Network-based Intrusion Detection System for Preventing the Cloud Computing Environment from Cyber-Attacks using Deep Learning Algorithms

MSc Research Project

MSc Cloud Computing

Shrikant Umakant Thombre

Student ID: 20205929

School of Computing

National College of Ireland

Supervisor: Jitendra Kumar Sharma

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Shrikant Umakant Thombre
Student ID:	2025929
Programme:	MSc Cloud Computing **Year:** 2022
Module:	MSc Research Project
Supervisor:	Jitendra Kumar Sharma
Submission Due Date:	15/08/2022
Project Title:	Network-based Intrusion Detection System for Preventing the Cloud Computing Environment from Cyber-Attacks using Deep Learning Algorithms
Word Count:	6081 **Page Count:** 19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Shrikant Umakant Thombre

Date: 15/08/2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Network-based Intrusion Detection System for Preventing the Cloud Computing Environment from Cyber-Attacks using Deep Learning Algorithms

Shrikant Umakant Thombre

20205929

Abstract

In the current era, cloud computing is considered the most widespread source of storage, computation and communication because of its reliability and considered a safe place to store user and business data. However, due to increase in the internet traffic, the chances of a malicious attack on the cloud networks have also increased. Therefore, providing security to such systems has become a thing of prominent significance. In this paper, we presented an Intrusion Detection System (IDS) for the cloud computing environment to detect any kind of network attacks. The major objective of this research is to process incoming data packets, which can accurately classify the Normal Traffic and Malicious attacks. Among the Malicious attacks, our system also should be able to detect the type of the attack, at the same time take/recommend appropriate action as per the type of attack. For accurate prediction of different types of attacks, in this research multiple deep learning models have been deployed and tested. After evaluating the performance of all the deep learning algorithms, it has been found that the LSTM model was correctly able to classify attack types with the highest accuracy score of 99% over the test data. Based on the LSTM model, a Web application for intrusion detection system has been developed, which is based on client-server architecture and can detect the network attack types and anomaly activities in the network system in real-time.

1 Introduction

With the rising data generation and articulation, various users and service providers have inclined toward technological advancement. The major concern for the rising data information was the optimal storage, accessibility, trustworthiness and foremost the security and privacy of the data. Considering the local storage system has several disadvantages associated with the service providers such as higher initial cost, limited resources, lack of reaching instant demand and more. To overcome these, the web-based computing module which is known as cloud computing is considered. Through cloud computing, one could expect feasible up-scaling of resources to reach demand anytime, better accessibility and trustworthiness. Furthermore, through this mode of computing, the initial investment could be limited with the wider access to the resources and the increasing demands. With the emerging networking scope, there is also a wider utilization and consideration for cloud computing. The cloud setting environment offers software, storage, infrastructure, hardware, and various other components on demand over the air or internet. Nowadays, various organizations, firms and individuals opt for the cloud system owing to various advantageous features such as reliability, cost-effectiveness, accessibility, flexibility, security, and privacy. The most known and considered feature of the cloud setting environment is instance resource availability on the demand. Although there are certain disadvantages in the cloud computing system which has been taken into the consideration in this paper.

Due to its open architecture, cloud computing is vulnerable to security and privacy breaches. This is a cloud cyber-security problem. Cloud computing is the greatest technological advancement for service providers and users. Although security risks may compromise user data privacy and confidentiality. These data might cause service providers and users enormous losses. Data can be used for criminal purposes. The system's integrity and user experience are also compromised. Cloud security intrusions might be known or unknown. The cloud environment has several intrusion-prevention measures. Signature-based and anomaly-based incursions exist. Signature-based intrusions are those whose type, trend, or pattern are known. The incursion is identified and delayed based on previous trends. Anomaly-based intrusions are another sort of innovative intrusion. Detecting an intrusion in the traditional way is tiresome, thus the intrusion is identified based on behavior deviation.

To retard the intrusion, there are various approaches utilized based on the type of intrusion. These approaches include identifying behavioural change, detecting intrusion, blocking the IP addresses, activating the firewall and more. Such activities undergo in the IDS. In the cloud setting architecture, IDS is one of the crucial factors. Although the conventional IDS are not robust enough to always hinder all types of intrusions. Furthermore, the distributed model of the cloud setting environment makes it a tedious task to perform the task of IDS modules. Therefore, the implementation of novel artificial intelligence (AI) techniques such as Deep Learning (DL) can be taken into the consideration. Through the implementation of DL models, a robust IDS can be built that can overcome various challenges in the conventional system. The attack on the cloud setting environment can be on either network mode or host mode. In this paper, we will be utilizing the deep learning algorithm for preventing intrusion into the cloud setting environment. Different types of algorithms are taken into the consideration and evaluated which are Convolutional Neural Network (CNN), Long Short-term memory (LSTM) and Convolutional LSTM (Conv-LSTM). For the implementation of these algorithms, our study considers the NSL-KDD dataset. In the next section, we have discussed various studies by the researchers on the context.

1.1 Research Question

- To what extent the Deep learning algorithm-based IDS can detect the Network penetrations in Cloud Computing Environment?

- Which algorithm can accurately identify the anomaly in the network system and also can classify the attack types accurately?

1.2 Research Objectives

- Development of IDS using deep learning algorithm for correctly classifying the types of network attack and reducing the false positive detection rate.
- To develop a web application for IDS, which can detect the network in real-time to strengthen the cloud security.

2 Related Work

In this part, we have discussed the different research papers and studies by several others. This section is further divided into Security Threats and Challenges in Cloud Computing, IDS in cloud framework, Machine Learning-based IDS and Deep Learning-based IDS.

2.1 Security Threats and Challenges in the Cloud Computing

In a paper by Nalini (Subramanian & Jeyaraj, 2018), the recent security challenges were discussed. Security vulnerabilities, dysfunctional login information, ruptured authentication tokens, security breach host controller and application programme interoperability, manipulated attack vectors, password cracking, data corruption, the sophisticated growing threat parasite, perpetual loss of information, insufficient rigour, virtualization violations, denial of service (DoS) intrusions, and integrated infrastructure consequences were twelve increasing threats described in the paper. The basic security requirement for the challenge is authentication, integrity, transparency, confidentiality, availability, and audits. Missing any of the basic requirements can deplete the security level and increase the threat. In the cloud setting environment, there is three level of challenges which are in communication, computation and Service Level Agreement (SLA). The communication level includes network, application and host whereas the computational level includes digitization and data level challenges. In SLA level, it includes the legal issues and coverages. In another paper, Khan surveyed the security-related expectations experienced in a cloud computing environment (Khan, 2016). The paper analyzed each security threat thoroughly. Network-, virtual machine-, storage-, and application-based assaults were identified. Port scanning, botnets, and spoofing assaults are network-based attacks. Cross-VM side-channel, VM creation, VM migration and rollback, and VM scheduler-based attacks are VM-based attacks. Data scavenging and deduplication are storage-based attacks. Malware injection and steganography, shared architectures and web services, and protocol-based assaults are also examples of application-based attacks. The paper also examined other issues.

A brief review of security issues for cloud computing systems was done by Iqbal in the paper (Ahmed, 2019). This research said several things affect cloud computing security. OS, network, load balancing, virtualization, memory management, database, transaction management, resource management, concurrency control, etc. Outsourcing, multi-latency, SLAs, heterogeneity, server downtime, backup, and data redundancy are other serious threats. The study covered several threats. Also, research papers were interpreted. The report explored issues and dangers through a detailed comparative study. Certain stages demand attention in real-world applications, though.On the other hand, Hamed (Tabrizchi & Kuchaki Rafsanjani, 2020) also surveyed the security challenges in the cloud computing environment. The paper evaluated the different security challenges and threats in different types of cloud computing systems. Generally, the security buoyancy remains in several factors such as confidentiality, integrity, availability, authentication, non-repudiation and more. Several research papers were also discussed in the study. To provide optimal security for the system, different paradigms were taken into consideration which was authentication, authorization, identification, and access management. It was also stated that threats and attacks were two different aspects which both indeed are alarming for the cloud computing system. To overcome these challenges, the author of the paper suggested different technological and methodological approaches which could be implemented in real-world implementations. Furthermore, the future scope of work was also discussed.

