Acta Scientific COMPUTER SCIENCES

Volume 4 Issue 12 December 2022

Research Article

Adaptive Trust-based Security Model for Intrusion Detection Using Deep Learning Technique in the Cloud

Khalid Al Makdi[1,2]*, Frederick T Sheldon[1] and Soule Terence[1]

[1]Computer Science Department, University of Idaho, Moscow, USA

[2]Computer Science Department, Najran University, Najran, Saudi Arabia

***Corresponding Author:** Khalid Al Makdi, Computer Science Department, University of Idaho, Moscow, USA.

Received: April 15, 2022

Published: November 29, 2022

Abstract

With the increasing numbers of Internet-connected devices, security and privacy issues are the biggest barriers to widespread cloud systems. Securing cloud systems has become a major concern for everyone, including consumers, businesses, and the government. While attacks on any system may never be completely stopped, real-time detection of threats is essential for efficient system defence. Limited research has been done on effective intrusion detection systems for IoT (Internet of Things) environments. In this paper, the authors provide a unique intrusion detection system that detects security anomalies in cloud networks using machine learning algorithms. This trust-based security paradigm acts as a service, allowing for interoperability between the many network communications protocols used in cloud systems. The authors present the system's framework and the intrusion detection procedure in detail. The proposed intrusion detection system is tested on real network traces for proof-of-concept and simulation for scalability using Deep Learning techniques and XGboost algorithm. The results shows 96% accuracy and proves that the suggested intrusion detection system is capable of effectively detecting real-world intrusions.

Keywords: Internet of Things (IoT); Cloud Computing; Software

Introduction

One of the world's advance technology is cloud computing. It is an internet-based computer system that provides clients with on-demand access to shared resources such as software, platform, storage, and information. Cloud computing refers to a technology that allows users to access dynamically scaled and virtualized resources via the internet. Customers who use cloud computing don't have to pay for physical infrastructure, which saves money [41]. They rent resources from a third-party supplier, use them as a service, and only pay for the resources they use. Because Small and mid-size enterprises [SMEs] cannot afford the massive capital expenditures required for traditional IT, cloud computing is becoming increasingly linked with them.

To support the pay-per-use business model of Cloud Computing, the Cloud infrastructure must adapt to changing client demands and operational conditions on a continuous basis. Service-oriented paradigms, multi-domains, multi-tenancies, on-demand flexibility, and multi-user autonomous administrative infrastructures, all of which are vulnerable to cyber-attacks, are all part of this model [39]. Specifically, Cloud may be vulnerable to a number of vulnerabilities at several architectural layers (infrastructure, platform, and application) as a result of developer and service provider design, programming, or configuration problems. Malicious users can take advantage of such flaws, jeopardising the assessment of the contracted Quality of Service (QoS). Cloud computing is both a technology for more efficient use of computing infrastructures and a business model for selling computing resources and services [10].

Intruders, on the other side, find such complicated and distributed architectures appealing targets. Cyber-attacks are a severe threat that might harm the quality of service provided to clients. In this paper, the most important research topics for distributed intrusion detection in cloud systems will be discussed.

Knowledge of ML among IDS in cloud system

Machine learning is divided into two categories: supervised and unsupervised learning. The use of useful information in labelled data underpins supervised learning. The most common goal in supervised learning (and therefore in IDS) is classification; yet, manually labelling data is expensive and time intensive [13]. As a result, the fundamental bottleneck to supervised learning is a lack of sufficient labelled data. Unsupervised learning, on the other hand, recovers valuable feature information from unlabeled data, making training material much easier to come by. Unsupervised learning approaches, on the other hand, typically perform worse in terms of detection than supervised learning methods [29].

Data is shared from a variety of sources and across a variety of networks. Because this information is transferred over the internet and is subject to a multitude of attacks, intrusion detection systems are required. Because intruders are continually altering their attack techniques, traditional intrusion detection systems are unable to identify newer attempts [20]. When it comes to data storage, intrusions and threats continue to evade traditional intrusion detection mechanisms [20]. Traditional approaches for observing, classifying, and recognising incursions in the cloud, on the other hand, are resource demanding in terms of time, money, and compute [18]. Furthermore, a relatively limited methodology has been proposed previously; thus, adaptable and scattered solutions for Cloud intrusion detection systems are required. Data can be corrupted as a result of an attack, resulting in data loss or the network being knocked down. As a result, a new intrusion detection device is required, one that can identify the most recent threats and update itself on a frequent basis [25,27,42].

Because of the system's sophistication and the difficulty of regulating each access attempt, vulnerabilities and incursions would be more prevalent [35]. Various rule-based techniques have been presented over the years, but they still have issues with high computer complexity or frequent rule modifications, limiting their applicability [8]. The authors proposed to provide cyber defence using a machine learning-based intrusion detection solution to address these concerns. The effectiveness of generally available attacks like DoS/DDoS, IP spoofing, and others can be compared to normal approaches utilising publicly available datasets. Online networks are frequently targeted by intruders and hackers. To fight these attacks, various approaches can be used, and these techniques will be examined and tested in this section. One of the most deadly network assaults is the Distributed Denial of Service attack (DDoS) [21].

Furthermore, providing a distributed architecture for delivering intrusion detection in Cloud Computing, allowing Cloud providers to provide security solutions as a service. It's a multi-layer, hierarchical architecture for collecting data in the cloud, with various dispersed security components that may be utilised to do complex event correlation analysis. However, cloud-based services and applications deliver rapid developments from both academia and industry; they also pose challenges to the security and privacy of cloud storage [22]. The most of the security challenges lead to the vulnerabilities in cloud architectural components and technologies like Internet communication, service-oriented architecture, web services, software, web-browsers, virtual machines, virtualisation, self-service management interfaces, multi-tenancy, hypervisors, etc. Hence, providing security to cloud data is a tedious process. Till date, there is a lack of effective approach for offering secure access to the data in cloud due to the open nature of the data [36,38].

