Design of Acceptance Sampling based Network Intrusion Detection system using Deep Learning Techniques

Dr. C. Siva Kumar[1]
Associate Professor *Dept. of School of Computing, Mohan Babu Univesity (ERSTWhileSree Vidyanikethan Engineering College)* Tirupati, Andhra Pradesh, India
sivakumar.c@vidyanikethan.edu

Mohd khalid[2]
Assistant Professor *Dept. of Animation, Chandigarh University Ghauan, Mohali, Punjab,* India
Khalid2088@yahoo.com

V. Sandeep Kumar Reddy[3]
Assistant Professor *Dept. of School of Computing, Mohan Babu Univesity (ERSTWhileSree Vidyanikethan Engineering College)* Tirupati, Andhra Pradesh, India
sandeepkumarreddy.v@vidyanikethan.edu

ABSTRACT*:* The goal of a NIDS is to keep abnormal network traffic out of networks. Inspections of all packets in network flows with the purpose of detecting malicious packets and, as a result, aberrant flows, are prohibitively expensive. To avoid a complete inspection, Acceptance Sampling can be used. A randomly selected sample of packets from a network flow is examined to see if it contains any anomalies. Using the Acceptance Sampling method, it reduces the computing work by a factor of ten. LSTM and CNN are the Deep Learning techniques used in this work. Acceptance sampling had a 70 percent accuracy for network intrusion detection, but utilizing deep learning techniques, the accuracy was boosted to 88 percent.

Keywords: Acceptance Sampling, Sequential Search, Sample Size, Convolutional Neural Network, Long Short- Term Memory.

I. INTRODUCTION

Cybercrime has increased as the internet has grown rapidly. For secured networks, traditional technologies such as firewalls and authentication systems provide a protective layer. They are never the less vulnerable to DOS attacks and probing attacks. Network intrusion detection systems (NIDS) help to ensure that networks are secure. The NIDS scans incoming packets for harmful patterns and alerts the network administrator if any are found. Because they allow network managers to detect policy violations.

Security systems now include NIDS. By keeping track of network flows at the network layer's fragment level, NIDS defends networks connected to the internet from malicious assaults. Inspection of all flows and each fragment is computationally prohibitive due to the massive number of flows on high-capacity networks.

II. RELATED WORK

To identify assaults, intrusion detection systems (IDSs) compare collected data to recognized signatures known to be malicious (misuse-based IDSs) or a model of legitimate activity (model-based IDSs) (anomaly-based IDSs). Bayesian networks assist in the aggregation of many model outputs and the smooth in corporation of additional input. We employed Bayesian networks to improve on the classic navie threshold-based approaches for combining model results. The challenge of establishing strong models of acceptable behavior in anomaly-based approaches can lead to a substantial number of false alarms.

Using "k-Means +c4.5," a way to cascade k-means clustering, and the C4.5 decision tree approaches, a method for distinguishing anomalous and normal in a computer network.

k-Means + c4.5, a technique for cascading k-means clustering, and C4.5 decision tree techniques, a technique for identifying abnormal and typical behaviour in a computer network. The proposed approach is used to the supervised data set to find abnormalities. In the experiment results, the suggested approach provides excellent detection accuracy. As the number of records

grows, the time complexity grows as well. This isn't the best method for dealing with continuous variables.

Many NIDS have been described in the literature. The detection technique and assessment datasets were mentioned as two well-known criteria for categorizing and evaluating NIDSs in the discussion. We categorize existing network anomaly detection methods and systems based on the underlying computational approaches used. Anomaly detection's main drawback is that it can be daunting and complicated. xIn accordance with the underlying computational strategies employed, we classify the current network anomaly detection techniques and systems. The biggest disadvantage of anomaly detection is that it can be intimidating and challenging.

The number of real-world and benchmark data sets for network flows that may be used to evaluate the performance of ANIDSs is restricted. The lack of assault information limits the utility of such data sets. Synthetic data sets for network flows at the fragment levels are required. Only False Positive Rates (FPR) and True Positive Rates are taken into account by the Receiver Operating Characteristic (ROC) curve (TPR). Type-I and Type-II error reduction are not taken into consideration.

