An IDS For Attacks Classification Using Recurrent Neural Network For Classification

Suresh Kumar Rulaniya
Computer Science
sobhasaria engineering college, sikar
Rajasthan Technical University
Sikar, India
sureshrulaniya@gmail.com

Gourav Mitawa
computer science
sobhasaria engineering college, sikar
Rajasthan Technical University
Sikar, India

***Abstract*— In modern life, networks play a significant role, and cybersecurity has been a core field of study. A cyber-security-important intrusion detection system (IDS) monitors the network's state of software & hardware. Despite decades of growth, there are still challenges for current IDSs to boost detector accuracy, reduce false alarm rates, and detect unknown attacks. Many researchers have been working to create IDSs that force machine learning (ML) methods to solve the above problems. The critical variations between normal data or abnormal data can be found through machine learning procedures with high precision. Deep learning (DL) is an area that has remarkable performance and has become a center of research. In this research, we have used an efficient Neural Network method as the taxonomy of the IDS attacks. The results are remarkable with Recurrent Neural Network (RNN) on the NSL-KDD dataset using python as the simulation tool.**

Keywords*— *Network Security; Intrusion Detection System Machine Learning; RNN; NSL-KDD.

I. INTRODUCTION

Over the years, people have relied on technology or their computer networks[1] for their regular tasks, such as messaging or shopping. These networks continuously face numerous online attacks, and so the privacy and functionality of these networks should be secured from violation and intrusion. Unknown intruders, extremist groups, or opposing firms have the opportunity or potential to perform complex features on information networks, making networks protection a big concern for many researchers. Network intrusion or attack means an unauthorized user (attacker) attempts to exploit the system's vulnerability to access network resources or cause functional network disruption [2]. If the attacker gets network resources, it can have illegal access to the network data, alter or corrupt the system and mislead or change the minds of the network.

Much effort was made to detect network interference automatically through the capabilities of network control software. The IDS manage a vast number of data & plays a strategic role in identifying multiple threats. Intrusion detection is based on a good classifier, which considers IDS as a classification challenge. The intrusion may be defined as malicious and is a gateway to compromising computer device access, credibility, and confidentiality. Classification models are particularly important techniques for intrusion detection in data mining [3].

NSL-KDD Dataset:

In this analysis, the data used is the data collection NSL-KDD Cup 99. NSL-KDD in many studies is a proposed dataset to resolve KDD Cup 1999 (KDD99) issue. The dataset KDD-99 is 15 years old, but it is still widely applied in the field of IDS for academic purposes as the public can openly access a shortage of data sets. Any issues in KDD-99 data collection, including the deletion of redundant data & merging of datasets, have been solved in NSL-KDD now [4].

Network layers attacking can be classified into four major categories to review of NSL-KDD CUP 99 Data Set:

1. Probing Attack: The offender tries to create data for the target host.

2. Denial of Service Attack (DoS): The prosecutor attempts to avoid using a facility by legitimate use.

3. User to Root Attack (U2R): Perpetrators have indigenous keys to the machine of the victim or work hard to achieve user access benefits.

4. Remote to Local Attack (R2L): offender has no Associate in the victim's nursing account, thus tries to achieve access.

II. LITERATURE REVIEW

This research [5] reveals that ANN-based ML with wrapper features selection (FS) outperforms an SVM technique during network traffic classification. NSL-KDD is used to test performance using Support vector machine (SVM) & Artificial Neural Network (ANN) ML strategies for classifying network traffic. A comparative study reveals that the model proposed is successful in terms of intrusion detection success rate than other current models.

In this paper [6], the author is automatically able to build or detect the distribution of the usual packet payload based on ML techniques, k-means clustering algos, and an alternative way of SVM classification algo. Our approach demonstrates the proposed hybrid algo that gives the most used Snort Open Source system significant detection precision.

Additionally, in the current paper [7], we present an unsupervised method of AD, combining Sub-Space Clustering (SSC) or one class SVM (OCSVM). The method suggested is tested using a well-known data set of NSL-KDDs. The test

results suggest that our approach functions better than any of the current strategies.

Amoli et al. [8] proposed a high-speed network IDS unsupervised to detect attacks by zero days on two different engines. 1st engine observed attacks on DBSCAN clustering, and the second engine found botnet under various protocols. Two publicly available datasets were used for the evaluation of the proposed model. The model proposed and then tested and contrasted with the outlining approaches centered on the K-means & DBSCAN.

