Computers, Materials & Continua
DOI:10.32604/cmc.2022.020141
Article

Tech Science Press

SIMAD: Secure Intelligent Method for IoT-Fog Environments Attacks Detection

Wided Ben Daoud[1] **and Sami Mahfoudhi**[2,*]

[1]NTS'Com Research Unit, Sfax, Tunisia
[2]Department of Management Information Systems and Production Management, College of Business and Economics, Qassim University, Buraidah, 51452, KSA
[*]Corresponding Author: Sami Mahfoudhi. Email: s.mahfoudhi@qu.edu.sa

Received: 11 May 2021; Accepted: 22 June 2021

Abstract: The Internet of Thing IoT paradigm has emerged in numerous domains and it has achieved an exponential progress. Nevertheless, alongside this advancement, IoT networks are facing an ever-increasing rate of security risks because of the continuous and rapid changes in network environments. In order to overcome these security challenges, the fog system has delivered a powerful environment that provides additional resources for a more improved data security. However, because of the emerging of various breaches, several attacks are ceaselessly emerging in IoT and Fog environment. Consequently, the new emerging applications in IoT-Fog environment still require novel, distributed, and intelligent security models, controls, and decisions. In addition, the ever-evolving hacking techniques and methods and the expanded risks surfaces have demonstrated the importance of attacks detection systems. This proves that even advanced solutions face difficulties in discovering and recognizing these small variations of attacks. In fact, to address the above problems, Artificial Intelligence (AI) methods could be applied on the millions of terabytes of collected information to enhance and optimize the processes of IoT and fog systems. In this respect, this research is designed to adopt a new security scheme supported by an advanced machine learning algorithm to ensure an intelligent distributed attacks detection and a monitoring process that detects malicious attacks and updates threats signature databases in IoT-Fog environments. We evaluated the performance of our distributed approach with the application of certain machine learning mechanisms. The experiments show that the proposed scheme, applied with the Random Forest (RF) is more efficient and provides better accuracy (99.50%), better scalability, and lower false alert rates. In this regard, the distribution character of our method brings about faster detection and better learning.

Keywords: Attack detection; Fog; IoT network; machine learning; distributed mechanism

1 Introduction

Nowadays, IoT networks are encountering enormous development and the number of IoT devices is expected to reach the number of 50 billion devices by the end of 2021 [1]. Hence, these networks deal with a large volume of data that needs to be put away efficiently, and which then should be allocated to consumers in a secure and protected manner [2,3].

A key solution for data treatment and storage is to the usage of cloud technologies [4]. These systems offer numerous functionalities for computing resources like effective information storage, high-speed connection to internet [5,6]. However, the enormous amount of processed, analyzed, and filtered information in the cloud can cause several issues in relation with latency, network traffic blockage, and also security and privacy concern [6–8]. These problems are manifested due to the inability of cloud to support heterogeneous systems and devices, mobility, delay, and geo-localization.

In this context, fog computing can be envisaged as a key solution for these limitations. In fact, fog paradigm is an emerging computing technology projected in order to reduce traffic congestion and delay production at the cloud network by moving several resources and functionalities nearer to the IoT devices and then to users. Thanks to these characteristics, these fog nodes form a network with distributed services like storage, data analysis, and computation in real time with reduced latency. Fig. 1 shows the fundamental architecture of IoT-Fog-Cloud layers.

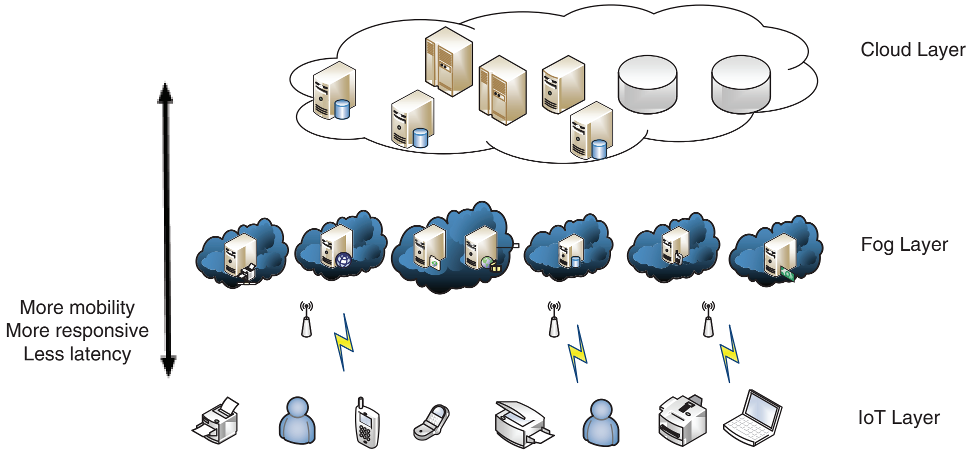

Figure 1: Architecture of IoT-Fog-Cloud environments

On the other hand, even with this technological development, the substantial number of data users still needs to be highly secured and protected. In Goasduff [9], Gartner has revealed that, during the security incidents that have taken place in recent years, 75 percent of behaviors' attacks were discovered at the application layer and that servers have been the major target for hackers. For these reasons, the security procedures should be taken into consideration when using an architecture since problems can occur in both data and networks. In fact, cyber security practices defend these networks against internal and external risks. However, the security and privacy mechanisms that exist in cloud cannot be directly applied to fog network due to its characteristics of geographical distribution, mobility, and heterogeneity [10,11].

To overcome these challenges, the fog system offers a powerful environment that provides added and supplementary resources for further improving data security. Accordingly, among the well-used techniques of data security, we have the Intrusion Detection System (IDS), which is a type of detection systems that protects networks by offering monitoring services for all systems [2]. Moreover, an IDS is an essential component of a network's security, where intrusion and other security gaps must be discovered rapidly and effectively [12]. In this regard, there are two main intrusion detection methods, namely anomaly-based methods and signature-based methods.