III. PROBLEM STATEMENT

The challenge statement is to use deep learning techniques to design a system for detecting abnormal flows with the competing goals of optimizing classification accuracy while minimizing computational effort.

In recent years, deep learning algorithms have proven to be beneficial for this function. The primary purpose of this study is to develop a cost-effective network intrusion detection system. We test three methods to determine which one is the most successful and efficient at detecting unusual flows.

We employed SLIQ with acceptance sampling, Convolutional Neural Networks, and Long Short Memory in this experiment. For the implementation of these algorithms, we used Google Colab.

IV. THE EXISTING SYSTEM

The existing system schemes will be presented in this section

It is self-evident that a complete assessment of network flow fragments ensures 100 percent accuracy in detecting unusual flows. However, the prohibitive computing effort has a negative impact on application response time. The selective sampling method is only useful for identifying port scan and host scan attacks on modest network flows.

As a result, the applicability of NIDS using acceptance sampling is thought to be worthwhile. Table 1 lists the acceptance sampling parameters for statistical quality control and network intrusion detection.

Parameter	Notation	NIDS
Lot Size	N	Number of fragments in a network flow
Sample size	n	Number of randomly chosen fragments for inspection
Acceptance number	c	Threshold of malicious fragments in a sample ought to be zero for a normal flow. Otherwise, the flow is detected is anomalous
% defective/anomalous	p	Percent malicious fragments in a network flow submitted for inspection

The following are the steps in the overall logic of acceptance sampling for detecting anomalous flows:

1. Assume that the network flow in question is a normal flow at first.

2. Select a random sample of size 'n' fragments from the network flow.

3. Examine the first / second piece.

4. If the fragment is found to be malicious, mark the flow as anomalous and terminate it.

5. If the next fragment is available, proceed to step 3.

V. MODULE DESCRIPTION

A NIDS keeps track of network traffic patterns to look for unusual behavior. At key checkpoints, such as the DMZ or behind a firewall, sensors are positioned, and every packet—both inbound and outbound—is scrutinised for malicious behavior. The location of the sensors must be carefully considered in order to provide them the best visibility. Despite the fact that a single

sensor can keep an eye on numerous hosts, the volume of traffic that passes between all network devices may need the deployment of multiple NIDS.

The administrator will look into it if the NIDS finds any anomalous traffic. Unexpected behavior can be detected by denial of service assaults on the network, port scanning, or a rapid rise in network traffic.

The first detects signals of known assaults, while the second searches for irregularities in normal behaviour.

A. Signature based Intrusion Detection System:

Network traffic is monitored by signature-based IDS, which makes an effort to compare it to a database of known IOCs (Indicators Of Compromise). The system administrator will be notified if any traffic activity matches a known attack signature, such as a malicious site, specialized network attack behavior, known malicious IP address, or email subject line. A major weakness of signature-based NIDS is that bad actors never rest and are constantly attempting to stay one step ahead of the game. A list of known indicators of compromise must be added to the signature database on a regular basis. Cybercriminals can potentially circumvent signature based IDS detection by changing threat filtration patterns or encrypting data.

B. Anomaly based Intrusion Detection System:

Signature-based NIDS differ from anomaly-based NIDS in that they operate in a different manner. Instead of searching for a known signature, it analyses network activity and employs machine learning and artificial intelligence to identify what "normal traffic" is through statistical analysis. It can more quickly recognize abnormal conduct and deliver a report once it has learned what constitutes typical activity.

Signature-based NIDS are more reliable because they base potential threats on known signatures. They generate less erroneous positive results. The advantage of anomaly-based NIDS, on the other hand, is their capacity to identify unknown threats, such as zero-day assaults, which signature-based systems would be unable to do. The majority of NIDS combine anomaly-based and signature-based detection to form a complete system.

