International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
©2021 INPRESSCO®, All Rights Reserved
Available at http://inpressco.com/category/ijcet

Research Article

Intelligent and Effective Intrusion detection system using Machine Learning Algorithm

Miss.Bhakti Nandurdikar and Prof.Rupesh Mahajan

Department of Computer Engineering Dr.D.Y.Patil Institute of Technology Savitribai Phule Pune University Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

A framework Network intrusion discovery framework (NIDS) helps the system admin to identify network security breaks in their own affiliation. Regardless, various troubles rise while developing a canny and powerful NIDS for unexpected and capricious attacks. In recent years, one of the preeminent focuses inside NIDS examines has been the application of machine learning knowledge of techniques. Proposed work present a novel deep learning model with support vector machine classifier to enable NIDS operation within modern networks. The model shows a combination of deep learning and machine learning, capable of correctly analyzing a wide-range of network traffic. This model increases the accuracy ,Precision and recall with reduced training time over existing system. Moreover, additionally proposes novel deep learning classification display built utilizing feature extraction techniques. The performance evaluated network intrusion detection analysis dataset, especially KDD CUP and NSL KDD dataset.

Keywords: *Deep and machine learning, intrusion detection, Auto-encoders, Network security.*

Introduction

One of the major challenges in network security is the provision of a powerful and effective Network Intrusion Detection System (NIDS). Regardless of the impressive advances in NIDS framework, the majority of solutions still operate using less-successful signature-based techniques, rather than anomaly recognition strategies. The present issues are the existing strategies prompts ineffectual and incorrect discovery of attacks. There are three fundamental limitations like, volume of network data, in-depth monitoring and granularity required to improve effectiveness and accuracy and finally the number of different protocols and diversity of data traversing. The main focus of NIDS research has been the application of machine learning and shallow learning techniques. The initial deep learning research has demonstrated that its superior layer-wise feature learning can better or at least match the performance of deep learning techniques. It is able to facilitating a deeper evaluation of network data and faster identification of any anomalies. In this paper, proposes a novel deep learning version to enable NIDS operation inside modern networks.

Despite expanding attention of network security, the existing solutions remain incapable of fully protecting inter- net applications and computer networks opposite the threats from ever-advancing cyber-attack method like as DoS attack and computer malware. Developing effective and adaptive security approaches, therefore, has become more critical than ever before. The traditional security techniques, as the first line of security defense, such as user authentication, firewall and data encryption, are insufficient to fully cover the hole landscape of network security while facing challenge from ever-evolving intrusion skills and technique [1]. Hence, other line of security defense is more suggested, like Intrusion Detection System (IDS). Currently, an IDS alongside with anti-virus software has become an important complement to the security infrastructure of most organizations. The combination of these two lines provides a more comprehensive defense against those threats and enhances network security. A significant amount of research has been conducted to develop intelligent intrusion detection techniques, which help achieve better network security. Bagged boosting-based on C5 decision trees [2] and Kernel Miner [3] are two of the earliest attempts to build intrusion detection schemes. Methods proposed in [4] and [5] have successfully applied machine learning techniques to classify network traffic patterns that do not match normal network traffic. The two frameworks were furnished with one classifiers to distinguish typical traffic and six different types of attacks.

However, current network traffic data, which are often huge in size, present a major challenge to IDSs [9]. These big data slow down the entire detection process and may lead to unsatisfactory classification

accuracy due to the computational difficulties in handling such data. Classifying a huge amount of data usually causes many mathematical difficulties which then lead to higher computational complexity. As a wellknown intrusion calculation dataset, KDD Cup 99 dataset is a typical example of more-scale datasets. This dataset contains of more than five million of training samples and two million of testing samples respectively. Such a large scale dataset check the building and testing procedure of a classifier, or form the classifier unable to do due to framework failures caused by low memory. Furthermore, large-scale datasets usually contain noisy, redundant, or uninformative features which present critical challenges to knowledge discovery and information modelling.

