International Journal of Computer Networks and Applications (IJCNA)
DOI: 10.22247/ijcna/2022/217701 Volume 9, Issue 6, November – December (2022)

RESEARCH ARTICLE

Scalable and Robust Intrusion Detection System to Secure the IoT Environments using Software Defined Networks (SDN) Enabled Architecture

Tahani M. Alshammari
Department of Computer Sciences, A-Jouf University, Saudi Arabia.
TahaniMAlshammari@gmail.com

Faeiz M. Alserhani
Department of Computer Engineering and Networks, Al-Jouf University, Saudi Arabia.
fmserhani@ju.edu.sa

Received: 03 September 2022 / Revised: 17 November 2022 / Accepted: 23 November 2022 / Published: 30 December 2022

Abstract – Due to the rapid development of smart devices with reduced costs and advanced sensing capabilities, the adoption of the internet of things has recently gained a lot of traction. However, such IoT devices are more vulnerable to being attacked or compromised. Moreover, traditional security mechanisms based on signatures and rules are no longer capable of detecting sophisticated intrusions. In the IoT context, the deployment of intelligent techniques in the control plane of the system architecture plays a vital role in identifying various attacks, including unknown ones. In this study, a software defined network (SDN)-based IoT anomaly intrusion detection system is proposed to detect abnormal behaviors and attacks. Five different machine learning techniques are investigated, including support vector machines, k-nearest neighbor, logistic regression, random forest, and decision trees. A scalable and robust intrusion detection system is designed based on machine learning models and placed at the SDN controller to observe and classify the behavior of IoT devices. A benchmark dataset, ToN-IoT, has been selected to test and evaluate the ML models by conducting several experiments. The obtained results have demonstrated that ML-based IDS can provide a reliable security system. Particularly, the random forest technique outperformed the other studied ML algorithms.

Index Terms – Intrusion Detection Systems (IDS), Attack Classification, Anomaly Detection, Machine Learning (ML), Internet of Things (IoT), Software-Defined Networks (SDN).

1. INTRODUCTION

1.1. Problem Overview and Motivation

The internet of things (IoT), with its rapid expansion across a multitude of sectors such as smart sensors, healthcare, home appliances, and wearable gadgets, has significantly impacted a wide range of our daily life aspects. Despite the fact that IoT technologies are essential for improving real-world smart systems, their huge scale and heterogeneous nature has presented new security challenges [1]. Also, due to the dynamic nature of IoT devices, the overall environment is subject to cyber-attacks like brute force, denial of service (DoS), distributed DoS attacks, and so on [2]. Compared to software-defined networks, traditional networks are in fact more complicated, error-prone, and time-consuming. Such traditional networks are composed of switches, routers, middleboxes, servers, and hosts that must be setup and maintained by network operators using manual entry of low-level device-specific terminology. In fact, this becomes a major cause of network downtime. To address these concerns, SDN emerges as a new paradigm in network management to fulfill the demand for programmatically managing networks [3]. Essentially, SDNs are highly useful in dealing with security concerns associated with IoT devices. These networks can effectively handle security threats in a flexible and dynamic fashion without imposing any extra load on IoT devices [3]. At this point, it is worth mentioning that the design and implementation of an SDN-enabled IDS to detect abnormal behaviors and cyber-attacks in IoT networks as early as possible, is indeed not straightforward.

1.2. Background

For over two decades, intrusion detection systems have been considered crucial to secure networks and information systems. However, traditional IDS techniques are hard to implement on IoT systems due to their unique characteristics, such as limited resource devices, protocol stacks, and standards [4]. In addition, traditional intrusion detection methods such as signature-based intrusion detection are ineffective against security threats due to the ambiguity, expansion, and complexity of IoT devices [4]. In light of the same, several studies have demonstrated that integrating machine learning technologies with IDS is an effective method for overcoming the limits of standard IDS when

RESEARCH ARTICLE

utilized for IoT [5].

Machine learning (ML) approaches may allow the detection system to enhance its automatic ability by learning from experience. Such approaches have been largely employed in traditional networks to classify malicious traffic and network attacks [6]. They are also commonly utilized for classification and prediction challenges and have shown substantial promise in network traffic classification [7]. In contrast to more localized policy implementation in conventional networks, the key benefit of applying ML approaches in SDN is related to their capability in influencing the network-wide security standards [6-7]. ML algorithms are usually classified as supervised, unsupervised, or semi-supervised. According to [8], the supervised machine learning approaches outperform unsupervised and reinforced learning in IDS. Besides, it is found that the data types and learning methods have an impact on machine learning technique performance. In light of existing studies that employ various ML algorithms and provide comparison results for different supervised algorithms in terms of performance and accuracy, this paper aims to investigate the capability of ML techniques in providing reliable security protection for IoT ecosystems. Specifically, five ML techniques are investigated in this paper, namely, KNN, SVM, RF, LR, and DT.

Despite the demand for reliable network traffic data for building efficient models, the majority of available IDS studies have relied on datasets created for legacy networks without IoT activity [9]. For this reason, ToN-IoT dataset is used in this study for testing and evaluating the system as it represents the heterogeneous nature of current IoT networks and, therefore, is suitable for IoT as compared to other datasets like NSL-KDD and UNSW-NB15 [10].

1.3. Contributions

Based on the above discussion, an anomaly IDS is proposed in this paper to defend IoT environments based on ML techniques and ToN-IoT dataset. The developed system is integrated into SDN architecture to provide robustness, flexibility, and scalability. The proposed system detects the attacks and abnormal behaviors in the network as early as possible using supervised ML approaches. Besides, the best practices available have been utilized to develop and assess IDSs in an SDN-based IoT environment. The key contributions of the paper can be highlighted as follows. (1) To design and implement an SDN-enabled IDS to detect cyber threats in IoT. (2) To apply intelligent and reliable detection mechanism using various learning techniques. (3) To analyze and evaluate the applied detection algorithms to provide a cost-effective solution to the IoT architectures. (4) To utilize the facilities of control plane of IoT architecture to provide an effective security mechanism. The general framework of the SDN-enabled and ML-driven for intrusion detection in IoT system is shown in Figure 1.