The manuscript is structured as follows. This chapter discusses the brief introduction of the concepts used in this paper and discusses the background of the Intrusion detection using deep learning. Chapter 2 discusses the concepts in detail, along with the related work. Chapter 3 discussed the implemented research methodology and the methods used. The results and discussion are analyzed and discussed in chapter 4. The final chapter concludes the paper with the limitation and gives the future work to the research.

Aim and Objective

The aim of the research is to detect intrusion attacks in cloud storage in order to enhance the security and privacy using pro-

posed a trust-based adaptive security system. The objectives of the research are to:

- To analyse the data available from the UGR-16 dataset in cloud system for possible associated risks.
- To process the cloud data by pre-processing the dataset and to detect various attacks for enhanced security over a period of time.
- To implement the framework of XGBoost in the application for feature selection.
- To use the deep learning algorithm to classify the intrusion data into a hierarchical structure to find the intrusion more easily and quickly.
- To evaluate the performance of the proposed framework by comparing the results with the existing techniques.

Related Work

The related work is the analysis of existing literature studies in the topic chosen. Here prior research is summarised and says how the current project is linked to it; the critical analysation of the research help to identify the gap from the prior research and it also helps for further enhancement. In the research, several models and methodologies have been designed for intrusion detection were analysed. Few recent methods have been signified in this section. In this section, have briefly reviewed some related machine learning technique used to detect intrusion in the cloud environment. The author also review some related intrusion detection system such as TSIDS, IDPS, BNID, etc.

A research by Aljamal., *et al*. [4] proposed hybrid intrusion detection system to identify attacks in cloud environment. The attacks can be known or unknown. The hybrid technique is the combination of the k-means and support vector machine classification techniques. The suggested technique is also known as network-based anomaly detection scheme. The researcher suggested the technique at the level of cloud hypervisor. In the research, UNSW-NB15 dataset is used. While comparing both the technique, the k-means clustering algorithms provide accuracy slightly higher than the SVM model. The obtained accuracy for the k-means algorithm is 88.6% and the SVM model is 84.7%. Hence the authors can conclude, the k-means algorithm is adequate when compared to the SVM model. The limitation is the precision and is not up to the mark. In future, the accuracy and precision need to improve further.

Research by Ghanshala., *et al*. [15] proposed BNID approach, that is behaviour based network intrusion detection system used to prevent the cloud environment form the network attacks. The BNID system captures the tenant virtual machine TVM communication behaviour and identifies the malicious network pattern on the network-layer. As a result, the BNID approach can identify the pre-identified and variant. The limitation is currently, the system can classify only attacks whose behavior is pre-identified and the variants of such attacks. But they intended to expand the system in the future to completely define the hidden attacks.

Moon., *et al*. [26] in the year 2016 proposed an approach to eradicate Advanced Persistent Threat attacks. This method is carried out by considering the 30 behavioural design pattern of the main user through an 83-dimensional vector. Each attribute from 83-dimensional vector signifies a particular manner of the user. To develop a database, the authors used a virtual machine environment to gather 8.7 million features from 4000 malicious and non-malicious programs. The authors Karatas., *et al*. (2018) explained that the system was developed based on the criteria that how many times a particular occurrence occur is estimated. C4.5 decision tree was utilized to develop a classifier for the gathered data, and each and every new occurrence was saved against the tree to be differentiated as malicious or normal occurrence. The system developed produces a false positive rate and false positive rate of 5.8% and 2.0%.

Abbes., *et al*. [1] proposed a method utilizing decision trees on top of the protocol analysis to develop intrusion detection systems. For every single layer of application, an adaptive decision tree was present and data occurrences were distinguished into benign and anomaly. The network was able to evidently classify the known attack types such as DoS, botnets, and scans but it was not able to handle the unidentified attack types.

Millar., *et al*. [26] proposed a method to decrease the cyber malicious attacks. A technique is developed to identify and classify such undesired traffic in the networks. The use of deep learning algorithms is in many domains as it is efficient enough to identify patterns from the huge datasets. The ways and means to classify using deep learning is always a big question. The author plans and identifies 3 different ways to introduce the data to a deep learning network for the classification of malicious traffic. Even though

in other research it is seen that deep learning was not efficient as machine learning techniques, this method uses novel deep packet inspection techniques. The level of predictions can be improvised if the packet's payload bytes have been limited to 50.

Sultana., *et al*. [40] presented machine learning based approaches that utilizes SDN to develop a network based intrusion detection system (NIDS). This study performed the detailed learning of the ways to develop a SDN-based NIDS. With the advancement in deep learning techniques, advantages of using deep learning techniques has proved its significance due to its efficacy in estimating the network security. Moreover, innovative techniques of deep learning are in play with faster capability and effective in data taxation. The authors Ponkarthika and Saraswathy [31] stated some of the challenges which are required to be considered for executing NIDS. AS most of the attacks are dynamic in nature. Henceforth, a detection method needs to be developed with the good adaptability. Building an efficient method such as selecting an optimal feature with classifier and also with minimizing the dimensionality of the dataset is a challenge. Deep learning technique is the emerging field of research in NIDS. In future to develop a significant SDN controller that can observe and implement real-time intrusion detection in high-speed networks would be a tougher task.

