A Review on Intrusion Detection using Deep Learning

Ashwini Vikas Solanke[1], Miss. Sweta Pandey[2]
[1]*M.E Student, SSBT's COET*
[2]*Assistant Professor, SSBT's COET*
[1]aashurajput236@gmail.com, [2]shwetapandey806@gmail.com

Abstract

In the field of information system, security is vital issue that arises because of huge amount internet traffic. For that purpose, many of the security techniques is used for avoiding threats. Intrusion detection is one of the techniques used for providing security to computer system. Intrusion detection technique is equivalent to binary and multi-class classification problem that check behavior of network, whether it is normal or anomalous. Here, various methods used for improving the performance of training and testing time. The experimental results of intrusion detection are to improve precision, recall, accuracy, detection rate and efficiency of system. Many of the researchers have worked on different techniques of deep learning for intrusion detection. Various learning mechanism are used for detecting intrusion in the system.

Keywords: *Deep Learning, Intrusion Detection System*

1. Introduction

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. These two methods are described briefly below:

1) Misuse Detection: It is also known as the signature-based detection. Here, the behavior of user is compared with existing pattern. But the problem with this technique is that only the known attack. Intrusion detection based on signature is not suitable for real time applications.

2) Anomaly Detection: It is relating to the normal behavior of user. If any action should pretend to be different with respect to the normal behavior of user or system. The anomaly detection consists of two different categories: Threshold- based and profile-based. In the threshold-based detection, it counts the number of occurrences of any event with some time interval.

In this process, intrusion detection system can check the behavior of network on the basis of binary classification and on the multiclass classification, means that identify whether it is normal or anomalous. The multiclass classification is categorized in following different category. They are as follows: It is either normal or from other category i.e. Denial-of-Service, User to Root, Root to Local and Probing.

2. Literature Survey

J. A. Khan et al. [1] described an intrusion detection systems and classification techniques. They present a brief idea on different intrusion detection techniques and compared them mathematically to give a essence of the best technique used for intrusion detection. They survey on applied classification techniques for ID such as, Support Vector Machine, Kernelized support vector machine, Extreme Learning Machine and Kernelized algorithm.

Y. LeCun et al. [2] described deep learning approaches. In that they proposed various deep learning approaches for checking the efficiency of different techniques. They also discussed about the deep learning applications and limitations by providing a short review.

Nathan Shone et al. [3] presented a Deep Learning approach to network intrusion detection specifically on Non-symmetric deep auto-encoder for unsupervised feature learning. The researchers can overcome some of the limitations of shallow learning. They can use KDD Cup'99 and NSL-KDD datasets. This approach offers high levels of accuracy, precision and recall together with reduced training time. They have to compare stacked NDAE model against the mainstream Deep Belief Network technique. The proposed novel classification model constructed from stacked NDAEs and the Random Forest classification algorithm.

Yuchen Liu et al. [4] described an Intrusion detection algorithm based on convolutional neural network using supervised feature learning. They can use KDD Cup 1999 dataset. Compared with other IDS classifiers, it has a highest detection rate and precision. The feasibility of applying convolutional neural network in highly –intruded detection has been proved. But the problem with this technique is that false alarm rate can't be improved.

Chuanlong Yin et al. [5] presented recurrent neural network for intrusion detection. They can use supervised learning feature of the deep network. The researchers have worked on recurrent neural network - Intrusion detection system model. And it has not only a strong modeling ability for intrusion detection, but also has high accuracy in both binary and multiclass classification. Compared with traditional classification methods, such as J48, naive Bayesian and random forest, the performance obtains a higher accuracy rate and detection rate with low false positive rate, especially for the task of multiclass classification on the NSL-KDD dataset. The researchers' can effectively improve both the accuracy of intrusion detection and the ability to recognize the intrusion type. But only the problem should arise in this model is that gradients are more occurring in the calculation of weight. So this problem can remove using LSTMs RNN technique effectively.

Muhamad Erza Aminento et al. [6] described deep learning approach for feature selection. This approach is worked on KDD99 dataset. Various machine learning algorithms are use for detecting intrusions in the network. But, the problem with this approach is that availability of important input feature. To overcome the problem of input features, the stack auto-encoder classifier is used as a feature selection algorithm. Using this approach, only the important input features are select from the huge number of input features.

Tuan A Tang et al. [7] presented a Network based Anomaly Detection using Deep Neural Network for SDN environment. This approach is worked with flow-based detection on NSL-KDD dataset. In that, they used only six basic features from the 41 features of NSL-KDD dataset.

Shan Sun et al. [8] described an intrusion detection based on Deep Belief Network. They work on feature selection factor. They can solve the problem of feature selection by using deep belief network algorithm. Because of this algorithm, detection accuracy of intrusion is performing better and reduces high missing rate in the traditional detection system. Because of this missing rate of data and dimension can accelerate the speed of data.

