

# A Survey on Intelligent and Effective Intrusion Detection system using Machine Learning Algorithm

Bhakti Nandurdikar<sup>1</sup>, Prof. Rupesh Mahajan<sup>2</sup>

Department of Computer Engineering

Dr. D. Y. Patil Institute of Technology, Pimpri, Pune

**Abstract-** A system Network intrusion discovery framework (NIDS) helps the system admin to identify network security breaks in their own association. Nonetheless, numerous difficulties emerge while building up an intelligent and effective NIDS for unexpected and capricious attacks. In recent years, one of the foremost focuses inside NIDS studies has been the application of machine learning knowledge of techniques. Proposed work present a novel deep learning model to enable NIDS operation within modern networks. The model shows a combination of deep learning, capable of correctly analyzing a wide-range of network traffic. Moreover, additionally proposes novel deep learning classification display built utilizing feature extraction techniques. The performance evaluated network intrusion detection analysis dataset, particularly KDD CUP dataset.

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

## 1. INTRODUCTION

One of the major challenges in network security is the provision of a robust and effective Network Intrusion Detection System (NIDS). Despite the considerable advances in NIDS system, the majority of solutions still operate using less-successful signature-based techniques, rather than anomaly detection strategies. The current issues are the existing techniques leads to ineffective and inaccurate detection of attacks. There are three main 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 increasing awareness 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 method [1]. Hence, other line of security defense is more recommended, 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. Both systems were equipped with five distinct classifiers to detect normal traffic and four different types of attacks (i.e., DoS, probing, U2R and R2L).

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 well-known 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.

### Objectives

- To devise a technique capable of providing reliable unsupervised feature learning, this can improve upon the performance and accuracy of existing techniques.
- To study stacked Non-Symmetric Deep Auto-Encoder for unsupervised features extraction.
- To analyze the experimental results of proposed stacked NDAE and SVM (Support Vector Machine) classifier algorithms for intrusion detection system.
- To reduce the training time.

### 2. RELATED WORK

The paper [1] focuses on deep learning methods which are inspired by the structure depth of human brain learn from lower level characteristic to higher levels concept. It is because of abstraction from multiple levels, the Deep Belief Network (DBN) helps to learn functions which are mapping from input to the output. The process of learning does not dependent on human-crafted features. DBN uses an unsupervised learning algorithm, a Restricted Boltzmann Machine (RBM) for each layer. Advantages are: Deep coding is its ability to adapt to changing contexts concerning data that ensures the technique conducts exhaustive data analysis. Detects abnormalities in the system that includes anomaly detection, traffic identification. Disadvantages are: Demand for faster and efficient data assessment.

The main purpose of [2] paper is to review and summarize the work of deep learning on machine health monitoring. The applications of deep learning in machine health monitoring systems are reviewed mainly from the following aspects: Autoencoder (AE) and its variants, Restricted Boltzmann Machines and its variants including Deep Belief Network (DBN) and Deep Boltzmann Machines (DBM), Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Advantages are: DL-based MHMS do not require extensive human labor and expert knowledge. The applications of deep learning models are not restricted to specific kinds of machines. Disadvantages are: The performance of DL-based MHMS heavily depends on the scale and quality of datasets.

Proposes the use of a stacked denoising autoencoder (SdA), which is a deep learning algorithm, to establish an FDC model for simultaneous feature extraction and classification. The SdA model [3] can identify global and invariant features in the sensor signals for fault monitoring and is robust against measurement noise. An SdA is consisting of denoising autoencoders that are stacked layer by layer. This multilayered architecture is capable of learning global features from complex input data, such as multivariate time-series datasets and high-resolution images. Advantages are: SdA model is useful in real applications. 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.

Proposes a novel deep learning-based recurrent neural networks (RNNs) model [4] for automatic security audit of short messages from prisons, which can classify

short messages (secure and non-secure). In this paper, the feature of short messages is extracted by word2vec which captures word order information, and each sentence is mapped to a feature vector. In particular, words with similar meaning are mapped to a similar position in the vector space, and then classified by RNNs. Advantages are: The RNNs model achieves an average 92.7% accuracy which is higher than SVM. Taking advantage of ensemble frameworks for integrating different feature extraction and classification algorithms to boost the overall performance. Disadvantages are: It is apply on only short messages not large-scale messages.

Signature-based features technique as a deep convolutional neural network [5] in a cloud platform is proposed for plate localization, character detection and segmentation. Extracting significant features makes the LPRS to adequately recognize the license plate in a challenging situation such as i) congested traffic with multiple plates in the image ii) plate orientation towards brightness, iii) extra information on the plate, iv) distortion due to wear and tear and v) distortion about captured images in bad weather like as hazy images. Advantages are: The superiority of the proposed algorithm in the accuracy of recognizing LP rather than other traditional LPRS. Disadvantages are: There are some unrecognized or miss-detection images.

In [6] paper, a deep learning approach for anomaly detection using a Restricted Boltzmann Machine (RBM) and a deep belief network are implemented. This method uses a one-hidden layer RBM to perform unsupervised feature reduction. The resultant weights from this RBM are passed to another RBM producing a deep belief network. The pre-trained weights are passed into a fine tuning layer consisting of a Logistic Regression (LR) classifier with multi-class softmax. Advantages are: Achieves 97.9% accuracy. It produces a low false negative rate of 2.47%. Disadvantages are: Need to improve the method to maximize the feature reduction process in the deep learning network and to improve the dataset.

The paper [7] proposes a deep learning based approach for developing an efficient and flexible NIDS. A sparse autoencoder and soft-max regression based NIDS was implemented. Uses Self-taught Learning (STL), a deep learning based technique, on NSL-KDD - a benchmark dataset for network intrusion. Advantages are: STL achieved a classification accuracy rate more than 98% for all types of classification. Disadvantages are: Need to implement a real-time NIDS for actual networks using deep learning technique.