Research by Shams., *et al*. [37] proposed and evaluated an Intrusion Detection System (IDS) model called Trust Aware SVM Based Intrusion Detection System (TSIDS) corresponding to the monitoring of the packet header. The amalgamation of data analysis support vector machine (SVM) and modified promiscuous data collection mode for a whole IDS in Vehicular Ad Hoc Networks (VANET). In the occurrence of attacks, TDIS shows the high performance by ending malicious nodes efficiently; it also avoids attacks from the same source in the future. The authors Chockwanich and Visoottiviseth [7] illustrated that by allowing each vehicle to track its next hop in the packet routing path for three critical limitations, an SVM can be trained effectively and classify invisible information in distinct situations with a high precision of over 98% in computer-generated simulations. TSIDS performance analysis demonstrates a notable improvement compared to non-intrusion detection scenarios in the network and other accessible techniques. The improvements are in terms of End-to-End Delay (EED) and Packet Delivery Ratio (PDR) to measure the network's reliability.

Research by Kumari and Varma, [24] proposed new semi-supervised hybrid mechanism for machines learning to develop an effective IDS, this research shows a hybrid semi-supervised machine learning method using Active Learning Support Vector Machine (ASVM) and Fuzzy C-Means (FCM) clustering. This algorithm is evaluated and discovered promising on the NSL KDD benchmark IDS data set. Compared to multi-classification, binary classification is fast. Only a few labels are used, so the labelling cost is low when compared to costly SVM. The authors Gurung., *et al*. [17] proposed that detection rate over other hybrid algorithms is quite similar. When a new label of data is added, training time is lower than full-SVM. Where the small amount of data can be retrained fasterly and easily. But in SVM huge amount of data need to retain. The proposed method is Strong and accurate. In future, similar work should be compared with other classification methods too. Where here only SVM and FCM is compared. Together with this technique of classification, features selection can be considered.

Research by Farnaaz and Jabbar, [12] proposed the research on network intrusion with random forest (RF) classifier to identify four kinds of attacks such as DOS, probe, U2R, and R2L. To overcome the non-linear and complicated problem in network traffic data this enhances network intrusion system is proposed by Farnaaz. In this experiment, NSL-KDD data set is used. The technique such as 10 cross-validations, feature selection and dimensionality reduction is used to improve accuracy. Further in terms of accuracy, False alarm rate (FAR), DC and MCC, the random forest classifier is compared with the j48 classifier. The experimental results show that the proposed method improved accuracy, DR and MCC for all four types of attacks. The limitation is only four attacks were detected using this proposed method. In future, all types of attacks need to be addressed and to improve accuracy the evolutionary computation algorithm can be used to improve the accuracy.

The researchers Taghavinejad., *et al*. [43] proposed that although the value of recall in the proposed method is lower than the corresponding values, the results generally indicate that the combination of multiple decision trees has a good effect on improving the performance of intrusion detection systems in the IoTBased SG. However, since in the proposed method uses a combination of multiple trees instead of a simple tree, the time required for modeling is increased to a negligible amount, which can be neglected, because the accuracy of operation and security of network is more important.

The authors Bhosale., *et al*. [5] explained that primary thought of the intrusion detection system is to see the pernicious attacks which scare the security from the information system's ordinary exercises. The intrusion detection system can be figured fundamentally as an issue of parallel classification, with the goal that it tends to be comprehended utilizing powerful classification technique. To correct this restriction, an altered form of SVM is presented in this work. In this work, classification is finished utilizing altered SVM and assessment of the proposed technique is finished utilizing KDD dataset by leading investigations. The exploratory outcome demonstrated that the broad time is diminished utilizing changed SVM by performing legitimate dataset.

In the ANN model, Taher., *et al*. [44], the authors experimented with different number of hidden layer and found that the detection success rate varies with the number of hidden layer. After several trial and error methods, we found best detection rate with 3 hidden layers and 0.1 learning rate.

The authors Chkirbene., *et al*. [6] explained that to perform oversampling to increase the training examples of the Analysis, Backdoor, DOS, Shellcode, and Worms attack classes. As a result, we would be balancing our dataset and hence improving the detection rate of those classes and consequently the overall performance of the proposed model will increase. In addition, the authors planned to experiment using the Principle Component Analysis (PCA) to transform the highly correlated features in our dataset into a set of uncorrelated features called the principal components and use those components as new added features to our dataset. This could improve the overall performance of the system significantly.

Research gap

[3] uses conventional shallow machine learning approach for security and privacy issue in cloud. However, Aldweesh find gap in security and privacy in cloud environment, so Aldweesh suggested deep learning approaches for future enhancement. Further, research by Kumar and Goyal [23] suggested trust-based adaptive security system for security and privacy challenges in cloud. Aljamal., *et al*. [4] used hybrid method, incorporating SVM and k-means, but the accuracy is only 88.6%. So Aljamal suggested high level machine learning approach to improve the accuracy. Similarly, researcher Ghanshala., *et al*. [15] found only predefined attacks, so the researcher further proposed a framework that predicts hidden attacks that are fully specified. Whereas, Farnaaz and Jabbar (Farnaaz and Jabbar, 2016) suggested model predict only four type attacks of intrusion in cloud environment. Farnaaz therefore intended to expand the system to predict all kinds of attacks and enhance security and privacy in the cloud environment. Comparing all the research have found that the cloud environment is still not completely secure and need privacy for cloud data. Therefore it is important to use high-level machine learning methodology to solve all of the above problems. So, have suggested deep learning method to solve security and privacy issues in cloud computing.

Research methodology

This section includes a brief description of the implementation flow and methodology used to secure and to provide cloud data privacy. In the present model uses four encoder layer and decoder layer and the architecture of the data is given below in figure 1.

Dataset

One of the challenging processes is to calculate the intrusions. It is, therefore, necessary to implement and validate a dataset that contains pre-collected intrusion data features. A dataset that includes various attacks such as DDoS, SQL attacks, etc. is chosen. In this study, UGR16 will be utilized as it is extensive and also a relatively new dataset [2]. It is a structured dataset that contains both intrusion packet information as well as regular packets. The UGR 16 dataset contains updated values which are gathered from real-time back traffic and some of them are not considered as the features of other standard datasets. Source IP address, protocol, packets, flow duration and data label are the main features that will be considered from the dataset [32].

