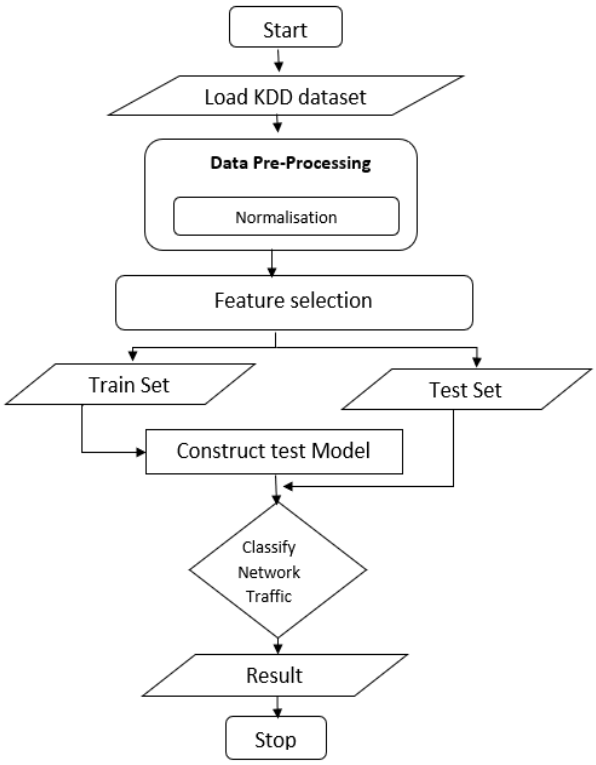


Fig 1: Flowchart of the System

3. METHODOLOGY

3.1 Architecture

The system proposed is composed of preprocessing which involves data cleaning, data normalization, feature selection and learning algorithm. 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 to monitor network traffic and compare it against an established baseline to determine what is considered normal and what is considered malicious for the network with respect to bandwidth, protocols, ports and other devices using machine learning algorithms. Then the learned intelligences are applied to the testing dataset to measure the accuracy of how much the model correctly classifies on unseen data. The classification report shows the main classification metrics precision, recall and f1-score on a per-class basis.

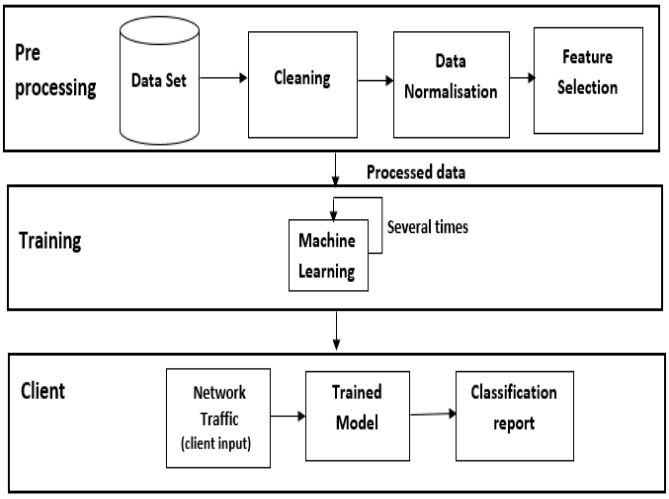


Fig 2: System architecture block diagram

3.2 Algorithms

ANN Classifier : A neural network consists of a collection of processing units called neurons that are highly interconnected according to a given topology. ANN has the ability to learning by example and generalize from limited, noisy, and incomplete data. They have been successfully employed in a broad spectrum of data-intensive applications. ANN is a system inspired by human brain system and replicates the learning system of human brain. It consists of input and output layers with one or more hidden layers in most cases. The ANN uses a technique called back propagation to adjust the outcome with the expected result or class. A neural network may contain the following 3 layers:

- **Input layer –** The activity of the input units represents the raw information that can feed into the network. Each of the nodes in input layer represents an individual feature from a given sample of our KDD dataset. When we pass a record to the model, each of the values contained in the record will be provided to a corresponding node in the input layer.
- **Hidden layer –** The input from Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layers can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by addition of learnable biases followed by activation function which makes the network nonlinear.
- **Output layer –** The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units The output from the hidden layer then fed into a logistic function like sigmoid. Sigmoid activation function is used whenever we need Probabilities of 2 categories or less, here it is either normal (+Ve) or (-Ve) attack

The implementation was carried out using Python programming language with Scikit-learn, Pandas, Numpy, Matplotlib and Keras libraries using ANN algorithm.

