IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Intrusion Detection for Computer Networks Using Deep Learning Approach

[1] TIKKADA GANESH KUMAR, [2] Dr. K RAJA KUMAR

[1]M. tech, Department of Computer Science and System Engineering(A), Andhra University, Visakhapatnam, AP, India,

[2]Assistant Professor, Department of Computer Science and System Engineering (A), Andhra University, Visakhapatnam, AP, India.

ABSTRACT: Within the previous decades, there has been increase in the technology which led to the usage of many devices thereby causing threat to personal data, corporate data, device and network itself. This gave rise to the advancements in **Intrusion Detection Systems (IDS)** which act as defense mechanism. The Intrusion Detection Systems helps to detect unintended access to network and device with the use of several strategies to identify threats in the network. The existing systems require human interaction to analyze and to identify the threats. However it's a major drawback as it is difficult to human to notice the intrusions from various sources of network traffic. Intrusion Detection System can automatically check for network intrusions without the use of human to interact with the systems. The main idea behind as IDS is to detect the threats based on analyzing and predicting the behaviors of users, these actions are going to be used by our system to train our IDS to enhance the detection of upcoming threats well in advance or at the time of attack happening. No existing system is not very much sure about the attacks and is prone to attacks and may fail to accurately detect the attack. The advancements in technology also led to the changing behavior of attacks and these attacks are to be monitored continuously to fine tune our systems for best results. A **Deep Neural Network (DNN),** which works in multilayered architecture, is used for the development of IDS for efficiency and accuracy to notify unknown and unforeseen attacks performed on the network. The continual updating and analysis of network behavior is required for the system to classify attacks within a short span of time. For this experiment we worked on preexisting dataset like **NSL-KDD, KDDCUP1999** for training our system and to test the working of our system. The Dataset has 41 different types of attack these attacks are classified and grouped into 4 categories like **Denial of Service (DOS), Unauthorized access to root Privileges (U2R), Unauthorized access from Remote System (R2L), Probing**. Based on the test results the proposed system model works better than previously existing systems the training time is lesser than that of existing once and loses are reduced with 96% accuracy test results, which is 3% more than the existing study.

KEYWORDS*:* Intrusion Detection System, Machine Learning Deep Neural Networks, KDDCUP dataset.

I. INTRODUCTION

The recent advancements in Technology lead to the increase in the IOT based attacks, Malware attacks, Internet Attacks, Phishing attacks, Dos attacks and many more types of attacks.

Denning's proposal for associate degree intrusion detection system targeted on a way to develop effective and correct ways for intrusion detection. The intrusion Detection Systems need to be continuously trained with the latest data set and monitor the behavior of its network user for its performance to be enhanced. The current day IDS are being trained with known Machine Learning algorithms for developing a most intelligent IDS system. Grouping, bunch and rule based generally procedures are normally utilized AI strategies. The IDS need to have high detection accuracy with less training time, and must be capable of fine tuning its parameters for improved accuracy and better performance. The Intrusions can be of 2 kinds Internal and External intrusions, the intrusions that are performed within the organization and From outside the organization respectively.

INTRUSION DETECTION SYSTEM

An IDS is an instrument utilized for automatic detection and expulsion of outer threats or admittance to the framework and takes a decision to see whether these are approved utilization of the framework. Figure one represents the organization of IDS in this fig solid arrows represent data/control flow and the dotted line represents the response.

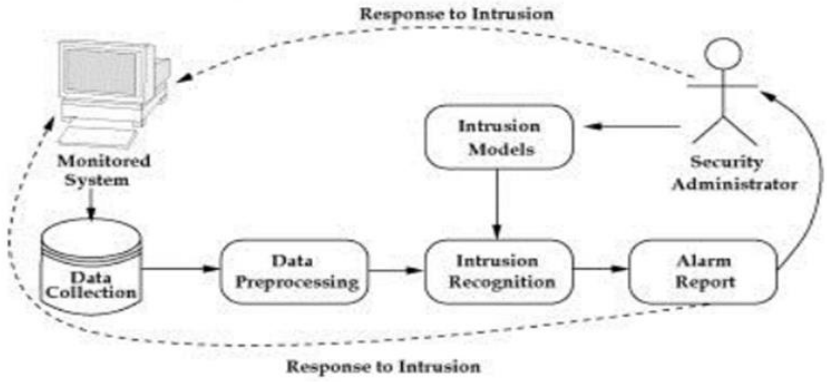


Figure 1. A general organization of a typical intrusion detection system

In general, this characterizes IDSs on the reason of identification methodologies they use into 2 classes, similar to (I) abuse recognition and (II) anomaly discovery. By coordinating found data, abuse identification recognizes interruptions with pre-characterized depictions of meddling conduct. Thus, outstanding intrusions may be detected in associate degree, economical manner utilizing a false positive rate. Therefore, this system is wide adopted within the majority of business systems. However, the categories of recent intrusions have evolved each moment and unceasingly, therefore. Abuse the past procedures for interruption finding can neglect to distinguish new obscure interruptions.

