International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 10 | Oct 2018 www.irjet.net p-ISSN: 2395-0072

Review on Intrusion Detection System using Recurrent Neural Network with Deep Learning

Miss. Autade Prajyot S[1], Prof. Kalavadekar P. N[2]

[1]*Student of Computer Engineering, SRES' College of Engg. Kopargaon, India*
[2]*Professor, Dept. of Computer Engineering, SRES' College of Engg. Kopargaon, India*

Abstract - *With the advancement of information and communication techniques, sharing information through online has been increased which leads to various security threats that we face are becoming more and more serious. Intrusion detection plays an important role in security field, which can accurately identify invasion in the network. As deep learning has the potential to extract better representations from the data to create much better models, this paper presents a Deep learning technique for Intrusion Detection using recurrent neural network. The critical part of building Intrusion Detection System is to select input network data. We used the NSL-KDD Dataset to train the IDS Model. The performance of the model in binary classification and multiclass classification is superior to that of tradition machine learning classification methods. The IDS model improves the accuracy of intrusion detection.*

Key Words: *Intrusion detection, Deep learning, Recurrent Neural Network.*

1. INTRODUCTION

The deep integration of the Internet and society is increasing day by day, the Internet is changing the way in which people live, study and work, but the various security threats that we face are becoming more and more serious. How to identify various network attacks, especially unforeseen attacks, is an unavoidable key technical issue. An Intrusion Detection System (IDS), a significant research achievement in the information security field, can identify an invasion, which could be an ongoing invasion or an intrusion that has already occurred.

According to an object of observation [2] there are two types IDS: 1) Host-based IDS (HIDS) and 2) Network-based IDS (NIDS). The first one, HIDS, watches the host system operation or states and detects system events such as unauthorized installation or access. It also checks the state of ram or file system whether there is an expected data or not, but it cannot analyze behaviors related to the network. The second one, NIDS is placed on choke point of the network edge which observes a real-time network traffic and analyzes it for detecting unauthorized intrusions or the malicious attacks. The detection can be a behavior-based intrusion detection called anomaly detection or knowledge based intrusion detection called misuse detection. Behavior-based intrusion detection catches attacks by comparing an abnormal behavior to a normal behavior. Knowledge based intrusion detection detects the attacks based on the known knowledge.

Existing machine learning methodologies have been widely used in identifying various types of attacks, and a machine learning approach helps the network administrator take the preventive measures for intrusions. However, most of the traditional machine learning methodologies belong to shallow learning and often emphasize feature engineering and selection; they cannot effectively solve the massive intrusion data classification problem that arises in the face of a real network application environment. With the dynamic growth of data sets, multiple classification tasks will lead to decreased accuracy. In addition, shallow learning is unsuited to intelligent analysis and the forecasting requirements of high-dimensional learning with massive data. In contrast, deep learners have the potential to extract better representations from the data to create much better models. As a result, intrusion detection technology has experienced rapid development after falling into a relatively slow period.

After Professor Hinton [3] proposed the theory of deep learning in 2006, there is remarkable achievements, especially in the fields of speech recognition, image recognition and action recognition. Deep learning theory and technology has had a very rapid development in recent years, which means that a new era of artificial intelligence has opened which offered a completely new way to develop intelligent intrusion detection technology. Due to growing computational resources, recurrent neural networks (RNNs) have recently generated a significant development in the domain of deep learning. In recent years, RNNs have played an important role in the fields of computer vision, natural language processing (NLP), semantic understanding, speech recognition, language modeling, translation, picture description, and human action recognition, among others.

In this study, we proposed a deep learning approach for an intrusion detection system using recurrent neural networks. In fact, intrusion detection is usually equivalent to a classification problem, which can be binary or a multi-class classification problem, i.e., identifying whether network traffic behaviour is normal or anomalous, or a five-category classification problem, i.e., identifying whether it is normal or any one of the other four attack types: Denial of Service (DOS), User to Root (U2R), Probe (Probing) and Root to Local (R2L). In short, the main motivation of intrusion detection is to improve the accuracy of classifiers in effectively identifying the intrusive behaviour.

2. RELATED WORK

Due to the increasing the importance of cyber security, researches about Intrusion Detection System (IDS) have been actively studying. Nathan Shone [1] proposed a network intrusion detection system using non-symmetric deep autoencoder (NDAE) for unsupervised feature learning. He constructed the proposed model using stacked NDAEs. The model is a combination of deep and shallow learning, capable of correctly analysing a wide-range of network traffic. He combined the power of stacking proposed Non-symmetric Deep Auto-Encoder (NDAE) (deep learning) and the accuracy and speed of Random Forest (RF) (shallow learning). The classifier has been implemented in graphics processing unit (GPU)-enabled TensorFlow and evaluated using the benchmark KDD Cup '99 and NSL-KDD datasets.