For the Signature-based techniques, it matches the pre-defined attack behaviors against the current patterns of the network [13]. The mechanism needs to store the behavior of each attacker in a database which means that the system will fail to detect hidden attacks. On the other hand, anomaly-based approaches try to construct a profile for normal behavior and subsequently any actions that do not fit this profile will be considered as anomalous task. This method is effective in the discovering of hidden attacks but it does so by producing high false positive rates. Furthermore, numerous researches have been concentrated on developing IDSs with greater detection rates and reduced rates of false alert [14,15]. However, because of the continuous and quickly changes in network environments, various new attacks appear continuously. Hence, another issue with current IDSs is that they cannot detect unknown attacks. For this reason, it is advised to improve IDSs quality [16]. Furthermore, network security weaknesses and limitations are reduced by the use of intelligent devices. Hence, it is imperative to deploy an intelligent platform to protect the systems against attacks. To address the abovementioned problems, Artificial Intelligence (AI) methods such as Machine Learning (ML) could be applied on the millions of terabytes of collected information to enhance and optimize the processes of IoT and fog systems [2]. For these environments, artificial intelligence stands to the addition and the application of specific studied procedures, for the control and the optimization of activities in the entire network. In addition, the smart analysis process improves the network awareness to the environment's changes, in order to enhance their performance, mainly regarding the latency and energy consumption metrics.

Within this framework, we propose in this article, a distributed attack detection scheme based on intelligent monitoring algorithm that takes advantage of the traditional IDS for the benefit IoT and fog environments. Firstly, our proposed scheme employs an algorithm that exploits the similarity of attack parameters to create a signature database. Secondly, it uses an intelligent learning algorithm to improve attack detection with high accuracy and low false-negative rates.

In this regard, we propose the division of fog network into small cells comprising a number of fog nodes. Each fog cell (FNcell) is managed by an Attack Detection Manager (ADM). Then, we place a Monitoring Agent (MA), called Attack Detection Controller (ADC), at each fog node.

Specifically, our work makes the following contributions:

- We propose a novel methodology for designing a scalable, self-adaptive, and autonomous attack detection and monitoring systems based on advanced artificial intelligence (AI) techniques.
- We take advantage of machine learning methodologies to identify malicious attacks and update alert signature databases.
- The deployment of fog cell architecture is proposed. This considerably reduces the latency of the anomaly detection at the fog layer by sending new attack signatures to neighboring ADC in the same FN-cell, and to neighboring ADM in other FN-cells. In a second stage, it enables the deactivation of malicious activities of such users. In a third stage, it provides attack classification that helps to determine precise prevention measures.
- We carried out an analysis and comparison of the proposed ADM/ADC with different solutions in the literature using NSL-KDD dataset.

Undoubtedly, the use of an intelligent method for network traffic ensures a quick attack detection, thereby reducing the operational overhead and the reaction time of the monitoring system to deactivate the malicious actions.

The paper is organized as follows; Section 2 will be reserved for describing and discussing some related works. In Section 3, we will present our proposed SIMAD contribution. The evaluation of the proposed model will be presented in Section 4. Finally, Section 5 concludes our contribution and offers the main findings of the article.

2 Related Works

The Internet of Things (IoT) offers tremendous occasions to the engineering and industrial activities [5,17]. Undoubtedly, this innovation is expected to be far more dynamic with the up and coming Fifth-Generation (5G) mobile network [18]. Nevertheless, the huge IoT utilization in basic areas, leads to the production of plentiful delicate and real time information.

Due to the aforementioned reason, these networks are considered as the most risky system. To resolve this matter, several research works have been suggested. Consequently, numerous security methods have been proposed such as, for example, cryptography procedures and authentication and access control schemes. However, regardless of the use of such solid privacy efforts, a network can even now be compromised by hackers utilizing innovative methods and high procedure assets. Top of FormHence, there must be an intrusion detection layer under any prevention layer. Accordingly, several intrusion detection systems (IDS) have been developed in the literature. The majority of the intrusion detection solutions deployed commercially implement signature-based approaches [14]. The authors in [14] proposed a signature-based IDS to detect Distributed DoS attacks in IoT networks. The proposed hybrid scheme includes two elements, which are the IDS routers and the IDS detectors. The former is placed in the edge gateway and it executes firewall functions. The latter unit supervises the internal traffic flow by deploying sensors. When detecting malicious devices, it sends corresponding information to the gateway. The presented results indicated that the proposed system detects attacks like hello flooding and version number modification.

In fact, the most significant benefit of anomaly traffic detection is the possibility of obtaining days without assaults. However, the proposed scheme in [14] is limited to a specific number of IoT devices.

In [19], Ullah et al. propose a two-level hybrid model for an anomalous activity detection scheme for intrusion detection in IoT networks. The suggested hybrid model comprises two phases for anomaly detection and for the identification of threat type. The first one uses the flow features and the decision tree classifier to sort normal and abnormal traffic according to the categories of attack. Then, these abnormal traffic will be sent to the second phase which will use Recursive Feature Elimination [19] in order to choose important parameters. The two-level hybrid scheme achieves satisfactory results in terms of recall, precision, and specificity.

In this regard, traditional intrusion detection systems (IDS) are becoming incompatible with the new network environment and with the emerging systems related with machine learning and deep learning. In Jan et al. [20], a lightweight IDS model is suggested for IoT networks, utilizing as a classifier the support vector machine (SVM) to distinguish between an interruption and an anomalous activity. In Pontevedra et al. [21], the authors suggested a real-time IDS model to identify wormhole attacks in IoT layers. This model uses the routing features and information in order to identify illegitimate users and nodes. authors in Anthi et al. [22] present a supervised IDS scheme to be deployed for smart home IoT devices.

Moreover, many efforts have been devoted to develop deep learning and machine learning attack detection approaches for IoT and fog systems, which lead to some algorithms accomplishing higher accuracy and adaptability to a wider range of situations. Unlike the signature-based IDS, the machine learning-based IDS are capable of detecting even unknown attacks. However, the fundamental challenge in this direction involves the designing of an efficient machine learning based IDS that performs well on real-time data. A Deep Learning Intrusion Detection System (DL-IDS) scheme is proposed in [22] to identify security risks in IoT. The authors use in their evaluation Minkowski distance and k Nearest Neighbor in order to generate missing data in the dataset at the preprocessing phase. The DL-IDS scheme achieves satisfactory results in terms of precision and accuracy. An Ensemble Learning-based Network Intrusion Detection System (ELNIDS) is suggested in [23]. It aims to recognize routing threats in RPL-based IoT networks. The authors utilize cooperative machine learning algorithms like Subspace Discriminant, Boosted trees, RUSBoosted Trees, and Bagged Trees to arrange network traffic into typical or anomalous.