C. Performance Evaluation:

In order to evaluate the effectiveness of the Acceptance Sampling Network Intrusion Detection (ASNID) technique, C. Madhusudhanrao and M. Naidu developed the Geometric Mean Accuracy Index. The geometric mean of True Positive Rate (TPR) and True Negative Rate (TNR) is used to determine the Global Mean Absolute Improvement (GMAI) (TNR).

GMAI = sqrt (TPR*TNR)

TPR is the proportion of correctly recognized anomalous flows, and TNR is the fraction of correctly detected normal flows.

VI. ALGORITHMS AND TECHNIQUES

We are using Deep Learning approaches to implement acceptance sampling for NIDS in this study. LSTM and CNN are the deep learning algorithms used. When compared to acceptance sampling based SLIQ, these deep learning techniques enable us to attain higher accuracy.

CNNs, or Convolutional neural networks, are one of the most promising ways for creating machine learning models. We require already categorized data for CNN to do text classification. As a result, we'll use the text as input and examine the data before assigning labels to it.

An artificial neural network, or LSTM, is a type of artificial neural network used in artificial intelligence and deep learning. The LSTM algorithm is ideal for classification, processing, and prediction. Labels 0 and 1 should be assigned to the data set. If we obtain label 0 for a certain data set, it means the packet is legitimate, and if we get label 1, it means the packet is malicious. We can detect normal and malicious packets by categorizing the data into label 0 and 1 and then using the LSTM algorithm to determine the accuracy for the given data set.

