ACTA SCIENTIFIC COMPUTER SCIENCES

Volume 3 Issue 8 August 2021

Review Article

# Intrusion Detection Using Deep Learning Techniques in the Cloud: A Survey

**Khalid Al Makdi[1,2]* and Frederick T Sheldon[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:** June 24, 2021

**Published:** July 13, 2021



## Abstract

The rapid growth of data and connectivity among computers has left the complex problem of information security. To protect data and the computer networks, *numerous* intrusion detection systems (IDS)have been developed that utilize machine learning (ML). However, many issues arise, especially since malicious attacks are constantly changing due to the huge volume of data stored in a distributed manner. This necessitates a scalable solution that incorporates effective feature extraction and a deep learning-based classification method. Due to the dynamic nature of malware and its continuously changing attack morphology,the malware signature datasets available publicly are updated systematically and benchmarked. This study presents a comprehensive review of IDS that uses deep learning and offers future research directions required to achieve a state-of-the-art IDS method: an objective with global security implications.

**Keywords:** Data Security; Deep Learning Technique; Intrusion Detection System; Cloud Database System

## Introduction

Web interaction and computer networks play a significant role in the features of affordability, resource availability, computation power, and scalable data storage. On the basis of our day-to-day activities, we depend increasingly on cloud platforms for storage and massive data collection [12]. To manage and access this enormous data volume is challenging, giving rise to critical threats, ranging from viruses to network intrusion that can cause significant data loss in numerous platforms. In this context, researchers suggest an investigation into intelligent frameworks that enhance the proficiency of the tools used in intrusion detection systems (IDS) be undertaken [13,21,31,56]. Conventional security methods, like firewalls and antivirus software, do not offer enough security over the network traffic. Threats often spread quickly, exploiting vulnerabilities within these environments [54]. IDS have been used to detect these intrusions and to prevent them from spreading throughout the data [66]. However, conventional IDS are not sufficient to manage the high volume of data now available. Thus, the massive volume of data requires more efficient frauddetection systems to manage these data efficiently and to find any intrusions [2]. To enhance system performance, self-learning capabilities that utilize machine learning (ML) have been reported that elucidate and differentiate patterns within data from multiple source streams [37]. ML is performed by example and is also able to adjust to obtain new concepts and knowledge, hence offering good potential to solve and optimize security and decision-making issues in a cloud environment. In the present study, we discuss IDS and review related deep learning techniques in a cloud environment in detail.

### Motivation for the study

The primary goal of IDS is to detect anomalous behavior, control and monitor network traffic information, and to detect network

misuse. After the evolution of Internet surveillance and threat monitoring, there was a sudden interest and importance placed on incorporating security infrastructure. The actions occurring in the system or network are measured by IDS. With the goal of better understanding and tracking network misuse over the network—more specifically, user behavior over the network—we were motivated to collect available studies, identify the most important issues in IDS from the last decades, and recommend directions for future research.

## Outline of the study

The structure of our study is as follows. Section 2 discusses the theoretical aspects of the model within the context of IDS in relation to the significant data volume, variety, integrity, visualization, velocity, and value, then summarizes the need for secure access control. Section 3 discusses the application of IDS especially in the field of IoT, smart cities, big data, fog computing, and mobile data. Section 4 summarizes important studies related to deep learning that address issues like efficiency. Finally, we have a holistic discussion through the comparison of methods in Section 5 and describe a solution and offer recommendations for future study in section 6.

## A trust-based access model of IDS: an overview

The growth of data stemming from different sources is rapid; however, unfortunately, simultaneously, the advancement of hacking tools and techniques is progressing even more quickly. In this way, there is a requirement for data security and information examination for shielding the information from the intrusion. Due to the enormous volume and rapid communication, conventional location frameworks cannot distinguish interruption more quickly. To deal with interruption proficiently, considerable information analysis methods are utilized. The extensive information from different sources is characterized as the seven *v*'s:

- **Volume:** The size of the data
- **Variety:** The varying types of the data
- **Integrity:** The trust worthiness of the data
- **Visualization:** The ease if accessibility or readability of the data
- **Velocity:** The speed at which the data are generated
- **Value:** The worth of the data
- **Variability:** The constant change in the meaning of the data.

The tremendous pace of information development makes data analysis through conventional data management frameworks obsolete because they consume time and assets to sift through this large load. The extensive volume of data makes dealing with such information profoundly challenging; remarkable advances in calculations are thus required. The IDS assume a significant role in recognizing hacker assaults by making up a framework that screens organizational traffic to discover any dubious movement and known dangers. It might likewise alert the administrator when such activity is found. To deal with and productively order the assaults, different ML calculations can be utilized.

Based on their action, intrusion detection is classified into two types, namely:

- **Passive IDS:** These IDS simply monitor and analyze the traffic and alert the administrator about the attacks and their vulnerability.
- **Active IDS:** These IDS monitor and analyze the traffic, alert the administrator, and take action against the attack by blocking the suspicious traffic.

This segment centers on other methods that are being used to recognize interruption. IDS can be an equipment framework or programming framework that naturally screens, distinguishes the assault or interruption, and alarms the PC or organization. This alarm report helps the director or client to discover and resolve the weaknesses present in the framework or organization. Some traditional methods of interruption discovery are anomaly-based intrusion detection, signature-based detection, and hybrid-based recognition. Anomaly-based intrusion detection, also called conduct-based location, models the behavior of the clients, organizations, and host frameworks. It then creates a caution or an alarm for the administrator immediately when conduct strays from standard behavior. Signature-based IDS are additionally called information-based discovery. This technique depends on an information base that contains past known assault signatures and known framework weaknesses. The hybrid-based detection system is a mix of anomaly-based intrusion detection and signature-based intrusion detection. The most crucial part of the IDS is that they correctly dis-

cover a specific interruption. Importantly, since both interruption elucidation methods have shortcomings, hybrid IDS are often used.

There are several types of intrusion detection techniques, such as host-based detection and network- based detection. Host-based intrusion detection is deployed in an individual host and employed to monitor the drive, incoming and outgoing traffic, and compare the result with a pre-created profile of the host flow activity. Host-based intrusion detection usually consists of a software agent who detects the intrusion by analyzing the host systems, system call, application logs, system directories, and other host-user activities. Network-based intrusion detection examines network traffic and monitors multiple hosts on the network to detect for any suspicious activity. This technique attempts to identify suspicious activity by analyzing the network, transport, application, and hardware layer protocols within the captured network traffic.