In [24], the authors proposed a model for 6LoWPAN-based IoT organizations called Compression Header Analyzer Intrusion Detection Scheme (CHA-IDS). They employed AI to identify and distinguish different risks categories. The CHA-IDS scheme is situated on the router and uses anomaly and signature-based recognition approaches.

As described above, the schemes presented in [19–24], which are based on anomaly detection, have used machine learning mechanisms. However, they could have high false-positive rates if bad decisions about normal traffic flow are made. The proposed hybrid schemes capture the advantages of the cooperation of various detection strategies. Yet, they may acquire the issues of the adopted techniques. In this regard and in order to resolve the problem of the absence of enough data for the anomaly detection mechanisms and to improve the administrations of IoT systems, the researchers in [25,26] utilized the fog computing paradigm with IoT networks. In fact, in [25], the authors proposed an Anomaly Detection IoT (AD-IoT) model to detect cybersecurity risks in smart city environment. This model utilizes the random forest (RF) classifier to classify network traffic flow into normal or anomalous. The results obtained show that the RF classifier achieved satisfactory accuracy in attack detection with a low false-positive rate. Fog computing-based schemes [25,26] employed anomaly-based detection methodology using machine learning or statistical techniques, which means the traffic will always be scanned to check if it is normal or abnormal. However, these schemes fail to utilize the full potential of the fog nodes in terms of storage even though storing the signatures of the previously detected attacks will improve attack detection accuracy, reduce computational overheads, and further decrease the detection response time.

In [13], the proposed IDS utilizes numerous base learners in various algorithms and executes diverse techniques in fog-to-things networks. It uses the NSL-KDD dataset, within which the learners accomplish higher precision than the single learner does. It identifies anomalies with reduced latency. Then, it gives the attack arrangement that facilitates the determination of the exact avoidance procedures. In [27], the authors proposed a two-case study that includes device-driven and human-driven intelligence. The first one employed machine learning to detect user comportments and to achieve low-latency in Medium Access Control (MAC)-layer scheduling among sensor devices. For the second case study, the authors employed a method for task offloading by using AI to identify the offloading decision among many adjacent fog nodes. In [12], Sadaf et al. proposed an intrusion detection scheme using deep learning methods like isolation forest and Autoencoder techniques. Their method aims to identify attacks from the normal traffic data at the fog layer rather than analyzing it in the cloud layer. In fact, many DL methods have been developed based on traditional ANNs [28]. Furthermore, the small size of the dataset for the training process was another factor that resulted in over-fitted models.

As can be seen, some of the traditional methods employed in the above studies like conventional attacks detection schemes based on machine learning has proved to be successful oftentimes. However, they still had many limitations such as; its lacking scalability against distributed and advanced attacks and low accuracy. With regard to our analysis, we observed that traditional solutions using machine learning methods are not effective and efficient in identifying new attacks or unknown attacks, and also small attack changes. Moreover, most approaches disregarded how to update their systems. Indeed, systems would not detect changing attacks without being updated in the real time. However, DL approaches deliver best accuracy and faster processing thanks to its self-learning capability. Besides, although the recent developments in the Internet of Things and in fog computing paradigm have brought new opportunities to improve their service, they are still facing several challenges in terms of achieving accurate security monitoring.

Unlike other existing works, in our work, we design our system by considering these challenges and limitations. In fact, some limits may be enhanced through the utilization of Machine Learning ML. By creating high-level representations of features, ML discovers complex functions that map input to output without the need for manual intervention by experts [29].

3 Proposed Secure Intelligent Method for Attacks Detection (SIMAD)

3.1 Overview of the Proposed Architecture

The distributed nature of IoT and Fog computing environments makes it more vulnerable to numerous internal and external attacks. For this reason, these networks must be monitored continuously. Hence, they require as a distributed solution a new service for cyber-security attack detection.

In this section, we describe our proposed approach based on a distributed model for attack detection and monitoring for the fog network. Moreover, considering the extended size of modern networks and the complexity of big network traffic data, the problem exceeds the limits of human managing capabilities. In this regard, we propose the use of an artificial intelligent algorithm to deliver a scalable, self-adaptive, and autonomous attack detection. Thus, we divide the fog computing network to small cells that contain a number of fog nodes (FNcell). Every cell is managed by an intelligent Attack Detection Manager (ADM). At every fog node, we place an intelligent local attack detection controller (ADC), which monitors the whole fog node to supervise the users' activities and to analyze the collected data. Within the fog environment, the monitoring is functional until the end of the user's activity. If the system detects any malicious behaviors, the ADC has the ability to deactivate their tasks, so that the IoT user would be rejected from the network. This procedure is repeated for each user. Consequently, for the purpose of collaboration, the incorporated monitoring agent ADC in the FN-cell reports the user's behaviors. Then, after detecting a new attack, identified by new signatures, this agent sends them to the ADM, which is responsible for updating the signature databases. Moreover, for homologous ADM at neighbor fog cells, the security information updates are exchanged. The suggested network model is illustrated in Figs. 2 and 3.

In Fig. 4, we present a detailed overview of our proposal, which comprises three steps realized by two components (ADC and ADM), as described below.