VII. RESULTS

Comparision graphs

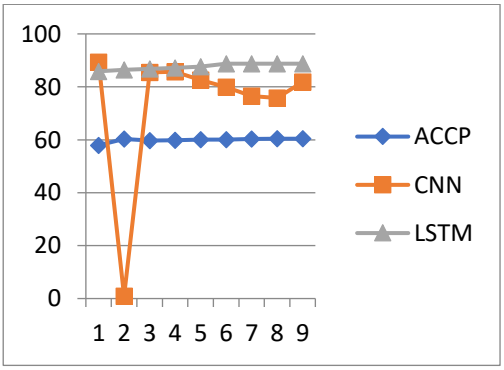

Fig 1: Comparison graph for 10% Anamolous packets

Here for 10% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

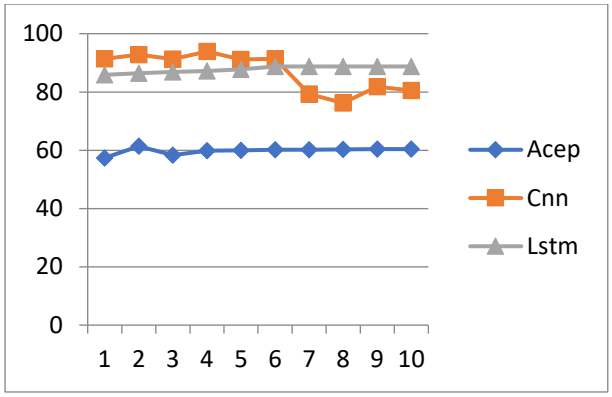

Fig2: Comparison graph for 20% Anamolous packets

Here for 20% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

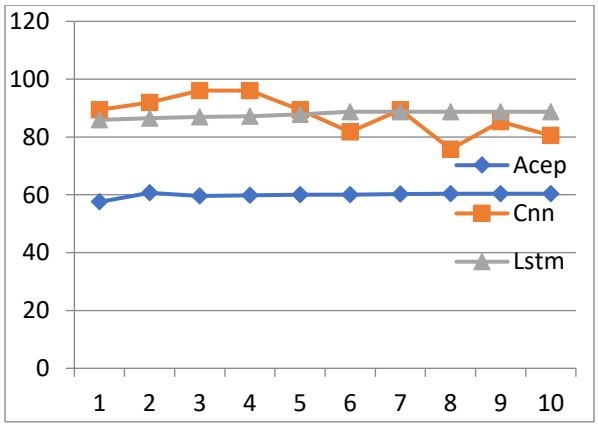

Fig3: Comparison graph for 30% Anamolous packets

Here for 30% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

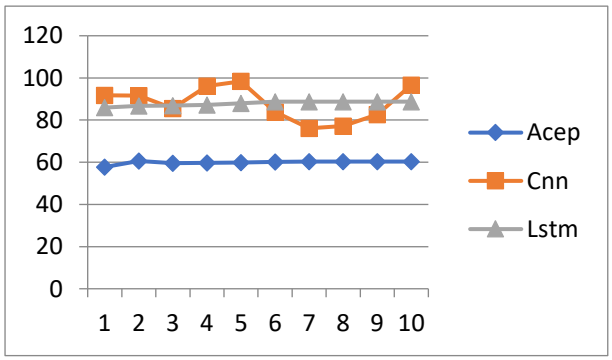

Fig4: Comparison graph for 40% Anamolous packets

Here for 40% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

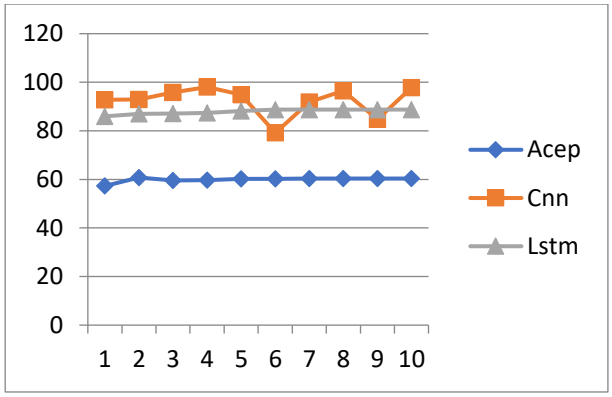

Fig5: Comparison graph for 50% Anamolous packets

Here for 50% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

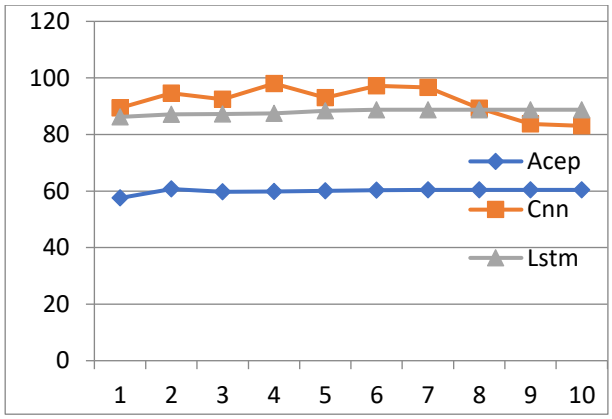

Fig6: Comparison graph for 60% Anamolous packets

Here for 60% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

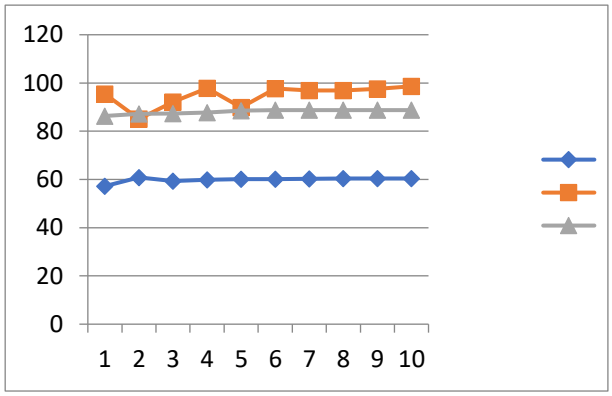

Fig7: Comparison graph for 70% Anamolous packets

Here for 70% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

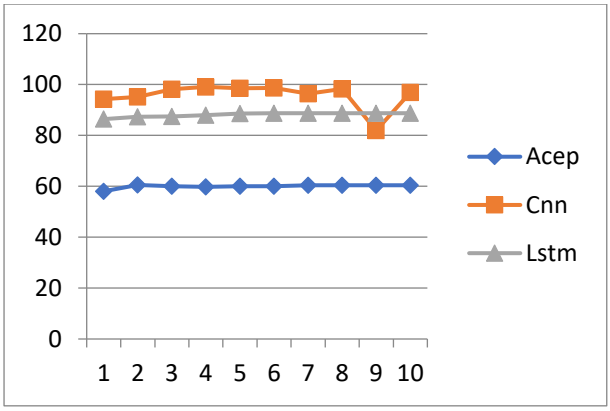

Fig8: Comparison graph for 80% Anamolous packets

Here for 80% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

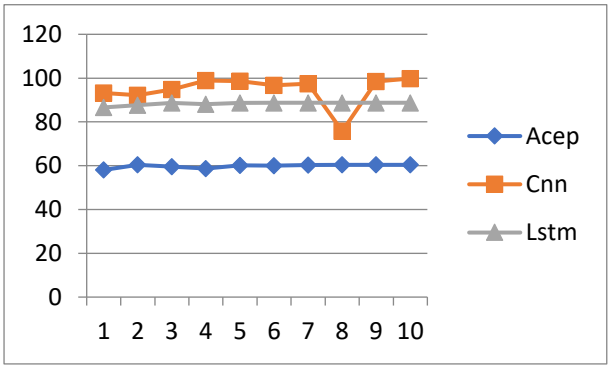

Fig9: Comparison graph for 90% Anamolous packets

Here for 90% Anamolous packets LSTM is chosen as more accurate because its accuracy is most constant than CNN and Acceptance Sampling.

VII.CONCLUSION AND FUTURE WORK

The use of Deep Learning techniques such as CNN and LSTM to identify network intrusion is proposed in this paper. Experiments have shown that the suggested LSTM plus CNN technique outperforms SLIQ. Acceptance sampling is a technique that is used to determine if When compared to other factors, accuracy has been proven to be high. Acceptance has been implemented by SLIQ. Sampling. Method for sampling when it comes to networks, detection of a breach The accuracy of acceptance sampling was 70%.

