International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

A Hybrid Model for Anomaly Detection in Cloud Using Deep Learning

Dhanujakshi[1], Dr. Ravinarayana B[2],

[1]Master's in Technology Student, Dept. of Computer Science and Engineering, Mangalore institute of technology and Engineering, Moodbidri,
[2]HOD & Associate Professor, Dept. of Computer Science and Engineering, Mangalore institute of technology and Engineering, Moodbidri,

---

Abstract - The exponential growth of cloud infrastructure and services has resulted in a sharp increase in the demand for effective anomaly detection methods. Traditional rule-based and statistical methodologies have a tough time keeping up with cloud systems because of their dynamic and complex nature. As a result, there is a growing need to apply deep learning techniques to address the issue. Here we provide a unique method for utilizing deep learning to identify anomalies in cloud environments. One of the main contributions of this research is the design and implementation of deep learning-based anomaly detection system that is specially tailored to the cloud's data properties. Convolution Neural Networks and auto encoders are used to extract the patterns and spatial representations from structured cloud data. Thorough experiments are conducted on a large scale cloud dataset with a range of usage scenarios and anomaly types to evaluate the effectiveness of the recommended technique. The deep learning model has benefits over more conventional anomaly detection methods when compared in terms of accuracy, sensitivity and false-positive rates, according to analyses. The findings of my analysis suggest that deep learning based LSTM and RSU algorithms significantly increase the security and reliability of network data by identifying abnormal behaviors. The RSU algorithm forecasts an accuracy of 99 comparable to the LSTM algorithm.

***Key Words*:** Convolution Neural Networks, LSTM algorithm, RSU algorithm, Deep Learning

1. INTRODUCTION

One of the strongest technological advances of the contemporary period, cloud computing, was developed as a result of the need to enhance processing capacity and on-demand services to meet customer expectations. The fundamental virtualized context they operate in presents a variety of operational and security challenges, rendering the transition from traditional client-server architecture to cloud computing challenging. On average, one form of denial-of-service attack against the servers of more than 20% of businesses has occurred frequently. Consequently, in order to ensure adaptability, the cloud must be able to respond to both known and newly emerging threats that may harm its foundational components. Experts frequently use intrusion detection systems as a preventative measure for security in the cloud to overcome these challenges. Cloud-based platforms make use of intrusion detection systems for exploitation detection, identifying anomalies, supervisory contemplation, virtual machines contemplation, and a mixture of these.

Although being excellent at identifying general irregularities, anomaly detection for heterogeneous traffic flow data provided by different application types is still an emerging method. Since immediate evaluation is required to find anomalies in networks within streams of data, the bulk of these systems have high rates of false alarms and are computationally challenging. As a result, they are not very effective at accomplishing it. Recent developments in deep learning have sparked research curiosity in identifying anomalies in networks. Since CNN may be organically taught and requires no preliminary processing, they are frequently used in the data classification and are suitable for finding abnormalities.

2. EXISTING SYSTEM

The vast majority of methods for identifying anomalies now in use operate based on recorded data, where there is no assumed causal relationship between data occurrences. Often, data sets may be linked to each other. Graphical data, spatial data, and sequential data are just a few instances. Numerous structures and approaches have been proposed with the goal of identifying abnormalities. Among these are approaches based on machine learning, mining data, and statistical detection of anomalies. The program keeps track on the activity and statistically builds the characteristics to characterize the way the subject acts. One of the earliest techniques for detecting intrusions is expert intrusion detection, which continually monitors user activity and identifies unusual occurrences. Anomaly detection systems are those that use machine learning techniques which can modify the layers of their application in accordance with newly gathered data. Furthermore, the current method takes a greater amount of time, and the models being used weren't trained on their own.

3. PROPOSED METHOD

Using the LSTM model architecture, we build a hybrid CNN and training it. The hyper-parameter value is maximized using the k-fold approach, with K set to 10. The proposed design has a novel ConvLSTM2D layer. The initial system call sequences that are passed into this layer are used for common operations. The layers of convolution use the convolution method followed by the pooling strategy to extract important properties from the sequences. We use a kernel that is (1, 2) size and has 64 filters. The LSTM layer then receives the extracted sequence and finds the recurring patterns in the sequence's natural behaviour. A data-distributed layer is employed to pass disguised output at each time step.

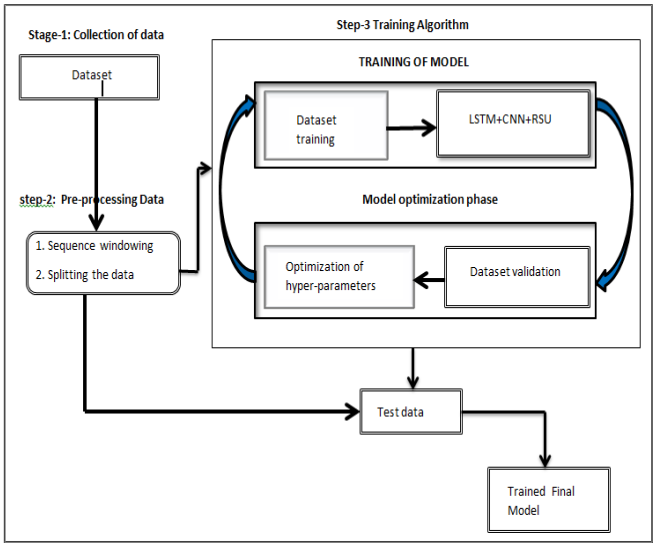

Fig -1: Proposed method

Pre-processing is done on the collected data in order to identify any gaps and fill them. Eighty percent of the information in the data set are taken for training, fed into the model to validate the data, and optimized for hyper parameters. Once the algorithm training is complete, the trained data is compared to the testing data to produce the final trained model, which can identify any abnormalities that may be present.