UGR'16

A Dataset for the Evaluation of Cyclostationarity-Based Network Intrusion Detection system. This data contains 80000 rows and 8 coulmns, features present in this dataset are Source_port, destination_port, duration_of_flow, and protocol, Flags, type_of_services, packets_exchanged and bytes and it is shown in figure 2 Few lines of the data is given below.

The Multi class target variable used in this dataset contains attacks such as - anamoly-spam, anamoly-udpscan, nerisbotnet,

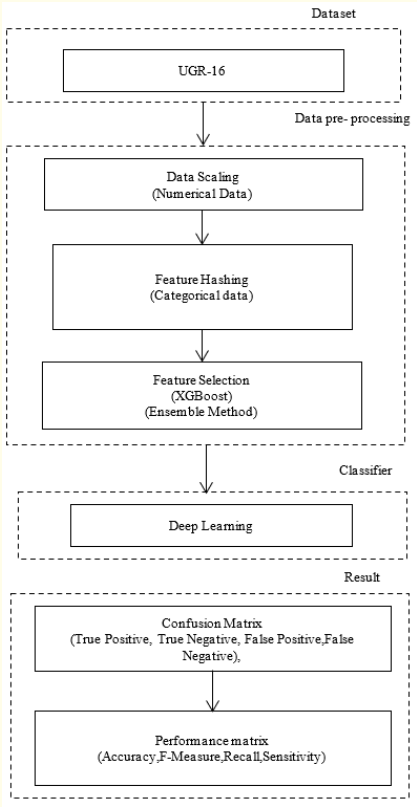

Figure 1: Implementation Flow.

data_

	source_port	destination_port	duration_of_flow	protocol	flags	type_of_Service	packets_exchanged	bytes
0	80.0	1533	0.233	1	15	0	5.0	420.0
1	1533.0	80	0.233	1	15	0	10.0	910.0
2	1567.0	80	45.365	1	14	0	18.0	2642.0
3	80.0	1567	45.365	1	10	0	4.0	190.0
4	6667.0	1586	0.291	1	6	0	1.0	48.0
...	...	...	...	...	...	...	...	...
79995	55529.0	22	1.004	1	1	0	2.0	120.0
79996	57398.0	22	1.004	1	1	0	2.0	120.0
79997	56450.0	22	1.004	1	1	0	2.0	120.0
79998	39691.0	22	1.004	1	1	0	2.0	120.0
79999	57564.0	22	1.004	1	1	0	2.0	120.0

80000 rows × 8 columns

Figure 2: List of UGR'16 Data.

scan44, scan11, anamoly-sshscan, blacklist and dos. One of these classes can be used for predicting the output.

Data pre-processing

The standard scalar will be applied for the data. The idea behind StandardScaler is that it will transform the data such that its distribution will have a mean value 0 and standard deviation of 1. In case of multivariate data, this is done feature-wise (in other words independently for each column of the data). Given the distribution of the data, each value in the dataset will have the mean value subtracted, and then divided by the standard deviation of the whole dataset (or feature in the multivariate case).

Data scaling is used to normalise the range of independent variables (Shi., *et al*. 2017). The main aim of data scaling is to achieve Gaussian with zero mean and unit variance. There are a lot of ways to do this, the two most popular are standardisation and normalisation. In the research the standardisation method will be used. If the variance of the order of magnitude of a feature is greater than the variance of other features, then that feature may dominate other features of the dataset; it should not occur in the model of learning. To normalise numerical data in all columns standard scalar technique is used [33]. Below can see the data changed due to data standardization with 80000 rows and 8 columns shown in figure 3.

dataset_2

	source_port	destination_port	duration_of_flow	protocol	flags	type_of_Service	packets_exchanged	bytes
0	-1.043166	-0.510366	-0.072673	-0.426132	1.774527	-0.616233	0.026339	-0.019892
1	-0.978375	-0.591899	-0.072673	-0.426132	1.774527	-0.616233	0.192033	-0.007064
2	-0.976859	-0.591899	0.907463	-0.426132	1.597662	-0.616233	0.457143	0.038275
3	-1.043166	-0.508458	0.907463	-0.426132	0.890201	-0.616233	-0.006799	-0.025912
4	-0.749446	-0.507392	-0.071414	-0.426132	0.182741	-0.616233	-0.106216	-0.029630
...	...	...	...	...	...	...	...	...
79995	1.429345	-0.595153	-0.055930	-0.426132	-0.701584	-0.616233	-0.073077	-0.027745
79996	1.512685	-0.595153	-0.055930	-0.426132	-0.701584	-0.616233	-0.073077	-0.027745
79997	1.470413	-0.595153	-0.055930	-0.426132	-0.701584	-0.616233	-0.073077	-0.027745
79998	0.723117	-0.595153	-0.055930	-0.426132	-0.701584	-0.616233	-0.073077	-0.027745
79999	1.520087	-0.595153	-0.055930	-0.426132	-0.701584	-0.616233	-0.073077	-0.027745

80000 rows × 8 columns

Figure 3: Data Standardization after change in data.

Below can see the Flags categorical features has more than 20 categories like APS, APRS, ARSF, APRSF, etc shown in figure 4.

Another important feature is Protocol, which is also a categorical variable using the features TCP, UDP, ICMP along with the range from 10000 to 70000 is shown in figure 5.