Similarly, Kriti and Bhushan studied the security threats in the cloud computing system (Bhushan & Gupta, 2017). The major roadblock to the service providers and the user is the Distributed Denial of Service (DDoS) attack. Due to the DDoS attack, the security, privacy, and availability of the cloud system are hindered. Therefore, the paper discussed the different approaches and studies to overcome the challenges of DDoS attacks. Various attributes and mechanisms of the DDoS attack were discussed in the paper. Several types of attack and threats other than DDoS attack was discussed in the paper. Furthermore, the paper scrutinized different methods to prevent or hinder the DDoS attack. Studies by different researchers were also discussed in the paper. Although, the mechanism to prevent the attack shall be interpreted and built for real-world application. Naresh (vurukonda & Rao, 2016) studied the security issues experienced in data storage cloud computing. Data privacy and integrity, data recoverability and vulnerability, poor data refining, and data backup are common issues. Malicious insiders, outside invaders, SLAs, and legal difficulties are other possibilities. To address security difficulties, threats, and attacks, researchers explored many features and methods. See Cloud, FADE, TimePre, and resident data security are proposed data storage tasks. For identity management and access control, SPICE, HASBE, a decentralized, role-based framework, and See agreement and Specs SLA were proposed. These research examined the pros and cons and advised the best approach.

2.2 Intrusion Detection System (IDS) in cloud computing

With the growing security attacks and threats, the most common mechanism to hinder these are the IDSs. In the paper by Yasir and the team (Mehmood et al., 2013), the challenges and the opportunities were discussed in IDS for cloud computing. Primarily, different types of intrusions were discussed in different cloud settings such as public, private, community, and hybrid. The attacks can be at different levels of a cloud setting such as network level attacks or host level attacks. The network-level attack includes address resolution spoofing, IP spoofing, DNS poisoning, port scanning, the man in the middle attack, routing information protocol attack, denial of service (DoS) attack and Distributed Denial of Service (DDoS) attack. Here, the intrusion is only considered when there is a void of confidentiality, integrity, and availability by bypassing the security mechanism of cloud computing. These challenges could be combated with the IDS. Although, the current conventional IDS has some drawbacks which are scalability and autonomic self-adaptation. Similarly, Patel (Patel et al., 2013) did a systematic analysis of the IDS and prevention system in cloud computing. Initially, the paper scrutinized the recent studies and research in the field. IDS mechanism was discussed in the paper. The functional layer includes levels for reaction, alarm processing, detection, and monitoring layer. It also has an infrastructure layer. Network-based, host-based, and application-based detection are in the functional layer. The wired or wireless structural layer follows. The module might be standalone, distributed, hierarchical, or mobile. In addition to these, different aspects of the IDS were discussed in the paper.

Kene in the paper did a systematic review of the IDS methods for the cloud computing security challenges (Kene & Theng, 2015). Different types of attacks with the recent types of advancement and modules were studied in the paper. The different attacks were VM attacks, user-to-robot attacks, insider attacks, denial of service (DoS) attacks, port scanning, and backdoor path attacks. This paper did a thorough analysis of different aspects of the cloud setting environment. Furthermore, conclusive suggestions were given regarding the future scope of work. On the other hand, Rajendra (Patil et al., 2019) designed robust security architecture for detecting network penetration in virtual machines (VMs) in the over-the-air architecture. Here, numerous publications on VM intrusions were studied. A VM must have certain moats, such as regulating large network packets, quick identification, sensing a range of vulnerabilities, minimal transmission cost, infiltration susceptibility, and good precision and false negatives and false alerts. VMs' signature-based intrusion detection can only detect established patterns and trends. Implement anomaly-dependent intrusion detection to detect unique patterns and trends. During the novel approach proposal, numerous experiments were implemented and tested.

2.3 Machine Learning-based Intrusion Detection System (IDS)

In a paper by Amir (Javadpour et al., 2017), a novel approach to intrusion detection in cloud framework using machine learning was proposed. The novelty of this paper is that it primarily implements the feature selection approach. Through feature selection, there is a certain aspect which helps in improving the output and the accuracy. For the classification challenge, the paper considered the implementation of four algorithms which were neural network, cart algorithm, id3 decision tree algorithm, and random forest algorithm. Furthermore, to interpret the performance the paper considers the precision, recall and accuracy score. In the final stage evaluation, it was stated that the neural network algorithm showed better performance than the other algorithms. On the other hand, Aljamal (Aljamal et al., 2019) proposed a hybrid IDS utilizing ML techniques in the over-the-air framework. The primary motive of this study was to build a robust mechanism to detect both signature-dependent attacks and anomaly-dependent attacks. Therefore, the paper considered the hybrid algorithms which detect both the signature attack and anomaly attack. Here, two different algorithms were considered which were combined to form a hybrid model. The algorithms were K-means clustering algorithms and SVM classification algorithms. To implement the model, the UNSW-NB15 dataset was considered in this paper. After the evaluation, the K-means performed better than the other models whereas the SVM algorithm performed poorly. Although both algorithms have their advantages and were combined to form a hybrid algorithm.

Balamurugan (Balamurugan & Saravanan, 2019) also proposed a novel and robust approach for IDS in cloud computing utilizing the machine learning approach. The paper describes the eight types of attacks which are DDoS, probing, U2R, R2L, zero-day attacks, distributed attacks, vulnerability reports and convert channel attacks. Considering the IDS, there are three types of models which are detection method, monitoring method, and behaviour pattern. The paper also discussed various research studies. The proposed model was compared and evaluated against the conventional modular systems such as anomaly detection, intrusion detection and prevention, fuzzy-based hybrid and context-aware anomaly detection. Furthermore, the evaluation of these models was done utilizing the evaluation metrics for instance PRF score.

In another paper, Singh and Ranga (Singh & Ranga, 2021), studied an ensemble learning approach for attack and intrusion detection in cloud computing. Initially, the paper considered different studies and evaluated the output. This study paper considered four different ensemble learning algorithms which were AdaBoost Classifier, Bagged Tree Classifier, Subspace discriminant, and RUS Boosted tree classifier. For the implementation, the CICIDS 2017 dataset was considered which was then pre-processed utilizing normalization, one-hot encoding and more. The dataset is then partitioned in a certain ratio. The considered algorithms are then implemented into the model and evaluated in the ensemble voting scheme to obtain the best and optimal performance. Certain evaluation metrics were also considered namely the PRF score and accuracy score. In the later stage, certain conditions were a drawback, but the implementation was promising. On the other hand, Preethi (Mishra et al., 2016) researched optimal algorithms for IDS for the cloud system. The primary motive of this study is to search for an efficient and optimal algorithm with a faster mode of detection. To achieve the goal, the paper incorporated machine learning algorithms and parallelization. The paper considered different methods such as Naïve Bayes, Neural Network, and Decision Tree. For the implementation, the KDD99 dataset was considered. This dataset contained the flow of different types of attacks such as DoS, Probe, U2R and R2L. In the evaluation of the model, it was sought that the decision tree algorithm outperformed all other algorithms. Although there was certainly room for improvement in the real-world applications.