This research focuses on the classification of various attacks using ML algorithms. ML can be explained as improving the learning process of computers based on their experiences without being programmed. The work compares the performance of various ML algorithms, such as J.48, decision table, logistic regression, artificial neural network (ANN), modified k-means, decision tree, support vector machine (SVM), and principal component analysis (PCA) for IDS. In addition to the above ML algorithms, this research work implements algorithms, like linear discriminant analysis (LDA), classification and regression trees (CART), and random forest (RF), for classifying the intrusion detection [50].

## Application of IDS

IDS are an essential technology to shield users from cyberattack. Every transaction and information processing step occurs through the Internet, which is prone to more varied types of malicious activity compared to standard networks. Thus, there is a need to provide more information security concentration [50]. The applications of IDS covered herein are:

- The Internet of things (IoT)
- Big data
- Mobile phones
- Smart cities
- Fog.

## IDS for the Internet of things (IoT)

The IoT is a network of devices and objects with distinctive identification that can accumulate, sense, and transfer data over the Internet without human-to-computer intervention or human-to-human intervention. The IoT uses low power devices and lightweight communication protocols. For example, [18] identified IoT devices in a smart grid that were highly vulnerable to spoofing attacks by modifying sensor data. The primary types of attack on the IoT devices are side-channel attacks, cryptanalysis attacks, Sybil attacks, physical attacks, environmental attacks, and black-hole attacks. For instance, [20] suggested attack and anomaly detection be implemented using ML algorithms, such as linear regression (LR), decision trees (DTs), ANN, SVM, and random forest (RF).

## IDS for smart cities

Research by [17] has applied IDS to smart cities. They did so by collecting data from intelligent water distribution plants to detect distributed denial-of-service (DDoS) attacks in smart city applications. Their model consists of two parts: the classifier model and the restricted Boltzmann machines (RBM) model. To normal and a variety of other DDoS attacks, they applied the classification model. To learn the high-level features, they used an RBM model in an unsupervised manner. Also, they adopted different classifiers, like the automated feed formal neural network (AFFNN), SVM, FFNN, and RF. They adopted up to five layers that offer five sub-versions of each network using the clustering method, then extracted the high- level features by *k* means clustering through RBM with various *k* values. For each five datasets generated from the clustering, four types of classifiers were applied; in total, 20 experiments were completed.

## IDS in big data

As mentioned, the massive volume of information from various sources comprises a huge measure of organized, unstructured, and semi-organized information in a heterogeneous arrangement. For such an enormous volume of digital knowledge, a conventional interruption detection framework is not sufficient to handle all the security issues. However, encouragingly, IDS in a large information climate can be just conceivable by utilizing ML calculations. [45] used an Apache Spark huge information stage to highlight determination and SVM to discover interruption recognition. A pre-handled model is normalized to a unit difference in sparkle Mllib. Chisqselector and SVM are utilized to highlight the choice, and the component determination model depends on the strategy

for numTopFeatures. To diminish the impact of a misclassification, the delicate SVM edge is employed. The client characterizes a variable, called the slack variable, to exchange among the edge and the misclassification. Their outcome shows that the interruption discovery on considerable information was accomplished better and with more speed.

### IDS for fog computing

Fog computing is another innovation of processing worldview, which carries explanatory support of the edge and improves the presentation by putting the assets closer to where they are required. Mist processinghas three layers, a cloud administration layer, a haze administration layer, and a client layer. The mistadministration layer has a topographically disseminated mist hub consisting of switches, a passage, and a worker at the edge that offers an extraordinary layer in haze figuring. Haze hubs uphold heterogeneous figuring that makes the haze hub more susceptible to assault, for example, in DDoS, remote-to-local (R2L), user-to-root (U2R), and PROBE. [3] contribute to the assault cycle of DDOS in mist registering and investigate the connection between the mist hub and DDOS dependent on a hyper diagram. The condition of the mist hub is registered by the heap factor. The state of the haze hub is contrasted and the edge load level of the hub/node. Their model is utilized to investigate the relationship of fog nodes experiencing DDoS assault.

### IDS for mobile phones

Mobile phones are becoming the predominant tool among people for communication and for storing sensitive information. Mobile phone vulnerabilities include application-device vulnerability, web and content vulnerability, vulnerability, and network vulnerability. To resolve these vulnerabilities and threats, the deviceshould have IDS. [41] proposed a 5G-oriented cyber defense architecture to identify cyber threats in 5G mobile networks by using a self-adaptive deep learning-based system. They designed their architecture for classifying the intrusion by arranging the detected anomaly into two levels: a network anomalydetection (NAD) module and an anomaly symptom detection (ASD) module. Primarily, the ASD module is implemented in a two-level supervised or semi-supervised way via a deep belief network (DBN) or stacked auto-encoders (SAE). The NAD is implemented in a supervised way using long short-term memory recurrent networks (LSTM).

### Deep learning techniques

The challenges of conventional cloud computing paradigms motivated the emergence of next- generation cloud computing architectures that generate a voluminous amount of data to process beyond the capability of shallow intelligent algorithms. Deep learning algorithms, with their ability to process large-scale datasets, have recently started garnering attention from researchers to solve problems in the emerging cloud computing architectures. However, no comprehensive literature review exists on the applications of deep learning architectures to solve complex problems in emerging cloud computing architectures. To fill this gap, we conducted an extensive literature survey on the applications of deep learning architectures in emerging cloud computing architectures.

Deep learning has gained popularity due to advancements in computing capabilities by the advent of graphics processing units (GPUs), reduced hardware cost, and improved network connectivity [68]. The proliferation of training data and current research progress in machine learning and information processing are also contributing factors to the prominence of deep learning [32]. Unlike in traditional machine learning, where a domain expert is needed to assist in feature extraction, deep learning can learn features automatically from a dataset. Instead of using a manually-generated collection of rules to obtain characteristics of data, deep learning possesses the ability to understand the essential features automatically at the training phase [57]. Deep learning uses a number (tens to even hundreds) of consecutive layers—with each layer providing a more significant representation of input data [57]. It has been applied in challenging machine learning fields, like image classification, voice recognition, handwriting transcription, natural language processing, and self-driving cars [27].