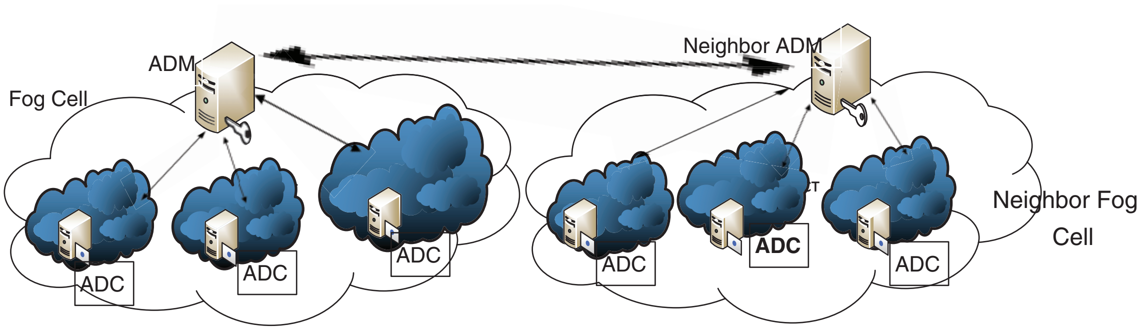

Figure 2: Proposed architecture

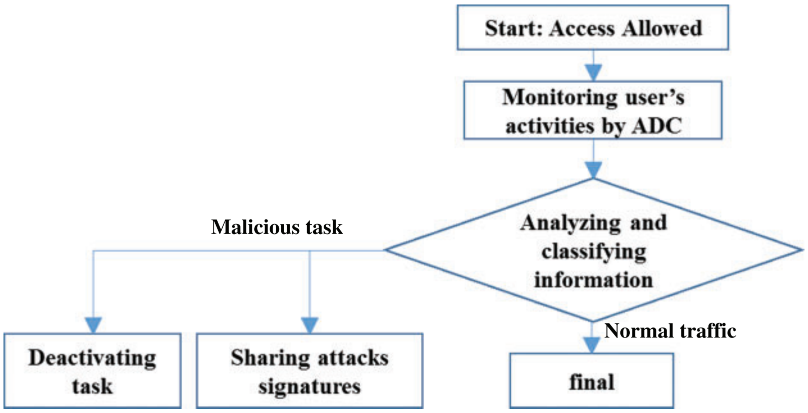
Start: Access Allowed
Monitoring user's activities by ADC
Analyzing and classifying information
Malicious task
Normal traffic
Deactivating task
Sharing attacks signatures
final

Figure 3: Flowchart of the proposed scheme

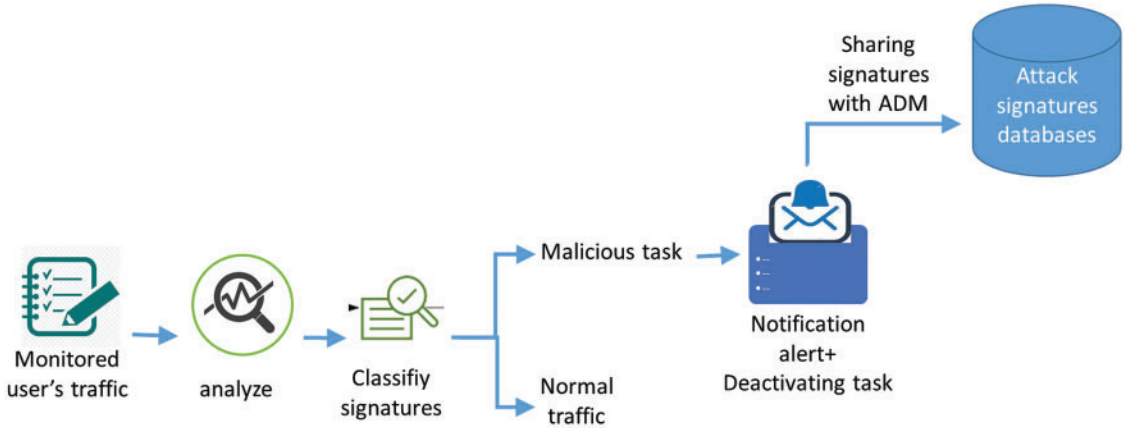
Sharing signatures with ADM
Attack signatures databases
Monitored user's traffic
analyze
Classifiy signatures
Malicious task
Normal traffic
Notification alert+ Deactivating task

Figure 4: Overview of the detection model

1) Monitoring user's activities: To protect and ensure efficient operation of IoT devices in the fog network, the suggested attack detection model can be employed to compensate for the lack of a monitoring process and to prevent possible attacks. This activity undertakes the task of coordinating the ADC for acquiring the basic knowledge that will be used for the learning step.

The network controllers should be able to identify the network features, such as its components, the architecture, the type, the available and the running services, the operating systems, the memories capacities, and even possible vulnerabilities. Each ADC monitors the whole fog node in order to supervise the user's behaviors. It collects information to be, after that, analyzed and classified.

2) Analyzing and classifying information: After collecting the necessary data, the ADC performs the analysis and the classification of traffic flows into normal tasks or malicious tasks. Through the use of the machine learning algorithm, the monitoring agent is able to detect the anomalous activity in comparison with the normal traffic. Then, the ADC utilizes network audit tools in order to generate alerts. This analyzer is able to analyze huge volumes of data for detection purposes. In fact, the ADC is able to determine any alteration action that must be detected and deactivated.

3) Sharing attacks signatures: After the detection of a malicious attack and the deactivation the user's task, the ADC sends attack signatures to the ADM, if it is a new attack. Then, the ADM updates its signature database and shares these information with other ADC placed on his fog cell. In addition, it sends these new signatures to the neighbor ADM placed in the neighbor fog cells.

3.2 Architecture of Intelligent Attack Detection Method

The presented smart attack detection scheme involves three steps comprising data preparation, feature extraction, and classifier attack phase. Firstly, the data preprocessing consists of a treatment phase to deliver good quality data. Secondly, the feature extraction is essential to extract diverse categories of characteristics from data as vectors. Moreover, this phase aims to produce classifiers that use the features attack to classify attacks from normal data. After this phase, the system will decide whether the input data is an attack or not. All these modules work together in collaboration to improve the mechanism of attack detection (Fig. 5).

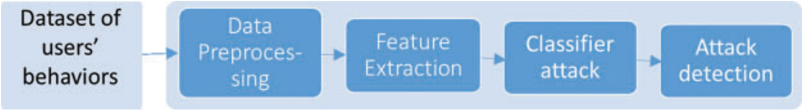

Figure 5: Proposed intelligent attack detection scheme

4 Experimental Evaluation

The goal here is to identify signature attacks and send information to computer managers in real time. In fact, our proposed ML aims to redefine how computers detect and identify attack signatures based on user's behavior.

Our method enables the ADC and the ADM to achieve high detection rates, even if it is applied to enormous environmental changes like fog and IoT networks. In this manuscript, based on distributed machine learning, we propose an attack detection system that takes advantage of an artificial intelligent algorithm. Therefore, we propose to evaluate our distributed approach by involving some ML/DL models in order to choose the best one.