2.4 Deep Learning-based Intrusion Detection System (IDS)

Sethi in the paper proposed the deep reinforcement learning-based IDS for the cloud computing system (Sethi et al., 2020). In the conventional models, there were sought to be certain fallacies such as reduced accuracy, slower detection, longer interpretational time and more. Therefore, to overcome the challenges experienced in the conventional approach, the paper here suggested the implementation of deep reinforcement learning in the cloud architecture. For the implementation, the UNSW-NB15 dataset is considered. The paper considers five different classifiers for the implementation which are Random Forest, Adaptive Boost, Gaussian Naïve Bayes, KNN and Quadratic Discriminant Analysis. In another paper, Jaron (Fontaine et al., 2020) proposed a log-based intrusion detection for cloud web applications using computational intelligence. The paper considered two algorithms which are neural network and decision tree. Upon evaluation, the neural network algorithms outperformed the different approaches.

On the other hand, Jan (Lansky et al., 2021) did a systematic review of the DL methods for the IDS in cloud computing and distributed computing. Several papers concerning the IDS were reviewed and analyzed. The interpretation from these papers was studied and discussed. Santhosh also studied the cloud-based real-time network IDS utilizing DL algorithms (Parampottupadam & Moldovann, 2018). For the implementation, the NSL-KDD algorithm was considered. The algorithms considered in this study are SVM, RF, LR and NB. In the final evaluation, the model achieved an accuracy score of 83%. Adel (Abusitta et al., 2019) proposed a DL method for a proactive multi-cloud cooperative IDS. The proposed DNN model is constructed in the architecture of Denoising Auto Encoder. For the implementation, several real-time datasets were obtained and implemented into the model. During the processing of the model, it was done in TensorFlow which is GPU enabled. The evaluation of metrics showed that the model could achieve accuracy by 95%.

3 Research Methodology

In recent years, cloud computing has become the most sought-after internet-based technology. The on-demand availability of computer resources like databases and increased computing power gives it an edge over the usual local systems. The clients/customers of the cloud system don't have to worry about the maintenance and upgradation of hardware and software. Moreover, several issues about the vulnerabilities of digital information in cloud computing architectures are associated.

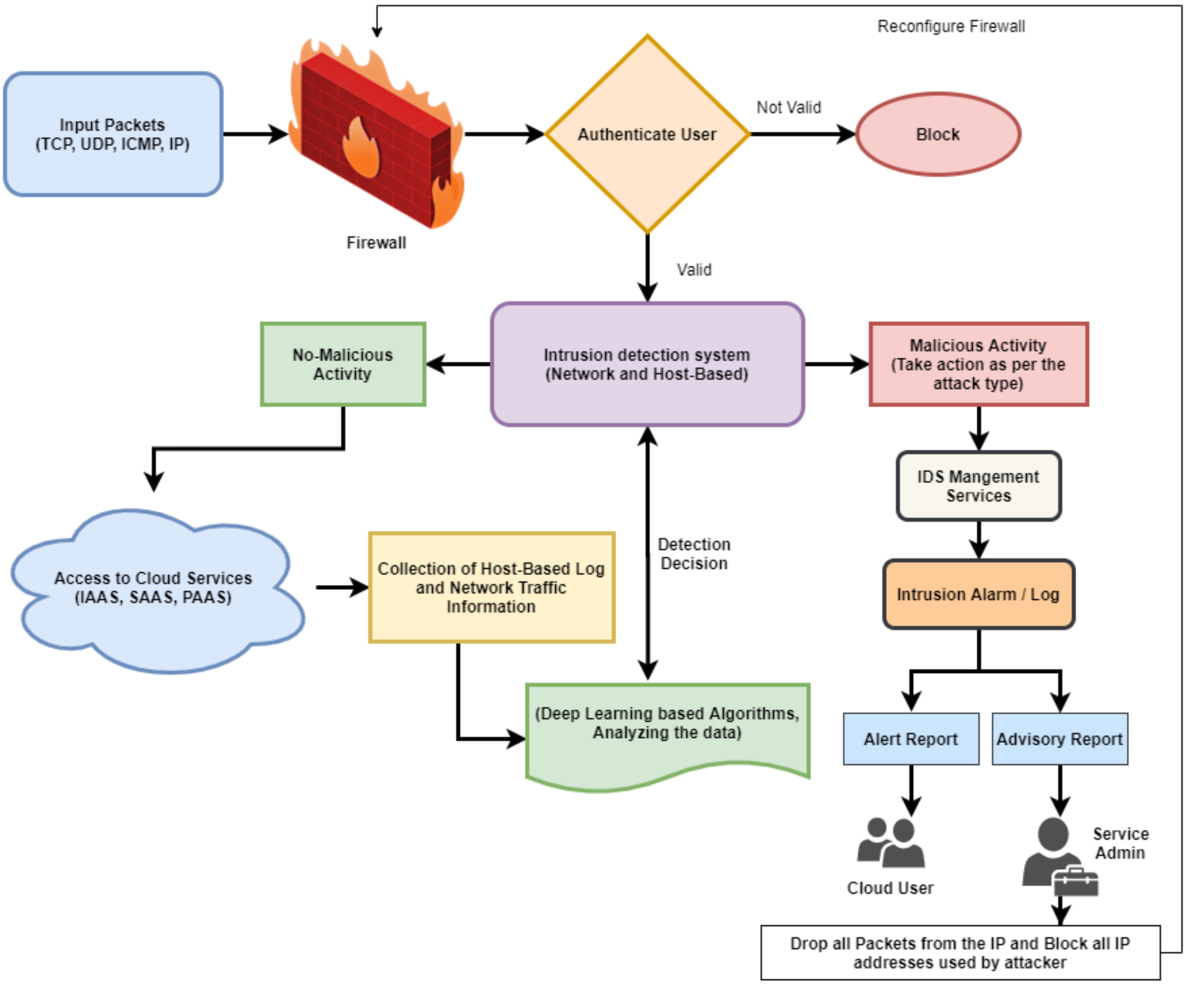

Reconfigure Firewall
Input Packets (TCP, UDP, ICMP, IP)
Firewall
Authenticate User
Not Valid
Block
Valid
No-Malicious Activity
Intrusion detection system (Network and Host-Based)
Malicious Activity (Take action as per the attack type)
IDS Mangement Services
Detection Decision
Access to Cloud Services (IAAS, SAAS, PAAS)
Collection of Host-Based Log and Network Traffic Information
Intrusion Alarm / Log
(Deep Learning based Algorithms, Analyzing the data)
Alert Report
Advisory Report
Service Admin
Cloud User
Drop all Packets from the IP and Block all IP addresses used by attacker

Figure 1. Intrusion Detection System in Cloud Computing Environment

In this section, the Cloud-based IDS along with important components will be discussed in detail. A typical cloud-based module works on the concept of virtualization of resources. A server is responsible for catering to the needs of clients. In a typical cloud system, a server is responsible for fulfilling the request of the client. Due to any malicious attack by the intruders, the server might face disruption of services and important data of users may be compromised. Therefore, the cloud requires a Network-based Intrusion detection system (NIDS) for handling malicious activities and identifying anomalies. In our work, we propose a framework which defends a system against multiple system penetrations in the virtual architecture, consisting of Denial of Service Attack (DOS), Remote to Local attack (R2L), Probing and User to Root Attack (U2R). To hinder the virtual system architecture from network-based penetrations, the IDS plays an important role. Major components of our Cloud system include a router, firewall and networks-based IDS (NIDS). In our proposed framework, the IDS utilizes DL-based methodologies to obtain a robust and precise output with the reduction of false positive inhibitions under various network attacks. The following Figure 1, better explains the architecture of our proposed Intrusion Detection System.