Work by [53] presented a hybrid binary classification approach as a combination of a deep neural network and a *k*-nearest neighbor (DNN-kNN) search for detecting intrusion in the IoT. The suggested model performance has been evaluated using the CICIDS2017 and NSL-KDD datasets. Furthermore, they selected the significant attributes based on the data gain ratio. The simulated results show that the suggested model gives better classification accuracy: in CICIDS2017 it is 99.85% and in the NSL-KDD dataset it is 99.77%. The presented model provides a higher precision ratio for detecting intrusion in the IoT systems while considering processing cost and memory—its functionality requires low overhead, too. However, there is a need for an effective model with an attribute selection module for specific kinds of attacks toward operation at second-level intrusion detection.

A study by [52] reviewed numerous studies related to deep learning and ML approaches for intrusion detection that included a DBN; however, it failed to give a complete overview of all studies pertaining to a DBN-based IDS model. Further, they stated that there is still a need for an effective model to resolve the issues of intrusion detection that includes the stages of pre-processing, fine-tuning, and optimization of structure in a detailed manner.

Work by [52] presented a new technique for intrusion recognition and detection that can model user behaviors and improve the capability of cloud service providers using a particle swarm optimization-based probabilistic neural network (PSO-PNN). During the intrusion recognition process, user behavior information converted into an understandable and readable format. Then, it recognized and classified the malicious behavior through a multilayer neural network. The performance of the suggested model has been tested using a UNSW-NB15 dataset that classifies the various kinds of malicious behaviors exhibited by users. The results show that the proposed approach provides a better performance, especially in recognition and security monitoring of malicious behavior. In the future, they have planned to extend this scheme by adoptinga deep learning approach for vulnerabilities and zero-day attacks. Also, they stated that to create an ideal recognition system there is a need to extract different features with fewer samples.

Work by [1] presented an IDS method on the basis of a graphical machine learning approach that effectively classifies and detects intrusion. Due to the performance of Microsoft Azure, the suggested model is scalable and distributed in nature. The proposed model gives better results in terms of detecting intrusion that is simple and easy to implement. They have tested the performance using spark ML libraries, which utilize structured streaming spark data that process intrusion in real time and provide multiple,significant functionalities. The results demonstrated that it gives better results with good processing speed via cluster data. In the future, the researchers plan to use various clusters that attain faster results, then validate the results by combining two different datasets. Additionally, they will enhance the system performance that will classify, detect, and recognize intrusion by developing a deep learning method.

[58] suggested an online cloud malicious detection method using a network analysis engine (NAE) and a system analysis engine (SAE) of denial of service (DoS) attacks. Subsequently, they classified the network attack data into three different levels: 1) system-level data comprised of peak memory usage that handled several threads; 2) network-level data containing bytes, packets, and flows per address; and 3) the statistical level meta-feature analysis measured the standard deviation, mean, and variance ratio. However, the drawback of the signature-based intrusion method alone is that it is not sufficient to deal with malware features; hence, there is a need for an intelligent defender model that requires monitoring user behavior. Fortunately, some researchers have already done analyses of malware behavior. These malicious behaviors included evasive behaviors [16], network scan behaviors [24], spyware-like behaviors [34], and environmentally-sensitive behaviors (also known as trigger-based behaviors) [48].

A study by Yang and Chen suggested a hybrid intrusion detection scheme that combines both anomalyand misuse detection [61]. Based on the feature selection and packet protocol, misuse detectioncould discriminate behavior and the anomaly detection phase could find the new attacks. However, the intrusion detection rates depended on the suspected connection. Nonetheless, the researchers tested the performance using the KDD dataset with an accuracy of 95.93% leading to 183 data items being wrongly classified.

Similarly, a study by [5] proposed a neural network with a fuzzy logic approach for detecting intrusion, then processed the network data by extracting the significant features. Importantly, it removed both anomaly and misuse detection initially. A snort system compared the input behavior for detection of the network data that has historically mined the information. The misuse detection phase is then only activated if the input is suspicious. The fuzzy association rules are upgraded to describe the overlapping categories from a large dataset to create more abstract patterns. Moreover, they utilized self-organizing maps (SOM) for host-based intrusion detection to build user profiles. Work by [17] suggested a centroid-based classification for providing composite indicators and intrusion detection toward evaluating the system. The performance of the suggested model was then classified into three classification schemes: LSCANN, CANN, and the CASMN, by using the NSL-KDD dataset. Finally, the outcome was compared with traditional methods to determine classification accuracy, which varied between 97.61 and 99.74%.

Work by Rehman., *et al*. presented a hybrid approach toward detecting malware in Android applications [49]. They completed

a heuristic and signature analysis of manifest .xml files. The performance of the suggested model was compared before installation of an Android package and then the malware analysis using manifest.xml files was completed. Work by [36] explored a deep learning method for detecting malicious codes hinged on a DBN-based auto encoders approach. The simulated results showed that the suggested model gives better accuracy than previously described. Also, this method merges various kinds of log files to achieve related goals, then analyzed the user behaviors [43]. Noteworthy, a distinct limitation of this approach is not updating the training data and the security system rules. To acquire the signature data used by misuse-based IDS, anomaly detection was used. This does not require security rules and identifies attacks even when the data are incomplete; however, it fails because defining the rules is difficult—it has a high false-positive rate and for it to determine the nature of the attack is impossible—which leads to a low detection rate.

Research by [43] explored a hybrid ML approach that combined misuse and anomaly detection in a cloud environment. The simulated results showed that the hybrid model provided better intrusion solutions with an accuracy ratio of 98.77%, thus reducing the error rate by 1.47%. However, the suggested model does not cover the significance of updating security rules. In the future, they should enhance the model by minimizing the number of features, predict the values of missing data in network traffic analysis by exploiting the data value method, then weigh the evidence to determine the overall success of this approach. Furthermore, an evaluation of the model using an outlier detection approach that describes the deviation probability during the investigation of unknown attack behavior would be highly beneficial.