4. IMPLEMENTATIONS AND RESULTS

Tensor flow, Keras, Pandas, and other libraries are imported into the Google Colab software, which is used to carry out the implementation. A total of 3,577,296 instances make up our dataset, Universal Del Cauca version 2-87 Atts intrusion detection dataset. There are 87 characteristics in this dataset. Each instance includes the source and destination IP addresses, nesting time, and other data. We apply the feature extraction approach to extract the necessary features from this large data collection. Following extraction, we have 500 rows with 72 characteristics, of which 80% are used for training and 20% for testing. In order to retain a small training set and train all of the network traffic samples, feature extraction was carried out. We use the LSTM algorithm to get an accuracy of 92%, and the RSU algorithm to achieve an accuracy of 99%.As a result, the recommended RSU algorithm makes predictions with more accuracy and is the most effective technique for anomaly identification.

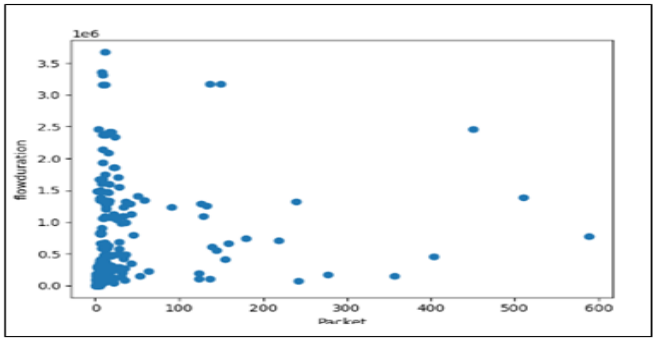

Fig-2: Flow duration and forward packets

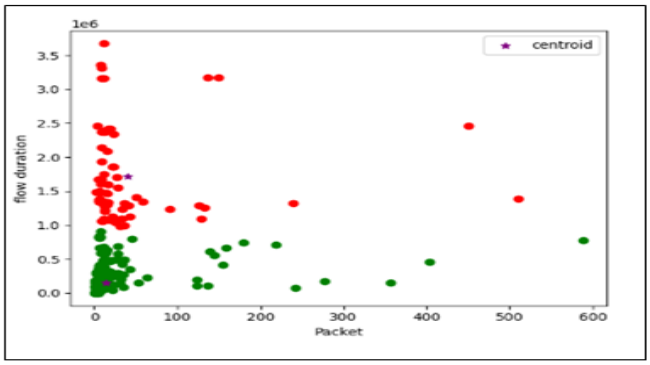

Fig-3: Centroids

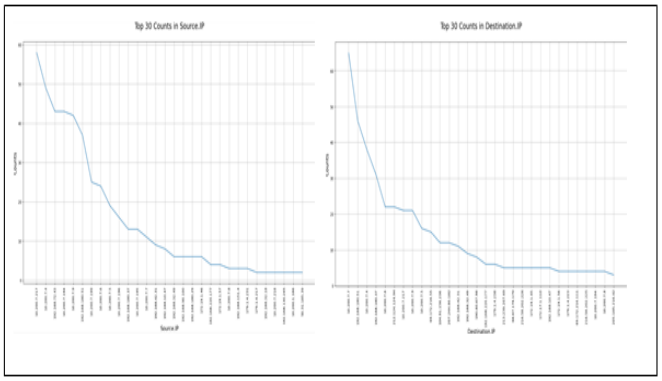

Fig-4: Source IP and destination IP histogram

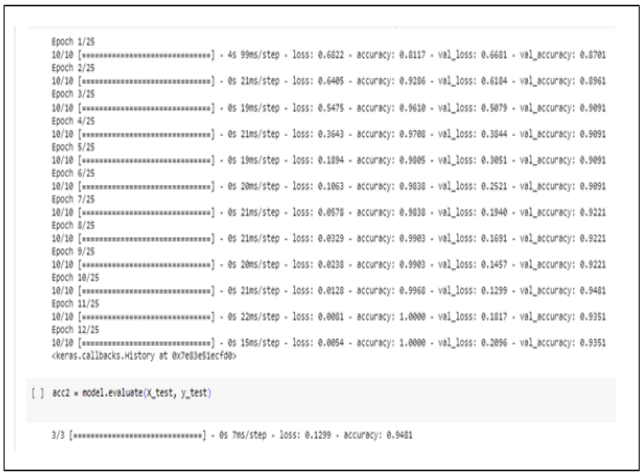

Fig-5: Accuracy reached using LSTM technique

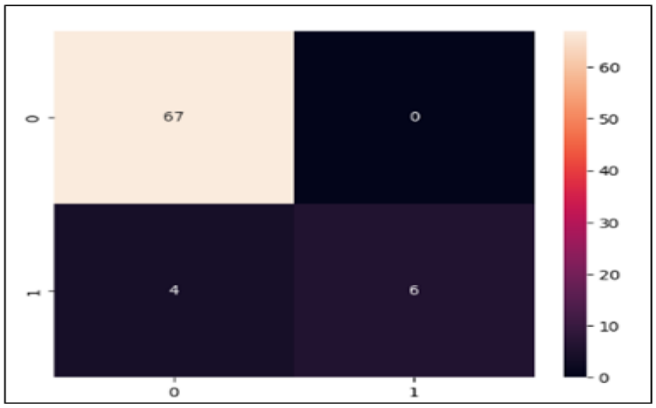

Fig-6: Confusion matrix

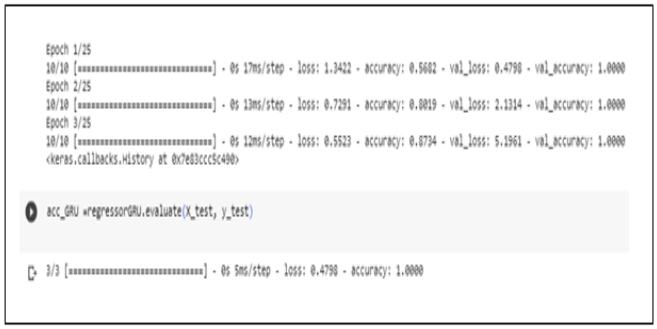

Fig-7: Accuracy reached using RSU technique

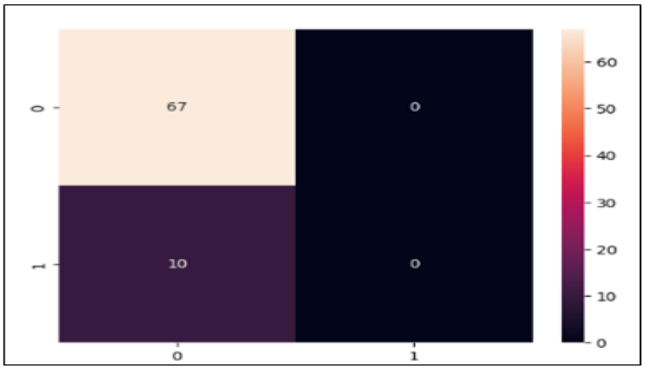

Fig-8: Confusion matrix

5. CONCLUSIONS

The core of the intrusion-based identification technique is the idea that normal events and abnormal events are separate from each other. Techniques for identifying errors pick up on specific behaviors whenever an application is functioning effectively. The way it displays of the network's instructions in