This paper [9] used KDD data set to train unsupervised algo of ML, Isolation Tree. The database is extremely imbalance in its absence but includes a wide variety of attacks, including DOS, Test, U2R, and R2L. This data set is inconsistent, which creates a class imbalance. This AD model for isolation forest was used to identify outliers but likely attacks using the anomaly score.

The focus of this paper [10] is on the efficient relationship between the DBN Algo & SPELM Algorithm of intruder detection methods. Scholar applied the proposed State Preserving Extreme Learning Machine (SPELM) algo as a classifier for ML but compared its output to the NSL KDD data set with Deep Belief Network (DBN). SPELM has demonstrated improved performance compared with 52.8 percent DBN; 69.492 SPELM precision compared to 66.836 DBN or 90.8 seconds SPELM processing time compared to 102 seconds DBN algo. The results indicated better performance of SPELM.

This work [11] indicates that the ELM is one of the common ML algos that can be easily implemented with excellent learning properties. The internal power (weight or core) parameters of ELM are therefore randomly initialized, which renders the algorithm unstable. PSO is a successful metaheuristic used for optimizing the ELM in this study. Our provided model was applied based on the NSL-KDD data set intrusion detection or validated. Our style was associated with a fundamental ELM. The simple research accuracy model PSO-ELM has outperformed.

III. RESEARCH METHODOLOGY

PROBLEM FORMULATION

One of the IDS' key challenges concerns creating useful behavioral patterns or estimates for analyzing typical behavior from abnormal behaviors. To overcome this issue, the IDSs previously used by security experts are widely used for data analysis or ML-based IDS formulation after all data volume, raise variable quickly; it's now a time-consuming or annoying challenge for ML to evaluate or extract attack signatures or identification rules from complex data or massive network data volumes.

PROPOSED APPROACH

The preprocessing, classification, or evaluation of the data set are the key phases of the model proposed. Each process is critical in the proposed method, which improves its efficiency. SVM and RNN are used as the methodologies where RNN is the proposed methodology, which is compared with the SVM, which was used in the existing research work.

A. Pre-Processing

The dataset includes symbolic features that cannot be processed by the classification. Therefore, there is a preprocessing. All non-numeric or symbolic characteristics are deleted or swapped in this stage. Non-numeric or symbolic characteristics are eliminated or replaced in the preprocessing step. The overall preprocessing process is important in order to delete or modify non-numeric or symbolic elements since they are of no direct importance in intrusion detection. Adjust or delete generic features such as protocol, service, and flag. In addition, the cases are classified as normal, DoS, Probe, and R2L.

B. Methodology

- A comparative comparison was made between SVM & RNN to evaluate their accuracy & misclassification rate for data set classification. The primary raw data set includes 24 forms of attack classified under four groups, and the class attribute includes them. Dos, Study, R2L are regular.

- Preprocessing labeling is then performed to transform the nominal value into binary. To achieve better performance, non-numeric features are omitted from the IDS.
- Methodologies such as CfsSubsetEval are used to obtain contrasting outcomes and to enhance the dataset efficiency. The given data set will be reduced and normalized after preprocessing.

- One method of attribute selection is CfsSubsetEval. It measures the value of attributes by considering the estimate of each function and the redundancy rate among them.

- The raw data set using SVM, SVM or SVM in different Normalization techniques or feature reduction was listed for the first 19,000 instances. The level of accuracy or Misclassification is also noted.

RNN: RNNs were built. In its design, it varies from other ANNs. During the feed-in loop or backpropagation loop, the RN "fly" in a linear direction. Simultaneously, the other networks overcome the recurrence relation and therefore use Back Propagation to learn over time.

RNN consists of several fixed function units, one per phase. Every unit has an internal state called the unit's hidden state (HS). This HS indicates the prior experience of the network at a given time. This HS will be revised to reflect a change in network knowledge of the past at all times. HSe is efficient with the repetition relationship:-

$$h_t = f_w(x_t, h_{t-1}) \quad (1)$$

h_t : new HS
f_w : old HS
x_t : current input
h_{t-1} : fixed function with trainable weights

The new HS is calculated at every step using the above recurrence ratio. This current HS is used to create a new HS.