A network consists of hardware such as a router and firewall. A firewall is a software deployed in a cloud network to defend against unauthorized access. If the attacker manages to bypass the firewall, then the Network IDS can be used to identify and mitigate the threat. The cloud-based intrusion detection system architecture would be able to process large data packets and will reduce packet loss by accurately detecting the attacks on the system. On identification of any malicious activity, the Intrusion Detection System will be detecting the attack type and accordingly, the actions can be taken. Later, IDS is also responsible for generating comprehensive reports on the attack detected and sending over the report to the cloud service provider and taking appropriate actions accordingly to prevent the system from such attacks. As per the types of attack, the pre-defined actions can be taken by Intrusion Detection System. If the incoming request or network packet is found to be legitimate (No-Malicious Activity) then, permission will be granted by the Intrusion detection system and the packet will be allowed to access the cloud services. IDS Continuously monitors the network activity and learns from the new data, before accessing the cloud service every packet/request has to pass through the intrusion detection system. For the cloud-based network IDS, there exists a mainstream IDS, that scrutinized the complete network system activities serving as the host-based IDS which eliminates the presence of multiple IDS.

4 Design Specification

Detection of any malicious activity through cloud-based network intrusion detection systems is a challenging task. The foremost issue experienced in our system is that the Network based IDS (NIDS), usually has a large false positive rate, so the classical Machine Learning models may not generalize well on the incoming packets such as TCP, IP etc. That’s why we implemented our detection system using deep learning models which through their sophisticated layer architecture handle the incoming data packets well. For our problem, we have utilized CNN, LSTM and Conv-LSTM architecture for predicting the different types of attacks. As discussed earlier the problem of false positive predictions, there has been an enormous amount of work done in the past to solve the problem of false positive rates, some of which have been discussed in the literature review part. Therefore, to reduce the FPR, we performed several steps which will be discussed in detail here. The framework for developing the Deep learning-based IDS is shown in Figure 2.

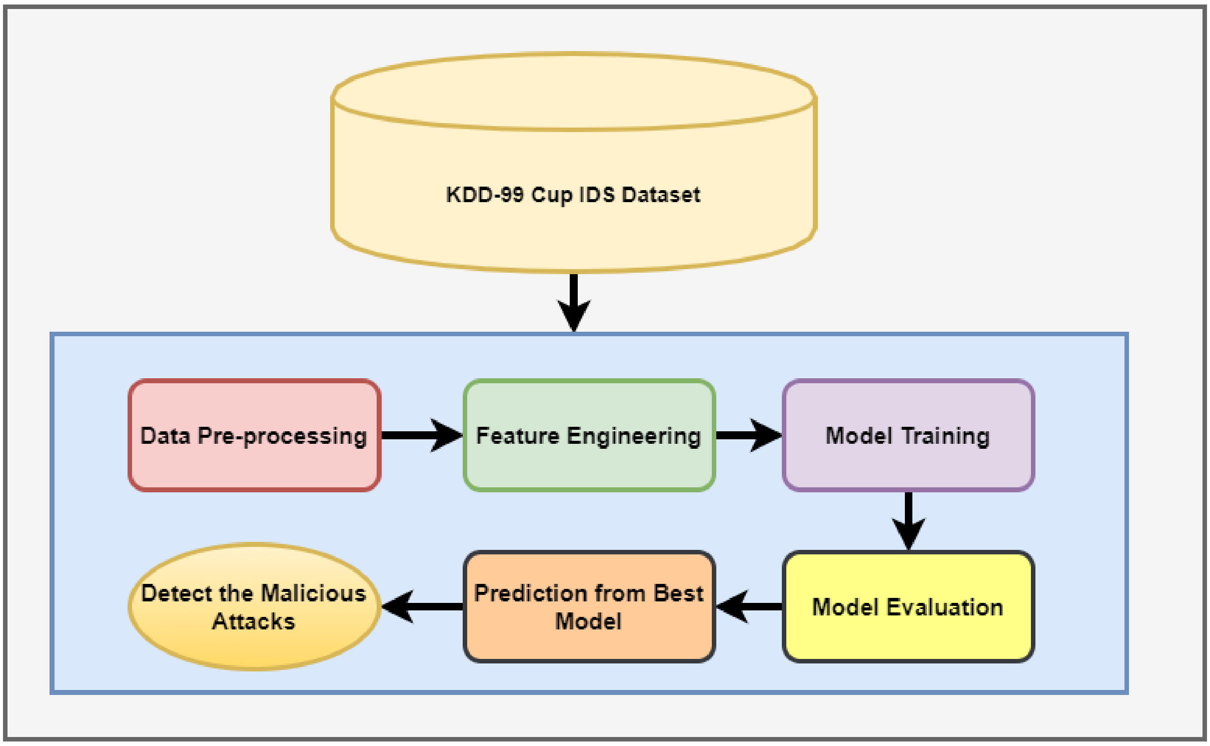


Figure 2 Intrusion Detection System Model Training Using Deep Learning

4.2 Data Collection

The dataset for the malware traffic analysis in the cloud computing environment is used as KDD-99 Dataset (KDD Cup 1999 Data, 2022). It is an updated and refined dataset specifically designed for malware analysis purposes. It is considered an ideal dataset for building an Intrusion Detection System as it contains all the latest information regarding the malware attacks such as DoS, R2L, U2R and probing. The dataset contains all the relevant features for deploying a realistic, error-prone IDS. Some of the features include protocol types, service information, data packet information, fragments, features specifically for the host and finally the target variable which contains unique attack names. The job of the deep learning model is to predict malware attacks based on some trained information. Although it is noted that at times, there are certain variations in the training and test data. The test data may include different or unique types of penetrations which makes the data more realistic to be evaluated. Furthermore, during the collection of data from the host-based network penetration, it is virtualized in the simulated setting and the information is obtained as the real-world scenario.

4.3 Data Pre-processing

Deep learning-based Intrusion Detection Systems are most likely to be vulnerable to the unprocessed dataset which can lead to the misclassification of incoming attacks. This calls for rigorous data cleaning and pre-processing so there are no errors which can affect our results. After collecting the initial KDD99 dataset, we further processed it by applying several pre-processing methods on it which we'll discuss in detail here. The major reason for data pre-processing is to generate a refined dataset on which deep learning models can be applied. Any mistake in this step can lead to several problems like high bias or low variance in our results/ predictions. In our code, numerous NumPy and pandas functions were applied to get the basic sense of the data and for manipulation purposes. One major method used was to group all the attack types present in the label column into 4 unique categories which were normal, probe, R2L and U2R. Every dataset taken from a real-world scenario has to have some deficiencies like the presence of null values or noisy data etc. Similar was the case with our dataset, that's why all the missing values in our dataset were dropped. Minmax Scaler was used for data normalization so that we have similar ranges for each feature and can avoid inaccurate results. Other than that, the usual dimensionality reduction techniques were applied to the features present and data balancing was achieved.

4.4 Feature Engineering

After the data pre-processing, the dataset is passed through the feature engineering process so that only those features are used which are contributing to predicting the attack label. A typical feature engineering process involves feature selection and feature extraction. Both are considered important for generating an ideal dataset for model training. For feature extraction, we applied principal component analysis methods to the features so that we could scale them without losing any important features. In addition to that, we observed that the majority of our dataset consisted of numerical features with only a few categorical columns. For encoding the categorical features into appropriate numerical labels, one hot encoding was applied. For selecting a sample for the model training process, a random under-sampling technique was used to filter training samples in our dataset. For feature selection, several methods could be applied such as missing values ratio, correlation filter, low variance filter and chi-square tests etc.