a sequence that controls the behavior of the procedure. We suggest a deep learning-based CNN and LSTM model combination to find abnormalities in the system call sequence. Though CNN was employed to identify important features from the condensed features, LSTM is used to detect sequential patterns.

REFERENCES

1. Jabez J, Gowri S, Vigneshwari S, Albert Mayan J, Srinivasulu S, ”Anomaly detection using CFS Subset and Neural Network with WEKA Tools”. In: Satapathy S., Joshi A. (Eds) Information and Communication Technology for Intelligent Systems. Smart Innovation, Systems and Technologies, vol 107. Springer, Singapore 2019

2. K. A Taher B, Mohammed Yasin Jisan , M.M. Rahman, ”Network Intrusion Detection Using Supervised Machine Learning Technique with Feature Selection”, International Conference on Robotics, Electrical and Signal Processing techniques(ICREST), Dhaka, Bangladesh, pp. 643-646,2019.

3. Ashlahi-Shahri, B.M, Rahmani, R, Chizari, M. et al. “Neural Compute & Applica”, 27: 1669, 2016.

4. Sumaiya Thaseen, Ikram, Aswani kumar, Cherukuri, “Intrusion Detection Model Using Fusion of Chi-Square Feature Selection and Multi class SVM”, Elsevier –Journal of king Saud University- Computer and Information Sciences, Volume 29, Issue 4, October 2017.

5. Kervic J, Jukic S, Subasi, A Neural Comput & Applic 28 (Suppl 1) **:** 1051, “An Effective Combining Classifier Approach Using Tree Algorithms for Network Intrusion Detection”, Springer , December 2017, Volume 28, Supplement 1, pp. 1051-1058.

6. M Ahmed, A.N. Mahmood, J.Hu, “A Survey of Network Anomaly Detection Techniques”, Journal of Network and Computer Application, vol.60, pp.19-13, 2016.

7. Garg K. Kaur, N. Kumar, S. Batra and M.S. Obaidat, “HyClass**:** Hybrid Classification Model for Anomaly Detection in Cloud Environment”, in IEEE International Conference on Communication (ICC), Kansas City, USA, May 2018.