[19] proposed a modern intrusion detection scheme based on a hybrid consisting of a network with multilayer perceptron, an artificial bee colony, and fuzzy clustering procedures. The multilayer perceptron recognized regular and irregular bundling of network traffic. Simultaneously, its foundation was derived from utilizing the ABC algorithm by adjusting the values for biases and linking weights. The NSL-KDD dataset and CloudSim simulator were utilized to authenticate the suggested method. Evaluation metrics included the kappa value, mean absolute error (MAE), and root mean square error (RMSE). The results signposted the dominance of the suggested method in comparison to the state-of-art methods. The amalgamation of metaheuristic methods with the genetic system can further expand the performance of the system.

[56] recommended an active detection structure against invasion using a support vector machine with amplified features. More precisely, a transformation with logarithm marginal density ratios was applied to the real parts to arrive at freshly transformed features of enhanced quality. The new features significantly expanded the capacity for detection of the SVM-centered detection model. The dataset adapted forthe work was NSL-KDD; it could be seen from the assessment that the suggested approach provided healthy performance in the aspects of rate of detection, accuracy, speed of training, and number of false alarms. The comparative outcomes against prevailing methods were also superior. We do note that the suggested scheme treated only binary cases of the problem related to detecting the intrusion. Even though various classification models have been designed for detecting network intrusion, every model has its merits and demerits together with the very generally applied SVM and clustering-centered on a self-organized ant colony network (CSOACN).

A study by [62] indicated a multiple level hybrid invasion-finding model that applied a support vector machine as well as an extreme learning machine to expand the competence of identifying "known" and "unknown" attacks. A modified *k*-means algorithm was proposed to construct a superior quality dataset of training, which was added to meaningfully refine the classifiers' performance. The amended *k*-means was utilized to form new minimum datasets for training that could describe a complete unique dataset of training, which extensively reduced the time of activity of the classifiers, and advanced the performance of the system for invasion detection. The famous KDD Cup 1999 dataset was taken for calculating the planned model.The efficiency of the designated model is on par or higher with that of existing models. The extension work would focus on building a useful model using superior classifiers for sorting the revamped attacks with excellent performance and in an efficient manner. The features of multiple agent schemes could be exploited for amplifying the speed of data analysis and ease the model reinstruction for fresh attacks by expanding the system's efficiency.

[55] suggested an innovative invasion exposure scheme using a genetic algorithm that chooses features and various classifiers with support vector machines for wireless mesh networks. The

recommended strategy determines helpful features from every division of attack more efficiently than traditional classification does; that is, the standard method chooses features for all of the common attacks. The datasets for intrusion have been mimicked with a mesh network with a wireless mode using a network simulator and was used to assess and validate the suggested method with delay and packet delivery ratios as factors. Through the analytical results, it was found that the recommended approach was effective in the detection schemes of intrusion in a wireless meshed network. The created dataset and benchmark datasets were used to test the other methods for selecting the features, and the results have been matched with the suggested method. The feature selection through the genetic approach showed high values of accuracy, the complexity of computation was reduced, and the overhead of communication was minimal. Thus, the suggested scheme can well be applied to WMN.

A KNN technique categorizes every remark allocated to the class label by means of calculating the maximum confidence among the *k* datapoints adjacent to the query datapoint [16]. A KNN- centered detection scheme takes a standard network outline and considers any deviance from the standard mark as an attack. It is a robust differential evaluation for the detection scheme since it does not ask for acclimatizing factors in the training phase. The KNN approach has been utilized to model a dependable system for detecting intrusion in the network (DIDS) decided by the weirdness and separation measures of its prospective tasks, which detects the network's attacks in an effective manner [7]. Even though KNNs are often slow and require massive quantities of storage to categorize network trafficking at high speed, KNNs are still highly useful. To demonstrate their utility, supplementary methods of classification, such as fuzzy logic, DT, and regression models [6,9,11], have also been utilized to model a net abnormality detection scheme. Yet, in general, classification-type schemes for invasion detection that depend on a strong guess require that every classifier be attuned individually and continuously, thus utilizing more resources than approaches associated with statistics. When the patterns built by the suggested methodologies are not standard, it is indicative that the schemes do not have the capacity to identify the fresh attacks [44].

The present study analyses 77 articles of which 44 articles are used to review the current state of the art. These studies were collected from reputed journals indexed in Google Scholar, Springer, Elsevier, IEEE Access, and ScienceDirect. In addition, the cited 77 articles have been published recently and are specific to IDS. From this body, we note the major limitations of our review are not providing an update on the training data and the security system rules. To acquire the signature data used by misuse-based IDS, anomaly detection has been used; thus, these data do not require security rules and identify attacks even when the data are incomplete. Anomaly detection fails due to its definition of rules, its high false-positive rate, and selection based on the nature of the attack is impossible, leading to a lower detection rate. Also, there is a lack of security in the IDS model. Hence, there is a need for an effective model at the system security level to recognize and detect malicious behavior [48].

## Discussion

Work by [10] presented a deep learning method for distributing intelligent video surveillance in edge computing. The model consists of two-level edge nodes with 200 monitoring terminals placed at the meeting points of 30 streets, 35 edge servers, and a cloud server. The connection between monitoring terminals and the edge nodes is achieved via the high-speed Gigabit Ethernet. The study employs ConvNet and LSTM on every edge node to conduct varying analyses concurrently. ConvNet was applied to extract features of vehicles from videos to make a classification in traffic monitoring, and the LSTM was used for traffic flow prediction. A dynamic data migration technique was also integrated into the algorithm to maximize load balancing. The experiment proved that LSTM and ConvNet were efficient in performing surveillance; however, the algorithm consumed large memory and bandwidth.

[39] proposed an LSTM for channel prediction in connected vehicles in the edge computing platform. The LSTM learns from past channel information and uses knowledge to predict future channel information. Channel prediction is essential to achieving high reliability and low latency in vehicle-to-vehicle (V2V) communication. The LSTM performance was compared with ARIMA and support vector regression (SVR). It was found that the LSTM performs better than the compared algorithms, but the algorithm can only handle short-term dependencies [27].