II. MOTIVATION AND BACKGROUND

The age old traditions of detecting malicious activities pose a serious threat and could damage the entire network and operations of a company or individual systems the old systems are to be maintained manually and this could be a burden some work for an individual and are prone to human errors which could result in integrity constraint of the company policies of its client data to be public without the knowledge of administrators. The recent attacks had put several organizations at risk for unauthorized access of data and also led to shutting down of many servers with DOS attack. So, for a system to be much more robust to network attacks it must be continuously monitored and fed with the latest data statistics so it could be able to detect if such similar behavioral activities are repeated once again in the same or different network with the same or different system. Machine Learning provides us with capability to train the system and network with data from time to time for efficient and accurate detection of attacks.

III. RELATED WORK

In this fragment analyzed about the researches finished in the district of interference acknowledgment system. The two procedures used to recognize interferences are Expert based and Statistical based philosophies. The expert based interference acknowledgment system recognizes the outstanding attacks. The disadvantage is it won't recognize the as of late attacking intrusion. The quantifiable based philosophy used to perceive the new interferences. There are a couple datasets open, the KDD cup 99 and NSL KDD cup 99 datasets applied for the going with investigation.

Kumar and Koyal implemented an IDS that masterminded the smurf attacked names using inherited estimations and achieved 0.2% low certain extent. Abdullah explained some gathering rules for perceiving interference by inherited computation. Ojugo et at. used the wellbeing take a shot at the genetic count for evaluating the standards. The AI strategies are made to recognize interferences.

Li et al utilized Genetic Algorithm for advancing standards for abuse recognition for DARPA interruption identification framework. The Chromosome contains the Source and target IP address, source and target port number, span of association, no of bytes, the convention utilized and the association state. Abraham et al. proposed a strategy dependent on Genetic Programming to arrange the assaults. The three procedures are utilized I) Linear Genetic Programming (LGP), ii) Multi-Expression Programming (MEP), iii) Gene Expression Program

The Roshani applied Artificial Neural Network (ANN) calculation for distinguishing interruptions. The fluffy grouping and ANN methods are consolidated and it beats the powerless solidness location depicted by Gaikwad et al. Fluffy grouping will create a few subclasses for preparing for diminishing the measure of subset size and trouble. Each subsection was talented by the distinctive sort of ANN strategies and get significant outcomes. Denning was built up an Intrusion Detection System utilizing the idea of Time arrangement, Markov chains, and insights. He estimated the inconsistency by the irregularity of the standard exhibition.

IV. DEEP NEURAL NETWORK (DNN)

Profound learning is a better-quality AI strategy for highlight reflection, discernment and learning of innovations. Profound learning calculations play out their tasks utilizing numerous back to back layers. The layers are interlaced and each layer acknowledges the yield of the previous layer as info. It is an incredible preferred position to utilize proficient calculations for separating progressive highlights that best speak to information instead of manual highlights in profound learning techniques. Multilayer Architecture was initially distributed in 1965

V. DATASET USED

The most moving stage to decide the exhibition of Intrusion Detection Systems is to locate the proper dataset. The data to be utilized for the information can be acquired by noticing the organization. Gathering data from the organization is expensive, along these lines engineers need to control their organization or frameworks utilizing accessible datasets. In this part, the most ordinarily utilized data sets for assault acknowledgment frameworks are announced. **KDD CUP99 and NSL-KDD** datasets are used, the NSL-KDD dataset is the decreased version of original KDDCUP99 dataset. In NSL-KDD all the duplicate records are removed. We use this dataset for assessing the results there are 41 types of attacks which are grouped into 4 categories as DOS, Probe, R2L and U2R

VI. EVALUATION METRICS

Assessment measurements are utilized figuring and noticing the exhibition of the IDS and for looking at the outcomes acquired from the dataset.

1) Accuracy: It is obtained by dividing total number of correct predictions by total number of predictions

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2) Precision: It is obtained by dividing correct predictions by total predicted positives

$$Precision = \frac{TP}{TP + FP}$$

3) F1-Score: F1-Score is additionally called as F1-Measure. On the off chance that theF1-Score is higher

$$F1 - Score = 2 \times \left(\frac{\Pr ecision \times Recall}{\Pr ecision + Recall} \right)$$

4) True Positive Rate (TPR): It is Recall. It appraises the proportion of the accurately ordered Attack association records to the complete number of Attack association records.

$$TPR = \frac{TP}{TP + FN}$$

5) False Positive Rate (FPR): It assesses the proportion of the Normal association records hailed as Attacks to the absolute number of Normal association records.

$$FPR = \frac{FP}{FP + TN}$$

VII. IMPLEMENTATION

The procedure of the experiment and the evaluation results obtained are discussed in this section.

