International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-6, August 2020

Intelligent and Effective Intrusion Detection System using Machine Learning Algorithm

Bhakti Nandurdikar, Rupesh Mahajan

Abstract: ***Intrusion Detection System observes the network traffic and identifies the attack and also inform the admin to corrective action. Powerful Intrusion Detection system is required for detection to various modern attack. There is need of efficient Intrusion Detection system .The focus of IDS research is the application of machine Learning and Deep Learning techniques. Projected work is combination of Deep Learning Technique in which Non Symmetric Deep Auto Encoder and Machine Learning Algorithm, Support Vector Machine Classifier is used to develop the Model. Stack power of the Non symmetric Deep Auto Encoder and Quickness with exactness of the SVM makes the Model very efficient. This Model not only improves the accuracy value but also improve recall and precision. It also cause the reduction of training time .To evaluate the performance of the Model and do the analysis the special Data set which are used are KDD CUP and NSL KDD Dataset.***

Keywords: Auto-encoders. Deep and Machine Learning, intrusion detection, Network security.

I. INTRODUCTION

Considering the current situation, Security plays a vital role in network and system. Firewall is used for the protection which acts as security Guard but the provision of efficient and effective Network intrusion Detection system is today's requirement. There are impressive advances in NIDS framework but maximum of solutions are based on Signature based techniques rather than anomaly recognition techniques. There are major issues like large volume of data, different protocols used in network, in depth monitoring, granularity need to improve the effectiveness and accuracy. Many existing strategies are incorrect and not able not take proper decision regarding the variety of attacks and Classification. For the research of NIDS the main attraction is of application of machine Learning and deep Learning Techniques. It provides best and greater assessment of network data and faster identification of attacks. Considering the current situation, there is lot of threat to the network and organization .Modern attacks are creating lot of problem. Intruders are causing great trouble they try to gain access to network or system data in unauthorized way. Instrusion Detection system is system which monitors the malicious activities taking place on the network and try to protect the network and secure the data. IDS is used in many organizations to protect the data. Large number of data is transfer on the network which leads to the increased in threat. These Intrusion affects the CIA model. A network system 's security policy should require that designated system and network administrator and response team members are train to use IDS.IDS can detect intrusions that circumvented or passed through firewall or that are occurring within the local area network behind the firewall.Various machine Learning algorithms are implemented for designing and developing intrusion Detection system like K-nearest Algorithm which is supervised algorithm ,it classify a given dataset through certain number of clusters . Naive Bayes classification technique which assume the presence of a particular feature in a class is unrelated to the presence of any other feature. Predictive model has also proven useful that are constructed by splitting a data set based on different conditions ,K-means Clustering that is used for grouping similar objects are also used in increasing the accuracy of the IDS. For Constructing IDS using machine learning application a famous dataset KDD Cup and NSL KDD are used which consist of 41 features and divided into five main groups like Normal and attacks: Probe, DOS, U2R and R2L.After evaluating and observing, the result obtain from our system are remarkable and are tough enough to deal with all harmful activities on network and system. In recent years machine Learning has become a current trend which is not only profitable but also acts as an all-rounder for detection and protection from intruders which are stealing valuable information of the system and industry.

II. REVIEW OF LITERATURE

The paper [1] is based on deep learning techniques which are motivated due to human mind gain from lower level to more elevated levels idea. The Deep Belief Network (DBN) has capacities which are mapping from input to the output. DBN makes use of an unsupervised learning algorithm, a Restricted Boltzmann Machine (RBM) for every layer is used. Benefits are: It has Deep coding capacity to adjust the changing settings of valuable information that guarantees the method performs comprehensive information analysis. Detects anomalies in the framework that incorporates peculiarity location, traffic recognizable proof. Disadvantages are: Need for quicker and effective information appraisal.

Revised Manuscript Received on July 20, 2020.
* Correspondence Author

Bhakti Nandurdikar*, Dr.D.Y.Patil Institute of Technology, Pune. India E-mail: bhaktiln@gmail.com

Rupesh Mahajan, Dr.D.Y.Patil Institute of Technology, Pune. India E-mail: mhjn.rpsh@gmail.com

Retrieval Number: F1231089620/2020©BEIESP
DOI: 10.35940/ijeat.F1231.089620
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.

In [2] paper, A Restricted Boltzmann Machine (RBM) and a deep belief network are executed for anomaly detection. This strategy utilizes a one-concealed layer RBM to do unaided component decrease. The resultant loads from this RBM are surpassed to some other RBM delivering a profound conviction organize.