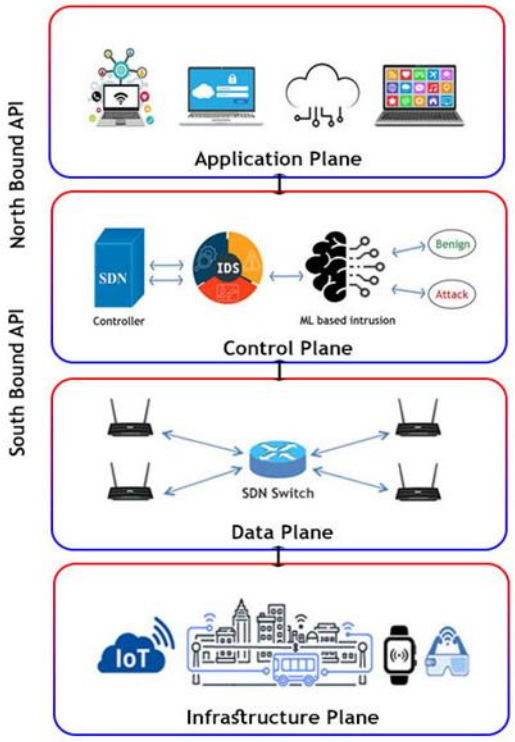

Figure 1 IoT Network Model

The rest of this paper is structured as follows. Section 2 reviews related studies exist in literature. In Section 3, the benchmark dataset is presented to illustrate the feature selection. The methodology and evaluation are introduced in Sections 4 and 5, respectively. The main results are discussed in Sections 6, and the paper is then concluded in Section 7.

2. RELATED WORK

It is known that traditional security mechanisms based on signatures and rules are no longer capable of detecting sophisticated intrusions. Therefore, several architectures and approaches have been developed in literature to detect abnormal behaviors and attacks in IoT environments based on ML and deep learning techniques using different datasets. In [11], an SDN-based framework called soft-things is developed to identify anomalies in IoT networks. The proposed IDS is evaluated via self-generated data and linear/nonlinear SVM approaches. Particularly, the results of nonlinear SVM outperformed the linear one in terms of precession and recall. To detect the attacks as early as possible, a real-time automated IDS is developed in [12] for SDN-enabled IoT networks based on automatic flow feature extraction and classification. The proposed scheme proved to be an efficient solution for mitigating the attacks in real-time with high accuracy in SDN-based IoT environment. It is worth mentioning that both [11] and [12] were conducted

RESEARCH ARTICLE

based on a specific dataset generated by the researchers, that is used then for evaluation of the intrusion detection systems.

In fact, there exist various standard datasets that are largely employed in literature like UNSW-NB15, NIMS and NSL-KDD. In [13] and [14], the UNSW-NB15 dataset was employed to evaluate the IDSs based on different techniques and ML approaches. In [13], an AdaBoost ensemble technique was used to develop an effective NIDS to detect the attacks in IoT environments. Such a technique is a well-known ML technique that integrates numerous base models to enhance the accuracy and reduce the false-positive rates as compared to a single model approach. Specifically, three approaches including KNN, DT, and Naïve Bayes (NB) techniques were employed in [13] to enhance the overall performance in terms of time processing and detection rate. Further, it was proved that the developed ensemble technique outperformed existing ensemble techniques as well as existing SVM and NB mechanisms. In fact, such superior performance is attributed to the proposed features characteristics of legal and suspicious incidents, in addition to the structure of developed ensemble approach-based NIDS, which has less overhead than other methods. Unlike [13], a classification model for IoT systems based on attack signatures was developed in [14] using machine learning approaches. Two scenarios i.e., without noise injection and with 10% noise filtering, were investigated, for which it was found that the KNN and Random Forest (RF) techniques performed well, with an accuracy of 100% and 99% for the first and second scenarios, respectively. On the other hand, the NB classifier achieved the worst results with an accuracy of 95.35% and 82.77% for both scenarios, Moreover, other evaluation matrices, such as recall and precision, also revealed that RF and KNN classifiers are more effective than Naïve Bayes. Similar to [13], an ensemble learning model was proposed in [15] for IoT anomaly detection using SDN. However, the most significant features were extracted utilizing a deep-auto encoder with a deployed learning model in the SDN controller. The suggested model was validated under a benchmark dataset and a real-time testbed. Consequently, it was observed that the proposed solution achieved more reliable and better performance than the existing models. To protect against DDoS attacks, various studies in the literature developed efficient solutions to detect and classify such attacks based on different approaches and ML techniques [16-17]. In [16], an ML-based detection and classification model was suggested to detect DDoS attacks based on KNN, Gaussian Naïve Bayes (GNB), and classification and regression tree (CART) techniques. Based on the conducted experiments, it was found that CART outperformed others algorithms in terms of robustness, training time, prediction speed, and accuracy. Another ML-based mechanism was proposed in [17] to detect low-rate distributed DoS attacks on switch nodes and SDN control under IoT networks. Various machine learning (ML) methods, including KNN, SVM, RF, and NB, were evaluated using a real-time dataset provided by the experimental environment. It is worth mentioning that RF, SVM, and KNN obtained equivalent performance and outcomes. However, NB was shown to be more appropriate for dealing with nominal than numeric data. In [18], an SDN-based detection system is proposed to also detect DDoS attacks. Four ML techniques were implemented for training and validating the model, namely artificial neural network (ANN), SVM, NB, and KNN classification models. Relative to other implemented algorithms, KNN outperformed and obtained the highest accuracy rate of 98.3%.