1 Data Preparation

KDD dataset is well known for benchmarking intrusion detection techniques. The dataset is a massive collection of data which is collected over months. The KDD dataset consists of 41 features and is labeled either normal or an attack grouped into four categories, DOS, U2R, R2L, and Probing.

2 Selection of evaluation metrics and Machine Learning Algorithm

For Intrusion Detection Algorithm it is important to have knowledge on Recall more important than that of precision, so we require F-score. The Algorithms were implemented in Python.

3 Calculation of Training time

The time taken to train the model will increase sequentially, which is known as training time, this training time is also observed.

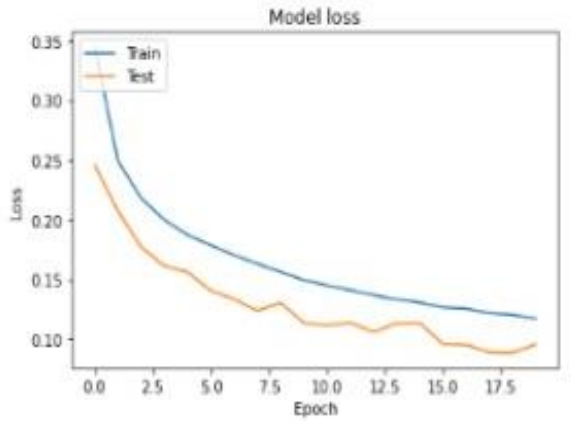
4 Increase in percentage of evaluation metrics

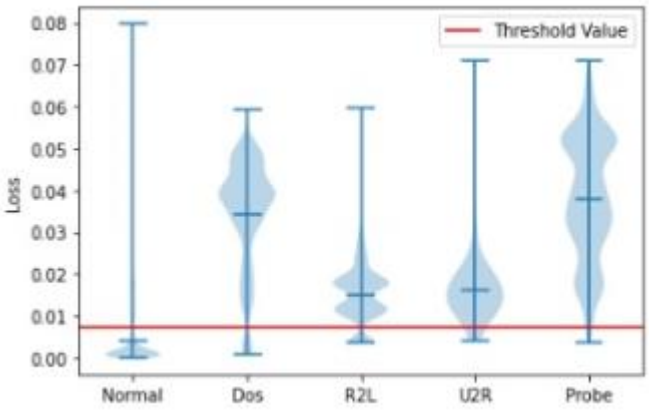
The values of evaluation metrics are increased accordingly, and these percentage increase in Accuracy, Precision, Recall, F-score is observed for all data samples is observed

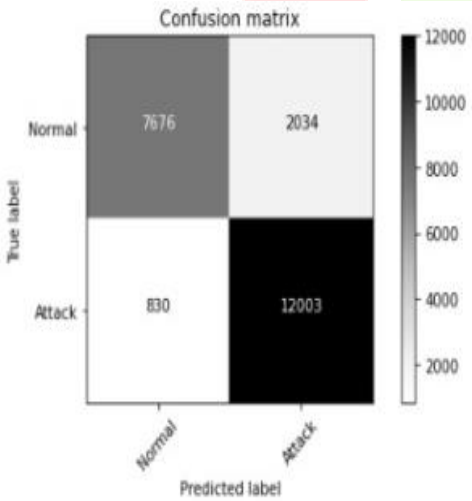
VIII. RESULTS

```
The training set has 5 possible outcomes

Normal, Dos, R2L, Probe, U2R.

The testing set has 5 possible outcomes

Dos, Normal, Probe, R2L, U2R.
```

```
Normal Detection Rate : 0.2094747682801236
Dos Detection Rate : 0.9405447878470403
R2L Detection Rate : 0.875968992248062
U2R Detection Rate : 0.9253731343283582
Probe Detection Rate : 0.9855431639818257
```








```
Epoch 10/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1495 - accuracy: 0.9499 - val_loss: 0.1138 - val_accuracy: 0.9617
Epoch 11/20
3543/3543 [==============================] - 12s 4ms/step - loss: 0.1451 - accuracy: 0.9510 - val_loss: 0.1113 - val_accuracy: 0.9632
Epoch 12/20
3543/3543 [==============================] - 12s 3ms/step - loss: 0.1410 - accuracy: 0.9523 - val_loss: 0.1139 - val_accuracy: 0.9579
Epoch 13/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1369 - accuracy: 0.9536 - val_loss: 0.1059 - val_accuracy: 0.9600
Epoch 14/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1332 - accuracy: 0.9548 - val_loss: 0.1132 - val_accuracy: 0.9553
Epoch 15/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1306 - accuracy: 0.9549 - val_loss: 0.1137 - val_accuracy: 0.9613
Epoch 16/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1265 - accuracy: 0.9569 - val_loss: 0.0957 - val_accuracy: 0.9674
Epoch 17/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1255 - accuracy: 0.9575 - val_loss: 0.0952 - val_accuracy: 0.9704
Epoch 18/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1218 - accuracy: 0.9585 - val_loss: 0.0891 - val_accuracy: 0.9679
Epoch 19/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1203 - accuracy: 0.9594 - val_loss: 0.0887 - val_accuracy: 0.9703
Epoch 20/20
3543/3543 [==============================] - 13s 4ms/step - loss: 0.1173 - accuracy: 0.9604 - val_loss: 0.0956 - val_accuracy: 0.9667
705/705 [==============================] - 2s 2ms/step - loss: 1.6850 - accuracy: 0.7364
```