[47] presented an LSTM-based DDoS attack detection algorithm in the fog environment. Their platform consisted of a Cent OS7 cloud server with a fog layer that had a few virtual machines with

an Apache server SDN. Multiple attackers and legitimate virtual machines running Windows and Linux OS form the application layer. The LSTM was employed because it is robust in managing time-dependent and sequential data. The LSTM is used in a fog server to detect and forward valid packets to the cloud server. Suspicious packets are prevented from reaching the cloud, which protects the entire fog environment from malicious attack. In comparison with stacked auto encoder detection (SAED) and LSTM-1, the proposed LSTM proved superior. Nonetheless, huge memory is devoured by the LSTM. In 2015, a machine learning algorithm [51] was applied in a network intrusion detection system, improving classification performance, making the system more intelligent and efficient. However, its accuracy needs to befurther enhanced; the false alarm rate was also high.

In [40], a distributed IDS framework is presented, where multiple instances cooperate to counter DoS and DDoS attacks. Specifically, the IDS exchange alerts and determine whether to accept the signs sent from the other IDS. However, the evaluation does not show a significant improvement in accuracy or detection time (conversely, it needs a little more computational effort compared with Snort), but focuses instead on improving system reliability to avoid a single point of failure.

A summary of the existing studies on IDS has been tabulated in table 1. The detection method, dataset,and the results obtained have been separated to better understand the study.

| Research | Detectionmethod | Method used for implementation | Dataset | Results |
|---|---|---|---|---|
| Mehmood (2016) | Intrusion detection likeDoS, U2R and R2L | SVM, Naive Bayes, decision tables, j.48. | KDD99 | The j.48 algorithm achieves better performance even under the redundant features among all other algorithms. |
| Yin., *et al*. (2017) | Intrusiondetection | RNN | NSL-KDD, binary, 5-class classification | 5-class classifier: accuracy = 81.29% |
| Le., *et al*.(2017) | Intrusiondetection | LSTM | KDD 99, 5-class classification | Accuracy = 97.54%; Rec = 98.95%; precision = 97.69%; FAR = 9.98% |
| Jan., *et al*.(2019) | DoS/DDoSattack | SVM | Simulated | Lightweight IDS for the IoT achieved. Experiments showthat packet arrival rate and SVM classifier is enough todetect intrusions |
| Maim., *et al*. (2018) | Anomaly detection | Deep belief network, stackedauto-encoder, L | Botnet | Anomaly detection accuracyachieved in 5G network by using a two-level deeplearning algorithm. |
| Elsaeidy., *et al*. (2019) | DDoS attack | Deep RBM, FFNN, SVM, *k*-means, FFNN, automatedRF. | Smart water distribution | Automated FFNN outperforms all other algorithms. |
| Hasan., *et al*. (2019) | Attack andanomaly detection | DF, ANN, Logistic Regression, SVM. | Kaggle.com | DF is an excellent technique to use in the IoT for IDS with accuracy of 99.4% |
| Jiang., *et al*. (2018) | IDS | LSTM | NSL-KDD, binary classification | Accuracy = 98.94%; FAR =9.86%; recall = 99.23%. |

| | | | | |
|---|---|---|---|---|
| Kolias., *et al*. (2016) | IDS | Adaboost, J48,OneR, RT, Hyperpipes, NB, RF, ZeroR | AWID | Introduced AWID dataset andgood accuracy in injection attack. However, this model has a lower detection rate of impersonation attack |
| Zhou., *et al*. (2019) | IDS | Ensemble classifier | KDDCUP'99, CIC-IDS2017, NSL-KDD. | Ensemble classifier achieved high accuracy. Unable to address atypical attacks from abundant network traffic. |
| Jan., *et al*.(2019) | IDS | SVM polynomial kernel-based | CICIDS2017, beget, a data-set via Poisson distribution | Lightweight IDS with a concomitant effect on traffic intensity; unable to recognizein-trusions that do not possess a concomitant impact on thetraffic rate |
| Rahman., *et al*. (2020) | IDS | SAE, CFS, information Gain,MLP, SVM, J48, OneR. | AWID | Achieved higher accuracy andthe CPU time is significantly reduced; can only detect impersonation attacks |

**Table 1:** Summary of previous studies related to IDS.

| Research | DL structure | Results | Limitation |
|---|---|---|---|
| Liu., *et al*. (2018) | ConvNet | ConvNet performed better than the C-System and D-System | Low response time |
| Pang., *et al*. (2019) | I-ConvNet | The I-ConvNet outperformed the compared algorithms, except for the Spindle | Does not minimize energyconsumption |
| Li., *et al*. (2018) | ConvNet | The ConvNet outperformed contour andpixel-based algorithms | The algorithm was only testedwith a single type of product |
| Blanco- Filgueira., *et al*. (2019) | ConvNet | ConvNet performed well in terms of latency, cost, and power consumption | The effectiveness of the proposed approach is challenging to measure due to the lack of comparison |
| Kijsipongse., *et al*. (2018) | ConvNet | ConvNet performed well | Lack of direct communication tovolunteer nodes |
| Zhang and Zheng (2019) | DQN | The DQN outperformed DP and NMA | Data security technique duringtask migration not employed |
| Huang., *et al*. (2019) | DQN | DQN performed better than the compared algorithms (local processingscheme, edge processing scheme, greedy scheme, and MUMTO algorithm) | Increase in learning rategenerates local optimum |

| | | | |
|---|---|---|---|
| Li., *et al.* (2019) | DDQN | The DDQN is more effective than the compared algorithms (RLU, FIFO, and LFU) | An increase in the volume of content decreases the number of hits |
| Ishakian., *et al.* (2018) | DNN | DNN tasks can be suitably run in aserverless platform | Delay at the cold start mightskew latency and affect SLA |
| Diro and Chilamkurti (2018) | DNN | DNN is more effective in attack detection than the compared DNN-Softmax | The algorithm has a slowlearning process |
| Jian., *et al.* (2019) | LSTM-ICBS | The LSTM-ICBS outperformed thecompared algorithms | Low accuracy is observed for tasks with a very small length |
| Liu., *et al.* (2019) | LSTM | LSTM outperformed the ARIMA and SVR | Limited to handling short termdependencies |