The pre-prepared loads are given into an outstanding tuning layer comprising of a Logistic Regression (LR) classifier with multi-class delicate max. This techniques achieves 97.9% exactness and delivers a low bogus negative pace of 2.47%.The strategy must improve the element decrease process in the profound learning system and for betterment of dataset. . The paper [3] proposes a deep learning based methodology for developing good NIDS. NIDS based on sparse auto encoder and soft-max regression was resolved. ,A deep learning based procedure Self-showed Learning (STL) and NSL-KDD dataset for organize interruption are used. Benefits are: STL obtain a characterization precision rate over 98% for a wide range of order. Weaknesses are: Need to execute a continuous NIDS for real systems utilizing profound learning strategy.

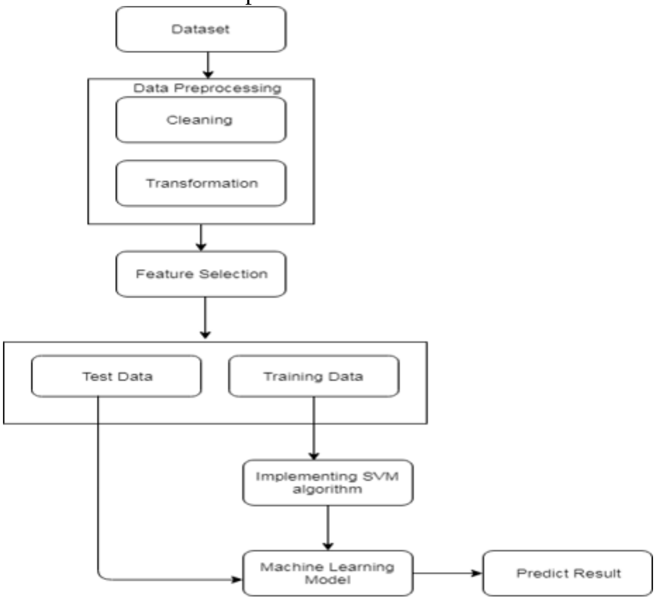
In [4] paper For huge network system data, pick Multi-center CPU's to assess the presentation of DNN based IDS is used. The parallel processing abilities of the neural system make the Deep Neural Network (DNN) to viably glance through the system traffic with a quickened presentation. Advantages are: The DNN based IDS is dependable and effective in interruption discovery

For recognizing the particular assault classes with required number of tests. Disadvantages are: Need to improve the discovery exact nesses of DNN based IDS.

In [5] Replicator Neural Networks (RNNs) Model for creating anomaly identification is proposed for discovering large network –wide attack. This methodology is unaided and do not need named information It also perfectly detects large network attack without insolent that the training data. Advantages are: The proposed procedure can effectively find all noticeable DoS attacks and SYN Port sweeps infused. Weakness are: Need to improve strategy by utilizing stacked auto encoder deep learning methods

In [6] paper. In SDN domain environment a deep Learning approach for anomaly identification is used. Advantages are: It finds an ideal hyper- parameter for DNN and affirms the location rate and bogus caution rate. The model gets the exhibition with precision of 75.75% which is very moderate from simply utilizing six basic system highlights. Improvement required is operating in real SDN environment. The paper [7]. The RNN model fundamentally has a stream of information in single direction. from the info units to the shrouded units, and the amalgamation of the single direction information stream from the past fleeting camouflage unit to the present planning concealing unit and it is used for intrusion detection. Advantages are: It has a solid intrusion recognition capacity. It has higher precision than the other AI techniques. Disadvantages are: Training time required for dataset is quite long. Applying all the 41 features in the NSL-KDD dataset to assess the noisy examples may prompts tedious and it likewise diminish execution debasement of the framework.

[8]. CFS Subset is used to diminish the dimensionality of the dataset.: The Random Forest calculation shows the most elevated exactness contrasted and remaining calculations by considering with and without highlight decrease. Arbitrary Forest is fast for grouping. Detriments are: Need to improve the Random Forest calculation requires to be more efficient to construct a proficient intrusion framework.



The paper[9] For feature extraction and classification a deep Learning algorithm (SDA) ie stacked denoising Autoencoder is used. This SdA model identify global and invariant features in the sensor signals for fault monitoring .It is robust against measurement noise. This multilayered architecture is capable of learning global features from complex input data, such as multivariate time-series datasets and high-resolution images. Merits are: The SdA model proposes effectively learn normal and fault-related features from sensor signals without preprocessing. Disadvantages are: Need to investigate a trained SdA to identify the process parameters that most significantly impact the classification results.

In [10]paper, For securing objective frameworks and systems against malicious activities anomaly-based network intrusion recognition plays vital role. The primary A-NIDS innovations, together with their general operational design, and gives a characterization to them as indicated by the kind of handling identified with the "conduct" model for the objective framework. Statistical, knowledge and machine learning-based techniques are three types of detection Advantages are: A strong global search method is the positive point. Demerits are: Requirement of large resources.