IX. CONCLUSION

The authors have analysed the class specific detection with the KDD dataset, using the supervised machine learning algorithm Random Forest for IDS and the test data and the training data is constructed for evaluating the performance to detect different types of attacks. The training time of the model is observed with the respective increase in the size of the dataset. The increase in the values of evaluation metrics (Accuracy, Precision, Recall, F-score) by increasing the size of the dataset in steps is observed. From the experiment conducted the authors obtained the results with an

accuracy of 96%. The criteria included accuracy, complexity, time for training a model, time for classifying a unknown data and understating final solution. It is difficult to identify a better approach of DNN method based on only one factor like accuracy. If the DNN methods compared based on accuracy, these methods should be trained on same accurate training data and tested on same accurate testing data. In this study several authors used same dataset for same methods but they have used subset of the same dataset (selected attributes), and they are not necessarily same.

Future Scope

Soft computing techniques are getting considerable attention from researchers in IDS. This is because this technique is easy to apply and often produce better result compared to single algorithm. Proper combination of multiple algorithms is the way forward. Most researchers are focusing on the classification of IDS, which is beneficial in determining known intrusion attacks. However, it may pose a problem in detecting anomalous intrusion, which may include new or modified intrusion attacks. Therefore to produce a more robust IDS, clustering algorithm should be considered for future development. KDDCup99 and its variant NSL-KDD datasets are the two most widely used datasets, although they are almost 20 years old. This continuous trend could result in static progress in IDS, while intrusion attacks continue to evolve together with new technologies and user behaviors. Ultimately, this situation will result in the obsolete use of IDS as part of a cyber-security tool.

REFERENCES & BIBLOGRAPHY

[1] G. C. Kessler, “Defenses against distributed denial of service attacks,” SANS Institute, vol. 2002, 2000.View publication stats

[2] H. A. Nguyen and D. Choi, “Application of data mining to network intrusion detection: classifier selection model,” in Asia-Pacific Network
Operations and Management Symposium. Springer, 2008, pp. 399–408.

[3] S. Paliwal and R. Gupta, “Denial-of-service, probing & remote to user (r2l) attack detection using genetic algorithm,” International Journal of Computer Applications, vol. 60, no. 19, pp. 57–62, 2012.

[4] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the kdd cup 99 data set,” in Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on. IEEE, 2009, pp. 1–6.

[5] P. Amudha, S. Karthik, and S. Sivakumari, “Classification techniques for intrusion detection-an overview,” International Journal of Computer Applications, vol. 76, no. 16, 2013.

[6] W. Alsharafat, “Applying artificial neural network and extended classifier system for network intrusion detection.” International Arab Journal of Information Technology (IAJIT), vol. 10, no. 3, 2013.

[7] S. D. Bay, “The uci kdd archive [http://kdd. ics. uci. edu]. irvine, ca: University of california,” Department of Information and Computer Science, vol. 404, p. 405, 1999.

[8] M. Al-Kasassbeh, “Network intrusion detection with wiener filter-based agent,” World Appl. Sci. J, vol. 13, no. 11, pp. 2372–2384, 2011.

[9] S. K. Pal and S. Mitra, “Multilayer perceptron, fuzzy sets, and classification,” IEEE Transactions on neural networks, vol. 3, no. 5, pp. 683–697, 1992.

[10] L. Breiman, “Random forests,” Machine learning, vol. 45, no. 1, pp. 5–32, 2001.

[11] J. R. Quinlan, C4. 5: programs for machine learning. Elsevier, 2014.

[12] M. S. Bhullar and A. Kaur, “Use of data mining in education sector,” in Proceedings of the World Congress on Engineering and Computer Science, vol. 1, 2012, pp. 24–26.

[13] P. Aditi and G. Hitesh, “A new approach of intrusion detection system using clustering, classification and decision table,” 2013.