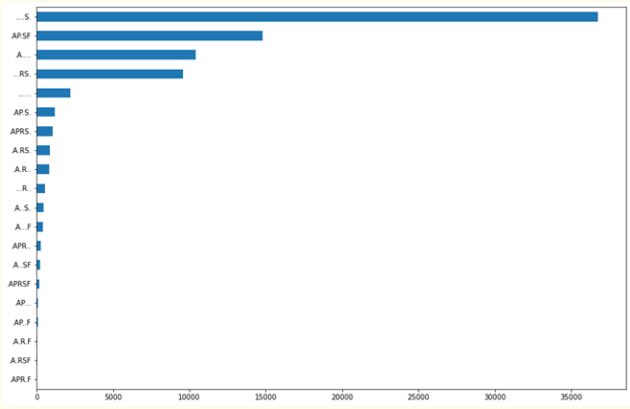

Figure 4: Flags categorical features.

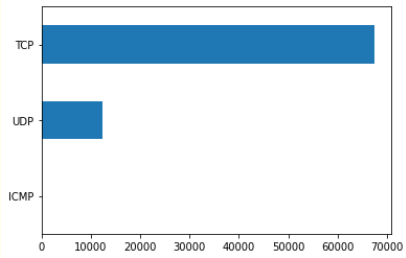

Figure 5: Categorical Variable Using the Features.

Feature selection

Feature selection technique is used to improve accuracy, reduce over fitting and to reduce training time [30]. Stacking Ensemble approach will be used for the features selection. The technique used as an ensemble feature selection is the XGBoost. The technique automatically selects the features from the data which contribute most to the prediction variable. The technique takes only a small subset of features rather than all features model.

XGboost algorithm is used as the feature selection using the standardized data. Then will be applying grid search CV for hyper parameter tuning to get better feature selection. After getting result from XGboost, have to run Feature importance function using the built model and it is shown in figure 6.

Feature hashing technique is used to pre-process the raw data effectively. The feature hashing is a smart way of modelling datasets which contain large quantities of character data and factor, where it utilises only less memory (Zhao., *et al*. 2018). Feature hashing is

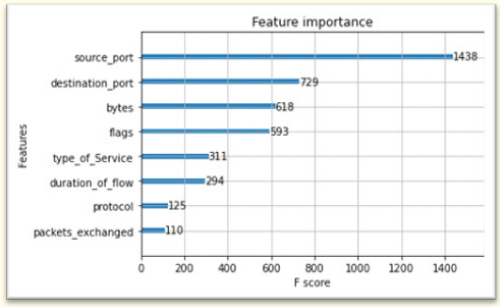

Figure 6: Feature Selection with its importance.

used to encode the categorical column. The feature hashing is also used to overcome the curse of dimensionality. This technique is called an encoding technique; it works on the exact strings, which is given as input. The categorical data provides great performance; hence it is used for feature hashing [9].

Now will remove the last 2 feature and then find accuracy of the model and also will be building another model using all the feature to see if the model increases in accuracy.

Data classifier

After the feature selection, the data traffic will pass through learning process that will collect all the relevant data, which has packet behaviour, and create a profile about the current data traffic, send and store it in the profile database. Anomaly detection will check the behaviour and properties of the packets and identify whether the packets are anomalous. If the properties of the packet are not matching to normal profile which is stored in a database, then it will be considered as anomaly. If the packet received contains normal data traffic, then it will be labelled and stored as ordinary traffic [31].

Real time implementation can be simulated with unsupervised implementation, however, we cannot measure any knid of performance metrics like accuracy in the work. Since validation is required, the performance metrics must be compared with existing work for identifying the effectiveness of the proposed algorithm. This is only possible in supervised implementation. Hence both supervised and unsupervised implementation has been performed in this work. For the unsupervised implementation, the same dataset is used, but the labels are removed in that version to act as unsupervised dataset [7].

The extracted data must be optimised for better intrusion detection. After the pre-processing technique, classification has been performed. Hence, in this work, a combination of Support Vector Machine (SVM) and Deep Neural Network (DNN) is performed for supervised implementation by stacking the classifiers. The use of random forest classifier prior to the classification (Ieracitano., *et al*. 2018). The DNN classifier is used as a supervised learning method whereas the resulted output is categorised based on the attacks. The primary objective of this approach is used to classify the new instance on the basis of training samples and number of attributes. The classifier helps to solve the classification issues via separating best data vectors into two separate classes for a given data [3]. In these regards the user can enhance the classification performance with respect to fast classification and potentially made an automated decisions for huge set of network data. After the training process, the pre-processed and feature extracted data is fed into the classifers for optimal prediction. Since the intrusions take place from different servers in realtime, the detection system has to have a high accuracy in face these live attacks [45].

Since deep networks have better computational accuracy and performance, an efficient deep network model is considered for intrusion detection in this work. The deep learning technique is used as a classifier. These methods of classification used to classify the selected features are normal features or features attacked by network intrusions.

Adam optmizer is used to compiling the model using and metric used for the model is accuracy. Next step the dataset has to be separated into training data and testing data, further the researchers have used epoch of 50, batch size of 64 and loss as mean squared error below, can see the training phase of the hidden layer is show in table 1.

DNN	Number of hidden layers=2, hidden1_ neurons=50, hidden2_neurons=50, learning_ rate=0.001 Epochs=50, Optimiser=Adam optimiser, Batch size = 64, hidden_layer1_activation=relu, hidden_layer2_ activation=relu

Table 1: Hyper Parameters for UGR'16.

Once the data is pre-processed, in order to effectively anticipate the outcomes, the data must be trained. For training the data, need a deep learning classification algorithm. The learned data is transmitted to the testing phase once it has been trained; the same two deep learning classification methods were utilized for testing. The splitting ratio in our dataset would be training 70% and testing 30%. Therefore the data is trained and tested in the ratio of 58000 and 24000. After Model training can see the accuracy rate of both training and testing dataset and is shown in figure 7.