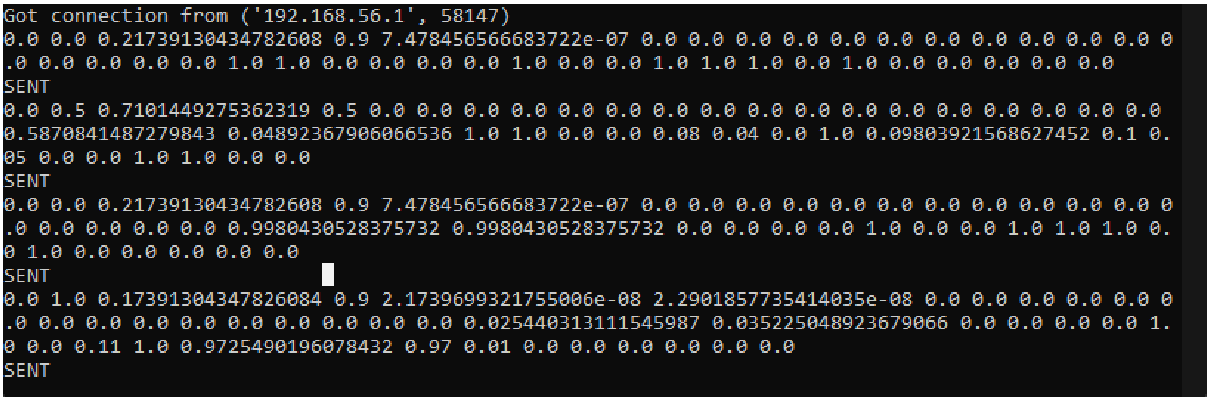
4.5 Model Training & Evaluation

Model training is considered one of the most important phases of our project as the deep learning models are solely responsible for malware detection which further leads to attack alarms which generate the alert and advisory reports. The model prediction is the core of the proposed Cloud based IDS. The primary phase is directed to extracting certain crucial and related factored data. In the next phase, the process of classification took place for the types of attacks based on severity or features such as DoS, R2L, U2R or probe. For better evaluation, the KDD99 dataset was divided into training and test sets with 70% and 30, respectively. The DL algorithms like CNN, LSTM, and CONV-LSTM were to the model utilizing the training dataset. Furthermore, the performance was evaluated utilizing the test dataset. To obtain a precise output, we trained several DL algorithms and compared their results with each other. The intrusion attack type prediction is a classification problem so the evaluation metrics used are accuracy, recall and precision. For an intrusion detection system, the model has high accuracy, precision and recall score will be considered most optimal. Loss is another metric which will be calculated over the test set, minimum loss is the better prediction by the model.

5 Implementation

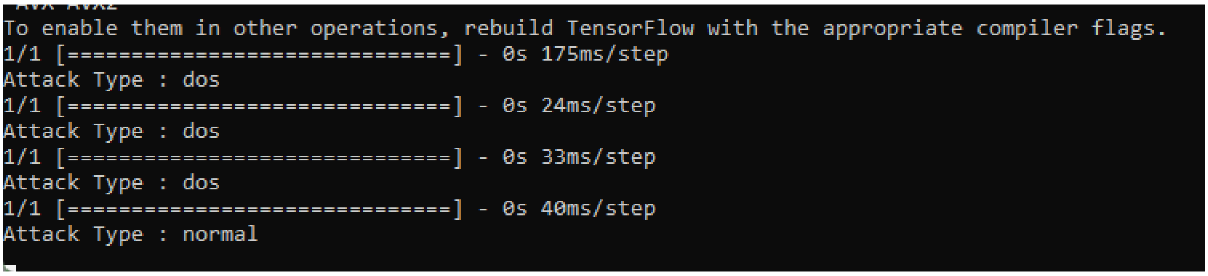
Our proposed model aims to predict the type of attack detected as well as generate an intrusion and advisory report to handle the intrusion. For this purpose, we have used three DL algorithms which are Convolutional Neural Network (CNN), Convolutional LSTM and Long Short-Term Memory (LSTM). But before reaching the stage of model training, several steps were performed which include exploratory data analysis to get to know the dataset even better. For creating visual aids, matplotlib, seaborn and plotly were used. After that, NumPy and pandas were used to clean and process the KDD99 dataset. In-built functions of NumPy and pandas were applied for manipulating the dataset according to our requirements. For implementing the deep learning models, Keras and TensorFlow libraries were used. The neural networks CNN, Convolutional LSTM and LSTM were used with appropriate activation functions. The activation functions were updated according to the requirements in the experimentation phases. The hyperparameters such as learning rate, batch size and the number of epochs were tuned for the best model performance. Finally, the sklearn scoring metrics were imported for generating the recall, precision and accuracy scores for evaluation of the models.

To develop the Web-UI for intrusion detection system, the client-server architecture has been utilized where the client sends the network packets to the cloud server. Before reaching the cloud server, the packet has to pass through the intrusion detection system based on which IDS will decide whether any kind of anomaly is there in the current packet or not. If any anomaly is identified then accordingly actions are taken or recommended. The Snapshot of the Backend, where the Client sends the network packet to the server is shown in Figure 3, Server predicting the type of attack based on the client packet from the backend is shown in Figure 4. Whereas, the snapshot of developed Web-UI detecting different types of attack is shown in Figure 5, Figure 6 and Figure 7. Training of this model was done on Google Colab. Link for the colab session and data set is as follow: Google Colab and Kdd-99 Data Set



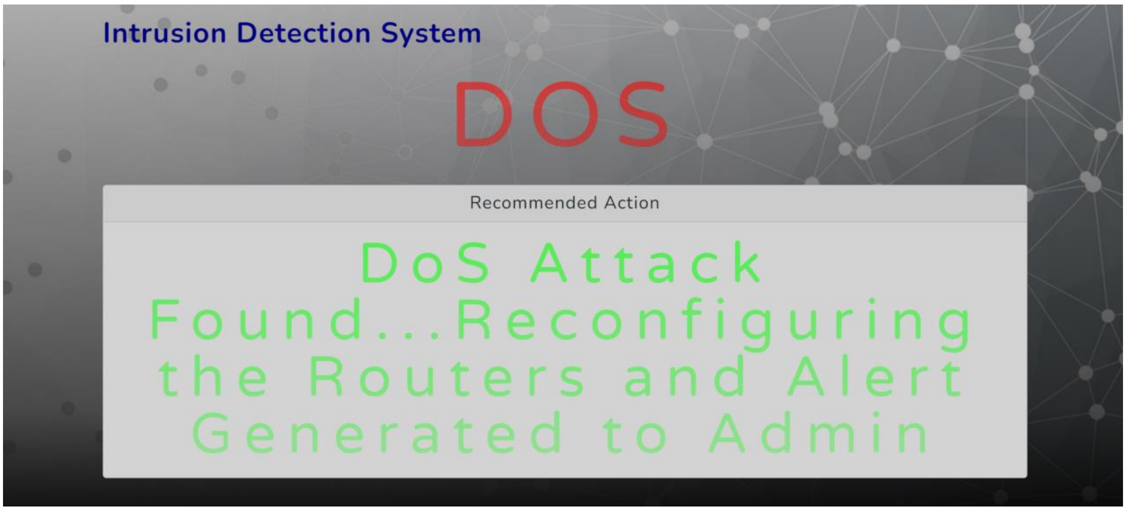
```
Got connection from ('192.168.56.1', 58147)
0.0 0.0 0.21739130434782608 0.9 7.478456566683722e-07 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0
.0 0.0 0.0 0.0 0.0 1.0 1.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 1.0 1.0 1.0 0.0 1.0 0.0 0.0 0.0 0.0 0.0
SENT
0.0 0.5 0.7101449275362319 0.5 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.5870841487279843 0.04892367906066536 1.0 1.0 0.0 0.0 0.08 0.04 0.0 1.0 0.09803921568627452 0.1 0.
05 0.0 0.0 1.0 1.0 0.0 0.0
SENT
0.0 0.0 0.21739130434782608 0.9 7.478456566683722e-07 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0
.0 0.0 0.0 0.0 0.0 0.9980430528375732 0.9980430528375732 0.0 0.0 0.0 0.0 1.0 0.0 0.0 1.0 1.0 1.0 0.
0 1.0 0.0 0.0 0.0 0.0 0.0
SENT
0.0 1.0 0.17391304347826084 0.9 2.1739699321755006e-08 2.2901857735414035e-08 0.0 0.0 0.0 0.0 0.0 0
.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.025440313111545987 0.035225048923679066 0.0 0.0 0.0 0.0 1.
0 0.0 0.11 1.0 0.9725490196078432 0.97 0.01 0.0 0.0 0.0 0.0 0.0 0.0
SENT
```