Gholam Reza Zargar et al. [9] described effective network parameters in attacks for intrusion detection. They can work on selection of effective network parameters to identify intrusion in the network. They use DARPA1998 dataset. Because of the effective selection of parameters training and testing time of detection system is reduce to maintain accuracy of detection in specified tolerable range. The problem of huge amount of data is flow on the internet that makes traffic, so the intrusion detection is impossible when we can work real time data set. For that purpose they use principle component analysis method used as the dimension reduction technique.

3. Methodology and Architecture

3.1 Hierarchy of Deep Network

Various approaches are used in deep learning for intrusion detection are shown in below fig.1

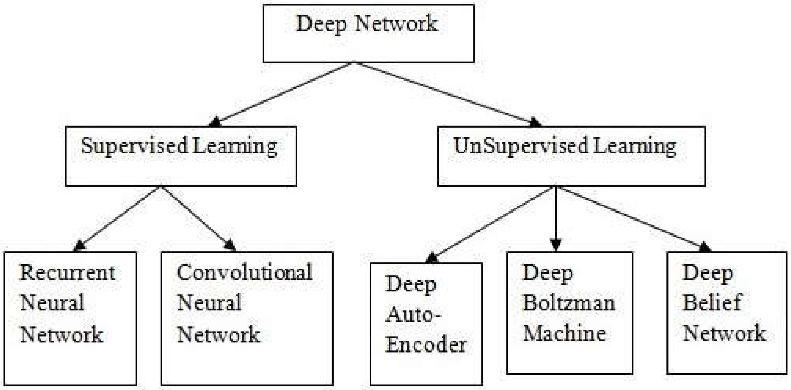

Figure 1. Hierarchy of Deep Network

All DL algorithms are based on Deep Neural Network (DNN), which are large networks organizes in many layers capable of representation learning.

1. Unsupervised Learning

- Deep Belief Network- It is modeled through composition of Restricted Boltzmann Machine, a class of neural networks with no output layer. DBN can be successfully used for pre-training tasks. Because they will exceptionally good at in the function of feature extraction. They require a training phase, but for non-labeled datasets.
- Stacked Auto-encoder- It is composed of multiple auto-encoders, where number of input and output neurons is same. It contains pre-training tasks similar to DBN. This presents better results on small datasets.

2. Supervised Learning

- Recurrent Neural Network- In the RNN, its neurons can send their output to previous layer. Recurrent neural networks can have their own internal memory. The name recurrent suggests that the same task is performing repeatedly for each sequence of output to the next of input in the computation.
- Convolutional Neural Network- In the CNN, where each neuron receives its input only from subset of neurons of the previous layers. Because of these characteristics makes CNN effective at analyzing spatial data, but sometimes their

performance decreases when applied to non-spatial data. It have a lower computation cost.

3.2 Architecture of Intrusion Detection using Deep Learning

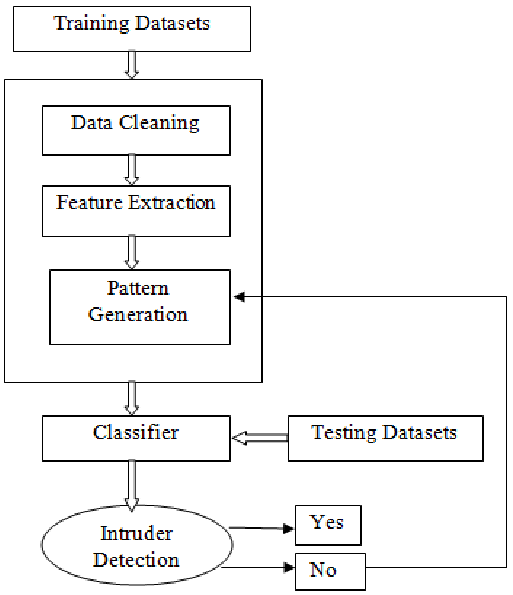

Figure 2. Architecture of Intrusion Detection using Deep Learning

Processing Steps**:** The above fig.2 shows that some processing steps. According to these preprocessing steps, it reduces some inflectional data from available training records. The brief discussion is carried out in following steps.

- Firstly, here the term data cleaning can express the redundancies in the datasets. Because of data cleaning we can reduces the noisy, unnecessary or unwanted data from the existing datasets.
- Then, feature extraction or the attribute selection is performing on datasets. In the feature selection, only relevant attribute should be select that are use for construction of model. For example, in the intrusion detection protocol_type, service etc.
- In the pattern generation phase, specific pattern is generated according to the incoming intruder. Then it will check the existing pattern with newly generated pattern.
- Using classifiers test that datasets and check for intrusion over a network for anomalous activity. If any abnormal situation is happens at that time intruder is detected. Else intrusion is not detected then it will go return to the pattern generation process. And process repeats until, intrusion detection done successfully.