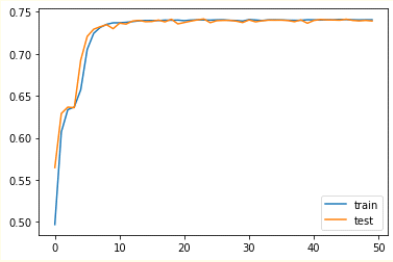

Figure 7: Rate of Training and Testing dataset.

Deep Learning is used for classification of intrusion attacks. Trusted based adaptive security model were examined in cloud services using the UGR-16 dataset and the performance metrics and confusion matrix were calculated. The Confusion matrix is a table; it is used to describe the classification model performance on the set of test data. Binary class confusion matrix calculates true positive, true negative, false positive and false negative. The Performance metric measures the performance, behaviour and activities of the intrusion in the network. Using confusion matrix the performance metric is calculated. The Accuracy, Precision, Recall, F1 Score and False positive rate are calculated using the performance metric.

Results and Discussion

The performance of suggested framework will be measured using four factors namely, false negatives (β), true negatives (δ), false positives (γ), and true positives (α) [18]. While the prediction of anomaly class, a correct classification which represents an intrusion is known as true positive (αA) and incorrect classification which represents an intrusion, i.e. there is no intrusion known as false positive (γA). Subsequently, a correct classification which

designates no intrusion is known as true negative (δA) and incorrect classification which shows no intrusion when there is an intrusion is known as false negative (βA). The probability of detecting intrusions that means true positive rate (TPR) is represented by

$$TPR = \frac{\alpha_A}{\alpha_A + \beta_A}$$

Similarly, the probability of incorrectly identifying normal behaviour as an intrusion is known as a false positive rate (FPR) which is mathematically written as,

$$FPR = \frac{\gamma_A}{\gamma_A + \delta_A}$$

The portion of total relevant records in a database that is retrieved by searching is known as recall (R), which is the same as the measure of TPR. The fraction of relevant records among the records retrieved is known as precision (P) which is represented by:

$$P = \frac{\alpha_A}{\alpha_A + \gamma_A}$$

F-score is represented by

$$F = \frac{2 * P * R}{P + R}$$

The overall accuracy is measured based on the proportion of correctly classified team data with incorrectly classified

$$Accuracy = \frac{\alpha_A + \delta_A}{\alpha_A + \delta_A + \gamma_A + \beta_A}$$

Further, the dataset will be trained with and without feature selection. First step using dataset along with 6 features the feature selection will be processed. After the process using XGBoost algorithm resulted an accuracy of approximately 85% using testing data. Results of each class from the algorithm is shown in figure 8.

```
] results
```

	Class	precision	recall	fscore
0	anomaly-spam	0.939204	0.999000	0.968180
1	anomaly-udpscan	0.864139	0.964667	0.911640
2	nerisbotnet	1.000000	1.000000	1.000000
3	scan44	0.991022	0.772667	0.868327
4	anomaly-sshscan	0.546349	1.000000	0.706631
5	blacklist	0.967960	0.191333	0.319510
6	dos	0.985790	0.994333	0.990043
7	scan11	0.986383	0.990000	0.988188

Figure 8: Results of each class from the algorithm.

Without feature selection

As the accuracy is less, so then have to proceed on to data with all the features. After training on XGBoost algorithm got an accuracy of approximately 96% using testing data. Hence without feature selection gives more accuracy than with feature selection. The results are far better without the Feature selection process and the comparison is given in table 2, figure 9.

Table 2: Comparison of the feature selection.

Feature	Accuracy %
With Feature Selection	85
Without Feature Selection	96

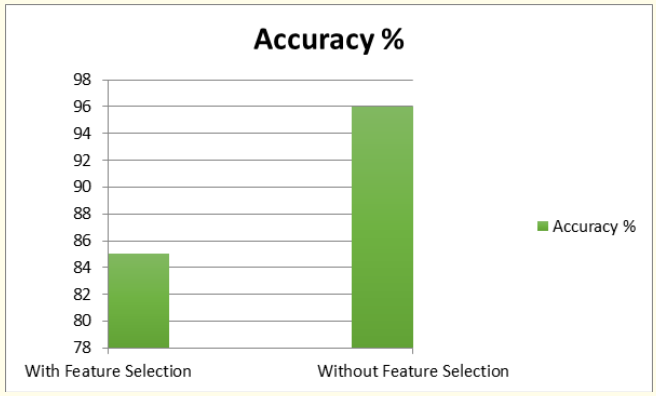

Figure 9: Bar graph for the comparison table 2.

Below are the result of each class variables along with the values class, precision, recall and f score. In which it resulted as 0.94 for precision, 0.99 for recall and 0.96 for f score and it is shown in figure 10.

The proposed technique has been evaluated using both the datasets and compared with the existing techniques and given in table 4. From the review, it has been found that few of the researchers presented a different classifier technique such as decision tree,

results

	Class	precision	recall	fscore
0	anomaly-spam	0.932272	0.995667	0.962927
1	anomaly-udpscan	0.916583	0.912000	0.914286
2	nerisbotnet	0.999334	1.000000	0.999667
3	scan44	0.987322	0.830667	0.902245
4	anomaly-sshscan	0.929656	1.000000	0.963546
5	blacklist	0.983826	0.932667	0.957563
6	dos	0.976409	0.993333	0.984798
7	scan11	0.941493	0.992333	0.966245

Figure 10: Result of Each Class Variables.

linear regression, naïve beyes, artificial neural network, GALR for intrusion detection and the obtained average accuracy rate is 85.56%, 83.15%, 82.07%, 81.34%, 81.42% respectively. The average accuracy rate for proposed method is 98.16% (UGR 16). This result clearly illustrates that the proposed method of intrusion detection gives better results with respect to the classification accuracy ratio and less false acceptance rate. Finally the comparison of the existing techniques is given in table 3.

Table 3: Comparison of the existing techniques.