RNN is a type of NN that provides an entry to the current stage to the output from the preceding example. Traditional NNs do not have both inputs and outputs, so whenever it is important to determine the next word of a sentence, the previous words are needed, so previous words have to be recalled. RNN was created that solved this problem by the support of the hidden layer (HL).

RNN achieves this:

- By supplying all layers with equal weights or preferences, RNN transforms the independent activations in reliant activations, reduces complexity to increase parameters & stores every preceding output, give every production as an entry into the next HL

- All three layers may be linked together in a single recurrent layer such that weights or bias of all HL are equal.

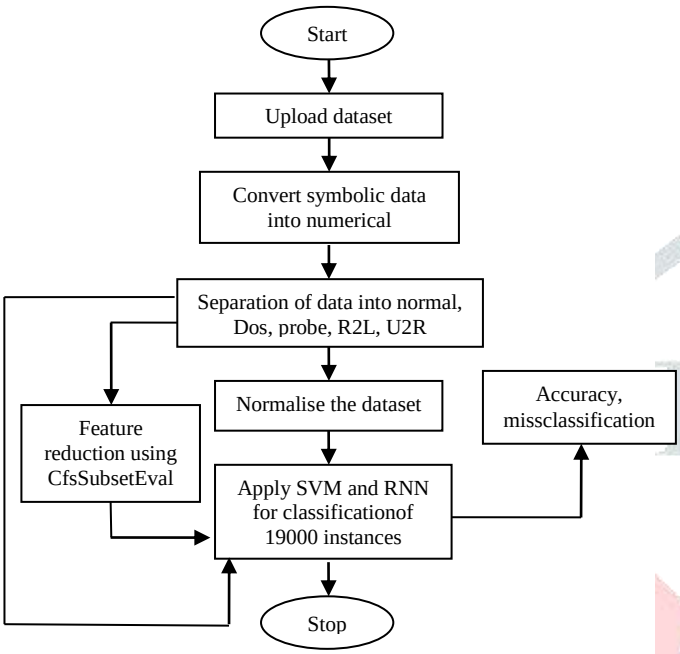

Fig 1: Data Flow Diagram of the proposed methodology

Fig 1 describes the data flow diagram of the proposed methodology, which visualizes the steps involved in the research methodology.

IV. RESULTS AND DISCUSSION

The output of SVM or NB (Navibe Bayes) algo is evaluated for 19,000 cases by assessing the precision or misclassification rate.

Evaluation

After applying methodologies like preprocessing or randomization, the model is tested based on NSL-KDD datasets. There are 19,000 samples in this dataset. Accuracy or Misclassification rate are called metrics.

Fig 2: SVM accuracy

Fig 3: RNN accuracy

The diagram above shows the contrast of the exact classification as well as the misclassification rate after preprocessing of the initial data set. It can be derived from this graph that SVM reaches 99.11 percent accuracy, or RNN reaches 97.96 percent accuracy rates for 19000 cases. SVM is high than RNN, which is particularly contentious in 19000 cases.

Fig 4: SVM after Normalization

Fig 5: RNN after Normalization

The graph above explains the classification accuracy comparison and the misclassification occurrence of the data set after normalization. The result shows that RNN hits 99.5 percent precision and that for 19000 instances, SVM achieves accurateness of 95.04. SVM has the RNN or 19000 cases high misclassification rate. For 19000 cases, the accuracy rate was decreased.

Fig 6: SVM after CfsSubsetEval (FS)

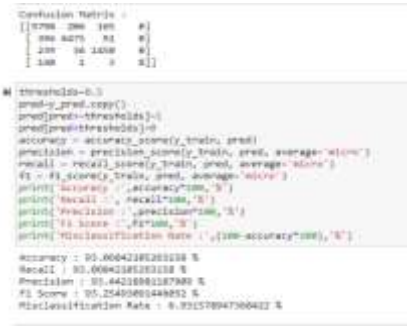

Fig 7: RNN with CfsSubsetEval for FS

The above graph shows how to compare the accuracy of classification with the error rate of the data set after the reduction of the function. The diagram will deduce that RNN achieves 93.06 percentage accuracy and that for 19000 cases, SVM achieves a precision rate of 86.37. The SVM output for 19000 instances is high than the RNN.