Figure 3 Client Sending the Network Packet



```
AVX AVX2
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
1/1 [==============================] - 0s 175ms/step
Attack Type : dos
1/1 [==============================] - 0s 24ms/step
Attack Type : dos
1/1 [==============================] - 0s 33ms/step
Attack Type : dos
1/1 [==============================] - 0s 40ms/step
Attack Type : normal
```

Figure 4 IDS on Cloud server Predicting the type of attack (Backend)



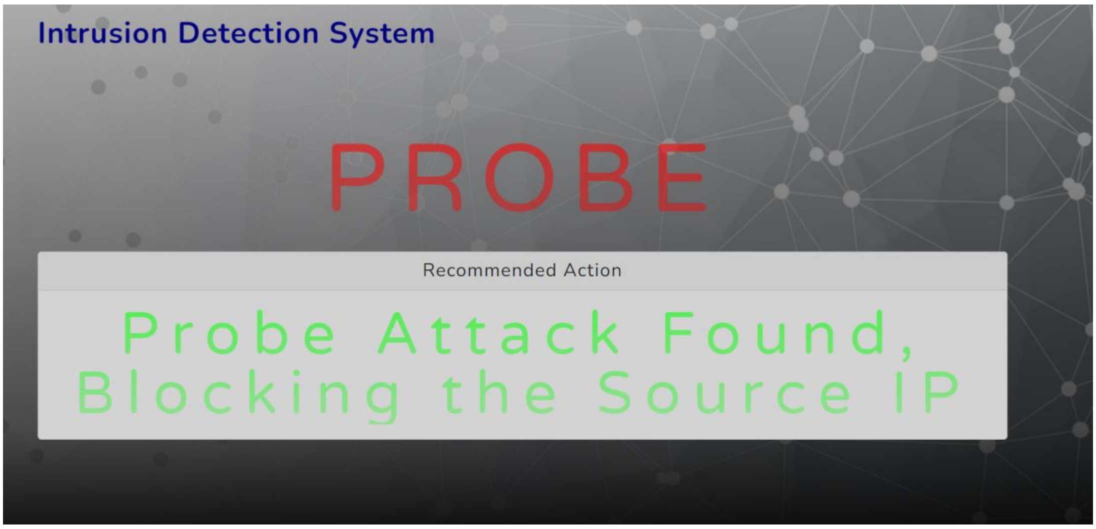
Intrusion Detection System

DOS

Recommended Action

DoS Attack Found...Reconfiguring the Routers and Alert Generated to Admin

Figure 5 Web-UI of Intrusion Detection System (Detecting DoS Attack)



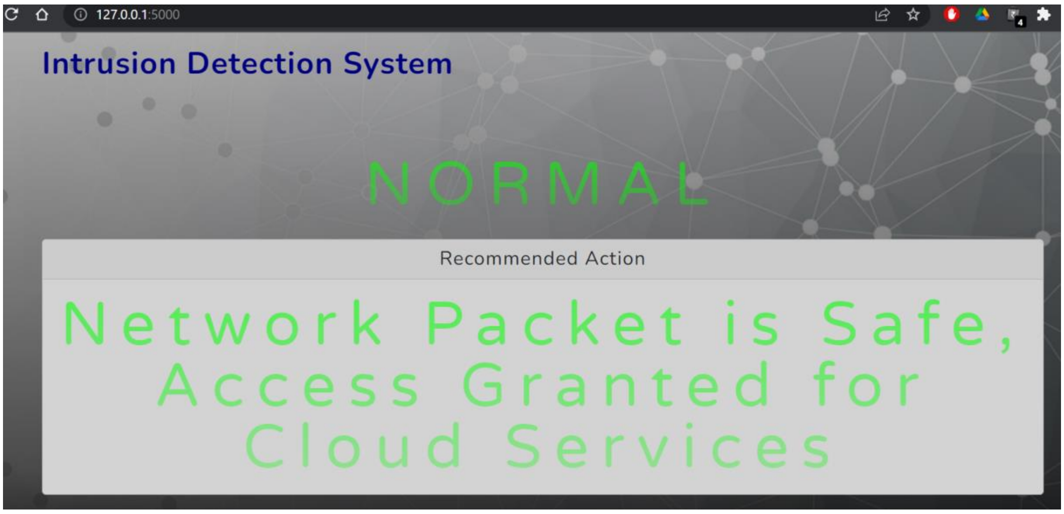
Intrusion Detection System

PROBE

Recommended Action

Probe Attack Found, Blocking the Source IP

Figure 6 Web-UI of Intrusion Detection System (Detecting Probe Attack)



Intrusion Detection System

NORMAL

Recommended Action

Network Packet is Safe, Access Granted for Cloud Services

Figure 7 Web-UI of the Intrusion detection system (Allowing legit packets to access cloud services)

6 Evaluation of Deep Learning Models

In this part, we have presented the classification results of our Intrusion detection system (IDS) using appropriate graphs of evaluation metrics such as accuracy, loss, recall and precision score. The results were generated on both the initial training and test sets. Comparative analysis between all the 3 DL models CNN, LSTM and Conv-LSTM is performed. The deep learning model with the highest accuracy, Precision and Recall score will be considered the most optimal algorithm for our virtual system-dependent IDS. In the following part, the results of all 4-evaluation metrics performed on the three deep learning models are discussed in detail. Each experimentation of performance metrics will be discussed in detail concerning each model.

6.1 Experiment 1 / Evaluation Based on Accuracy

Accuracy is the primary benchmark for a model that obtains the thorough and overall performance of the model. Although, in our study, the accuracy score is calculated utilizing the test data. In our case, accuracy can be defined as the percentage of intrusion attacks which are correctly classified. The formula for the accuracy is defined as:

Accuracy = 100 × (correctly classified attacks/total no. of records)

For all the 3 models (CNN, LSTM and Conv-LSTM) the accuracy has been calculated over the test data. As the model was trained over the 15 epochs, on every epoch the performance of each model will be calculated. The accuracy over every epoch for all the 3 models is shown in Figure 8.