Ozgur Depren [2] proposed a novel Intrusion Detection System (IDS) architecture utilizing both anomaly and misuse detection approaches. This hybrid Intrusion Detection System architecture consists of an anomaly detection module, a misuse detection module and a decision support system combining the results of these two detection modules. For proposed anomaly detection module he used a Self-Organizing Map (SOM) structure to model normal behavior. Deviation from the normal behavior is classified as an attack. He used J.48 decision tree algorithm to classify various types of attacks. The principle interest of his work was to benchmark the performance of the proposed hybrid IDS architecture by using KDD Cup 99 Data Set.

Professor Hinton [3] proposed the theory of deep learning in 2006, deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.

Jihyun Kim [4] applied Long Short Term Memory (LSTM) architecture to a Recurrent Neural Network (RNN) with deep learning approach. He used KDD Cup 1999 dataset to train the IDS model and measured the performance. Through the experiments, he found an optimal hyper-parameter for LSTM-RNN and confirmed the detection rate and false alarm rate.

Bhupendra Ingre [5] analyzed the performance of NSL-KDD dataset by evaluating it using Artificial Neural Network and obtained result for both binary class as well as five class classification (type of attack). Results were analyzed based on various performance measures and better accuracy was found. The detection rate obtained was 81.2% and 79% for intrusion detection and attack type classification task respectively for NSLKDD dataset. The performance of the proposed scheme has been compared with existing scheme and higher detection rate has been achieved in both binary class as well as five class classification problems.

Mahbod Tavallaee [6] analyzed KDD CUP 99. During the last decade, anomaly detection has attracted the attention of many researchers to overcome the weakness of signature based IDSs in detecting novel attacks, and KDDCUP99 was the mostly widely used data set for the evaluation of those systems. After conducting a statistical analysis on this data set, he found two important issues which highly affects the performance of evaluated systems, and he proposed a new data set, NSL-KDD, which consists of selected records of the complete KDD data set and does not suffer from any of mentioned shortcomings.

R. Ravinder Reddy [7] proposed theory on intrusion detection using SVM. SVM is remodeled to increase the detection rate. The underlying principle of SVM is structural risk minimization. It attracts many fields because it builds on strong mathematical proofs. The support vector machine based classification algorithm is used to classify the intrusions accurately by using the discriminant function. The effective discriminant function will be accurately identifies the data into intrusion and anomaly. The evaluation of the discriminant is important in the evaluation of the intrusion detection system. Performance of intrusion detection system depends on the choice of the discriminant function.

N. Farnaaz [8] built a model for intrusion detection system using random forest classifier. Random Forest (RF) is an ensemble classifier and performs well compared to other traditional classifiers for effective classification of attacks. Intrusion Detection System (IDS) attempts to identify and notify the activities of users as normal (or) anomaly. IDS is a nonlinear and complicated problem and deals with network traffic data. Many IDS methods have been proposed and produce different levels of accuracy. To evaluate the performance of model, system conducted experiments on NSL-KDD data set. Empirical result show that proposed model is efficient with low false alarm rate and high detection rate.

Anna Buczak [9] survey paper described a focused literature survey of machine learning (ML) and data mining (DM) methods for cyber analytics in support of intrusion detection. Short tutorial descriptions of each ML/DM method are provided. Based on the number of citations or the relevance of an emerging method, papers representing each method were identified, read, and summarized. Because data are so important in ML/DM approaches, some well-known cyber data sets used in ML/DM are described. The complexity of ML/DM algorithms is addressed, discussion of challenges for using ML/DM for cyber security is presented, and some recommendations on when to use a given method are provided.

3. PROPOSED WORK

Recurrent neural networks include input units, output units and hidden units as shown in figure 1. Hidden unit completes the most important work, and hidden units

are the storage of the whole network, which remember the end-to-end information.

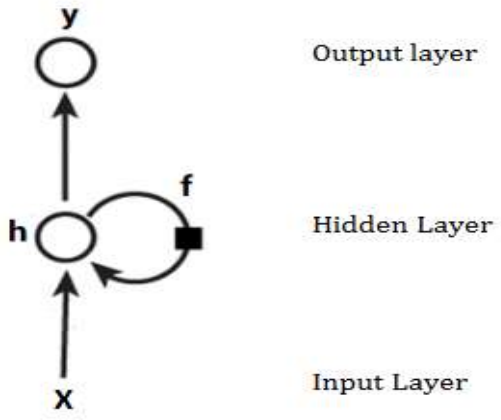

Fig -1: Recurrent Neural Network

Figure 2 shows the architecture of the system. . NSL-KDD is the dataset used to train recurrent neural network for Intrusion detection. Preprocessing is applied to the given input dataset which includes numericalization and normalization. Recurrent neural network is trained for both binary and multiclass classification. Accuracy is the evaluation method to check the performance of the model.