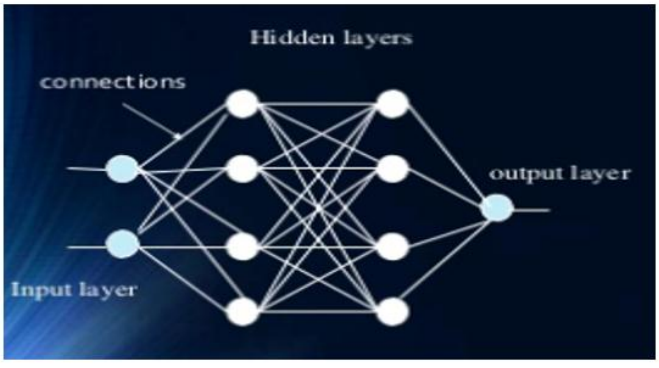


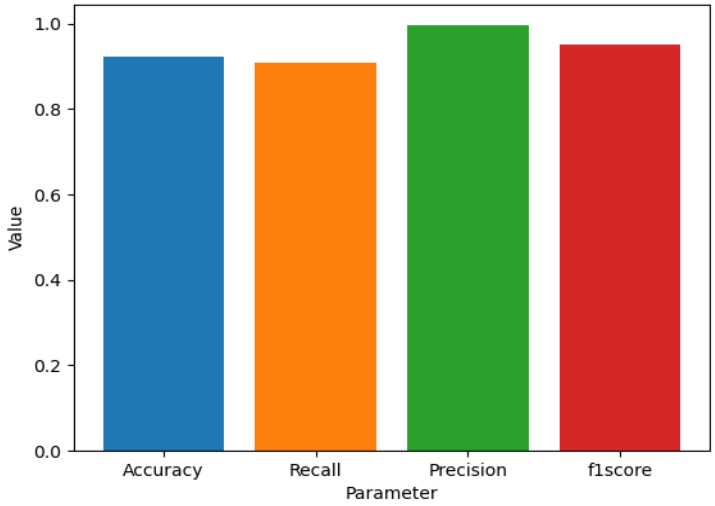
Fig III: Artificial Neural Network

Artificial Neural Network is capable of learning any nonlinear function. Hence, these networks are popularly known as Universal Function Approximators. ANNs have the capacity to learn weights that map any input to the output.
One of the main reasons behind universal approximation is the activation function. Activation functions introduce nonlinear properties to the network. This helps the network learn any complex relationship between input and output.

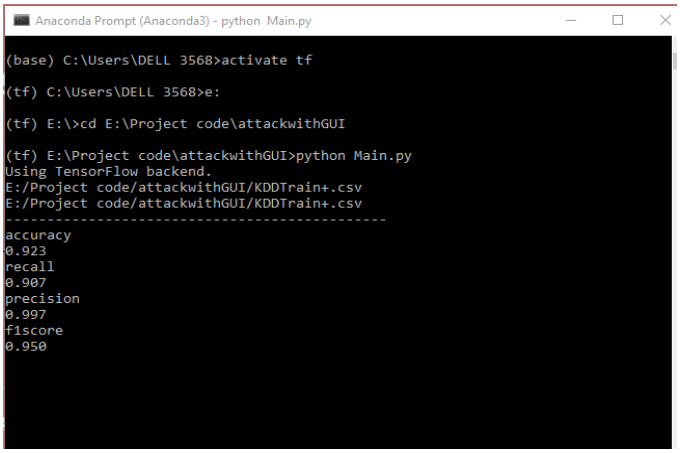
4. EXPERIMENTAL RESULTS

4.1 ANN Metric Values

A bar a graph is displayed on the window showing the parameters like accuracy rate, precision, recall and f1-score.



4.2 Analytical Information of ANN Algorithm



Command prompt displays accuracy rate, precision, recall and f1-score. The accuracy rate calculated by ANN is 92%. It is the highest accuracy rate compared to other algorithms.

5. RESULT

As discussed above the algorithm used is ANN with wrapper feature selection for evaluation of network intrusion detection system. The NSL-KDD dataset is divided into training and testing data. The result presented by using ANN algorithm in the model shows the accuracy rate of 92%. The comparative study shows that it is the highest accuracy rate calculated as compared to other algorithms.

6. CONCLUSION

In this paper, machine learning algorithm along with data preprocessing and feature selection is used to classify network traffic as either malicious or benign. There are many network intrusion detection systems which classify network traffic but fail to return appropriate and accurate results. It is a challenging task to develop a system which would classify new attacks with high accuracy and precision. The proposed model is built using different machine learning algorithms and feature selection methods to find a best model. The results show that the model built using ANN and wrapper feature selection transcends all other models in classifying network traffic correctly with a detection rate of 90-95%

REFERENCES

[1] 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.
[2] 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.
[3] 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
[4] 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.
[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] 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.