International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 5 May 2021, pp: 890-995 www.ijaem.net ISSN: 2395-5252

Intrusion Detection Using Deep Learning Techniques

Prof. R.Aarthy,[1] M.Tech, Assistant Professor, S.Mammutha,[2] M.Jayapriya,[3] N.Gunabharathi,[4] M.Deepika[5]

Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur.

Submitted: 05-05-2021 Revised: 17-05-2021 Accepted: 20-05-2021

ABSTRACT: Intrusion Detection System (IDS) is meant to be a software application which monitors the network or system activities and finds if any malicious operations occur. Tremendous growth and usage of internet raises concerns about how to protect and communicate the digital information in a safe manner. Nowadays, hackers use different types of attacks for getting the valuable information. As the internet emerging into the society, new stuffs like viruses and worms are imported. The malignant so, the users use different techniques like cracking of password, detecting unencrypted text are used to cause vulnerabilities to the system. Hence, security is needed for the users to secure their system from the intruders. Firewall technique is one of the popular protection techniques and it is used to protect the private network from the public network. IDS are used in network related activities, medical applications, credit card frauds, Insurance agency. Many intrusion detection techniques, methods and algorithms help to detect these attacks. This main objective of this project is to provide a comparative study about intrusion detection using various machine learning and deep learning techniques. Various machine learning techniques have been used to develop IDs, such as Back Propagation, Feed Forward, Recurrent neural network and Multilayer Perceptron (MLP) in real time network datasets such as Intrusion Detection System (IDS) datasets and UNSW datasets. MLP is widely used neural network classifier based on number of classes (output) and number of hidden layers, MLP uses weights for every node at neural network, most effective attributes will get large weights conversely attributes not affect in predictive class. The proposed system can be analysed in terms of error rate and accuracy values and implement in python tool for performance analysis.

I. INTRODUCTION

NETWORKING

Networking is the exchange of information and ideas among people with a common profession or special interest, usually in an informal social setting. Networking often begins with a single point of common ground. Networking is used by professionals to expand their circles of acquaintances, to find out about job opportunities in their fields, and to increase their awareness of news and trends in their fields or in the greater world. (The term computer networking refers to linking multiple devices so that they can readily share information and software resources.) A computer network is a group of computers that use a set of common communication protocols over digital interconnections for the purpose of sharing resources located on or provided by the network nodes. The interconnections between nodes are formed from a broad spectrum of telecommunication network technologies, based on physically wired, optical, and wireless radio-frequency methods that may be arranged in a variety of network topologies. The nodes of a computer network may include personal computers, servers, networking hardware, or other specialized or general-purpose hosts. They are identified by hostnames and network addresses. Hostnames serve as memorable labels for the nodes, rarely changed after initial assignment. Network addresses serve for locating and identifying the nodes by communication protocols such as the Internet Protocol. Computer networks may be classified by many criteria, for example, the transmission medium used to carry signals, bandwidth, communications protocols to organize network traffic, the network size, the topology, traffic control mechanism, and organizational intent.Computer networks support many applications and services, such as access to the World Wide Web, digital video, digital audio, shared use of application and storage servers, printers, and fax machines, and use

DOI: 10.35629/5252-0305890895

of email and instant messaging applications. Most modern computer networks use protocols based on packet-mode transmission. A network packet is a formatted unit of data carried by a packet-switched network.

ADVANTAGES OF NETWORKING

Main benefits of networks include:

- **File sharing** - you can easily share data between different users, or access it remotely if you keep it on other connected devices.
- **Resource sharing** - using network-connected peripheral devices like printers, scanners and copiers, or sharing software between multiple users, saves money.
- **Sharing a single internet connection** - it is cost-efficient and can help protect your systems if you properly secure the network.
- **Increasing storage capacity** - you can access files and multimedia, such as images and music, which you store remotely on other machines or network-attached storage devices.