Review of Literature

The paper [1] focuses on deep learning strategies which are motivated by the structure profundity of human mind gain from lower level trademark to more elevated levels idea. It is a direct result of reflection from various levels, the Deep Belief Network (DBN) takes in capacities which are mapping from input to the output. The method of becoming more acquainted with doesn't subject to human-made features. DBN makes use of an unsupervised learning algorithm, a Restricted Boltzmann Machine (RBM) for every layer. Advantages are: Deep coding is its capacity to adjust to changing settings concerning information that guarantees the method conducts comprehensive information analysis.Detects anomalies in the framework that incorporates peculiarity location, traffic recognizable proof. Disadvantages are:Demand for quicker and effective information appraisal.

In [2] paper, a deep learning approach for anomaly detection using a Restricted Boltzmann Machine (RBM) and a deep belief network are executed. This strategy utilizes a one-concealed layer RBM to perform 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. Favorable circumstances are: Achieves 97.9% exactness. It delivers a low bogus negative pace of 2.47%. Disservices are: Need to improve the strategy to boost the element decrease process in the profound learning system and to improve the dataset.

The paper [3] proposes a deep learning based methodology for building an efficient NIDS. A sparse autoencoder and soft-max regression based NIDS was resolved. U Utilizations Self-showed Learning (STL), a deep learning based procedure, on NSL-KDD - a benchmark dataset for organize interruption Advantages are: STL accomplished 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 pick multi-center CPU's just as GPU's to assess the presentation of the DNN based IDS to deal with huge network system data.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 for preparing. The multicore CPU's was faster than the serial training mechanism. Disadvantages are: Need to improve the discovery exactnesses of DNN based IDS.

In [5] paper, proposes a system for identifying huge scale network-wide attacks using Replicator Neural Networks (RNNs) for creating anomaly identification models. Our methodology is unaided and requires no named information It also accurately detects network-wide anomalies without presuming that the training data is completely free of attacks. Advantages are: The proposed procedure can effectively find all noticeable DoS atacks and SYN Port sweeps infused. Proposed system is versatile against learning within the sight of assaults, something that related work needs. Disadvantages are: Need to improve proposed strategy by utilizing stacked autoencoder deep learning methods.

Based on the stream based centrality of SDN, author propose a stream based anomaly identification system using deep learning. In [6] paper, apply a deep learning approach for stream based anomaly recognition in a SDN domain environment. Advantages are: It finds an ideal hyperparameter 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. Disadvantages are: It will not work on real SDN environment.

The paper [7] proposes a deep learning approach for intrusion detection using recurrent neural networks.The RNN model essentially has a single direction stream of information 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. Advantages are: It has a solid demonstrating capacity for intrusion recognition. It has higher precision than the other AI techniques. Disadvantages are:spend more time for training dataset.

Utilizing all the 41 highlights in the NSL-KDD dataset to assess the nosy 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. Focal points are: The Random Forest calculation shows the most elevated exactness contrasted and rest of the calculations by considering with and without highlight decrease. Arbitrary Forest has the fast for grouping. Detriments are: Need to improve the Random Forest calculation to construct a proficient intrusion framework.

The paper [9] proposes a deep learning based approach to implement such an effective and flexible

NIDS. We use Self-taught Learning (STL), a deep learning primarily based approach, on NSL-KDD - a benchmark dataset for network intrusion. Self-taught Learning (STL) is a deep learning technique that consists of two stages for the classification. First, a good feature representation is learnt from a large collection of unlabeled data, known as Unsupervised Feature Learning (UFL). In the second stage, this learnt representation is applied to labeled data, and used for the classification task. Although the unlabeled and labeled data may come from different distributions, there must be relevance among them. Advantages are: NIDSs based on this approach achieved very high-accuracy and less false-alarm rates. Disadvantages are: Need to work on stacked auto-encoders in deep belief network.

In [10] paper, anomaly-based network intrusion recognition methods are a significant innovation to secure objective frameworks and systems against malicious activities. 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. There are three types of Anomaly detection: statistical, knowledge and machine learning-based techniques. Advantages are: Machine learning A-NIDS is the use of a flexible and robust global search method. Disadvantages are: High resource consumption.