III. PROPOSED METHODOLOGY

In this paper we have proposed the model which is very smart and is able to identify large range of network traffic. We have develop this Model by using deep Learning algorithm and Machine Learning Algorithm. In Deep Learning Algorithm we have used Non symmetric Deep Auto Encoder which has multiple hidden Layer. It is used as Feature Extractor. It can contain large huge dimensional Inputs.

It provides layer wise unsupervised representation which allows us to show complex relationships between Different features .It has feature extraction capabilities and help the model in prioritizing the most descriptive Feature. In Machine Learning we have used SVM algorithm as classifier. It accuracy and speed is used to build the Model.

Network Intrusion Detection System develop using combining stacked NDAEs and SVM classification algorithm increases the strength and reduces the analytical overhead and reduces the training time.

A. Architecture

In Data Processing cleaning and Transformation takes place then Feature selection further testing and Training of Data then implementing SVM which shows result.

Mathematical Model

1. **PREPROCESSING:**

In this step, training data source (T) is normalized to be equipped for processing by using following steps:

$$T_{norm} = \{\frac{T-\mu T}{} \text{ , } \sigma_T \quad 0 \, and \, T - \mu_T, \, \sigma_T = 0 \qquad (1)$$

Where,

$T = \{x_{i,j} | i = 1, 2, ..., m \text{ and } j = 1, 2, 3, ..., n\}$

$\mu_T = \{\mu_j | j = 1, 2, 3, ..., n\}$

$\sigma_T = \{\sigma_j | j = 1, 2, 3, ..., n\}$

T is m samples with n column attributes; x_{ij} is the jth column attribute in ith sample, $_T$ and $_T$ are 1 n matrix which are the training data mean and standard deviation respectively for each of the n attributes. Test dataset (TS) which is used to measure detection accuracy is normalized using the same $_T$ and $_T$ as follows:

$$TS_{norm} = \frac{\sigma T}{} (x)/\sigma_T \text{ , } \sigma_T \, f= 0 \text{ and } TS - \mu_T \text{ , } \sigma_T = 0 \text{ (2)}$$

2. **FEATURE SELECTION:**

Non Symmetric Deep Auto Encoder, which is Deep Learning Algorithm consist of large hidden layers and takes input vector $x \in R^d$ and step-by-step maps it to the latent representations $h_i \in R^d$ (here d represents the dimension of the vector) using a deterministic function shown in (3) below:

$$h_i = \sigma(W_i.h_{i-1} + b_i); \; i = 1, \overline{\;} n, \text{ (3)}$$

Here, $h_0 = x$, σ is an activation function (in this work use sigmoid function $\sigma(t) = 1/(1 + e^{-t})$ and n is the number of hidden layers., The proposed NDAE does not contain a Decoder and its output vector is calculated by a similar formula to (4) as the latent representation.

$$y = \sigma(W_{n+1}.h_n + b_{n+1}) \qquad (4)$$

The estimator of the model $\theta = (W_i, b_i)$ can be obtained by minimizing the square reconstruction error over m training samples $(x^{(i)}, y^{(i)})^m$, as shown in (5)

$$E(\theta) = \Sigma m (x^{(i)}, y^{(i)})^2 \text{ (5)}$$

IV. ALGORITHM

Support Vector Machine:

Machine Learning is classified into supervised and unsupervised Learning algorithm. One of the crucial algorithm is Support Vector Machine. which is supervised Algorithm. The SVM is used to solve classification problems and it generate hyperplane in repetitive manner to reduce the error. It also divide the datasets .It represents the different class in hyperplane in multidimensional space.

Steps:

Step 1: Read the test features and trained features.

Step 2: Check the all test features of dataset and also get all train features.

Step 3: Consider the kernel.

Step 4: Train the SVM using both features and show the output.

Step 5: Classify an observation using a Trained SVM Classifier.

V. RESULTS AND DISCUSSION

The experimental result evaluation, we have notation as follows:

– True positive (TP): Attack data (Intrusion) is correctly identify as an attack.

-False positive (FP): Normal data or behavior that is incorrectly classified as attack by the IDS.

-True Negative (TN): Normal or non-intrusive behavior that is rightly identify as normal by the IDS.

-False Negative (FN): Intrusions (attack data) that are incorrectly classified as normal.

On the basis of this parameter, we can calculate four measurements

Accuracy = TP+TN/TP+FP+TN+FN

Precision = TP/ TP+FP

Recall= TP/TP+FN

F1-Measure = 2*(Precision*Recall)//Precision+ Recall.