In IDS literature, several architectures and techniques were developed to build a reliable the detection systems based on deep and machine learning approaches [19-20]. A hierarchical ML architecture was proposed in [19] with two classifier stages in the SDN-controller and the processing device. Various ML algorithms were evaluated including KNN, NB, SVM, LDA and CART. Similar to [16], the CART technique also proved to be the best solution among the other implemented techniques. In [20], IDS in the context of IoT environment was investigated using deep learning and ML techniques to identify the privacy and security aspects of IoT networks. Specifically, LSTM and KNN algorithms were implemented, and bot-IoT dataset was utilized to analyze the developed algorithms in the detection module. Essentially, LSTM proved to be an effective solution for attack identification in IoT networks. For the same purpose, a centralized signature-based IDS was developed in [21]. However, RF classifier was employed to train and validate the model using CICIDS2017 dataset and achieved promising results and high performance. Botnet traffic in an IoT environment was analyzed in [22] using three different ML algorithms i.e., SVM, LR and RF. All implemented classifiers achieved an accuracy above 99% which is indeed promising for such kind of application. An ML-SDN model was proposed in [23] for flow-based anomaly detection to accelerate the detection process and achieve the highest accuracy in SDN, wherein NSL-KDD dataset was utilized to verify the proposed model. Also, a similar dataset was employed in [24] and [25]. In [24], a combined approach was developed to prevent the saturation of the control plane and enhance the accuracy and scalability of the anomaly detection system. By using an ensemble learning method, a high detection accuracy was achieved under the suggested model. Relative to previous studies, two different methods were investigated in [25] based on ML and deep neural network (DNN). NSL-KDD dataset was employed and obtained an accuracy rate of 82% in the first method which was based on RF classifier. However, the second approach which was integrated with a DNN-based IDS achieved an accuracy of 88%.

RESEARCH ARTICLE

The majority of the aforementioned studies were conducted using ML systems trained on outdated and unreliable data sets. To handle this issue, a more recent publicly released dataset was generated in [26]. Such a dataset, known as 'ToN-IoT', reflects the heterogeneous nature of IoT. Although ToN-IoT is more appropriate for IoT environments, it is found that there is still a lack of implementation of this dataset in the literature.

Table 1 Summarization of Relevant Studies in Literature

Ref.	Algorithms	Dataset	Domain
[11]	Linear and non-linear SVM	Generate by researchers	Dynamic anomaly early detection of IoT traffic.
[12]	AdaBoost with DT, NB, and ANN	UNSW-NB15 and NIMS	IDS for botnet assaults against MQTT, HTTP, DNS in IoT systems.
[13]	RF	SDN-specific generated by researchers	Feature extraction and classification at SDN application layer.
[14]	RF, KNN and Naïve Bayes	UNSW-NB15	Attacks and anomaly in IoT Networks
[15]	LDA, KNN, CART, NB SVC	CICIDS2017	SDN security
[16]	GNB, kNN and CART	SDN-specific generated by researchers	ML-based detection and classification model for DDoS attacks.
[17]	Ensemble classifiers	Real-time testpad and N-BaIoT	Anomalies detection in SDN-enabled framework.
[18]	KNN, NB, RF and SVM	Real-time generated by experiments.	DDoS attack on switch nodes and SDN control
[19]	LSTM, KNN	Abot-IoT	IDS in IoT networks
[20]	RF	CICIDS2017	Signature-based IDS
[21]	SVM, NB, ANN and KNN	Real-time generated by experiments.	SDN-based detection systems for DDoS attacks
[22]	LR, SVM and RF	Botnet Traffic	Botnet traffic in IoT environment
[23]	RF	NSL-KDD	ML-based SDN networks
[24]	SVM, NB and RD	NSL-KDD and generate data	Scalability enhancement of detection system
[25]	RF	NSL-KDD	OpenFlow Controller
[26]	RF,GBM and NB	ToN_IoT	AI-based security solutions

Table 1 summarizes relevant studies exist in literature with emphasis on the utilized algorithms, employed datasets and the domain of these studies.

3. DATASET

In this study, 'ToN-IoT' [26] data collection was employed as it contains diverse data sources collected from IoT telemetry sensors, network traffic datasets, and datasets for Windows 7 and 10, Ubuntu 14. As compared to other existing datasets like NSL-KDD and UNSW-NB15, the ToN-IoT dataset is more suitable for IoT systems as it represents the heterogeneous nature of current IoT networks [27]. Moreover, ToN-IoT is presented in a CSV-format with a labeled column identifying the behavior and type of the attack which includes DoS, distributed DoS, back door, ransomware, data injection, cross-site scripting (XSS), man-in-the-middle (MITM), scanning and password attacks. These attacks were performed against a range of IoT devices, and sensors, and gathered across IoT networks. More information about the dataset may be found in [28]. The identified attacks in 'ToN-IoT' can be categorized into one of the following categories:

1) Scanning Attack

This attack aims to capture information about testbed network victim systems, including open ports and active IP addresses. Such attack represents the initial step in the penetration testing or cyber killing chain model which is commonly known as reconnaissance or probing.

2) DoS Attack

Any attempted sabotage of IoT network services and resources is known as DoS attack. The end of such an attack is to render IoT services unavailable.

3) DDoS Attack

This type of attacks is generally conducted by a group of infected machines, commonly known as bots. Such attack works by flooding and depleting the victim IoT resources with a large number of connections.

4) XSS Attack

Using XSS technique, malicious code can be injected into trusted online applications, for example, Web Pages for IoT services. In XSS attacks, the attacker sends malicious codes, typically in a browser-side script form, to several end-users through an online application.

RESEARCH ARTICLE

5) Password Cracking Attack

This type of attacks refers to any hacking technique like dictionary and brute-force attacks, employed to guess potential password combinations until the exact password is identified. Such attack may be utilized to hack the passwords of IoT services, operating systems, and web applications installed on the testbed.

6) MITM Attack

This type of attacks might have occurred when the hackers establish themselves in the middle, between users and applications, in order to monitor or pretend to be one of them, giving a misleading illusion that a routine information flow is taking place. Information regarding IoT services, web applications and networks could be stolen under such hacking scenarios.

7) Injection Attack

Injection attacks include injecting or inserting fictitious input data from clients into targeted systems, for example, SQL injection for attacking ASP and PHP applications.

8) Ransomware Attack

This is a sophisticated sort of malware attack that encrypts systems or services and prevents normal users from accessing them unless they pay a ransom. As IoT apps and devices perform critical activities, they are considered potential targets for ransomware attacks. For example, when the access is blocked, it might have disastrous consequences, such as significant financial losses for stakeholders [29].

9) Backdoor Attack

The attacker may get unauthorized remote access to the hacked IoT systems via backdoor malware. Such attack may be employed to gain control of the compromised IoT devices and perform botnet-based DDoS attacks.