**Table 2:** Summary of deep learning techniques in cloud computing.

| Algorithm | Suitability | Strength | Limitation |
|---|---|---|---|
| Convolutional neural network | Image | Requires fewer neuronconnections | Sometimes requires many layersto extract a hierarchy of features |
| Deep reinforcement learning | Decisionmaking | Labelled data is not required | Trade-off between exploitationand exploration; policy evaluation and comparison is challenging |
| Recurrent neural Network | Speech/video | Memorizes sequential events;models time dependencies | Vanishing and exploding gradient; extremely difficult totrain |
| Long short-term memory | Time series | Suitable for data when long time interval memory cells aresecured by gates | Only handles short term dependencies; high memory andbandwidth requirement |
| Deep neuralnetwork | Regression | Dimension reduction | The learning process is sometimes slow; computationally intensive |
| Deep beliefnetwork | variety | Labelled data are not requiredfor training; avoids overfittingand underfitting; adopts layer-by-layer greedy learning approach for initializingnetwork | Initialization and sampling process render training to becomputationally expensive |

**Table 3:** Summary of deep learning strength, limitations, and suitability.

de Souza CA., *et al.* [53] presented a DNN-kNN for detection of intrusion in the IoT and attained better results than had been previously reported. Nonetheless, improvement is needed in performance by integrating the attribute selection module for detecting specific kinds of attacks toward operating at second-level intrusion. A study by [52] reviewed numerous studies related to deep learning and ML in intrusion detection that included the DBN; however, it failed to give a complete overview of the tasks

| Research | DBN structure | Scope/aim | Dataset | Results |
|---|---|---|---|---|
| Ding., *et al*. (2016) | Input: 20~400; hidden: 200-200-50; output: 2 | Applied DBN-IDS to malwaredetection | Real benign; maliciousfiles | Accuracy = 96.1% using 400features |
| Kang and Kang (2016) | unknown | Applied DBN-IDSto in-vehicular network | OCTANE data set | Accuracy = 97.8% |
| Zhao., *et al*. (2017) | Input: 122; hidden: 90-21-17; output: 2 | Used PNN for detection module;used PSO to find optimal # of units | KDD Cup 99 | Data detection rate = 99.14%; accuracy = 93.25%; FAR = 0.615% |
| Qu., *et al*. (2017) | Input: 122; hidden: 60-60-100; output: 2 | Detailed analysis of DBN-IDS withNSL-KDD | NSL-KDD | Accuracy = 95.25% |
| Huda., *et al*.(2018) | Input: NA; hidden 1: 30~44; hidden 2: 30~44; output: 2 | Applied DBN-IDSto the IoT ICS | Real data Vx-heaven | Accuracy = 99.75%; FPR = 0.29%; FNR = 0.22% |
| Yang., *et al*.(2019) | NSL-KDD: 122-40-20-10-5; UNSW-NB15: 196-80-40-20-10 | Used MDPCA to optimize trainingdata using fuzzy algorithm with *k* sub-DBNs | NSL-KDD UNSW- NB15 | Accuracy = 90.21%; detection rate = 96.22%;precision = 87.3%; FPR = 17.15% for NSL-KDD |
| Wei., *et al*.(2019) | Input: 122; hidden: 75-33-18-12 output: 2 or 5 | Used PSO, AFSA, and GA to optimize # of hidden layers and # of units | NSL-KDD | KDDTest + Accuracy = 82.36%; KDDTest-21 Accuracy = 66.25% for five category case |
| Zhang., *et al*.(2020) | Input: 41; hidden: 30-20-10; output: 5 | Used GA for real-time attack detection; used DBN for attack type classification | CICIDS2017 | Precision: 97.74%; recall: 97.67% |

**Table 4:** Summary of previous work related to deep learning.

related to the DBN-based IDS model. Also, they have stated that there is still a need for an effective model to resolve the issues of intrusion detection that include stages of pre-processing, fine-tuning, and optimization of structure in a detailed manner. [52] presented PSO-PNN for intrusion detection. They suggested that they will extend this scheme by adopting a deep learning approach for vulnerabilities and zero-day attacks. Also, they stated that PSO-PNN needs to extract different features with fewer samples, which would lead to the ideal recognition system.

Work by [1] presented an IDS method on the basis of a graphical ML approach that effectively classifies and detects intrusion. The results demonstrated that it gives good processing speed via cluster data. In the future, they plan to use multiple clusters to attain faster results, then validate the results by combining two different sets of data. Additionally, they will enhance the system performance to classify, detect,and recognize intrusion by developing a deep learning method. Work by [36,43] explored a deep learning method in the detection of malicious codes on the basis of a DBN- based auto-encoders approach. The simulated results show that the suggested model offers better accuracy; however, a limitation to this approach is the failure to update the training data and the security system rules. To acquire the signature data used by misuse-based IDS, anomaly detection has been used, which does not require security rules and identifies the attacks even when the data are incomplete. Nonetheless, it fails due toa lack of definition of the rules, a high false-positive rate, and selecting the nature of the attack is impossible—leading to a reduced detection rate.

From our survey here, we observed that there are still gaps in security in the IDS model. Thus, there isa need for an effective model at the system security level to recognize and detect malicious behavior [48]. For instance, the earlier available security systems offered a lack of preventive capabilities, self- optimization in self-configuration, and learning [4,58]. Even though some of theresearchers attained high-performance results for some particular types of attacks, like DoS, they are not applicable to every single type of attack [11]. Also, there is a lack of consensus among detectors in determining which patterns qualify as anomalous. There needs to be a universal model that can recognize zero-day attacks and new attacks [48].

## Conclusion and Future Scope

This review presents a detailed overview of the underlying security models for IDS frameworks that use artificial intelligence, including ML and deep learning techniques in the context of cloud computing. Most studies we surveyed utilized ML techniques, while only a few employed deep learning methods. However, thereare still issues during the detection of intrusion, especially while analyzing and classifying the massive access to network data. There is an overwhelming need for an operational framework that can handle a massive event,extract the feature effectively, then classify it to enhance the system prediction performance.