Proposed Methodology

In this paper, propose a novel deep learning model to enable NIDS operation within modern networks. The model proposes is a combination of deep and machine learning, capable of correctly analyzing a wide-range of network traffic. More specifically, we combine the power of stacking our proposed Non-symmetric Deep Auto-Encoder (NDAE) (deep learning) and the accuracy and speed of Support Vector Machine (SVM). This paper introduces our NDAE, which is an auto-encoder featuring non-symmetrical multiple hidden layers. NDAE can be used as a hierarchical unsupervised feature extractor that scales well to accommodate high dimensional inputs. It learns non-trivial features using a similar training strategy to that of a typical auto-encoder. Stacking the NDAEs offers a layer-wise unsupervised representation learning algorithm, which will allow our model to learn the complex relationships between different features. It also has feature extraction capabilities, so it is able to refine the model by prioritizing the most descriptive features.

Advantages are:

- Due to deep learning technique, it improves accuracy of intrusion detection system.
- The network or device is continuously monitored for any invasion or attack.
- The system may be modified and modified in step with desires of unique client and can help outside as well as inner threats to the system and network.
- It presents user friendly interface which allows easy protection management systems.
- Any alterations to files and directories on the machine can be easily detected and reported.

A. Architecture

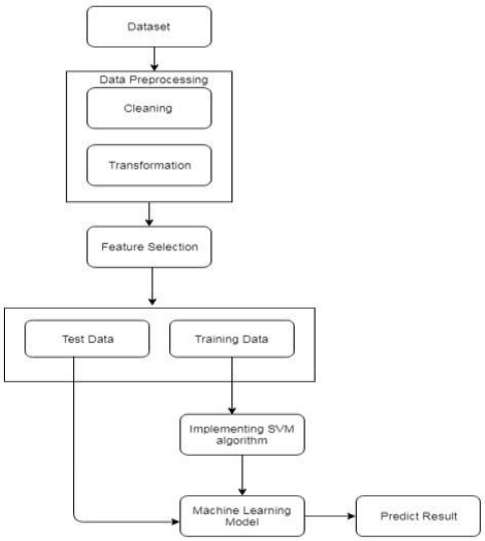

Fig. 1. Proposed System Architecture

B. Mathematical Model

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

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

Where,
$T = \{x_{ij} | i = 1,2,...,m \text{ and } j = 1,2,3,...,n\}$ $\mu_T = \{\mu_j | j = 1,2,3,...,n\}$ $\sigma_T = \{\sigma_j | j = 1,2,3,...,n\}$
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:

$$TS_{norm} = \frac{\sigma_T}{\sigma_T}(x)/\sigma_T,\ \sigma_T \neq 0\ and\ TS-\mu_T, \sigma_T = 0 \quad (2)$$

Feature Selection:

NDAE is an auto-encoder featuring non-symmetrical multiple hidden layers. The proposed NDAE takes an 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: $h_i = \sigma(W_i.h_{i-1} + b_i)$; $i = 1,n,^{-}$ (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. Unlike a conventional Auto-Encoder

and Deep Auto-Encoder, the proposed NDAE does not contain a Decoder and its output vector is calculated by a similar formula to (4) as the latent representation.

$y = \sigma(Wn{+}1.hn + bn{+}1)$ (4)

The estimator of the model $\theta = (W_i, b_i)$ can be obtained by minimizing the square reconstruction error over m training samples $(x^{(i)}, y^{(i)})_{i=1}^{m}$, as shown in (5).

$E(\theta) = \sum_{i=1}^{m} (x^{(i)}, y^{(i)})^2$ (5)

C. Algorithms

1. Support Vector Machine:

Support Vector Machine (SVM) is used to classify the fruit quality. SVM Support vector machines are mainly two class classifiers, linear or non-linear class boundaries.

The idea behind SVM is to form a hyper plane in between the data sets to express which class it belongs to.

The task is to train the machine with known data and then SVM find the optimal hyper plane which gives maximum distance to the nearest training data points of any class 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.

Results and Discussion

The experimental result evaluation, we have notation as follows:

TP: True positive (correctly predicted number of instance)

FP: False positive (incorrectly predicted number of instance), TN: True negative (correctly predicted the number of instances as not required)

FN false negative (incorrectly predicted the number of instances as not required),

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

Accuracy = TP+TNTP+FP+TN+FN

Precision = TP TP+FP

Recall= TPTP+FN

F1-Measure = 2PrecisionRecall Precision+ Recall.