4.1 Methodology

The proposed intelligent scheme is presented in Fig. 5. The methodology of our work is detailed as follows.

4.1.1 Data Preprocessing/ Preparation

This step comes after the collection of the handled data. Data preparation, generally named as "pre-processing", is the stage during which the data is cleansed and well-structured for the following step of information process. Throughout this step, the raw dataset is fastidiously checked for any kinds of faults. The aim of this preparation is to exclude low-quality information which can either be incomplete, redundant, or incorrect, and to begin making the information that can guarantee the high quality of the adopted intelligent model.

During this preparation phase, we transformed the input data into a matrix that can be vectorized. Among the known operations, we cite:

(a) Outlier's replacement: by computed medians, to avoid disappeared data point.
(b) Replacement of the missing values by computed medians.
(c) Data normalization: this step leads to the organization of the data in order to ensure the coherence between the records in the dataset.
(d) Deletion of the constant features and the low variance features by the Variance Threshold function.
(e) Feature extraction through PCA: Since it comes from diverse sources, network traffic generally has different types of parameters. Here, the features are collected together in the form of vectors, and they are then allocated to different channels for training. For this reason, we will reduce the dimension of the selected features using Principal Component Analysis (PCA).

4.1.2 Classification

In this step, the classifier returns whether the network traffic is an attack or not. Therefore, we will apply our distributed approach by using some classifiers, in order to choose the best one.

4.2 Dataset Description

In the security domain, there are numerous types of datasets that have been used by researchers to train and evaluate their machine learning systems and deep learning systems. The most commonly used one is NSL-KDD dataset, which is mainly based on KDDCUP'99 dataset [30]. The dataset was thoroughly processed in order to remove the redundancy and/or the incomplete records. Tab. 1 shows the traffic distribution of the NSL-KDD dataset [30].

Table 1: Traffic distribution of the NSL-KDD in two classes with statistics on the attack types

Traffic	Training	Test
Normal	67343	9711
Dos	45927	7458
Probe	11656	2754
R2L	995	2421
U2R	52	200
Total attack	58630	12833
Total traffic	125973	22544

The selected learnt features were applied to the labelled test dataset to classify them into attack and normal traffic.

The used NSL-KDD dataset of [30] is pre-processed and prepared, following the next steps:

Step 1: Collecting and receiving data frame from the dataset in both train and test and having column for labels csv files.

Step 2: Merging all the collected data frames into one data frame (df). Then, searching for missing, duplicated, and mistaken values in all columns, and dropping them from the (df).

Step 3: Normalizing numeric data with a min-max scalar (between 0 and 1). Thus, adding a target column to the data frame. Then, deleting constant features to select the best ones.

- Initial number of features = 122
- Deleted Features are Index(['num_outbound_cmds'], dtype='object')
- Number of features now is = 121

Afterwards, we aim to reduce the number of the selected features while conserving the total variance at a great value. Consequently, we plotted, in Fig. 6, the number of components in function of the cumulative sum of variance, in order to choose the desired number of features.

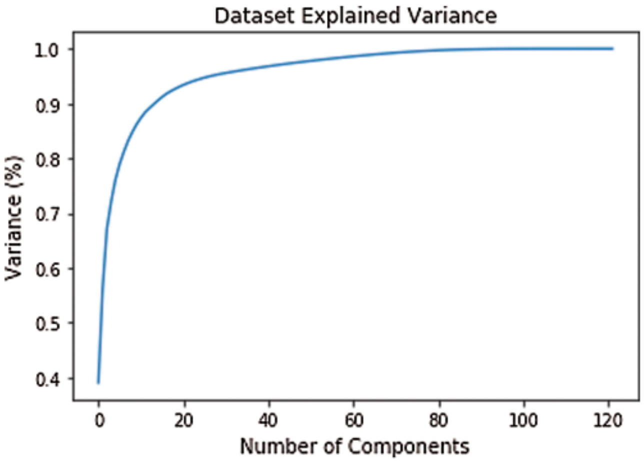

Figure 6: Dataset explained variance

Step 4: Reducing the dimension of the selected features using Principal Component Analysis (PCA), to have the following final configuration.

- Number of components = 10
- Total variance from PCA components = 0.8621156371639519

After these steps, the dataset is prepared and ready to be used by applying our approach with the integration of some classifiers from the literature.

4.3 Statistical Evaluations

To evaluate our work, we test the performance of the RF algorithm, in the context of attack detection, in comparison with other methods tested using the NSL-KDD dataset. There are diverse metrics that can be used to evaluate this performance, like accuracy (Ac), sensitivity, detection rate (DR), positive predictive value (PP), and negative predictive value (NP). These factors are assessed through the use of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Using these factors, the performance evaluation measurements can be defined as [31]:

Confusion matrix: which is a matrix that envisages the work of a tested classification technique versus actual classification.

Accuracy (Ac): is the percentage of the proximity between the predicted value and the actual value. Specifically, it represents the ratio of the sum of the TP and the TN against the total number of population. Eq. (1) depicts the mathematic expression of Ac:

$$\mathrm{Ac} = (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}) \tag{1}$$

Recall: also called Sensitivity, represents the proportion of the actual positives that are properly predicted. It shows how many attacks the detection scheme may return. The mathematic expression of recall is as follows:

$$Recall = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{2}$$

Positive predictive (PP) value/ Precision: is the percentage of proximity between the predicted positive results compared to the value when the true condition is positive. Its mathematical formulation is as follows:

$$\mathrm{Precision} = \mathrm{PP} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \tag{3}$$

F1 score: represents the weighted harmonic mean of recall and precision. It is formulated as follows:

$$\mathrm{F1\ score} = 2 * \mathrm{Recall} * \mathrm{Precision}/(\mathrm{Recall} + \mathrm{Precision}) \tag{4}$$

Negative predictive (NP) value: represents the ratio of the predicted negative values compared to the value when the true condition is negative. Its mathematical expression is:

$$\mathrm{NP} = \mathrm{TN}/(\mathrm{FN} + \mathrm{TN}) \tag{5}$$

4.4 Results

In this section, we provide the results achieved from applying our proposed intelligent scheme and other DL/ML models for attack detection using NSL-KDD dataset. The evaluation metrics are summarized in Tab. 2, which displays that the proposed RF classifier for the application of our method is in fact better than others traditional models with the highest accuracy. In fact, the accuracy value of the proposed model has been perceived to be **99, 50** percent (Fig. 7).

