International Journal of Research
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 05
May 2020

Network Intrusion Detection Using Deep Learning

Mrs J. Jeeshitha, Asst. Professor
Department of Computer Science and Engineering
Anurag Group of Institutions, Hyderabad, Telangana, India
jeeshithacse@cvsr.ac.in

Vasavi Sai Nunna, Alekhya Muppirishetty, Aniruth Gullapelly and Sandesh Pattem
Anurag Group of Institutions, Hyderabad, Telangana, India
16h61a0598@cvsr.ac.in, 16h61a0595@cvsr.ac.in, 16h61a0583@cvsr.ac.in, 16h61a05a2@cvsr.ac.in,

Abstract – Network intrusion detection is one among the foremost important parts for cyber security to guard computer systems against malicious attacks. With the emergence of various sophisticated and new attacks, however, network intrusion detection techniques face several significant challenges. We propose a completely unique network intrusion model by stacking autoencoders and evaluate our method on intrusion detection datasets. The auto-encoder is one among the foremost interesting models to extract features from the high-dimensional data within the context of deep learning. Our proposed model provides an accuracy which is more efficient than machine learning techniques like Random forest and Naive Bayes.

I. INTRODUCTION

Intrusion Detection System (IDS) are software or hardware systems that automate the process of monitoring and analyzing the events that occur in a computer network, to detect malicious activity. Since the severity of attacks occurring in the network has increased drastically, Intrusion detection systems have become a necessary addition to security infrastructure of most organizations.

Intrusion detection allows organizations to protect their systems from the threats that come with increasing network connectivity and reliance on information systems. Given the level and nature of modern network security threats the question for security professionals should not be whether to use intrusion detection but instead which intrusion detection features and capabilities can be used. Intrusions are caused by: Attackers accessing the systems, Authorized users of the systems who attempt to gain additional privileges for which they are not authorized, Authorized users who misuse the privileges given to them.

Intrusion detection systems (IDS) take either network or host based approach for recognizing and deflecting attacks. In either case, these products look for attack signatures (specific patterns) that usually indicate malicious or suspicious intent. When an IDS looks for these patterns in network traffic then it is network based. When an IDS looks for attack signatures in log files, then it is host based. Various algorithms have been developed to identify different types of network intrusions; however there is no heuristic to confirm the accuracy of their results. The exact effectiveness of a network intrusion detection system's ability to identify malicious sources cannot be reported unless a concise measurement of performance is available.

The rest of the paper is organized as follows. Proposed algorithm is explained in section II. Related work is presented in section III. Experimental results are presented in section IV. Concluding remarks are given in section V.

II. PROPOSED ALGORITHM

The proposed system is composed of three phases a) Pre-Processing, b) Training Phase, c) Anomaly Detection

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 07 Issue 05
May 2020

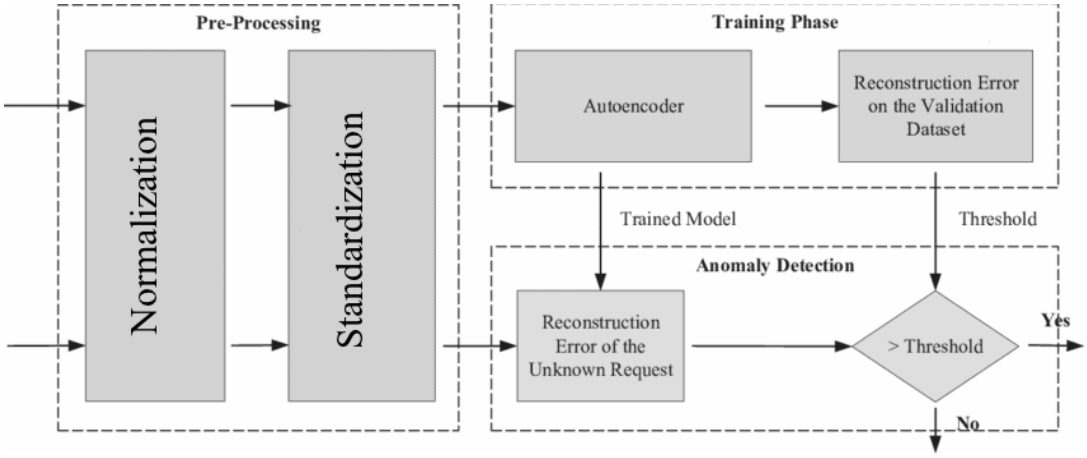


Figure 1. Architecture of our proposed model

a) *Preprocessing*

We performed the following pre-processing procedures on the KDD-CUP'99 dataset.