In fact, there exist 44 features associated with each data entity in ToN-IoT, in addition to the type label (normal or attack). The statistics and distributions of the 'Normal' and 'Abnormal' records in the training-testing of ToN-IoT with multi-class categories are shown in Table 2.

Table 2 Distribution of 'Normal' and 'Abnormal' Labels

Label	Category	Percentage
Normal		65%
Abnormal	DoS	5%
	DDoS	5%
	XSS	4%
	MITM	0.0002%
	Scanning	5%
	Injection	4%
	Password	4%
	Ransomware	4%
	Backdoor	4%

4. EXPERIMENTAL METHODOLOGY

In this paper, Ton-IoT dataset is employed to conduct several experiments. This dataset is publicly released as a network-based intrusion detection system (NIDS) dataset that reflects current modern network behavior. The dataset is initially prepared for efficient machine and deep learning algorithms. Then, the classifiers' predictions are collected, and certain evaluation metrics are further determined statistically. The experiments are implemented and evaluated using Anaconda Navigator and Python, and the ML models are built using the SciKitLearn modules. The methodology is represented by the flowchart diagram in Figure 2. In the following, the methodology steps are explained in detail.

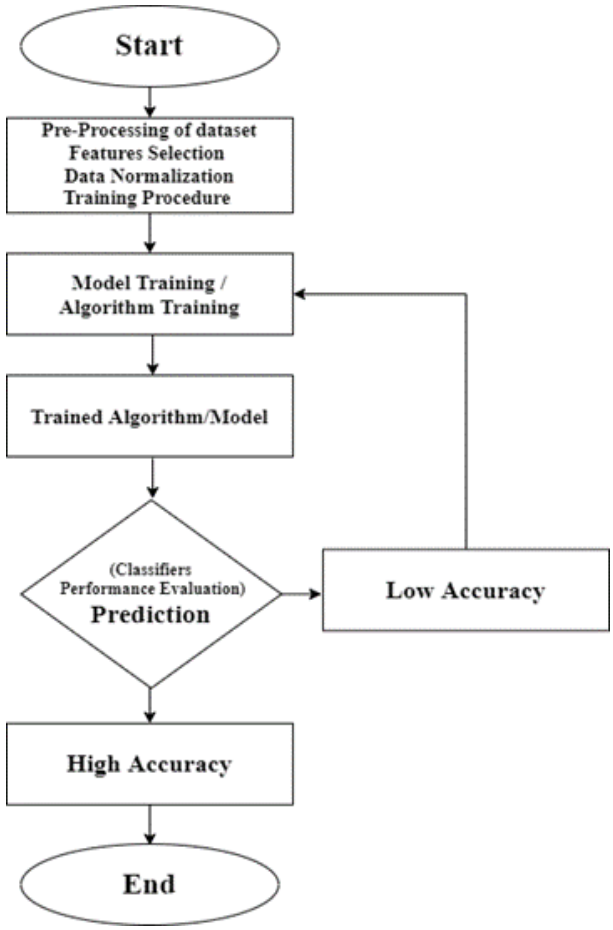

Figure 2 Flowchart of the Experimental Methodology

RESEARCH ARTICLE

4.1. Dataset Preprocessing

To obtain a high accuracy rate and acceptable performance, cleaning and preparation of the employed dataset are an essential task before feeding the data into machine learning algorithms. In fact, there exist several challenges associated with the dataset, for example, the missing values, categorical characteristics, and class imbalance. The performance of the selected machine learning algorithm may be impacted by unnecessary features. Essentially, the selected algorithm is tested using several normalization and preprocessing techniques.

1) Dealing with Missing Values

The ToN-IoT is a huge dataset in which the missing values are widespread. Such values must be properly treated to develop further relevant analysis. In essence, the most-frequent value in each feature with missing value is used to replace these missing values.

2) Empty and Dash Values

The empty and dash values are replaced with 0 in order to obtain a numerical-only dataset that is suitable for the following phases.

3) Transforming Categorical Features to Numerical

The categorical features are converted to numerical values. As there exist several categorical features in ToN-IoT dataset, one-hot encoding was employed to transform these features and accomplish the task.

4.2. Features Selection

NIDS datasets are composed of a number of network data features that mainly indicate the information characterized by the datasets. Such features are used to reflect a reasonable number of security incidents in order to achieve successful classification objectives [30]. The quality of any ML-based NIDS is heavily influenced by NIDS dataset features [31]. As a result, feature selection is vital task to identify intrusions using ML-based IDS systems. The process of feature selection entails assigning a score for every feature and choosing the best k-features. For intrusion detection systems, several features shall be investigated in detail for which some of them are relevant and others might be useless. Removing the unnecessary features improves the performance and the accuracy of the detection process as it eliminates overfitting and reduces computation time, and improves accuracy. In this work, correlation-based feature selection approach is employed using Pearson correlation [32]. This filtering method takes into account a set of features that are strongly associated with the target class but not with another. As a result, this filtering process is successful at removing redundant and irrelevant features since they have a poor association with the target class and are connected to at least one other feature [31]. In fact, the degree to which two variables are linearly connected is expressed by a Pearson correlation, which ranges from -1 to 1. Moreover, the correlation coefficients for multiple variables are generally presented in a table known as a correlation matrix. Such matrix is a useful tool for summarizing a large dataset and identifying and visualizing patterns, in which the variables are represented by rows and columns and the correlation coefficient is contained in each cell of the matrix.

The binary correlation matrix is illustrated in Figure 3. Several positively or negatively correlated features are indicated based on the color of the cells. Two traits are considered negatively (or positively) correlated if the Pearson correlation coefficient among them is relatively large, close to -1 (or 1). If such coefficient is near 0, the features are not related.