Networking computers can also help you **improve communication**, so that:

- staff, suppliers and customers can share information and get in touch more easily
- your business can become more efficient - eg networked access to a common database can avoid the same data being keyed multiple times, saving time and preventing errors
- staff can deal with queries and deliver a better standard of service as a result of sharing customer data
- Cost benefits of computer networking

Storing information in one centralised database can also help you reduce costs **and** drive efficiency**. For example:**

- staff can deal with more customers in less time since they have shared access to customer and product databases
- you can centralise network administration, meaning less IT support is required
- you can cut costs through sharing of peripherals and internet access

1.4 DISADVANTAGES OF NETWORKING

- Purchasing the network cabling and file servers can be expensive.
- Managing a large network is complicated, requires training and a network manager usually needs to be employed.
- If the file server breaks down the files on the file server become inaccessible. Email might still work if it is on a separate server. The computers can still be used but are isolated.
- Viruses can spread to other computers throughout a computer network.
- There is a danger of hacking, particularly with wide area networks. Security procedures are needed to prevent such abuse, eg a firewall.

LIMITATION

- Noise can severely limit an intrusion detection system's effectiveness. Bad packets generated from software bugs, corrupt DNS data, and local packets that escaped can create a significantly high false-alarm rate.
- It is not uncommon for the number of real attacks to be far below the number of false-alarms. Number of real attacks is often so far below the number of false-alarms that the real attacks are often missed and ignored.
- Due to the nature of NIDS systems, and the need for them to analyse protocols as they are captured, NIDS systems can be susceptible to the same protocol-based attacks to which network hosts may be vulnerable. Invalid data and TCP/IP stack attacks may cause an NIDS to crash.

II. LITERATURE SURVEY

2.1 Title: Anomaly-Based Intrusion Detection System Through Feature Selection Analysis And Building Hybrid Efficient Model

Author: Shadi Aljawarneh

Efficiently detecting network intrusions requires the gathering of sensitive information. This means that one has to collect large amounts of network transactions including high details of recent network transactions. Assessments based on meta-heuristic anomaly are important in the intrusion related network transaction data's exploratory analysis. These assessments are needed to make and deliver predictions related to the intrusion possibility based on the available attribute details that are involved in the network transaction. We were able to utilize the NSL-KDD data set, the binary and multiclass problem with a 20% testing dataset. This paper develops a new hybrid model that can be used to estimate the intrusion scope threshold degree based on the network transaction data's optimal features that were made available for training. The experimental results revealed that the hybrid approach had a significant effect on the minimisation of the computational and time complexity involved when determining the feature association impact scale. The accuracy of the proposed model was measured as 99.81% and 98.56% for the binary class and multiclass NSL-KDD data sets, respectively. Intrusion detection systems (IDS) are generally divided into two types (see Fig. 1): misuse and anomaly intrusion detection systems. For a misuse IDS, instructions are identified based on parameters of system weaknesses and known attack signatures. However, it does not recognise attacks that are new or unfamiliar. On the other hand, anomaly IDS is based on normal behaviour parameters and utilizes them to pinpoint any action that deviates significantly from normal behaviour. The misuse intrusion detection mechanism identifies intrusions by matching existing intrusion patterns in consideration for examination with previously identified patterns.

DISADVANTAGES

- Accuracy rate is less

2.2 TITLE: SURVEY ON SDN BASED NETWORK INTRUSION DETECTION SYSTEM USING MACHINE LEARNING APPROACHES
AUTHOR: NASRIN SULTANA

Network Intrusion Detection systems (NIDS) have been developed rapidly in academia and industry in response to the increasing cyber-attacks against governments and commercial enterprises globally. The annual cost of cybercrime is continuously raising. Organizations can lose their intellectual property with such malicious software crept into the system which may lead to disruptions to a country’s critical national infrastructure. Organizations deploy a firewall, antivirus software, and an intrusion detection system (NIDS) to secure computer systems from unauthorised access. Software Defined Networking Technology (SDN) provides a prospect to effectively detect and monitor network security problems ascribing to the emergence of the programmable features. Recently, Machine Learning (ML) approaches have been implemented in the SDN-based Network Intrusion Detection Systems (NIDS) to protect computer networks and to overcome network security issues. A stream of advanced machine learning approaches – the deep learning technology (DL) commences to emerge in the SDN context. In this survey, we reviewed various recent works on machine learning (ML) methods that leverage SDN to implement NIDS. More specifically, we evaluated the techniques of deep learning in developing SDN-based NIDS. In the meantime, in this survey, we covered tools that can be used to develop NIDS models in SDN environment. This survey is concluded with a discussion of ongoing challenges in implementing NIDS using ML/DL and future works. Software-defined network is an emerging architecture that decouples network control and forwarding functions so that the network control can be directly programmable. The segregation of the control plane from the data plane enables easy network management.