(1) *Feature Normalization:* the numeric features must be normalized for removing the effect of original feature value scales.

(2) *Auxiliary procedures:* now some of the auxiliary procedures are applied like min-max scalers. MinMaxScaler Transform features by scaling each feature to a given range. This estimator scales and translates each feature individually such that it is in the given range on the training set, e.g. between zero and one.

(3) *Redundancies reduction:* one of the main problems of the KDD-CUP'99 data is the large number of duplicate records that leads to the bias towards more frequent records. To solve this problem, we removed all duplicate records in data, and kept only one copy of each record.

b) *Training Phase*

After a thorough preprocessing i.e normalization and redundancy reduction on the dataset, we define the model. In our project we are using an autoencoder approach. An autoencoder is a type of artificial neural network used to learn efficient data codings in an unsupervised manner.

The aim of an autoencoder is to learn a representation (encoding) for a set of data, typically for dimensionality reduction, by training the network to ignore signal noise. Along with the reduction side, a reconstructing side is learnt, where the autoencoder tries to generate from the reduced encoding a representation as close as possible to its original input, hence its name.

The encoder and decoder functions are each fully-connected neural layers. The encoder function uses a ReLU activation function, while the decoder function uses a sigmoid activation function. The encoder layer encodes the input image as a compressed representation in a reduced dimension. The compressed image typically looks garbled, nothing like the original image. The decoder layer decodes the encoded image back to the original dimension. The decoded image is a lossy reconstruction of the original image.

We extract the encoder model from the first layer of the autoencoder model. The reason we'd want to do this is to examine what an encoded image looks like. The training data is iterated in batches of 256 in 500 epochs.

c) *Anomaly Detection*

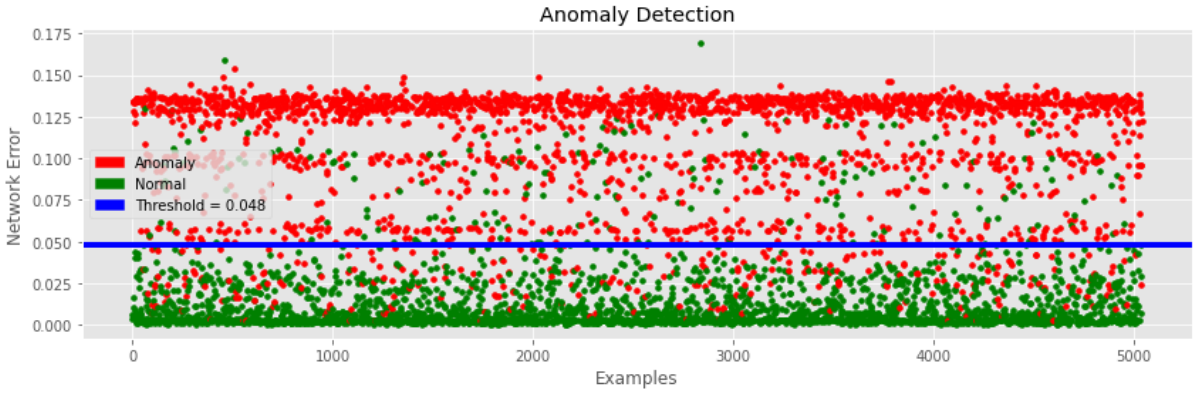
As the model is defined and trained based on the autoencoder approach, now we evaluate the model and make predictions. Based on the threshold value, if the new unknown request is above the threshold then the network is malicious and unreliable. Or if the unknown request is below the threshold then the network is reliable and non malicious. Finally the accuracy of the model is evaluated. Our proposed model gave an accuracy of 91 % which in comparison is better than some of the previous proposed models.