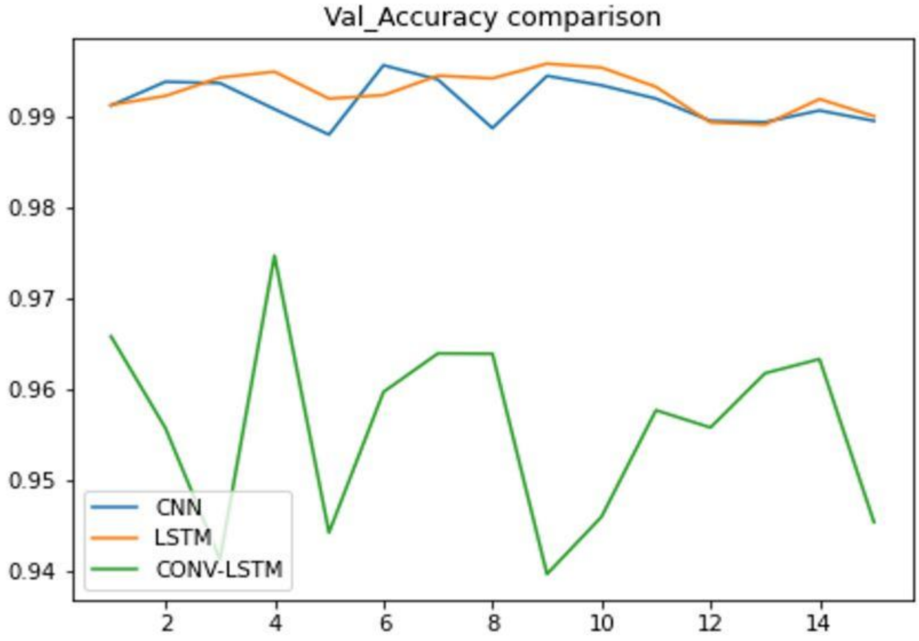


Figure 8 Validation Accuracy Comparison

On analyzing the line graph over every epoch as shown in Figure 8, it can be stated that using LSTM the highest accuracy score achieved is 99%. On the other hand, the CNN model has achieved an accuracy score of 98.95%. Whereas, Conv-LSTM has achieved an accuracy score of 94.54%. Thus, after analyzing the accuracy score, it can be said that LSTM provides the highest accuracy score of 99%, followed by the CNN model.

6.2 Experiment / Evaluation Based on Loss

In this section, the models will be further evaluated based on their loss at each epoch. higher loss means that the model does not generalize well to the dataset and is deviating from truth values which are the correct attack classifications. So, the model with the lowest loss is considered the optimal model. The loss obtained over the test data for all the models is shown in Figure 9.

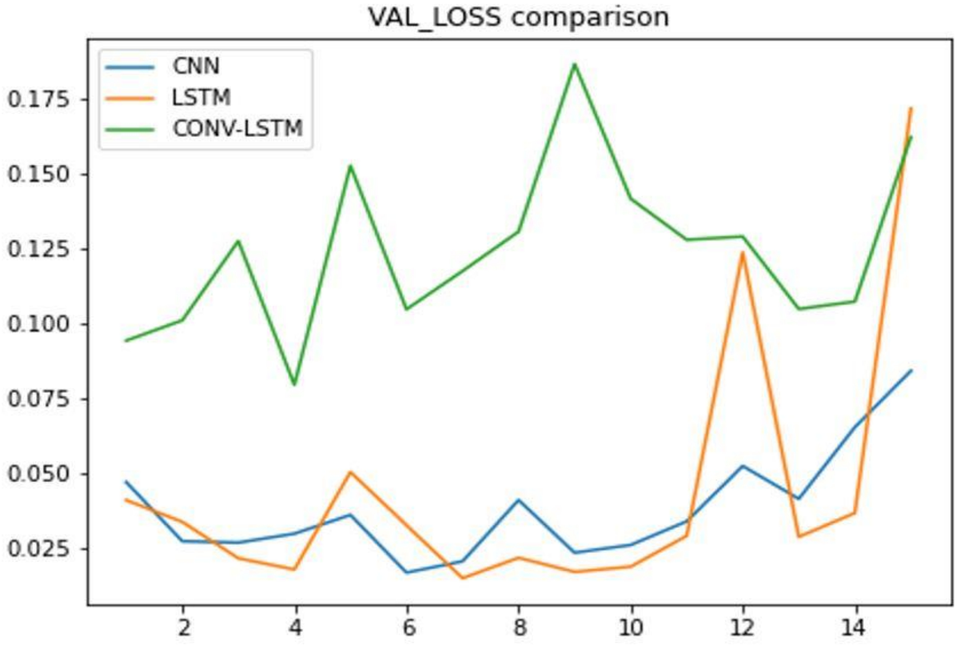


Figure 9 Validation Loss Comparison

The Validation loss for all three deep learning models generated an interesting line plot. Here, the validation loss of LSTM on random validation sets is the highest making it the worst performing model in this experiment. The Validation loss of LSTM reaches almost 0.17. Whereas, the loss of CONV-LSTM was at 0.16 approximately at the 15th epoch. The CNN model performed best in this case with a validation loss of 0.075. Thus, in terms of loss, it can be said that the CNN model outperforms as compared to the other models.

6.3 Experiment / Evaluation Based on Precision & Recall

Precision is an important indicator for finding out the performance of a DL model. it represents the quality of the true predictions made. The precision score of all the models has been compared and the results are examined. Whereas, the recall is calculated for identifying the false negative outputs. The methods with the highest precision and recall score are considered the optimal model. The line graph of Precision and Recall scores over the test data is shown in Figure 10 and Figure 11.

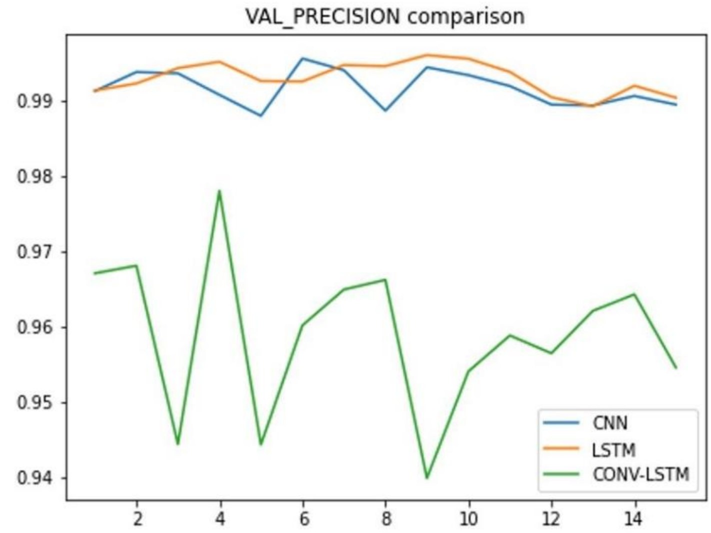


Figure 10 Precision Score of all models over the test data

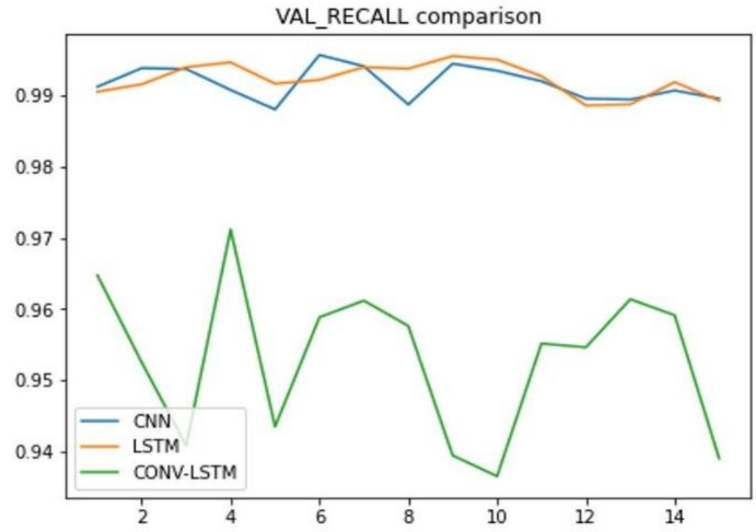


Figure 11 Recall Score of all models over the test data

On analyzing the graph of precision and recall scores as shown in Figure 10 and Figure 11. It has been observed that the Conv-LSTM model performs very poorly as compared to the CNN and LSTM. The highest precision and recall score has been obtained using the LSTM model with a score of 99.04 and 98.92.