In addition, the experimental results show that the proposed distributed scheme has good performance in attack detection at the fog layer, without the need for having an attack detector for each fog node. Our proposed model is evaluated under a number of classification factors and it is proved that its attack detection rates with the application of the RF classifier rises up to 99%, and that FPR = 0.00847, which means that this method is performing well. Moreover, in Tab. 2 we present the overall statistical evaluation values achieved by the RF in comparison with other classifiers including [32]: Decision tree, Gradient Boosting, SVM, ANN, Ada Boost, Logistic Regression, Quadratic Discriminant Analysis, Linear Discriminant Analysis, and Gaussian NB. The obtained results show that, without exception, the RF has the best results in terms of precision, recall, and F1 score. Furthermore, RF has the highest TP number of 17548, which are predicted as abnormal samples and are indeed abnormal (Figs. 8 and 9). Thus, the number of FN and TP can indicate that the attacks detection rates of the RF is the highest.

Table 2: Performance evaluation values

Classifier metrics	TP	FN	FP	TN	Precision	Recall	Accuracy (%)	F1
Random forest	17548	**187**	**163**	**19079**	**0.9950**	**0.9950**	**99,50**	**0.9950**
Decision tree	17484	251	272	18970	0.9858	0.9858	98,58	0.9858
Gradient boosting	16966	769	354	18888	0.9698	0.9696	96,96	0.9696
SVM	16933	802	336	18906	0.9694	0.9692	96,92	0.9692
ANN	18853	389	871	16864	0.9662	0.9659	96,59	0.9659
Ada Boost	16722	1013	561	18681	0.9576	0.9574	95,74	0.9574
Logistic Regr.	16126	1609	941	18301	0.9315	0.9315	93,10	0.9309
Quad. Discr. Anal.	16267	1468	1201	18041	0.9278	0.9278	92,78	0.9277
Linear Discr. Anal.	15697	2038	986	18256	0.9193	0.9182	91,82	0.9180
Gaussian NB	14712	3023	1591	17651	0.8771	0.8752	87,52	0.8748

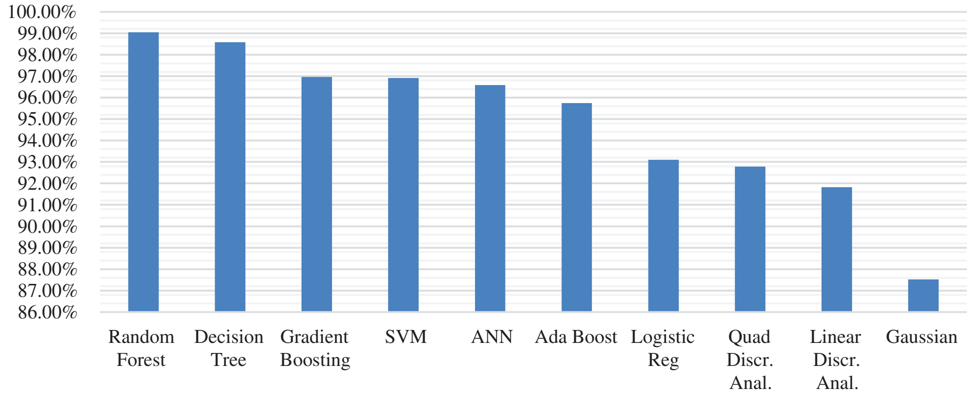

Figure 7: Accuracy comparison

$$\begin{bmatrix} 17548 & 187 \\ 163 & 19079 \end{bmatrix}$$

Figure 8: Confusion matrix

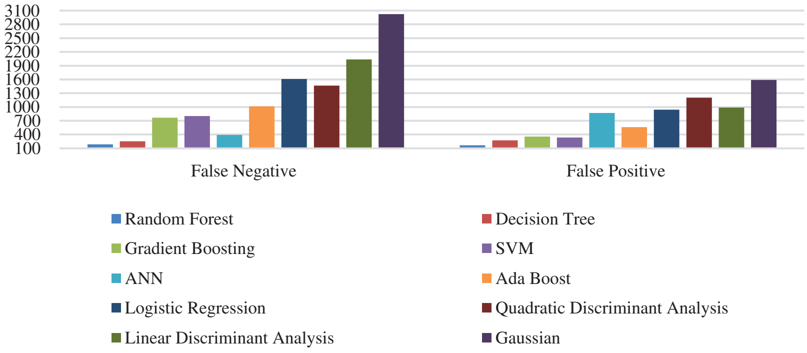

Figure 9: FN and FP statistical comparison

Fig. 10 shows that our distributed scheme outperforms the centralized model when we applied the RF classifier. In fact, our intelligent distributed detection method confirms its reliability and its scalability with a significant number of fog nodes. This indicates that when we integrate a large number of fog nodes, the accuracy increases and the delay reduces. Moreover, we deduce that these substantial results are a consequence of the distribution aspect of our method. Furthermore, the choice of RF as classifier to be applied in our model is greatly confirmed. Potentially, the combination of many prediction values issued from numerous decision trees into one single method, is essential to have these significant results. This is explained by the fact that even if a model is created out of many mediocre methods, it will always be superior to a single good model.

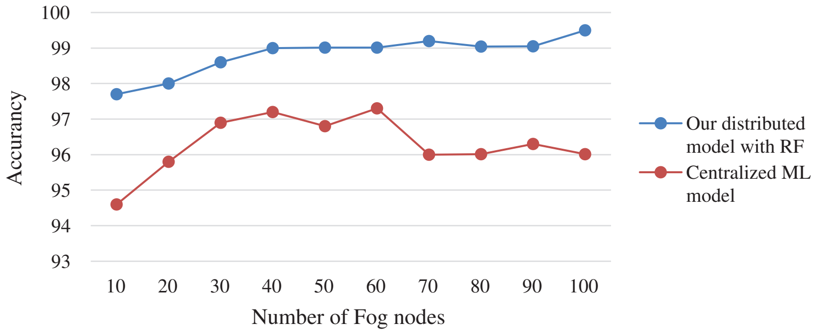

Figure 10: Comparison between our distributed scheme with the application of RF method against centralized ML model

Furthermore, the learning and the optimization of our system are done simultaneously, at the ADC and then, distributing attack signatures to others nodes. This improves the performance of our proposal due to the sharing of attack parameters in the IoT-Fog distributed environment.