III. RELATED WORK

Many data mining techniques have been used for intrusion detection. In 1980; James P. Anderson [1] classified the threats and introduced a system which can detect the anomalies in the user's behaviour. Later on many researchers used different techniques i.e., SVM (Support Vector Machine), Principal Component Analysis (PCA) to make an efficient intrusion detection system, genetic network programming (GNP), Levenberg Marquardt (LM) Learning, etc., to make an efficient intrusion detection system. In 2007, Shai Rubin and Barton P. Miller [2] introduced a technique called protomatching that combines protocol analysis, normalization and pattern matching into a single phase. In 2009, Meng Jianliang and Shang Haikun [3] , used the K-Means cluster algorithm for intrusion detection. Later in 2010, Mohammaderza, Sara, Fatimah and Lilly [4] used two techniques i.e., C4.5 and SVM for detecting network intrusion and found that the C4.5 algorithm performs better than SVM in detecting network intrusion. Zubair A. Baig [5] , in his AODE-based NIDS, suggested that the Naive Bayes does not accurately detect network intrusion. In 2012, Yogendra Kumar Jain [6], compared four machine learning algorithms i.e., J48, BayesNet, OneR and, NB for intrusion detection and results shows that the J48 decision tree gives more accuracy than the other three algorithms. In the same year, R Rangaduari [7] introduced an Adaptive NIDS using a Hybrid Approach which uses a two stage approach: in the first stage, a probabilistic classifier is used whereas in the second stage, a HMM based traffic model is used. V. Jaiganesh used Kernelized [8] SVM with Levenberg Marquardt Learning for intrusion detection. Gholam Reza Zargar [9] introduced a category based IDS using PCA. Christopher and Justin [10] described the application of carefully selected nonparametric, semi-supervised learning algorithms to the network intrusion problem, in their study they compared the performances of different model types using feature-based data derived from operational networks. Chitrakar et al. [11] proposed a hybrid approach of combining k-means clustering techniques with Naive Bayes classification.

IV. EXPERIMENT AND RESULT

The proposed system is trained on KDD CUP'99 dataset. The whole setup has run on the Google Collab IDE for a faster training process. Keras 2.0 was used to implement the deep learning model with the AutoEncoder approach. Tensorflow is a deep learning library developed by Google installed as a backend for the Keras framework for creating and training deep neural networks. After training the model using the AutoEncoder approach, the results are shown as below.



Model Accuracy:
0.9128795395911887

Confusion matrix:
[[2499 177]

Classification report:

	precision	recall	f1-score	support
0.0	0.91	0.93	0.92	2676
1.0	0.92	0.89	0.91	2363
accuracy			0.91	5039
macro avg	0.91	0.91	0.91	5039

[262 2101]] weighted avg 0.91 0.91 0.91 5039

V. CONCLUSION

In this project, we presented a deep auto-encoder approach for improving the intrusion detection system. We proposed a novel network intrusion model by stacking autoencoders and evaluating our method on intrusion detection datasets. The auto-encoder is one of the most interesting models to extract features from the high-dimensional data in the context of deep learning. Our proposed model provides an accuracy of 90% which is more efficient than machine learning techniques like random forest and naive bayes.

REFERENCES

[1] Vaishali V. Khandagale, Yoginath Kalshetty, 2013, Review and Discussion on different techniques of Anomaly Detection Based and Recent Work, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 02, Issue 10 (October 2013)

[2] Rubin, Shai & Jha, Somesh & Miller, Barton. (2006). Protomatching network traffic for high throughput network intrusion detection. 47-58. 10.1145/1180405.1180413.

[3] Jianliang, Meng & Haikun, Shang & Ling, Bian. (2009). The Application on Intrusion Detection Based on K-means Cluster Algorithm. Information Technology and Applications, International Forum on. 1. 150-152. 10.1109/IFITA.2009.34.

[4] http://ijcsit.com/docs/Volume%205/vol5issue02/ijcsit20140502113.pdf

[5] Baig, Zubair & Shaheen, Samir & AbdelAal, Radwan. (2011). An AODE-based intrusion detection system for computer networks. World Congress on Internet Security, WorldCIS-2011. 28-35. 10.1109/WorldCIS17046.2011.5749877.

[6] http://www.ijsrp.org/research_paper_jan2012/ijsrp-jan-2012-21.pdf

[7] https://www.atlantis-press.com/journals/ijcis/25868734/view

[8] Jaiganesh, V. & Sumathi, P.. (2012). Kernelized Extreme Learning Machine with Levenberg-Marquardt Learning Approach towards Intrusion Detection. International Journal of Computer Applications. 54. 38-44. 10.5120/8638-2577.

[9] Zargar, Reza & Baghaie, Tania. (2012). Category-Based Intrusion Detection Using PCA. Journal of Information Security. 03. 259-271. 10.4236/jis.2012.34033.

[10] Symons, Christopher & Beaver, Justin. (2012). Nonparametric semi-supervised learning for network intrusion detection: Combining performance improvements with realistic in-situ training. Proceedings of the ACM Conference on Computer and Communications Security. 49-58. 10.1145/2381896.2381905.

[11] Chitrakar, Roshan & Huang, Chuanhe. (2012). Anomaly Based Intrusion Detection Using Hybrid Learning Approach of Combining k-Medoids Clustering and Naïve Bayes Classification. 2012 International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM 2012. 1-5. 10.1109/WiCOM.2012.6478433.