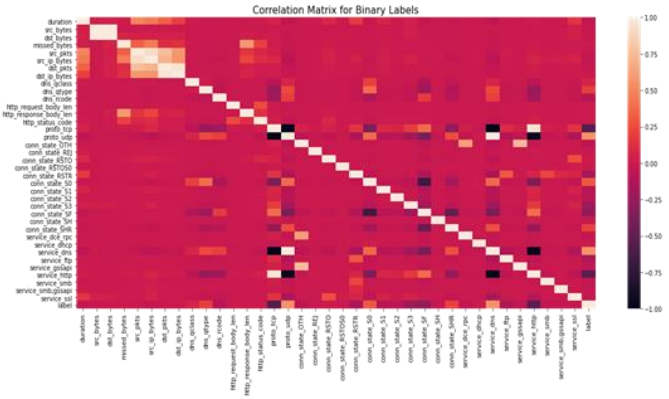

Figure 3 Binary Correlation Matrix

4.3. Data Normalization

In 'ToN-IoT' dataset, some features have relatively larger values than other features, while others might also have lesser values. Furthermore, out-of-range values might produce incorrect results since the classification algorithm could be biased in favor of features with higher values. In view of the same, data normalization is essential to avoid outweighing problem which would favor features with larger values over those with lesser ones. There exist different normalization techniques such as min-max and standard scalar approaches. In fact, min-max technique is employed in this study to scale the values of features between zero to one as given in Eq. (1).

$$Z = \frac{X - Xmin}{Xmax - Xmin} \quad (1)$$

Where Z denotes the normalized value and X represents the value of the feature. Xmax and Xmin indicate the maximum and minimum feature quantities, respectively.

4.4. Training Procedure

The dataset, which is provided in CSV format, is initially divided into two main sets. The first one comprises 70% of

RESEARCH ARTICLE

the entire dataset employed for training and validation purposes, while the second set represents the testing set used for assessing the effectiveness of the selected machine-learning techniques. Various evaluation metrics are mainly employed to assess the performance and effectiveness of particular machine learning techniques as explained in the following section.

4.5. Performance Evaluation

To provide a comprehensive description of the obtained results of ML-based IDS, different evaluation methods are selected to assess the efficiency of the used ML techniques. Particularly, precision, recall, F-measure, and accuracy metrics are employed to assess the performance of the detection rate using the confusion matrix, shown in Table 3, as explained in the following.

Table 3 Confusion Matrix

	Predicted as 'Normal'	Predicted as 'Attack'
Actual Normal Class	TP: True Positive	FP: False Positive
Actual Attack Class	FN: False Negative	TN: True Negative

4.5.1. Accuracy

The accuracy, defined in Eq. (2), is characterized by the ratio of the correct data of the model to the entire data as given below in Eq. (2).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

4.5.2. Precision

The precision, defined in Eq. (3), is characterized by the proportion of real cases amongst all positive cases identified by the model (TP). In other words, it represents the percentage of classified attack occurrences which are actually categorized as acksatt.

$$Precission = \frac{TP}{TP+FP} \quad (3)$$

4.5.3. Recall or Sensitivity

Sensitivity, commonly known as recall, is characterized by the ratio of the number of attacks identified as an attack by the model to the total number of attack traffic instances as defined in Eq. (4). It represents the number of true cases revealed in relation to the total number of true instances.

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

4.5.4. F1-Score

This metric, defined in Eq. (5), delivers a harmonic average measurement of sensitivity and precision of an estimator [33].

$$F1 - Score = \frac{2*Sensitivity*Precision}{Sensitivity+Precision} \quad (5)$$

5. DATASET EVALUATION

In this section, the proposed model using the selected machine learning algorithms is evaluated.

5.1. Random Forest (RF)

Due to its capabilities in tackling both classification and regression problems, RF technique has become very popular and largely utilized in today's machine learning professions. It is a supervised machine learning approach that generates excellent results even when no parameters are adjusted. Based on a limited number of records and features, each tree is deemed to be a poor classifier. Further, the trees' predictions were used in order to produce a final classification of attack and normal data [34]. The precision, recall, F1-score, and the overall accuracy metrics, are presented in Table 4. Moreover, the macro-averaging and weighted-averaging are also provided. In fact, the weighted-averaging is commonly utilized to score each prediction equally, while the macro-averaging is utilized to examine the performance of the classifier.

Table 4 RF Performance Evaluation

	Precision	**Recall**	**F1-score**
Abnormal	0.99	0.99	0.99
Normal	0.98	0.96	0.97
Accuracy			0.99
Macro Avg.	0.98	0.98	0.98
Weighted Avg.	0.99	0.99	0.99

5.2. K-Nearest Neighbor

Table 5 KNN Performance Evaluation

	Precision	**Recall**	**F1-score**
Abnormal	0.99	0.99	0.99
Normal	0.97	0.96	0.96
Accuracy			0.98
Macro Avg.	0.98	0.97	0.98
Weighted Avg.	0.98	0.98	0.98

KNN is a fundamental approach for classifying samples from a testing group to the nearest sample in the training group using certain criteria. It is a non-parametric method that does

RESEARCH ARTICLE

not impose any restriction on the distribution of the data. In the training set, KNN locates a set of k- observations that are nearest to the testing observation. The group is then labeled according to the k neighbors' most common class. Essentially, the most crucial aspects of KNN approach are the distance and number of neighbors [35]. The precision, recall, F1-score metrics, in addition to the overall accuracy, are presented in Table 5. Moreover, the macro-averaging and weighted-averaging are also provided.

5.3. Decision Tree Classifier (DT)

DT is a well-known building technique that utilizes the leaves and branches to imitate the decision tree in which the classification rule is represented by the internal node, and the class label is represented by the leaves. Further, the branch represents the results. In the training phase, the best qualities for the branches and core node are picked by using the information gain. Then, the decision node is built on the basis of the highest score of information gain. Consequently, a new sub-tree is created under the decision node. This procedure will be terminated only if all items in the selected sub-groups have a similar value, wherein the final value will be calculated and used as the output value. The cycle might also be terminated in case of one node only in the subgroup and no further possibilities. Similar to [36], a linear decision tree classifier has also been used in the experiments of this study. The training-testing set of 'ToN-IoT' has been employed in this context. The precision, recall, F1-score metrics, in addition to the overall accuracy, are presented in Table 6. The macro-averaging and weighted-averaging are also provided.