6.4 Discussion

In our project, the major goal was the implementation of a cloud-based Intrusion Detection System. For this purpose, we introduced three deep learning models which are CNN, LSTM and CONV-LSTM which were used for the classification of malware attacks present in the dataset. There were 4 attacks to be classified termed as DoS, R2L, U2R and probe. Our deep learning models were able to classify all attack labels with an accuracy of more than 99%. The obtained data set was split into training and tests with a ratio of 70:30. The architecture of the models was almost the same with 4 hidden layers for CNN, 3 hidden layers for LSTM and 5 hidden layers used for CONV-LSTM. Different techniques were applied to the incoming data packets such as batch normalization and flattening etc. The activation functions were changed for almost every layer. For our models ReLu, tanh and SoftMax activation functions were applied. Adam was used as the optimizer for every model. After every deep learning model was processed, comparative analysis techniques were applied for the evaluation of the optimal model. After all the analysis it has been found that the LSTM model is the most accurate model for the classification of Network attacks.

7 Conclusion and Future Work

In cloud computing, the preventive measures to hinder network penetrations and intrusions are a crucial priority. The main goal of our project was to implement an IDS in a virtual system setting. IDS would be responsible for detecting any security breach, let it be malware or an anomaly. Our developed Intrusion detection is based on the client-server architecture where the client sends the network packet and the intrusion detection system available on the cloud server correctly identifies the type of network attack and accordingly the actions are taken on priority. Our developed intrusion detection system is 99% accurate mainly for the detection of DoS attacks, Probe attacks and Normal Traffic. After a certain set of experiments, we have identified that the LSTM model is the most optimal model for the classification of network attacks. Along with the model, a web application also has been developed, which provides the monitoring facility in real-time. When any kind of malicious attack is detected, the system generates the alarm and shows it on the user screen and also advises the most appropriate action based on the type of attack. Thus, we can say that our developed system not only enhances the security of the cloud computing environment but also helps to detect any kind of anomaly in the system available in the cloud network. Our current research mainly focuses on the detection of network attacks. However, there are various types of attacks, which are performed at the server level, called Host-based IDS. In future work, host-based log data from various cloud servers can be collected and advanced deep learning models can be deployed, to detect host-based attacks such as viruses, Trojan Horse, Worms etc. Network-based and Host-based IDS can prevent the network as well as the server of the cloud from cloud almost all attacks and make the cloud environment safer for the users.

References

Abusitta, A., Bellaiche, M., Dagenais, M., & Halabi, T. (2019). A deep learning approach for proactive multi-cloud cooperative intrusion detection system. *Future Generation Computer Systems, 98*, 308–318. https://doi.org/10.1016/j.future.2019.03.043

Ahmed, I. (2019). A brief review: Security issues in cloud computing and their solutions. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *17*(6), 2812–2817. http://telkomnika.uad.ac.id/index.php/TELKOMNIKA/article/view/12490

Aljamal, I., Tekeoğlu, A., Bekiroglu, K., & Sengupta, S. (2019). Hybrid Intrusion Detection System Using Machine Learning Techniques in Cloud Computing Environments. *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)*, 84–89. https://doi.org/10.1109/SERA.2019.8886794

Balamurugan, V., & Saravanan, R. (2019). Enhanced intrusion detection and prevention system on cloud environment using hybrid classification and OTS generation. *Cluster Computing*, *22*(6), 13027–13039. https://doi.org/10.1007/s10586-017-1187-7

Bhushan, K., & Gupta, B. B. (2017). Security challenges in cloud computing: State-of-art. *International Journal of Big Data Intelligence*, *4*(2), 81. https://doi.org/10.1504/IJBDI.2017.083116

Fontaine, J., Kappler, C., Shahid, A., & Poorter, E. D. (2020). Log-Based Intrusion Detection for Cloud Web Applications Using Machine Learning. In L. Barolli, P. Hellinckx, & J. Natwichai (Eds.), *Advances on P2P, Parallel, Grid, Cloud and Internet Computing* (pp. 197–210). Springer International Publishing. https://doi.org/10.1007/978-3-030-33509-0_18

Javadpour, A., Kazemi Abharian, S., & Wang, G. (2017). Feature Selection and Intrusion Detection in Cloud Environment Based on Machine Learning Algorithms. *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, 1417–1421. https://doi.org/10.1109/ISPA/IUCC.2017.00215

Kdd.ics.uci.edu. 2022. KDD Cup 1999 Data. [online] Available at: <http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html>

Kene, S. G., & Theng, D. P. (2015). A review on intrusion detection techniques for cloud computing and security challenges. *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, 227–232. https://doi.org/10.1109/ECS.2015.7124898

Khan, M. A. (2016). A survey of security issues for cloud computing. *Journal of Network and Computer Applications*, *71*, 11–29. https://doi.org/10.1016/j.jnca.2016.05.010

Lansky, J., Ali, S., Mohammadi, M., Majeed, M. K., Karim, S. H. T., Rashidi, S., Hosseinzadeh, M., & Rahmani, A. M. (2021). Deep Learning-Based Intrusion Detection Systems: A Systematic Review. *IEEE Access*, *9*, 101574–101599. https://doi.org/10.1109/ACCESS.2021.3097247

Mehmood, Y., Shibli, M. A., Habiba, U., & Masood, R. (2013). Intrusion Detection System in Cloud Computing: Challenges and opportunities. *2013 2nd National Conference on Information Assurance (NCIA)*, 59–66. https://doi.org/10.1109/NCIA.2013.6725325

Mishra, P., Pilli, E. S., Varadharajan, V., & Tupakula, U. (2016). Efficient approaches for intrusion detection in cloud environment. *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 1211–1216. https://doi.org/10.1109/CCAA.2016.7813926

Parampottupadam, S., & Moldovann, A.-N. (2018). Cloud-based Real-time Network Intrusion Detection Using Deep Learning. *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 1–8. https://doi.org/10.1109/CyberSecPODS.2018.8560674

Patel, A., Taghavi, M., Bakhtiyari, K., & Celestino Júnior, J. (2013). An intrusion detection and prevention system in cloud computing: A systematic review. *Journal of Network and Computer Applications*, *36*(1), 25–41. https://doi.org/10.1016/j.jnca.2012.08.007

Patil, R., Dudeja, H., & Modi, C. (2019). Designing an efficient security framework for detecting intrusions in virtual network of cloud computing. *Computers & Security*, *85*, 402–422. https://doi.org/10.1016/j.cose.2019.05.016

Sethi, K., Kumar, R., Prajapati, N., & Bera, P. (2020). Deep Reinforcement Learning based Intrusion Detection System for Cloud Infrastructure. *2020 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 1–6. https://doi.org/10.1109/COMSNETS48256.2020.9027452

Singh, P., & Ranga, V. (2021). Attack and intrusion detection in cloud computing using an ensemble learning approach. *International Journal of Information Technology*, *13*(2), 565–571. https://doi.org/10.1007/s41870-020-00583-w

Subramanian, N., & Jeyaraj, A. (2018). Recent security challenges in cloud computing. *Computers & Electrical Engineering*, *71*, 28–42. https://doi.org/10.1016/j.compeleceng.2018.06.006

Tabrizchi, H., & Kuchaki Rafsanjani, M. (2020). A survey on security challenges in cloud computing: Issues, threats, and solutions. *The Journal of Supercomputing*, *76*(12), 9493–9532. https://doi.org/10.1007/s11227-020-03213-1

vurukonda, N., & Rao, B. T. (2016). A Study on Data Storage Security Issues in Cloud Computing. *Procedia Computer Science*, *92*, 128–135. https://doi.org/10.1016/j.procs.2016.07.335

Google Colab: https://drive.google.com/file/d/1P9K2dUK3MDUp_ifAsGGJ-rW7Q2C8AoCj/view?usp=sharing