Based on our thorough analysis, the directions for future work are now availed:

- To use a better feature selection method (e.g., random classifier-RF, wrapper model, principal component analysis) that classifies the intrusion by extracting significant features from a vast collection of network data.
- To exploit the numerous features of a massive set of data to facilitate model retraining on intrusion or attack classification, then speed up data analysis by enhancing the efficiency of the network attack detection methods. Also, we await improvement in classification performance by obtaining an optimal solution (by integrating the meta-heuristic optimization method) that offers fast execution time and high classification accuracy using an effective optimization method.
- There is a need for effective deep learning techniques with DT or a naive bias classifier that can handle data overfitting and generally resolve multi-attribute decision-making problems.
- To increase the number of test systems and runtimes linearly by creating an effective optimization method, while considering a meta-heuristic optimization method. Further, we aim to prove models' validity by considering more experiments and efficient deep learning models for dealing with numerous attributes with lower classification error rates.
- To explore a U-Net-based deep learning model to learn the weights of the loss functions and the network's parameter separately. In this regard, we may improve the system prediction performance through less execution time. Finally, we seek to integrate security rules that will minimize detection of false positives, then determine the nature of the attack.

## Bibliography

1. Abid A and Jemili F. "Intrusion Detection based on Graph oriented Big Data Analytics". *Procedia Computer Science* 176 (2020): 572-581.

2. Alsafi HM., *et al*. "IDPS: An Integrated Intrusion Handling Model for Cloud" (2012).

3. An X., *et al*. "Hypergraph clustering model-based association analysis of DDOS attacksin fog computing intrusion detection system". *EURASIP Journal on Wireless Communications and Networking* 1 (2018): 1-9.

4. Bai J., *et al*. "Malware detection method based on the control-flow construct feature of software". *IET Information Security* 8.1 (2014): 18-24.

5. Bashah N., *et al*. "Hybrid intelligent intrusion detection system". *World Academy of Science, Engineering and Technology* 11 (2005): 23-26.

6. Bhuyan MH., *et al*. "Network Anomaly Detection: Methods, Systems and Tools". *IEEE Communications Surveys and Tutorials* 16.1 (2014): 303-336.

7. Bhuyan MH., *et al*. "Network Traffic Anomaly Detection and Prevention". Computer Communications and Networks. Cham: Springer International Publishing (2017).

8. Blanco-Filgueira B., *et al*. "Deep Learning-Based Multiple Object Visual Tracking on Embedded System for IoT and Mobile Edge Computing Applications". *IEEE Internet of Things Journal* 6.3 (2019): 5423-5431.

9. Chandola V., *et al*. "Anomaly detection". *ACM Computing Surveys* 41.3 (2009): 1-58.

10. Chen J., *et al*. "Distributed deep learning model for intelligent video surveillance systems with edge computing". *IEEE Transactions on Industrial Informatics* (2019).

11. Corona I., *et al*. "Adversarial attacks against intrusion detection systems: Taxonomy, solutions and open issues". *Information Sciences* 239 (2013): 201-225.

12. da Costa K., *et al*. "Internet of Things: A survey on machine learning-based intrusion detection approaches". *Computer Networks* 151 (2019): 147-157.

13. 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.

14. Ding Y., *et al*. "Application of deep belief networks for opcode based malware detection". In: 2016 International Joint Conference on Neural Networks (IJCNN) IEEE (2016): 3901-3908.

15. Diro AA and Chilamkurti N. "Distributed attack detection scheme using deep learning approach for Internet of Things". *Future Generation Computer Systems* 82 (2018): 761-768.

16. Dua S and Du X. "Data Mining and Machine Learning in Cybersecurity". 1st Ed. Auerbach Publications.

17. Elsaeidy A., *et al*. "Intrusion detection in smart cities using Restricted Boltzmann Machines". *Journal of Network and Computer Applications* 135 (2019): 76-83.

18. Ghasempour A. "Internet of Things in Smart Grid: Architecture, Applications, Services, Key Technologies, and Challenges". *Inventions* 4.1 (2019): 22.

19. Hajimirzaei B and Navimipour NJ. "Intrusion detection for cloud computing using neural networksand artificial bee colony optimization algorithm". *ICT Express* 5.1 (2019): 56-59.

20. Hasan M., *et al*. "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches". *Internet of Things* 7 (2019): 100059.

21. Hosseini Bamakan SM., *et al*. "An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization". *Neurocomputing* 199 (2016): 90-102.

22. Huang L., *et al*. "Deep reinforcement learning-based joint task offloading and bandwidth allocation for multi-user mobile edge computing". *Digital Communications and Networks* 5.1 (2019): 10-17.

23. Huda S., *et al*. "A malicious threat detection model for cloud assisted Internet of things (CoT) based industrial control system (ICS) networks using deep belief network". *Journal of Parallel and Distributed Computing* 120 (2018): 23-31.

24. Inoue D., *et al*. "Automated Malware Analysis System and Its Sandbox for Revealing Malware's Internal and External Activities". *IEICE Transactions on Information and Systems* (2018): E92-D (5) 945-954.

25. Ishakian V., *et al*. "Serving deep learning models in a serverless platform". In: 2018 IEEE International Conference on Cloud Engineering (IC2E). IEEE (2018): 257-262.

26. Jan SU., *et al*. "Toward a Lightweight Intrusion Detection System for the Internet of Things". *IEEE Access* 7 (2019): 42450-42471.

27. Jauro F., *et al*. "Deep learning architectures in emerging cloud computing architectures: Recent development, challenges and next research trend". *Applied Soft Computing* 96 (2020): 106582.

28. Jian C., *et al*. "An Improved Chaotic Bat Swarm Scheduling Learning Model on Edge Computing". *IEEE Access* 7 (2020): 58602-58610.

29. Jiang F., *et al*. "Deep learning based multi-channel intelligent attack detection for data security". *IEEE transactions on Sustainable Computing* (2018).

30. Kang MJ and Kang JW. "Intrusion detection system using deep neural network for in-vehicle network security". *PloS one* 11.6 (2016): e0155781.

31. Karsligil E., *et al*. "Network intrusion detection using machine learning anomaly detection algorithms". In: 25th Signal Processing and Communications Applications Conference (SIU). IEEE (2017): 1-4.