When compared to LSTM, CNN takes longer to execute. While the accuracy produced by CNN is higher than that of existing systems, the execution time is longer. As a result, this problem must be rectified in the future so that the computing effort is reduced and good results can be achieved.

VIII. REFERENCES

[1] C. Madhusudhan rao, M. M. Naidu, “A model for generating synthetic network flows and Accuracy index for evaluation of anomaly network intrusion detection systems”, Indian Journal of Science and Technology, Vol.10 No.14, 2017, pp 1-16.

[2] C. Madhusudhana rao, “Design of Optimal Acceptance Sampling Plan for Network Intrusion Detection”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019.

[3] S Jiang, X Song, H Wang, J J Han, Q H. Li, “A Clustering Based Method for Unsupervised Intrusion Detections,” Pattern Recognition Letters, Vol. 27, no. 7, 2006, pp. 802–810.

[4] N. G. Duffield, P. Haffner, B. Krishnamurthy, and H. Ringberg, “Rule-Based Anomaly Detection on IP Flows,” Proc. 28th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies. Rio de Janeiro, Brazil, April, 19-25, 2009, pp 424–432.

[5] Source fire, Inc., Snort: The Open Source Network Intrusion Detection System. http://www.snort.org, 2007.

[6] *Dr.Sivakumar. C (2022)." Dynamic Tree Routing Protocol with Convex Hull Optimization for Optimal Routing Paths”, jour of Proceedings of theInternational Conference on Electronics and Renewable Systems (ICEARS2022) IEEE Xplore Part Number: CFP22AV8-ART; ISBN: 978-1-6654-8425-1*

[7] *Dr.Sivakumar .C (2022)." An Improvised Method for Anomaly Detection in social media using Deep Learning”, jour of Proceedings of the International Conference on Electronics and Renewable Systems (ICEARS 2022) IEEEXplore Part Number: CFP22AV8-ART; ISBN: 978-1-6654-8425-1*