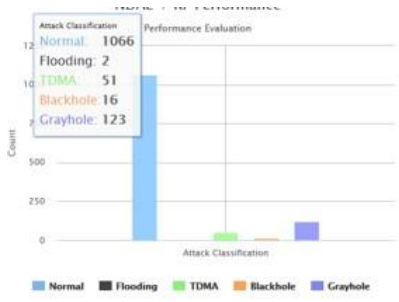

Fig. 2. Performance analysis graph to count the attacks

Conclusion

Considering Current scenario, intrusion detection remains critical for network security and machine learning based applications which have given a major boost in finding novel attacks. In this paper, we have mentioned the problems confronted by previous NIDS techniques and have proposed our novel NDAE Method for unsupervised feature learning. Proposed Model is constructed from stacked NDAEs and SVM .A novel algorithm for intrusion detection system which would detect the attacks in the dataset has been successfully designed and implemented. The result shows that given approach offers high levels of accuracy, precision and recall together with reduced training time. This method has high accuracy, and it can also solve the high requirement of intrusion detection timely. Still there is need for further work on real-time network traffic and to handle zero-day attacks.

References

[1]. Nathan shone , tranngguyenngoc, vu dinhphai , and qi sh, "a deep learning approach to network intrusion detection",ieee transactions on emerging topics in computational intelligence, vol. 2, no. 1, february 2018

[2]. B. Dong and X. Wang, "Comparison deep learning method to traditional methods using for network intrusion detection," in Proc. 8th IEEE Int.Conf. Commun. Softw. Netw, Beijing, China, Jun. 2016, pp. 581–585.

[3]. K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," in Proc. 15th IEEE Int. Conf. Mach. Learn. Appl., Anaheim, CA, USA, Dec. 2016, pp. 195–200.

[4]. A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in Proc. 9th EAI Int.Conf. BioInspired Inf. Commun. Technol., 2016, pp. 21–26. [Online]. Available: http://dx.doi.org/10.4108/eai.3-12-2015.2262516

[5]. S. Potluri and C. Diedrich, "Accelerated deep neural networks for enhanced intrusion detection system," in Proc. IEEE 21st Int. Conf. Emerg. Technol. Factory Autom., Berlin, Germany, Sep. 2016, pp. 1–8.

[6]. C. Garcia Cordero, S. Hauke, M. Muhlhauser, and M. Fischer, "Analyzing flow-based anomaly intrusion detection using replicator neural networks," in Proc. 14th Annu. Conf. Privacy, Security. Trust, Auckland, New Zeland, Dec. 2016, pp. 317–324.

[7]. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for network intrusion detection in software defined networking," in Proc. Int. Conf. Wireless Netw. Mobile Commun., Oct. 2016, pp. 258–263.

[8]. C. Yin, Y. Zhu, J. Fei and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," in IEEE Access, vol. 5, pp. 21954-21961, 2017.

[9]. Revathi, S Malathi, A. (2013). A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection. International Journal of Engineering Research Technology (IJERT). 2. 1848-1853.

[10]. N. Shone, T. N. Ngoc, V. D. Phai and Q. Shi, "A Deep Learning Approach to Network Intrusion Detection," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 4150, Feb. 2018.

[11]. Anomaly-based network intrusion detection: Techniques, systems and challenges Garcia-Teodoro P. Diaz-Verdejo J., Macia-Fernandez G., Vazquez E. (2009) Computers and Security, 28 (1-2), pp. 18-28.

[12]. Claude Turner*, Rolston Jeremiah, Dwight Richards, Anthony Joseph, "A Rule Status Monitoring Algorithm for Rule-Based Intrusion Detection and Prevention Systems", Procedia Computer Science, ISSN: 1877-0509, Vol: 95, Page: 361-368, 2016

[13]. R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring: A survey," Submitted to IEEE Trans. Neural Netw. Learn. Syst., 2016. [Online]. Available: http://arxiv.org/abs/1612.07640

[14]. H. Lee, Y. Kim, and C. O. Kim, "A deep learning model for robust wafer fault monitoring with sensor measurement noise," IEEE Trans. Semicond. Manuf., vol. 30, no. 1, pp. 23–31, Feb. 2017.

[15]. L. You, Y. Li, Y. Wang, J. Zhang, and Y. Yang, "A deep learning based RNNs model for automatic security audit of short messages," in Proc. 16th Int. Symp. Commun. Inf. Technol., Qingdao, China, Sep. 2016, pp.225-229.