Techniques	Accuracy
SVM/Naïve Bayes [16]	93.95%
ANN/SVM [44]	94.02%
Classification and Regression Trees (CART) [6]	87.74%
Hybrid three decision Tree [43]	83.14%
MNBIDS [5]	95%
Proposed Technique	96%

In order to give the clear view of the proposed work, the dataset, data pre-processing, Feature Selection Techniques, Classification algorithm and evaluation metrics in the proposed work and existing works are clearly compared in table 4.

The proposed trust-based adaptive security system is the defensive approach. It provides one of the efficient solutions for cloud data security and privacy problems [4]. The sensitive and confidential information from the data user and service providers can be secured using the proposed method. A proposed system is a security tool that captures and monitors the network traffic and system logs and scans the system/network for suspicious activities. It further alerts the system or cloud administrator about the attacks [15]. A

Table 4: Comparison of evaluation of different classification algorithms performance.

Reference	Data set	Data preprocessing Techniques	Feature Selection echniques	Classification Algorithm	Evaluation Metrics
[16]	NSL - KDD	Convert nominal attribute to binary attribute non-numeric, dimension reduction, Normalization	CfsSubsetEval	SVM Naïve Bayes	SVM -accuracy of 93.95
[44]	NSL - KDD	Reduce features	Correlation Chi-Square	ANN SVM	ANN with Wrapper (correlation) accuracy 94.02%
[6]	UNSW - NB15	Categorical features remove redundant and irrelevant features	Random Forest	Classification and Regression Trees (CART)	Accuracy 87.74
[43]	NSL - KDD	Data Normalization	CART tree	Hybrid three decision tree	Accuracy 83.1485, Precision 97.2193, recall 72.4694, F-score 83.0394
[5]	Real Time Data KDD Cup 99 dataset	Data normalization and feature extraction	Hybrid feature selection	MNBIDS	Accuracy 95%, Precis ion 98%, recall 99%
Proposed work	UGR'16	Standardization	XGboost	Deep learning	Accuracy - 96%, Precision 98- , Recall -99 , F-score - 98

feature of the proposed system is periodic monitoring and it provides behavioural patterns both for normal and abnormal activities [45].

Conclusion

The authors proposed a trust-based security model for intrusion detection systems in this study, which uses a highly scalable architecture on commodity hardware servers and can evaluate network and host-level activities. In order to handle and analyse very large scale data in real time, the framework used a distributed deep learning model. On several benchmark IDS datasets, the deep learning model was chosen after a thorough evaluation of their performance against classical machine learning classifiers [14]. Furthermore, the researchers used the suggested model with the XG boost method to detect attacks and intrusions by collecting host-based and network-based information in real-time. When comparing deep learning to traditional machine learning classifiers, have found that deep learning consistently outperformed them. In NIDS, the suggested architecture outperforms classical machine learning classifiers that have previously been deployed. To the best of knowledge, the framework could really collect the network-level and host-level actions in a distributed manner utilising deep learning to deliver more accurate protection [34]. The suggested framework's performance can be improved even more by including a module for monitoring network intrusion events. The suggested system's execution time can be improved by adding more nodes to the existing cluster.

The results show that the method provides excellent levels of accuracy, precision, and recall while also requiring less training time. We examined the model's capabilities using both benchmark datasets, which revealed a consistent degree of classification accuracy, unlike most earlier studies. Despite the fact that the model has produced the above encouraging results, has to be acknowledged that it is far from flawless and that there is still potential for improvement. The first avenue of improvement in the future work will be to examine and extend the model's capability to handle zero-day attacks. The enhanced model's merits will next be demonstrated using real-world backbone network traffic, which will be used to supplement the existing evaluations.

Bibliography

1. Abbes T., *et al*. "Efficient decision tree for protocol analysis in intrusion detection". *International Journal of Security and Networks* 5.4 (2010): 220-235.

2. Aburomman AA and Reaz MBI. "A Survey of intrusion detection systems based on ensemble and hybrid classifiers". *Computers and Security* 65 (2017): 135-152.

3. Aldweesh A., *et al*. "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues". *Knowledge-Based Systems* (2019): 105124.

4. Aljamal I., *et al*. "Hybrid Intrusion Detection System Using Machine Learning Techniques in Cloud Computing Environments". In: 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA). May 2019, IEEE (2019): 84-89.

5. Bhosale KS., *et al*. "Modified Naive Bayes Intrusion Detection System (MNBIDS)". In: 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). December 2018, IEEE (2018): 291-296.

6. Chkirbene Z., *et al*. "Hybrid Machine Learning for Network Anomaly Intrusion Detection". In: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). February 2020, IEEE (2020): 163-170.

7. Chockwanich N and Visoottiviseth V. "Intrusion Detection by Deep Learning with TensorFlow". In: 2019 21st International Conference on Advanced Communication Technology (ICACT). February 2019, IEEE (2019): 654-659.

8. Dey S., *et al*. "A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks". *Information Fusion* 49 (2019): 205-215.

9. Dhaliwal S., *et al*. "Effective Intrusion Detection System Using XGBoost". *Information* 9.7 (2018): 149.

10. Elmasry W., *et al*. "Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic". *Computer Networks* 168 (2020): 107042.

11. Elrawy MF., *et al*. "Intrusion detection systems for IoT-based smart environments: a survey". *Journal of Cloud Computing* 7.1 (2018): 21.

12. Farnaaz N and Jabbar. "Random Forest Modeling for Network Intrusion Detection System". *Procedia Computer Science* 89 (2016): 213-217.

13. Ferrag MA., *et al*. "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study". *Journal of Information Security and Applications* 50 (2020): 102419.

14. Ge M., *et al*. "Deep Learning-Based Intrusion Detection for IoT Networks". In: 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC). December 2019, IEEE (2019): 256-25609.