3.3 Attacks Classes

In the attacks classes there are four categories use for intruder. These are follows:
1. Denial-of-Service (DOS) - Denial-of-Service attack can use up complete resources or reserves of target system and stop servers from accepting to provide services. For e.g. Syn flooding, land etc.
2. Remote to local (R2L) - This attack can allow unauthorized remote access. Means attacks inhibits into remote machine and gain access of that machine. For e.g. Password guessing, imap etc.
3. User to Root (U2R) - This attack can try to acquire superuser permission. Attacker uses normal account to login into harmed system and tries to gain root privileges by exploiting some vulnerability in the victim. For e.g. Buffer-overflow.
4. Probing (Probe) - This attack only gain information about remote victim. For e.g. port scanning.

3.4 Performance Parameters

According to the raw metrics, we can visualize the performance parameters. These are following:

- True Positive (TP): It is correctly rejected. It consists of the number of anomaly record that is identified as an anomaly.
- False Positive (FP): It is incorrectly rejected. It consists numbers of records are identified as an anomaly.
- True Negative (TN): It is correctly admitted. It consists numbers of normal records are identified as normal.
- False Negative (FN): It denotes incorrectly admitted. It consists number of anomaly records are identified as normal.

Based on above metrics, the equation of measurement is follows in different terms. Here, the following measures are used to evaluate the performance.
1. Accuracy: The percentage of number of correct records versus total records.

$$\text{Accuracy}= (TP+TN)/ (TP+TN+FP+FN)$$

2. Precision: It measures number of correct records penalized by number of incorrect records.

$$\text{Precision}= TP/ (TP+FP)$$

3. Recall: It measures number of correct records as a number of missed entries.

$$\text{Recall}= TP/ (TP+FN)$$

4. False Alarm Rate: Number of normal patterns classified as attack divided by total number normal patterns.

$$FAR= FP/ (FP+TN)$$

4. Conclusion

Intrusion detection plays an important role in information security. Using deep learning mechanism, we not only improve the performance of system but also reduce some kind of redundancies in the dimension of data. From that we can achieve higher accuracy and detection rate with low false positive rate. From the review of paper using different learning mechanism we can easily detect multiclass classification problem with respect to their category.

ACKNOWLEDGMENT

First and foremost, I would especially like to thank my adviser and guide Miss. Sweta Pandey, for being a wonderful instructor and always being in bright spirits. I am incredibly grateful to Principal K. S. Wani, Head of the department Prof. Dr. Girish K. Patnaik and the Professors in the M.E. Computer Engineering Program at SSBT's COET, Jalgaon. Without their time and support, I would not have been able to complete this work. Last, but not least, I like to thank my family and friends who also supported me throughout my studies.

5. References

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, *Nature*, vol. 521, no. 7553, pp. 436_444, May 2015.

[2] J. A. Khan and N. Jain, “A survey on intrusion detection systems and classification techniques”, Int. J. Sci. Res. Sci., Eng. Technol., vol. 2, no. 5, pp. 202208, 2016.

[3] Shone, N, Tran Nguyen, N, Vu Dinh, P and Shi,Q, “A Deep Leaning Approach to Network Intrusion Detection”,IEEE Transactions on EmergingTopics in Computational Intelligence,2(1)2018.

[4] Yuchen Liu, Shengli Liu & Xing Zhao, “Intrusion Detection Algorithm Based on Convolutional Neural Network”, 4th International Conference on Engineering Technology and Application (ICETA 2017).

[5] Chuanlong Yin , Yuefei Zhu, Jinlong Fei, And Xinzheng He, “A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks”, volume 5,Nov 2017.

[6] Muhamad Erza Aminanto, Kwang Jo Kim, “Deep Learning-based Feature Selection for Intrusion Detection System in Transport Layer”, Summer Conference of Korea Information Security Society (CISC-S'16), June 26, 2014. National Pukyong National University, Busan.

[7] Tuan A Tang, Lotfi Mhamdi, Des McLernon, Syed Ali Raza Zaidi and Mounir Ghogho,“Deep Learning Approach for Network Intrusion Detection in Software Defined Networking”, International Conference on Wireless Networks and Mobile Communications (WINCOM). International Conference on Wireless Networks and Mobile Communications (WINCOM'16), 26-29 Oct 2016.

[8] Peyman Kabiri and Gholam Reza Zarger , “Category-Based Selection of Effective Parameters for Intrusion Detection”, IJCSNS International Journal of Computer Science and Science and Network Security, VOL .9 No.9, September 2009.

[9] BaoyiWang, Shan Sun and Shaomin Zhang, “Research on Feature Method of Intrusion Detection Based on Deep Belief Network”, 3rd International Conference on Machinery, Material and Information Applications 2015.