3.1 Dataset Description

NSL-KDD is a benchmark dataset, which covers training as well as testing sets. KDDTrain+ is used for training and KDDTest+ and KDDTest-21 is used for testing. NSL-KDD dataset can be used for binary as well as multiclass classification as has normal records and records for four different types of attacks. There are 41 features and 1 class label for each record. In order to provide more realistic theoretical basis for intrusion detection, some specific attack types that disappear in the training set are added to testing set.

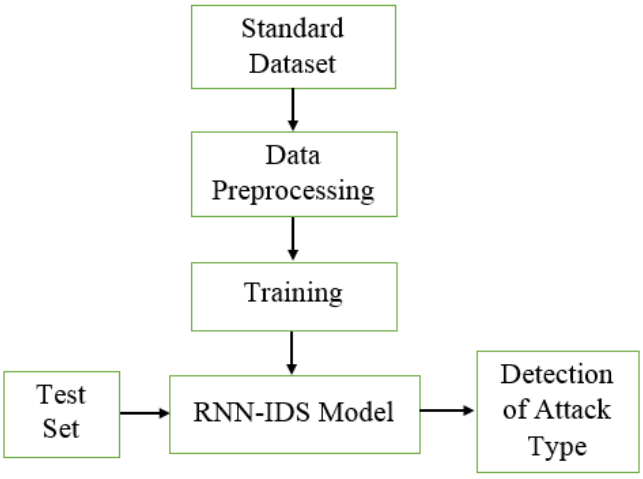

Fig -2: System Architecture

3.2 Preprocessing

Preprocessing includes numericalization and normalization. In numericalization, nonnumeric features are converted into numeric features as recurrent neural network model requires numeric matrix as an input. In normalization, logarithmic scaling is applied to features the difference between the maximum and minimum values has a very large scope and every feature to is mapped to [0, 1] range using MinMax Scaling.

3.3 Training

Training includes Forward Propagation and Weights Update (Back propagation). In neural networks, forward propagation is to predict the output values and compare it with the real/actual value to get the error or loss. The loss is calculated by using actual value and predicted values. The calculation flow is going in the natural forward direction from the input through the neural network to the output hence it is called forward-propagation. In forward propagation, there is activation function at each layer. Back propagation means backward propagation of errors. In back propagation, compute the partial derivatives of loss with respect to weight matrices and bias vectors, then propagate backward to update the weights.

3.4 Testing

This is the final step where model performance is evaluated. In NSL-KDD dataset, KDDTest + and KDDTest −21 sets are used for predicting the performance of model. Confusion matrix can be used to check the accuracy of model. Accuracy is most important performance indicator used to measure performance of RNN model.

4. CONCLUSION

This paper proposes approach for an intrusion detection system using recurrent neural networks with deep learning. NSL-KDD dataset is used to train the model for binary and multiclass classification. To evaluate, standard parameter such as accuracy, detection rate can be used. In near future we plan to apply feature selection to improve model performance.

ACKNOWLEDGEMENT

I would like to thank my esteemed guide Prof. P. N. Kalavadekar whose interest and guidance helped me to complete the work successfully. Also the valuable help of Dr. D. B. Kshirsagar (HOD Comp. Dept.) who provided facilities to explore the subject with more enthusiasm.

REFERENCES

[1] Nathan Shone, Tran Nguyen Ngoc, Vu Dinh Phai, and Qi Shi, "A Deep Learning Approach to Network Intrusion Detection", vol. 2, pp. 41-50 no. 1, feb 2018.

[2] Depren, Ozgur, et al., "An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks, Expert systems with Applications" 29.4, pp. 713-722, 2005

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", Nature, vol. 521, no. 7553, pp. 436-444, May 2015.

[4] Jihyun Kim, Jaehyun Kim, Huong Le Thi Thu, and Howon Kim, "Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection", Platform Technology and Service (PlatCon), 2016 International Conference on. IEEE, 2016.

[5] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN", in Proc. Int. Conf. Signal Process. Commun. Eng. Syst., Jan. 2015, pp. 92-96.

[6] M. Tavallaee, E. Bagheri,W. Lu, and A. A. A. Ghorbani, "A detailed analysis of theKDDCUP 99 data set", in Proc. IEEE Symp. Comput. Intell. Secur. Defense Appl., Jul. 2009, pp. 1-6.

[7] R. R. Reddy, Y. Ramadevi, and K. V. N. Sunitha, "Effective discriminant function for intrusion detection using SVM", in Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI), Sep. 2016, pp. 1148-1153.

[8] N. Farnaaz and M. A. Jabbar, "Random forest modeling for network intrusion detection system", Procedia Comput. Sci., vol. 89, pp. 213-217, Jan. 2016.

[9] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection", IEEE Commun. Surveys Tuts, vol. 18, no. 2, pp. 11531176, 2nd Quart, 2016.