Table 6 DT Performance Evaluation

	Precision	Recall	F1-score
Abnormal	0.99	0.99	0.99
Normal	0.96	0.97	0.96
Accuracy			0.98
Macro Avg.	0.97	0.98	0.97
Weighted Avg.	0.98	0.98	0.98

5.4. Logistic Regression (LR)

The LR technique is widely employed for classification tasks since it is capable of estimating the likelihood that a given observation belongs to a particular group. LR is a variant of the linear regression and may be used in different useful applications effectively like intrusion detection and spam filtering. Based on a defined threshold, a particular instance might be predicted as an 'Attack' if the estimated probability is larger than the threshold. Otherwise, it is considered a 'Normal' instance. Inspired from [37], a logistic regression technique is employed in the experiments of this study. Again, the training-testing set of 'ToN-IoT' is used in this process. The precision, recall, F1-score metrics, in addition to the overall accuracy, are presented in Table 7. Moreover, the macro-averaging and weighted-averaging are also provided.

Table 7 LR Performance Evaluation

	Precision	Recall	F1-score
Abnormal	0.97	1.00	0.99
Normal	0.99	0.91	0.95
Accuracy			0.98
Macro Avg.	0.98	0.96	0.97
Weighted Avg.	0.98	0.98	0.98

5.5. Linear Support Vector Machine (SVM)

SVM is a well-known classification approach capable of dealing with both linear and nonlinear datasets. It is based on a separating hyperplane principle, in which the primary objective of SVM is to search for the best hyperplane which widens the difference between the groups. Essentially, there exist several kernel functions that may be used to characterize the hyperplane, varying from a linear kernel to a nonlinear kernel like the radial basis function (RBF). Similar to [38], a linear SVM has been employed in this study. The precision, recall, F1-score metrics, in addition to the overall accuracy, are presented in Table 8. Moreover, the macro-averaging and weighted-averaging are also provided.

Table 8 SVM Performance Evaluation

	Precision	Recall	F1-score
Abnormal	0.97	1.00	0.99
Normal	0.99	0.91	0.95
Accuracy			0.98
Macro Avg.	0.98	0.96	0.97
Weighted Avg.	0.98	0.98	0.98

6. DISCUSSION

6.1. Machine Learning Algorithms

Based on the employed methodology, five of the most reliable algorithms are used to train the system using 'ToN-IoT' dataset. The test has been conducted on 1000000 records included in the dataset with two labels, namely, 'Normal' and

RESEARCH ARTICLE

'Abnormal' (which represents 9 types of attacks). The preprocessing step is first implemented to remove irrelevant/missing data, and then the feature engineering process has been utilized to extract the most important features. Several trials have been conducted under two scenarios wherein all features are included in the first scenario. However, in the second scenario, some of these features are excluded by applying the feature engineering procedure. The obtained results under the first and second scenarios are provided in Table 9 and Table 10, respectively. It is evident that the accuracy in the second scenario is significantly improved as compared to the first scenario when all features were included. Essentially, the integration of the feature selection process played an essential role in enhancing the performance of the ML algorithms and evaluation results. The most important features obtained after applying the feature engineering process resulting in the highest accuracy which are highlighted in Table 11.

Table 9 Summarized Results under the First Scenario

Technique	RF	KNN	DT	LR	SVM
Accuracy	80.7%	76.3%	80%	80.6%	80.6%

Table 10 Summarized Results under the Second Scenario

Technique	RF	KNN	DT	LR	SVM
Accuracy	99%	98%	98%	98%	98%

Table 11 The Selected Features in the Second Scenario

ID	Feature	Type	Description
1	ts	Time	Timestamp of connection
2	src_ip	String	Source IP addresses
3	src_port	Number	Source ports
4	dst_ip	String	Destination IP addresses
5	dst_port	Number	Destination ports
6	proto	String	Transport layer protocols
7	service	String	Dynamically detected protocols

Based on the obtained results of the second scenario, it can be observed that RF algorithm outperformed KNN, DT, LR, and linear SVM algorithms. As summarized in Table 10, RF achieved 99% accuracy, while the other techniques achieved a comparable accuracy of 98%. Using the training-testing set, which is a part of the entire dataset, the five machine learning techniques achieved a great performance. These results were achieved by the generated features in the dataset, which significantly vary between 'Attack' and 'Normal' instances. The models performed quite well in terms of detection accuracy and false alarm rates. Essentially, the classification decision was supported by the integration of ports and IP-addresses features with the employed training/testing subgroups.

6.2. Integration of ML IDS in SDN

The intelligent IDS based on ML techniques can improve the operation of the traditional IDS based on signatures and rules. Secured access to the SDN controller can be implemented by authentication techniques, which is out of the scope of this study. However, the proposed architecture aims to integrate the ML-based IDS into the control plane for several reasons summarized as follows: (1) It is a programmable device that provides flexibility and easy maintenance, (2) It is scalable as any IoT device can be added to extend the network, (3) The essential task for the control plane is to perform routing and performance management in a dynamic way, which makes it perfect for traffic analysis and intrusion detection. OpenFlow protocol is employed for communication between different layers to manage the network. It can be used to transfer the detected activities to the IDS deployed in the controller. This protocol provides some other traffic characteristics based on aggregated traffic which can be used as flow features to feed the IDS system. The communication of the proposed architecture in the context of IoT is illustrated in Figure 3.