To test the efficiency of the proposed attack detection method in terms of latency, we implemented our model in iFogSim [33], using distributed and centralized fog architectures.

The response time is measured 10 times and we computed the average value for different network architectures. The obtained results show that the response time of the proposed distributed scheme in IoT-fog systems is less than that of the centralized architecture, since the fog nodes are closer to the IoT layer, and hence can discover attacks with a short delay (See Fig. 11).

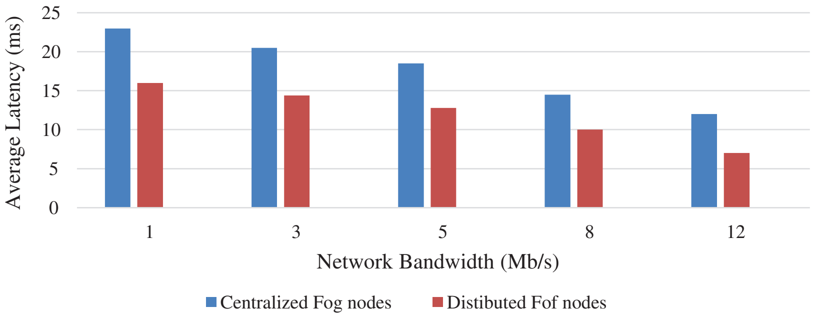

Figure 11: Average response time for fog-based centralized and distributed detection architectures

Thereby, IoT devices, organizations, and applications can obtain a better and faster reply in a secure way, due to the deactivation of malicious users/applications that might execute operations that may diminish their efficiency and stretch their resources.

4.5 Performance Analysis

The performance measurements of the proposed model show a significant difference compared to other models. This could be explained by the environmental characteristics surrounding our model. Accordingly, our scheme allows a greater bandwidth consumption and low latency in data processing.

Data Security: the IoT fog networks are not secure from malicious threats. In this regard, our intelligent model is able to analyze the network traffic and detect intruders in order to ensure the security and privacy of users' credentials.

Latency: most of IoT applications necessitate substantially severe latency constraints. Thus, this parameter is very important to evaluate certain mechanisms. In fact, by realizing the processing and the traffic analysis at the fog layer instead of the cloud, the delay could be reduced considerably. Consequently, the servicing in fog and cloud network would be performed more efficiently and rapidly. Moreover, the concordance between the proposed components of our model makes the attack detection considerably faster. At the point when a an ADC produces an attack alert, these information should be communicated to the ADM, therefore, to decrease the risks in other fog nodes, a part of the detection processing will be reduced, and consequently, we gain time. In this respect, our proposal achieves low-latency in attack detection and can thus reduce its own energy objectives.

Network Availability: the volume of data exchanged in the IoT network is enormous. Here, the proposed model ensures an autonomous, intelligent, and distributed system for dealing with this amount of information. Therefore, we avoid the unavailability of IoT-fog networks.

Network Bandwidth: by using blacklist, a repeated attack can be prevented. For instance, the IP address, email address, and similar other information would be recorded in a blacklist in order to prevent the hacker from repeating the same attack. Since the ADC would monitor the behavior of such user, we cannot detect false positives. Therefore, the network throughput will be well alleviated.

5 Conclusion

Despite the fact that recent advances in the IoT and fog paradigms have brought about new opportunities to improve the service, achieving accurate security monitoring still faces several challenges, which make cyber security an important field of research. Therefore, in this paper, we presented an intelligent method to deliver a scalable and autonomous attack detection and monitoring system. Our methodology aims to divide fog computing into small cells situated at the edge of the network, to operate the services of detection, monitoring, evaluation, and administration. The proposed scheme enables to achieve high attack detection rates, even if it is imposed on successive environmental variations like fog computing. Furthermore, our method focuses on segregating attacks from the normal network traffic data at the fog layer. Subsequently, we have chosen the most accurate classifier, which is the random forest RF, to finally detect attacks. The optimal achievements of the chosen method are that it eliminates the complexity of the system in the IoT-fog networks and that it rises the classification accuracy, stability, and the reliability. Indeed, by using our scheme, the latency and the response times of the attacks detection will be highly reduced. Additionally, the experimental results revealed that our intelligent scheme outperforms diverse other models with high accuracy, sensitivity and positive and negative detection rates.

As a future work, in order to address the matter of energy efficiency for the proposed scheme in IoT-fog environments, we will apply intelligent parameters to improve its performance in terms of energy consumptions. Thus, we will aim to address the issue by developing a predictive model for energy consumption in IoT fog environments using machine learning. We will try to use and compare different classifiers in order to choose the more accurate one.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] M. Zhong, Y. Zhou and G. Chen, "Sequential model based intrusion detection system for IoT servers using deep learning methods," *Sensors*, vol. 21, no. 1113, pp. 1–21, 2021.

[2] M. Qiu, S. Kung and K. Gai, "Intelligent security and optimization in edge/fog computing," *Future Generation Computer Systems*, vol. 107, no. 6, pp. 1140–1142, 2020.

[3] A. A. Alli and M. Mahbub, "Internet of things the fog cloud of things: A survey on concepts, architecture, standards, tools, and applications," *Internet of Things*, vol. 9, no. 100177, pp. 1–20, 2020.

[4] W. Ben Daoud, A. Meddeb-Makhlouf and F. Zarai, "A model of role-risk based intrusion prevention for cloud environment," in *14th Int. Wireless Communications & Mobile Computing Conf.*, Limassol, Cyprus, IEEE, pp. 530–535, 2018.

[5] M. A. Lawal, R. A. Shaikh and S. R. Hassan, "Security analysis of network anomalies mitigation schemes in IoT networks," *IEEE Access*, vol. 8, no. 2976624, pp. 43355–43374, 2020.

[6] B. Susilo and R. F. Sari, "Intrusion detection in IoT networks using deep learning algorithm," *Information—An International Interdisciplinary Journal*, vol. 11, no. 5, pp. 279, 2020.