32. Khan M., *et al*. "Deep Learning: Convergence to Big Data Analytics". SpringerBriefs inComputer Science. Singapore: Springer Singapore (2019).

33. Kijsipongse E., *et al*. "A hybrid GPU cluster and volunteer computing platform for scalable deep learning". *The Journal of Supercomputing* 74.7 (2018): 3236-3263.

34. Kirda E., *et al*. "Behavior-based Spyware Detection". In: Usenix Security Symposium 2006 (2006): 694.

35. Kolias C., *et al*. "Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset". *IEEE Communications Surveys and Tutorials* 18.1 (2016): 184-208.

36. Li L., *et al*. "Deep Learning for Smart Industry: Efficient Manufacture Inspection System With Fog Computing". *IEEE Transactions on Industrial Informatics* 14.10 (2018): 4665-4673.

37. Li W., *et al*. "Edge caching for D2D enabled hierarchical wireless networks with deep reinforcement learning". *Wireless Communications and Mobile Computing* (2019).

38. Liu C., *et al*. "A New Deep Learning-Based Food Recognition System for Dietary Assessment on An Edge Computing Service Infrastructure". *IEEE Transactions on Services Computing* 11.2 (2018): 249-261.

39. Liu G., *et al*. "Deep Learning-Based Channel Prediction for Edge Computing Networks Toward Intelligent Connected Vehicles". *IEEE Access* 7 (2019): 114487-114495.

40. Lo CC., *et al*. "A cooperative intrusion detection system framework for cloud computing networks". In: 2010 39th International Conference on Parallel Processing Workshops IEEE (2018): 280-284.

41. Maim LF., *et al*. "A self-adaptive deep learning- based system for anomaly detection in 5G networks". *IEEE Access* 6 (2018): 7700-7712.

42. Mehmood T and Rais HBM. "Machine learning algorithms in context of intrusion detection". In: 3rd International Conference on Computer and Information Sciences (ICCOINS). IEEE (2016): 369-373.

43. Meryem A and Ouahidi B EL. "Hybrid intrusion detection system using machine learning". *Network Security* 5 (2020): 8-19.

44. Moustafa N., *et al*. "A holistic review of Network Anomaly Detection Systems: A comprehensive survey". *Journal of Network and Computer Applications* 128 (2019): 33-55.

45. Othman SM., *et al*. "Intrusion detection model using machine learning algorithm on Big Data environment". *Journal of Big Data* 5.1 (2018): 34.

46. Pang S., *et al*. "An Improved Convolutional Network Architecture Based on Residual Modeling for Person Re-Identification in Edge Computing". *IEEE Access* 7 (2019): 106748- 106759.

47. Priyadarshini R and Barik RK. "A deep learning based intelligent framework to mitigate DDoS attack in fog environment". *Journal of King Saud University-Computer and Information Sciences* (2019).

48. Rahman MA., *et al*. "Scalable machine learning-based intrusion detection system for IoT-enabled smart cities". *Sustainable Cities and Society* 61 (2020): 102324.

49. Rehman ZU., *et al*. "Machine learning-assisted signature and heuristic-based detection of malwares in Android devices". *Computers and Electrical Engineering* 69 (2018): 828-841.

50. Saranya T., *et al*. "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review". *Procedia Computer Science* 171 (2020): 1251-1260.

51. Singh R., *et al*. "An intrusion detection system using network traffic profiling and online sequential extreme learning machine". *Expert Systems with Applications* 42.22 (2015): 8609-8624.

52. Sohn I. "Deep belief network based intrusion detection techniques: A survey". *Expert Systems with Applications* (2020): 114170.

53. de Souza CA., *et al*. "Hybrid approach to intrusion detection in fog-based IoT environments". *Computer Networks* 180 (2020): 107417.

54. Subramanian N and Jeyaraj A. "Recent security challenges in cloud computing". *Computers and Electrical Engineering* 71 (2018): 28-42.

55. Vijayanand R., *et al*. "Intrusion detection system for wireless mesh network using multiple support vector machine classifiers with genetic-algorithm-based feature selection". *Computers and Security* 77 (2018): 304-314.

56. Wang H, *et al*. "An effective intrusion detection framework based on SVM with feature augmentation". *Knowledge-Based Systems* 136 (2017): 130-139.

57. Wani MA., *et al*. "Advances in deep learning". Springer (2020).

58. Watson MR., *et al*. "Malware Detection in Cloud Computing Infrastructures". *IEEE Transactions on Dependable and Secure Computing* 13.2 (2016): 192-205.

59. Wei P., *et al*. "An optimisation method for intrusion detection classification model based on deep belief network". *IEEE Access* 7 (2019): 87593-87605.

60. Yang J., *et al*. "HIDS-DT: An effective hybrid intrusion detection system based on decision tree". In: 2010 International Conference on Communications and Mobile Computing IEEE (2010): 70-75.

61. Yang Y., *et al*. "Building an effective intrusion detection system using the modified density peak clustering algorithm and deep belief networks". *Applied Sciences* 9.2 (2019): 23.

62. Al Yaseen WL., *et al*. "Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system". *Expert Systems with Applications* 67 (2017): 296-303.

63. Yin C., *et al*. "A deep learning approach for intrusion detection using recurrent neural networks". Ieee Access 5 (2017): 21954-21961.

64. Zhang C and Zheng Z. "Task migration for mobile edge computing using deep reinforcement learning". *Future Generation Computer Systems* 96 (2019): 111-118.

65. Zhang H., *et al*. "A real-time and ubiquitous network attack detection based on deep belief network and support vector machine". *IEEE/CAA Journal of Automatica Sinica* 7.3 (2020): 790-799.

66. Zhang Z and Meddahi A. "Intrusion Prevention and Detection in NFV". In: Security in Network Functions Virtualisation Elsevier (2017): 157-172.

67. Zhao G., *et al*. "Intrusion detection using deep belief network and probabilistic neural network". In: 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). IEEE (2017): 639-642.

68. Zhao R., *et al*. "Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing* 115 (2019): 213-237.

69. Zhou Y., *et al*. "An efficient intrusion detection system based on feature selection and ensemble classifier". *arXiv preprint arXiv:1904.01352* (2019).

**Volume 3 Issue 8 August 2021**