V. CONCLUSION

The paper proposes an IDS taxonomy to present the ML algos working in the field using data sources as the key line. We examine or discuss IDSs that have been applied to the NSL-KDD dataset based on this taxonomy. IDSs aim to detect attacks, so it is necessary to select the correct data set based on the attack property. DL models play an ever more critical role & have become an excellent direction of research. DL techniques include many deep networks to improve IDS efficiency. Using RNN as the primary approach of Deep Learning, we found that this algorithm has outperformed compared to SVM that, too, with three different conditions and reached 99% as the highest accuracy.

REFERENCES

[1] [1] G. C. Kessler, "Defenses against distributed denial of service attacks," SANS Institute, vol. 2002, 2000.View publication stats

[2] H. A. Nguyen and D. Choi, "Application of data mining to network intrusion detection: classifier selection model," in Asia-Pacific Network Operations and Management Symposium. Springer, 2008, pp. 399–408.

[3] S. Paliwal and R. Gupta, "Denial-of-service, probing & remote to user (r2l) attack detection using genetic algorithm," International Journal of Computer Applications, vol. 60, no. 19, pp. 57–62, 2012.

[4] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on. IEEE, 2009, pp. 1–6.

[5] Taher, K. A., Mohammed Yasin Jisan, B., & Rahman, M. M. (2019). Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection. 2019 International Conference on Robotics, Electrical, and Signal Processing Techniques (ICREST).

[6] Zhang, H., Lin, K.-Y., Chen, W., & Genyuan, L. (2019). Using Machine Learning techniques to improve Intrusion Detection Accuracy. 2019 IEEE 2nd International Conference on Knowledge Innovation and Invention (ICKII).

[7] Pu, G., Wang, L., Shen, J., & Dong, F. (2021). A hybrid unsupervised clustering-based anomaly detection method. Tsinghua Science and Technology, 26(2), 146–153.

[8] P. V. Amoli, T. Hamalainen, G. David, M. Zolotukhin, and M. Mirzamohammad, Unsupervised network intrusion detection systems zero-day fast-spreading attacks and botnets, International Journal of Digital Content Technology and Its Applications, vol. 10, no. 2, pp. 1–13, 2016.

[9] Vikram, A., & Mohana. (, 2020). Anomaly detection in Network Traffic Using Unsupervised Machine learning Approach. 2020 5th International Conference on Communication and Electronics Systems (ICCES).

[10] Singh, K., & Mathai, K. J. (2019). Performance Comparison of Intrusion Detection System Between Deep Belief Network (DBN)Algorithm and State Preserving Extreme Learning Machine (SPELM) Algorithm. 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)

[11] Ali, M. H., Fadlizolkipi, M., Firdaus, A., & Khidzir, N. Z. (2018). A hybrid Particle swarm optimization -Extreme Learning Machine Approach for Intrusion Detection System. 2018 IEEE Student Conference on Research and Development (SCOReD).

[12] Hongyu Liu, Bo Lang, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey" Appl. Sci. 2019, 9, 4396;

[13] Amar Meryem Bouabid EL Ouahidi, "Hybrid intrusion detection system using machine learning Network Security" Volume 2020, Issue 5, May 2020, Pages 8-19

[14] J. Ryan, M. Lin, and R. Miikkulainen, "Intrusion Detection with Neural Networks," Advances in Neural Information Processing Systems, NIPS conference, pp. 943-949, 1997.

[15] J. Cannady, "Artificial Neural Networks for Misuse Detection," National Information Systems Security Conference, pp. 443 – 456,1998.

[16] P. L. Nur, A.N. Zincir - Heywood and M. I. Heywood, "Host Based Intrusion Detection Using Self Organizing Maps," Proceedings of the IEEE International Joint Conference on Neural Networks, pp, 1714 – 1719, 2002.

[17] K. Labib and R. Vemuri, "NSOM: A Real-Time Network-Based Intrusion Detection System using Self Organizing Maps," 2000.

[18] FE Heba, A. Darwish, A. E. Hassanien and A. Abraham, "Principal Components Analysis and Support Vector Machine based Intrusion Detection System," ISDA, pp. 363 – 367, 2010.

[19] Fan Li, "Hybrid Neural Network Intrusion Detection System using Genetic Algorithm" 978-1-4244-7874-3/10/$26.00 ©2010 IEEE

[20] Mahdi Rabbani, Yong Li Wang, Reza Khoshkangini, Hamed Jelodar, Ruxin Zhao, Peng Hu "A hybrid machine learning approach for malicious behavior detection and recognition in cloud computing" Journal of Network and Computer Applications 2019 Published by Elsevier Ltd.