6.3. Comparison with Relevant Studie

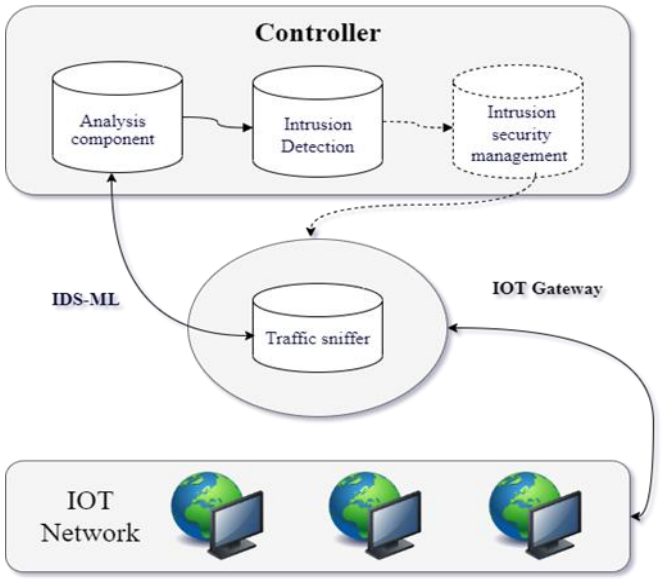

Figure 3 IDS-ML in IOT

RESEARCH ARTICLE

Table 12 Comparison with Relevant Studies

Ref.	Detection	SDN-based	ML-based	Algorithms	Accuracy
[16]	√	√	√	GNB	96.2%
				KNN	96.0%
				CART	98.9%
[18]	√	√	√	KNN	97.0%
				NB	89.0%
				RF	97.0%
				SVM	97.0%
[21]	√	√	√	SVM	92.5%
				NB	95.7%
				ANN	92.3%
				KNN	98.3%
[25]	√	√	√	RF	81.9%
Proposed	√	√	√	RF	99.0%
				DT	98.0%
				LR	98.0%
				KNN	98.0%
				SVM	98.0%

To assess the effectiveness of the proposed architecture for IDS functionalities in the IoT context, the proposed approach is compared with other existing approaches in literature. Relevant studies that are based on SDN and ML techniques for detection application are presented and summarized in Table 12. It can be observed that the proposed solution significantly improves the performance of the overall systems. The nature of the used dataset and the process of feature selection have contributed effectively to this improvement. However, due to time constraints, the integration of the proposed IDS in the SDN layer was not implemented in this study; however, it is anticipated to achieve higher levels of performance and accuracy.

7. CONCLUSION

In this study, an SDN-based IoT anomaly IDS is proposed to detect abnormal behaviors and attacks in IoT systems. ToN-IoT dataset is employed to analyze and evaluate the machine learning techniques implemented in the detection module. The obtained results demonstrate the detection accuracy of the five machine learning algorithms. Deployment of such system in SDN controllers improves the detection functionality by offering dynamic analysis and management processes. Future works include the implementation of the proposed system in SDN architecture, maximizing the detection coverage using combined traditional and ML-based IDS systems, in addition to investigating the possibility of zero-day detection and sophisticated multi-stage attacks using deep inspection of network traffic.

REFERENCES

[1] Bhunia, S. S., & Gurusamy, "Dynamic attack detection and mitigation in IoT using SDN," In 27th International telecommunication networks and applications conference (ITNAC), 2017,pp. 1-6.
[2] Moustafa, N., Turnbull, B., & Choo, K. K. R, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things," in IEEE Internet of Things Journal, vol. 6(3), 2018, pp.4815-4830.
[3] Sarica, A. K., & Angin, "Explainable security in SDN-based IoT networks," in Sensors, vol. 20(24), 2020, pp. 7326.
[4] Al-Akhras, M., Alawairdhi, M., Alkoudari, A., & Atawneh, S, "Using machine learning to build a classification model for iot networks to detect attack signatures," in Int. J. Comput. Netw. Commun.(IJCNC), vol. 12, 2020, pp. 99-116.
[5] Amangele, P., Reed, M. J., Al-Naday, M., Thomos, N., & Nowak, M, "Hierarchical machine learning for IoT anomaly detection in SDN," In 2019 International Conference on Information Technologies (InfoTech), 2019-September, pp. 1-4.
[6] Sangodoyin, A. O., Akinsolu, M. O., Pillai, P., & Grout, "Detection and Classification of DDoS Flooding Attacks on Software-Defined Networks: A Case Study for the Application of Machine Learning," in IEEE Access, vol. 9, 2021, pp. 122495-122508.
[7] Tsogbaatar, E., Bhuyan, M. H., Taenaka, Y., Fall, D., Gonchigsumlaa, K., Elmroth, E., & Kadobayashi, Y, "Sdn-enabled iot anomaly detection using ensemble learning," In IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, Cham, 2020, June, pp. 268-280.
[8] Cheng, H., Liu, J., Xu, T., Ren, B., Mao, J., & Zhang, "Machine learning based low-rate DDoS attack detection for SDN enabled IoT networks," in International Journal of Sensor Networks, vol. 34(1), 2020, pp. 56-69.
[9] Amangele, P., Reed, M. J., Al-Naday, M., Thomos, N., & Nowak, "Hierarchical machine learning for IoT anomaly detection in SDN," In International Conference on Information Technologies (InfoTech), IEEE, 2019-September, pp. 1-4.
[10] Sugi, S. S. S., & Ratna, S. R, "Investigation of machine learning techniques in intrusion detection system for IoT network," In 3rd International Conference on Intelligent Sustainable Systems (ICISS), IEEE, 2020-December, pp. 1164-1167.
[11] Bhunia, S. S., & Gurusamy, "Dynamic attack detection and mitigation in IoT using SDN," In 27th International telecommunication networks and applications conference (ITNAC), 2017,pp. 1-6.