DISADVANTAGES

- Scalability is less

III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

The IDS can be distinguished on the basis of where the detection is taking place and how or by which technique it is being detected. The IDS is classified into two niche segment one being Network Intrusion Detection System (NIDS) and the other being Host Intrusion Detection System (HIDS). The first system mentioned helps in the analysis the incoming networking traffic whereas the HIDS functioning is based on the activity of the operating system. The main aspects of data mining on IDS that were dealt with originally were termed as clustering and classification. Since there exist no label for the initial data set for clustering issue, the object created for the clustering algorithm was allocated the same class with similar data records. The behavior of the packet was termed as a normal class or abnormal class according to the features and characteristics of already existing data. In Classification, this works on mining from the already clustered data. This implies that the data is labelled. Classification is a data mining technique which is used for examining a data set. In this world of continuous streaming data, classification plays an important role in classifying the data. Many algorithms such as decision tree, rule-based induction, Bayesian network, genetic algorithm etc are used to classify the data. In existing framework implement, machine learning techniques such as Random forest, Naives Bayes, Support Vector machine algorithms are implemented to detect the intrusion from network datasets. In existing framework can be provide high false alarm and low accuracy.

3.1.1 DISADVANTAGES

- High level false positive rate
- Computational complexity is high
- Time complexity can be occurred
- Difficult to handle streaming of data
- Need hardware to detect the intrusion

3.2 PROPOSED SYSTEM

Deep learning is an emerging trend in the area of machine learning. It is sub-field of machine learning in artificial neural networks. Using deep learning approach in the application area, we can process on large amount of items in order to be trained. Process is placed on millions of data points. Deep learning is learns features from the data. If large amount of data is available, it can reduce the performance of system. For achieving better accuracy in terms of performance deep learning is well suited learning mechanism. Learning is varies in three major categories i.e. supervised, semi-supervised and unsupervised. Here, the intrusion detection is carried out with respect to the deep learning approach. Intrusion is the term that can violate security of computer system or network. And another is intrusion detection is the process to identify intrusion. Intrusion detection technique is classified in two methods i.e. anomaly detection or misuse detection. With the rapid expansion of computer networks during the past decade, security has become a crucial issue for computer systems. Different machine learning based methods have been proposed in recent years for the development of intrusion detection systems. This project presents a neural network approach to intrusion detection. A Multi-Layer Perceptron (MLP) is used for intrusion detection based on an off-line analysis approach. While most of the previous studies have focused on classification of records in one of the two general classes - normal and attack, this research aims to solve a multi class problem in which the type of attack is also detected by the neural network. MLP is a layered feed forward network typically trained with static back propagation (BP). Such networks have found their way into countless applications requiring static pattern classification. The MLP model is a flexible type of ANN composed of one input layer, one or more hidden layers, and one output layer

3.2.1 ADVANTAGES

- The success of the Deep learning architectures lies in using fast learning algorithms and efficient solutions.
- The development of GPU accelerated computing has led to the increase in their development and lead to faster convergence of the algorithms.
- Reduce the false positive rate and improve the accuracy
- Time complexity can be reduced

IV. SYSTEM IMPLEMENTATION

1. DATASETS ACQUISITION

The KDD Cup dataset, utilized for benchmarking intrusion detection issues, is used in our experiments. The dataset is a gathering of simulated crude TCP dump data over a time of 9 weeks on a LAN. The training data was processed to about 5 million connections records from seven weeks of network traffic and two weeks of testing data yielded around 2 million connection records.

2. PREPROCESSING

Data pre-processing is an important step in the [data mining] process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. Thus, the representation and quality of data is first and foremost before running an analysis. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. In this module, eliminate the irrelevant and missing values in uploaded datasets

3. FEATURES EXTRACTION

Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy. Many machine learning practitioners believe that properly optimized feature extraction is the key to effective model construction.Determining a subset of the initial features is called feature selection.The selected features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data. In this module, we can select the many attributes from pre-processed datasets.