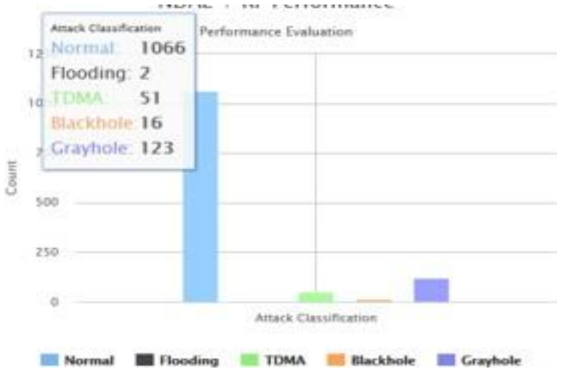


Fig. 2. Performance analysis graph to count the attacks

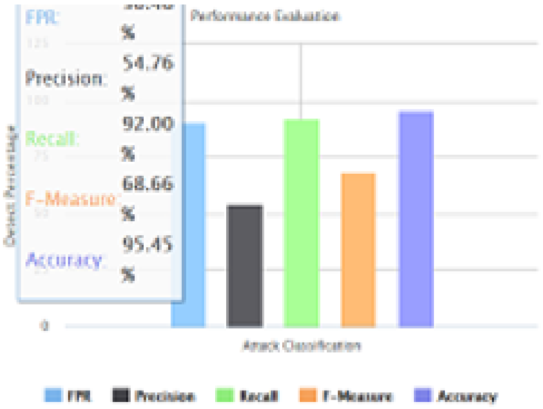


Fig 3. Improved Accuracy, precision, Recall.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.

VI. CONCLUSION & FUTURE WORK

Intrusion Detection System (IDS) identifies successfully the real time attacks on a network and alarm to Admin. The main focus of NIDS research has been the application of machine learning and Deep learning techniques which has given a great encouragement to large number of network attack In this paper, we have mentioned the problems of the previous NIDS techniques and have proposed model constructed from stacked NDAEs and SVM.Avice algorithm which would find the attacks in the dataset correctly has been successfully designed and developed for Intrusion Detection system. This approach provides high levels of accuracy, precision and recall together with reduced training time. It can also solve the high requirement of intrusion detection timely. In future work, we need to improve the capacity to handle the faults and Vulnerabilities of system or network and to manage the time at which this attack (zero day attack) is discover. Also to expand our current evaluations for real-world traffic of the Network.

REFERENCES

1. ”Comparison Deep Learning Method to traditional Methods Using for Network Intrusion Detection ”Bo Dengue Wang .IEEE conference 2016.
2. https://medium.com/cuelogic-technologies/evaluation-of-machine -learning-algorithms-for-intrusion-detection-system
3. www.google.com
4. Unsupervised learning techniques for an intrusion detection system, Stefano Zanero, Sergio M. Savaresi, Piazza L. da Vinci, 32; 20133, Milan, Italy
5. T. Kohonen, "Things you haven't heard about the Self-Organizing Map,"in Neural Networks, 1993., IEEE International Conference on, 1993, pp.1147-1156.
6. Machine Learning Based Network Intrusion Detection” Chie-Hong Lee, Yann-Yean Su, Yu-Chun Lin and Shie-Jue Lee,IEEE conference 2017.
7. Role of Machine Learning in Instrusion Detection system: Review”,L Hripriya,M A.Jabbar,IEEE conference 2018.
8. A Machine Learning Approach to IDS:A Comprehensive Review”,Kunal and Mohit Dua,IEEE conference 2019.
9. A comparative study of using several artificial intelligence algorithims on intrusion detection system “, Ahmad Yoosofan , Fatemeh Ghovanlooy Ghajar, Masoud Moghadasian and Reza Babaee ,IEEE 2019.
10. lightweight network intrusion detection system using chi-square and cuckoosearch optimization algorithms with decision tree classifier”, D.Kayathri Devi, Dr.R.Sukumar,Dr.R,suresh Babu,IEEE 2019

AUTHORS PROFILE

Bhakti Nandurdikar is Pursuing Master’s in computer engineering from Pad. Dr.D.Y.Patil Institute of Technology, Pune .She has completed her Engineering in IT from Pune University in 2011.She is involve in Teaching and area of Interest is Machine Learning, IOT and Cloud Computing

Prof Rupesh Mahajan is pursuing PhD in Computers from Sandip Foundation, Nasik. He has received Master’s Degree in Computers from Pune university in 2008.He is currently working as Assistance Professor in Pad Dr.D.Y.Patil Institute of Technology,Pune. He is involve in Research Project and teaching .His Research Interest include Data Mining, Cloud Computing and Machine Learning.

Retrieval Number: F1231089620/2020©BEIESP
DOI: 10.35940/ijeat.F1231.089620
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.