In [8] paper choose multi-core CPU's as well as GPU's to evaluate the performance of the DNN based IDS to handle huge network data. The parallel computing capabilities of the neural network make the Deep Neural Network (DNN) to effectively look through the network traffic with an accelerated performance. Advantages are: The DNN based IDS is reliable and efficient in intrusion detection for identifying the specific attack classes with required number of samples for training. The multicore CPU's was faster than the serial training mechanism.

Disadvantages are: Need to improve the detection accuracies of DNN based IDS.

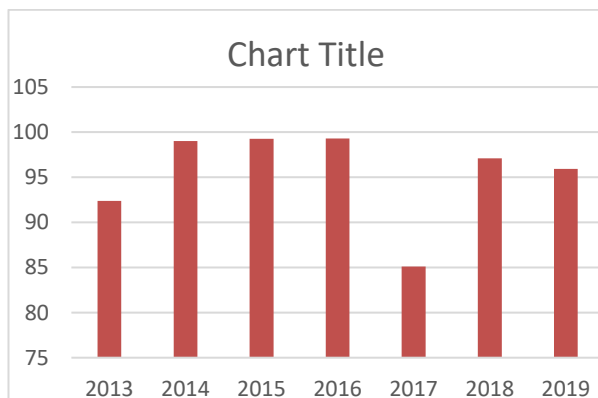
In [9] paper, proposes a mechanism for detecting large scale network-wide attacks using Replicator Neural Networks (RNNs) for creating anomaly detection models. Our approach is unsupervised and requires no labeled data. It also accurately detects network-wide anomalies without presuming that the training data is completely free of attacks. Advantages are: The proposed methodology is able to successfully discover all prominent DDoS attacks and SYN Port scans injected. Proposed methodology is resilient against learning in the presence of attacks, something that related work lacks. Disadvantages are: Need to improve proposed methodology by using stacked autoencoder deep learning techniques.

Based on the flow-based nature of SDN, we propose a flow-based anomaly detection system using deep learning. In [10] paper, apply a deep learning approach for flow-based anomaly detection in an SDN environment. Advantages are: It finds an optimal hyper-parameter for DNN and confirms the detection rate and false alarm rate. The model gets the performance with accuracy of 75.75% which is quite reasonable from just using six basic network features. Disadvantages are: It will not work on real SDN environment.

### Overview

Lot of work has been done on this Intrusion Detection system as it is basic building block for the detection of various network attacks. Variety of Machine Learning and Deep Learning algorithms are implemented to develop an efficient and useful IDS system.

Sr No	Publishing Year	Dataset	Algorithm	Detecti on Rate
1	2013	KDDCUP99, NSL-KDD	Self organized MAP	92.37
2	2014	NSL-KDD99	Genetic Algorithm	99
3	2015	KDD-CUP99	CANN and KNN	99.28
4	2016	ISCX 2012	K-Means Clustering with Decritization and Naïve Bayes	99.3
5	2017	KDD-CUP99	Initial centroid selection algorithm based on density-K-means	85.1
6	2018	NSL-KDD	Clustered Self organized map.	97.09
7	2019	KDD-CUP99	Naïve bayes,Random tree comparison	95.92



### 3.EXSTING SYSTEM

The current network traffic data, which are often huge in size, present a major challenge to IDSs. 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. Machine learning technologies have been usually used in IDS. However, most of the traditional machine learning technologies refer to shallow learning; they cannot effectively solve the enormous intrusion data classification issue that arises in the face of a real network application environment. Additionally, shallow learning is incompatible to intelligent analysis and the predetermined requirements of high-dimensional learning with enormous data.

#### Disadvantages:

Computer systems and internet have become a major part of the critical system. The current network traffic data, which are often huge in size, present a major challenge to IDSs. 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.

### 3. SYSTEM OVERVIEW

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 shallow 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 Random Forest (RF) (shallow learning). 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 computer is constantly monitored for any invasion or attack.
- The system can be modified and changed according to needs of specific client and can help outside as well as inner threats to the system and network.
- It effectively prevents any damage to the network.
- It provides user friendly interface which allows easy security management systems.
- Any alterations to files and directories on the system can be easily detected and reported.

#### 4. CONCLUSION

In this paper, we have discussed the problems faced by existing NIDS techniques. In response to this we have proposed our novel NDAE method for unsupervised feature learning. We have then built upon this by proposing a novel classification model constructed from stacked NDAEs and the RF classification algorithm. Also we implemented the Intrusion prevention system. The result shows that our approach offers high levels of accuracy, precision and recall together with reduced training time. The proposed NIDS system is improved only 5% accuracy. So, there is need to further improvement of accuracy. And also further work on real-time network traffic and to handle zero-day attacks.

#### *Future Scope*

- In our future work, the first avenue of exploration for improvement will be to assess and extend the capability of our model to handle zero-day attacks.
- We look to expand upon our existing evaluations by utilizing real-world backbone network traffic to demonstrate the merits of the extended model

#### REFERENCES

- [1] 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.
- [2] 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>
- [3] 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.
- [4] 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.
- [5] R. Polishetty, M. Roopaei, and P. Rad, "A next-generation secure cloud based deep learning license plate recognition for smart cities," in Proc. 15th IEEE Int.
- [6] 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.
- [7] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in Proc. 9th EAI Int.Conf. Bio-Inspired Inf. Commun. Technol., 2016, pp. 21–26. [Online]. Available: <http://dx.doi.org/10.4108/eai.3-12-2015.2262516>
- [8] 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
- [9] 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 Zealand, Dec. 2016, pp. 317–324.
- [10] 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.