RESEARCH ARTICLE

[12] Sarica, A. K., & Angin, "Explainable security in SDN-based IoT networks," in Sensors, vol. 20(24), 2020, pp. 7326.
[13] Moustafa, N., Turnbull, B., & Choo, K. K. R, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things," in IEEE Internet of Things Journal, vol. 6(3), 2018, pp.4815-4830.
[14] Al-Akhras, M., Alawairdhi, M., Alkoudari, A., & Atawneh, S, "Using machine learning to build a classification model for iot networks to detect attack signatures," in Int. J. Comput. Netw. Commun.(IJCNC), vol. 12, 2020, pp. 99-116.
[15] Tsogbaatar, E., Bhuyan, M. H., Taenaka, Y., Fall, D., Gonchigsumlaa, K., Elmroth, E., & Kadobayashi, Y, "Sdn-enabled iot anomaly detection using ensemble learning," In IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, Cham,, 2020, June, pp. 268-280.
[16] Sangodoyin, A. O., Akinsolu, M. O., Pillai, P., & Grout, "Detection and Classification of DDoS Flooding Attacks on Software-Defined Networks: A Case Study for the Application of Machine Learning," in IEEE Access, vol. 9, 2021, pp. 122495-122508.
[17] Cheng, H., Liu, J., Xu, T., Ren, B., Mao, J., & Zhang, "Machine learning based low-rate DDoS attack detection for SDN enabled IoT networks," in International Journal of Sensor Networks, vol. 34(1), 2020, pp. 56-69.
[18] Polat, H., Polat, O., & Cetin, "Detecting DDoS attacks in software-defined networks through feature selection methods and machine learning models," in Sustainability, vol. 12(3), 2020, pp. 1035.
[19] Amangele, P., Reed, M. J., Al-Naday, M., Thomos, N., & Nowak, "Hierarchical machine learning for IoT anomaly detection in SDN," In International Conference on Information Technologies (InfoTech), IEEE, 2019-September, pp. 1-4.
[20] Sugi, S. S. S., & Ratna, S. R, "Investigation of machine learning techniques in intrusion detection system for IoT network," In 3rd International Conference on Intelligent Sustainable Systems (ICISS), IEEE, 2020-December, pp. 1164-1167.
[21] Zeleke, E. M., Melaku, H. M., & Mengistu, F. G, "Efficient Intrusion Detection System for SDN Orchestrated Internet of Things," in Journal of Computer Networks and Communications, vol.2021, pp. 14.
[22] Bagui, S., Wang, X., & Bagui, S, "Machine Learning Based Intrusion Detection for IoT Botnet," in International Journal of Machine Learning and Computing, vol. 11(6), 2021, pp. 399-406.
[23] Satheesh, N., Rathnamma, M. V., Rajeshkumar, G., Sagar, P. V., Dadheech, P., Dogiwal, S. R., ... & Sengan, S, "Flow-based anomaly intrusion detection using machine learning model with software defined networking for OpenFlow network," in Microprocessors and Microsystems, vol. 79, 2020, pp.103285.
[24] Jafarian, T., Masdari, M., Ghaffari, A., & Majidzadeh, K, "Security anomaly detection in software- defined networking based on a prediction technique," in International Journal of Communication Systems, vol. 33(14), 2020, pp. e4524.
[25] Dey, S. K., & Rahman, M, "Effects of machine learning approach in flow-based anomaly detection on software-defined networking," in Symmetry, vol. 12(1), 2020, pp. 7.
[26] Moustafa, N, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets," in Sustainable Cities and Society, vol. 72, 2021, pp. 102994.
[27] Moustafa, Nour, "New Generations of Internet of Things Datasets for Cybersecurity Applications based Machine Learning: TON_IoT Datasets," in Proceedings of the eResearch Australasia Conference, Brisbane, Australia, 2019.
[28] https://research.unsw.edu.au/projects/toniot-datasets.
[29] Gad, A. R., Nashat, A. A., & Barkat, T. M, "Intrusion Detection System Using Machine Learning for Vehicular Ad Hoc Networks Based on ToN-IoT Dataset," in IEEE Access, vol. 9, 2021, pp. 142206-142217.
[30] Sarhan, M., Layeghy, S., Moustafa, N., Gallagher, M., & Portmann, M, "Feature Extraction for Machine Learning-based Intrusion Detection in IoT Networks," in arXiv:2018.12722 v1, N1, 2021.
[31] Binbusayyis, A., & Vaiyapuri, T, "Identifying and benchmarking key features for cyber intrusion detection: An ensemble approach," in IEEE Access, vol. 7, 2019, pp. 106495-106513.
[32] Chen, P., Li, F., Wu, C, "Research on intrusion detection method based on Pearson correlation coefficient feature selection algorithm," J. Phys. Conf. Ser, vol. 1757(1), 012054, 2021, pp.10.
[33] Precision and recall definition | deepai, https://deepai.org/machine-learning-glossary-and-terms/precision-andrecal
[34] Negandhi, P., Trivedi, Y., & Mangrulkar, R, "Intrusion detection system using random forest on the NSL-KDD dataset," In Emerging Research in Computing, Information, Communication and Applications, Springer, vol. , 2019, pp. 519-531.
[35] Almseidin, M., Alzubi, M., Kovacs, S., & Alkasassbeh, M, "Evaluation of machine learning algorithms for intrusion detection system," In 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), 2017-September, pp. 000277-000282.
[36] B. Ingre, A. Yadav, and A. K. Soni, "Decision tree based intrusion detection system for NSL-KDD dataset," in Information and Communication Technology for Intelligent Systems (ICTIS), vol. 2. Cham, Switzerland:Springer, 2018, pp. 207–218.
[37] C. Ioannou and V. Vassiliou, ''An intrusion detection system for constrained WSN and IoT nodes based on binary logistic regression,'' in Proc. 21st ACM Int. Conf. Modeling, Anal. Simulation Wireless Mobile Syst. Oct. 2018, pp. 259–263.
[38] Mohammadi, M., Rashid, T. A., Karim, S. H. T., Aldalwie, A. H. M., Tho, Q. T., Bidaki, M., ... & Hosseinzadeh, M, "A comprehensive survey and taxonomy of the SVM-based intrusion detection systems," in Journal of Network and Computer Applications, vol. 178, 2021, pp. 102983.

Authors

Tahani M. Alshammari: Received her Bachelor's degree in Computer Sciences from Northern Border University, KSA. Master's student in the Department of CyberSecurity at Al-Jouf University, KSA. Her interests cover several aspects, including: CyberSecurity, Open Source Programming, Network Security, Penetration Testing. It is also concerned with the impact of technology on society.

Faeiz M. Alserhani: Received his Bachelor's degree in Computer Engineering from King Saud University, Riyadh, SA, M.Ss degree in Computer and Information Networks from University of Essex, UK and the Ph.D. degree in Network and Information Security from University of Bradford, UK . He is currently a Professor assistance in The Department of Computer Engineering and Networks, Jouf University, SA. His interests cover several aspects across Network Security, CyberSecurity, Intrusion Detection Systems and Application of AI in Cybersecurity.

How to cite this article:

Tahani M. Alshammari, Faeiz M. Alserhani, "Scalable and Robust Intrusion Detection System to Secure the IoT Environments using Software Defined Networks (SDN) Enabled Architecture", International Journal of Computer Networks and Applications (IJCNA), 9(6), PP: 678-688, 2022, DOI: 10.22247/ijcna/2022/217701.