15. Ghanshala KK., *et al*. "BNID: A Behavior-based Network Intrusion Detection at Network-Layer in Cloud Environment". In: 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). December 2018, IEEE (2018): 100-105.

16. Gulla KK., *et al*. "Machine learning based intrusion detection techniques". In: Handbook of computer networks and cyber security. Springer (2020): 873-888.

17. Gurung S., *et al*. "Deep Learning Approach on Network Intrusion Detection System using NSL-KDD Dataset". *International Journal of Computer Network and Information Security* 11.3 (2019): 8-14.

18. Idhammad M., *et al*. "Distributed Intrusion Detection System for Cloud Environments based on Data Mining techniques". *Procedia Computer Science* 127 (2018): 35-41.

19. Ieracitano C., *et al*. "Statistical Analysis Driven Optimized Deep Learning System for Intrusion Detection". In: (2018): 759-769.

20. Iqbal S., *et al*. "On cloud security attacks: A taxonomy and intrusion detection and prevention as a service". *Journal of Network and Computer Applications* 74 (2016): 98-120.

21. Jansen W and Grance T. "Guidelines on security and privacy in public cloud computing" (2011).

22. Karatas G., *et al*. "Deep Learning in Intrusion Detection Systems". In: 2018 International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism (IBIGDELFT). December 2018, IEEE (2018): 113-116.

23. Kumar R and Goyal R. "On cloud security requirements, threats, vulnerabilities and countermeasures: A survey". *Computer Science Review* 33 (2019): 1-48.

24. Kumari V and Varma RK. "A semi-supervised intrusion detection system using active learning SVM and fuzzy c-means clustering". In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). February 2017, IEEE (2017): 481-485.

25. Lin H., *et al*. "Adaptive security-related data collection with context awareness". *Journal of Network and Computer Applications* 126 (2019): 88-103.

26. Millar K., *et al*. "Deep learning for classifying malicious network traffic". In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. 2018, Springer (2018): 156-161.

27. Mohammadi M and Al-Fuqaha A. "Enabling Cognitive Smart Cities Using Big Data and Machine Learning: Approaches and Challenges". *IEEE Communications Magazine* 56.2 (2018): 94-101.

28. Moon D., *et al*. "Host-based intrusion detection system for secure human-centric computing". *The Journal of Supercomputing* 72.7 (2016): 2520-2536.

29. Otoum S., *et al*. "On the Feasibility of Deep Learning in Sensor Network Intrusion Detection". *IEEE Networking Letters* 1.2 (2019): 68-71.

30. Peker M., *et al*. "A novel hybrid method for determining the depth of anesthesia level: Combining ReliefF feature selection and random forest algorithm (ReliefF+RF)". In: 2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA). September 2015, IEEE (2015): 1-8.

31. Ponkarthika M and Saraswathy VR. "Network intrusion detection using deep neural networks". *Asian Journal of Science and Technology* 2.2 (2018): 665-673.

32. Rajagopal S., *et al*. "A Stacking Ensemble for Network Intrusion Detection Using Heterogeneous Datasets". *Security and Communication Networks* (2020a): 1-9.

33. Rajagopal S., *et al*. "A predictive model for network intrusion detection using stacking approach". *International Journal of Electrical and Computer Engineering (IJECE)* 10.3 (2020b): 2734.

34. Riyaz B and Ganapathy S. "A deep learning approach for effective intrusion detection in wireless networks using CNN". *Soft Computing* 24.22 (2020): 17265-17278.

35. Sahmim S and Gharsellaoui H. "Privacy and Security in Internet-based Computing: Cloud Computing, Internet of Things, Cloud of Things: a review". *Procedia Computer Science* 112 (2017): 1516-1522.

36. Selvakumar B and Muneeswaran K. "Firefly algorithm based feature selection for network intrusion detection". *Computers and Security* 81 (2019): 148-155.

37. Shams E., *et al*. "Trust aware support vector machine intrusion detection and prevention system in vehicular ad hoc networks". *Computers and Security* 78 (2018): 245-254.

38. Shanhong Liu. "Size of the cloud computing and hosting market market worldwide from 2010 to 2020 (in billion U.S. dollars)*". (2018).

39. Shone N., *et al*. "A Deep Learning Approach to Network Intrusion Detection". *IEEE Transactions on Emerging Topics in Computational Intelligence* 2.1 (2018): 41-50.

40. Sultana N., *et al*. "Survey on SDN based network intrusion detection system using machine learning approaches". *Peer-to-Peer Networking and Applications* 12.2 (2019): 493-501.

41. Susilo B and Sari RF. "Intrusion Detection in IoT Networks Using Deep Learning Algorithm". *Information* 11.5 (2020): 279.

42. Szilagyi I and Wira P. "An intelligent system for smart buildings using machine learning and semantic technologies: A hybrid data-knowledge approach". In: 2018 IEEE Industrial Cyber-Physical Systems (ICPS). May 2018, IEEE (2018): 20-25.

43. Taghavinejad SM., *et al*. "Intrusion Detection in IoT-Based Smart Grid Using Hybrid Decision Tree". In: 2020 6th International Conference on Web Research (ICWR). April 2020, IEEE (2020): 152-156.

44. Taher KA., *et al*. "Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection". In: 2019 International Conference on Robotics,Electrical and Signal Processing Techniques (ICREST). January 2019, IEEE (2019): 643-646.

45. Wang Y., *et al*. "Privacy-Preserving Cloud-Based Road Condition Monitoring With Source Authentication in VANETs". *IEEE Transactions on Information Forensics and Security* 14.7 (2019): 1779-1790.

46. Wang Z. "Deep Learning-Based Intrusion Detection With Adversaries". *IEEE Access,* 6 (2018): 38367-38384.