4. CLASSIFICATION

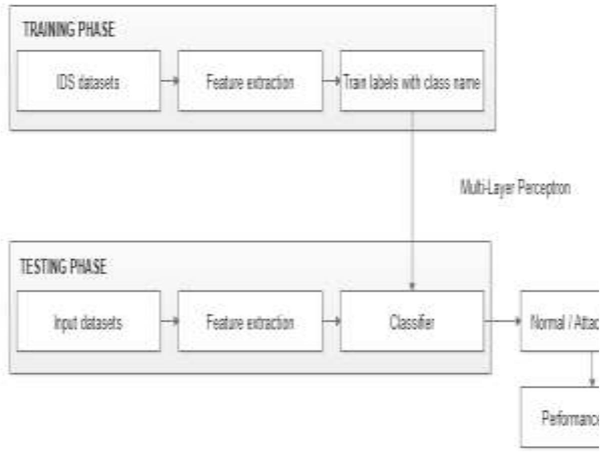
As the proliferating growth of computer network activities and sensitive information on network systems increases, more and more organizations are becoming susceptible to a wider variety of attacks. The question of how to protect network systems from intrusion, disruption, and other anomalous activities from unwanted attackers becomes paramount. In this module, implement machine learning and deep learning techniques to detect the intrusion. A multilayer perceptron (MLP) is a class of feed forward artificial neural network. An MLP consists of at least three layers of nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function.

5. PERFORMANCE EVALUATION

In this module, performance can be evaluated in terms of accuracy rate. Proposed work provide improved accuracy rate than the existing systems

6.SYSTEMDESIGN

6.1 SYSTEM ARCHITECTURE



V. SOFTWARE DESCRIPTION

FRONT END: PYTHON

Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. In July 2018, Van Rossum stepped down as the leader in the language community. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library. Python interpreters are available for many operating systems. CPython, the reference implementation of Python, is open source software and has a community-based development model, as do nearly all of Python's other implementations. Python and CPython are managed by the non-profit Python Software Foundation. Rather than having all of its functionality built into its core, Python was designed to be highly extensible. This compact modularity has made it particularly popular as a means of adding programmable interfaces to existing applications. Van Rossum's vision of a small core language with a large standard library and easily extensible interpreter stemmed from his frustrations with ABC, which espoused the opposite approach. While offering choice in coding methodology, the Python philosophy rejects exuberant syntax (such as that of Perl) in favor of a simpler, less-cluttered grammar. As Alex Martelli put it: "To describe something as 'clever' is not considered a compliment in the Python culture."Python's philosophy rejects the Perl "there is more than one way to do it" approach to language design in favour of "there should be one—and preferably only one—obvious way to do it".

VI. CONCLUSION AND FUTURE ENHANCEMENT

Intrusion detection plays an important role in the network security as the applications and their behaviour are changing day to day. Network intrusion detection has extensively researched in recent years and many techniques have been proposed including machine learning and deep learning techniques. As a result there increased the need for accurate classification of the network flows. Here we have proposed deep learning model using Multi-layer perceptron with feature selection for the accurate classification of intrusion detection. In this project we have demonstrated the construction of a lightweight neural network capable of real-time network intrusion detection. In the process, we have also provided greater insight into methodologies used by different classification schemes. We discussed potential procedures for both data processing and optimization which are generalizable to other supervised machine learning methods. We also outlined a fast method of identifying key attributes in the neural network based on the connection weights. And compared the deep learning algorithm (MLP) with BPN, FNN and RNN algorithm. Comparison done based Error metrics and Accuracy metrics. From the above comparison, MLP can be provided less error metrics and highest accuracy than the existing machine learning algorithms

FUTURE ENHANCEMENTS

In future we can also use intrusion detection using advanced neural network algorithm such as convolutional neural network using MATLAB Toolbox to improve the accuracy in intrusion detection and also reduce the false alarm rate.

REFERENCES

[1]. Aljawarneh, Shadi, MontherAldwairi, and MuneerBaniYassein. "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model." Journal of Computational Science 25 (2018): 152-160.

[2]. Sultana, Nasrin, et al. "Survey on SDN based network intrusion detection system using machine learning approaches." Peer-to-Peer Networking and Applications 12.2 (2019): 493-501.

[3]. Peng, Kai, Victor CM Leung, and Qingjia Huang. "Clustering approach based on mini batch kmeans for intrusion detection system over big data." IEEE Access 6 (2018): 11897-11906.

[4]. Farahnakian, Fahimeh, and JukkaHeikkonen. "A deep auto-encoder based approach for intrusion detection system." 2018 20th International Conference on Advanced Communication Technology (ICACT).IEEE, 2018.