[7] A. A. Elsaeidy, N. Jagannath, A. G. Sanchis, A. Jamalipour and K. S. Munasinghe, "Replay attack detection in smart cities using deep learning," *IEEE Access*, vol. 8, no. 2020, pp. 137825–137837, 2020.

[8] H. Mahmood, D. Mahmood, Q. Shaheen, R. Akhtar and W. Changda, "S-DPS: An SDN-based DDoS protection system for smart grids," *Security and Communication Networks*, vol. 2021, no. 6629098, pp. 1–19, 2021.

[9] L. Goasduff, "Gartner says 5.8 billion enterprise and automotive IoT endpoints will be in use in 2020," 2019. [Online]. Available: https://gtnr.it/35hq94q [Accessed: 28-Apr-2021].

[10] V. H. Osmanaj and A. Al-ahmad, "Fog computing security and privacy for the Internet of Thing applications: State-of-the-art," *Security Privacy*, vol. 4, no. e145, pp. 1–26, 2021.

[11] W. Saeed, Z. Ahmad, A. I. Jehangiri, N. Mohamed and A. I. Umar, "A fault tolerant data management scheme for healthcare Internet of Things in fog computing," *KSII Transactions on Iinternet and Information Systems*, vol. 15, no. 1, pp. 35–57, 2021.

[12] K. Sadaf and J. Sultana, "Intrusion detection based on autoencoder and isolation forest in fog computing," *IEEE Access*, vol. 8, no. 3022855, pp. 167059–167068, 2020.

[13] P. Illy, G. Kaddoum, C. M. Moreira, K. Kaur and S. Garg, "Securing fog-to-Things environment using intrusion detection system based on ensemble learning," *IEEE Wireless Communications and Networking Conf.*, Marrakesh, Morocco, pp. 1–7, 2019.

[14] P. P. Ioulianou, V. G. Vassilakis, I. D. Moscholios and M. D. Logothetis, "A signature-based intrusion detection system for the internet of things," in *Information and Communication Technology Forum*, Graz, Austria, pp. 1–7, 2018.

[15] S. Sumathi and N. Karthikeyan, "Detection of distributed denial of service using deep learning neural network," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 6, pp. 5943–5953, 2021.

[16] D. Papamartzivanos and G. Kambourakis, "Introducing deep learning self-adaptive misuse network intrusion detection systems," *IEEE Access*, vol. 7, no. 2019, pp. 13546–13560, 2019.

[17] G. Caiza, M. Saeteros, W. Oñate and M. V. Garcia, "Fog computing at industrial level, architecture, latency, energy, and security: A review," *Heliyon*, vol. 6, no. e03706, pp. 1–7, 2020.

[18] N. N. Khumalo, L. Mfupe and O. O. Oyerinde, "Reinforcement learning-based resource management model for fog radio access network architectures in 5G," *IEEE Access*, vol. 9, no. 3051695, pp. 12706–12716, 2021.

[19] I. Ullah and Q. H. Mahmoud, "A two-level hybrid model for anomalous activity detection in IoT networks," in *16th IEEE Annual Consumer Communications & Networking Conf.*, Las Vegas, USA, pp. 1–6, 2019.

[20] S. U. Jan and S. Ahmed, "Toward a lightweight intrusion detection system for the Internet of Things," *IEEE Access*, vol. 7, no. 2019, pp. 42450–42471, 2019.

[21] V. Pontevedra and S. Sonavane, "A real-time intrusion detection system for wormhole attack in the RPL based Internet of Things," *Procedia Manufacturing*, vol. 32, no. 2, pp. 840–847, 2019.

[22] E. Anthi, L. Williams, M. Słowi, G. Theodorakopoulos and P. Burnap, "A supervised intrusion detection system for smart home IoT devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9042–9053, 2019.

[23] A. Verma and V. Ranga, "ELNIDS: Ensemble learning based network intrusion detection system for RPL based Internet of Things," in *Proc. of the 4th Int. Conf. on Internet of Things: Smart Innovation and Usages*, Ghaziabad, India, pp. 1–6, 2019.

[24] M. N. Napiah, M. Yamani, I. Idris, R. Ramli and I. Ahmedy, "Compression header analyzer intrusion detection system (CHA-IDS) for 6LoWPAN communication protocol," *IEEE Access*, vol. 3536, no. 6, pp. 16623–16638, 2018.

[25] I. Alrashdi, A. Alqazzaz, E. Alouf, R. Alharthi, M. Zohdy *et al.,* "AD-IoT: Anomaly detection of IoT cyberattacks in smart city using machine leaming," in *IEEE 9th Annual Computing and Communication Workshop and Conf.*, Las Vegas, USA, pp. 305–310, 2019.

[26] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," *Future Generation Computer Systems*, vol. 82, no. 6, pp. 761–768, 2018.

[27] Q. Duy, M. V. Ngo, T. Quang, T. Q. S. Quek and H. Shin, "Enabling intelligence in fog computing to achieve energy and latency reduction," *Digital Communications and Networks*, vol. 5, no. 1, pp. 3–9, 2019.

[28] S. Mohamed and B. Abdellah, "Deep neural models and retrofitting for arabic text categorization," *International Journal of Intelligent Information Technologies*, vol. 16, no. 2, pp. 74–86, 2021.

[29] H. Liu and B. Lang, "Applied sciences machine learning and deep learning methods for intrusion detection systems: A Survey," *Applied sciences*, vol. 9, no. 4396, pp. 1–28, 2019.

[30] S. Kavitha, N. Uma Maheswari and R. Venkatesh, "Network anomaly detection for NSL-KDD dataset using deep learning," *IT in Industry*, vol. 9, no. 2, pp. 821–827, 2021.

[31] A. Pradesh, "A comparative study of machine learning algorithms using quick-witted diabetic prevention," *Annals of R.S.C.B*, vol. 25, no. 4, pp. 4249–4259, 2021.

[32] J. Rahman, H. S. Suri and M. Abedin, "Accurate diabetes risk stratification using machine learning: Role of missing value and outliers," *Journal of Medical Systems*, vol. 42, no. 92, pp. 1–17, 2018.

[33] W. Ben Daoud, M. S. Obaidat, A. M. Makhlouf, F. Zarai and K. F. Hsiao, "TACRM: Trust access control and resource management mechanism in fog computing," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, pp. 82, 2019.