مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Application of Machine Learning Techniques for Traffic Anomaly Detection in IoT Environments

[1]**Osamah Waleed Abdulqahar**
Halab Intermediate School,
Directorate of Nineveh Education,
Ministry of Education
Mosul, 41001, Iraq
oms-wa82@nan.epedu.gov.iq

[2]**Jalal Qais Jameel**
[3]**Asmaa Waleed Abdulqahar**
[4]**Lamia Hassan Rahef**
Physiology Department, College of Medicine,
Al-Mustansiriyah University
Baghdad,Iraq
[2] jalalalqaisy1@uomustansiriyah.edu.iq
[3]physiology_aw@uomustansiriyah.edu.iq
[4]physiology_l@uomustansiriyah.edu.iq

Abstract

The current work proposed a modular solution based on data analysis from IoT networks operating with the MQTT protocol through intrusion detection agents linked to Machine Learning techniques for data streams. This research proposes an anomaly detection system in IoT environments of different infrastructures using the MQTT protocol and two machine learning algorithms. It also evaluates the integration between machine learning techniques and signature-based intrusion detection systems. This solution can detect unknown attacks by integrating machine learning techniques and a signature-based intrusion detection system. To evaluate the proposal in different networks, data collection initially used a pre-existing dataset containing several classes of MQTT attacks. These data were generated in a network simulation in the study. Communication data were also collected from a simulated network. Network information was captured by a dedicated collector agent programmed in Python. The model was evaluated using five metrics: Precision, Recall, F1-score, Accuracy and AUC ROC. The proposed models achieved in standard data, which becomes interesting in the defense of IoT networks because several SDIs suffer from many false positives. The accuracy of the applied models maintained an average between 75% and 82%, and standard data recall was obtained above 80% and attacks above 60%

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

using the online approach. Unlike related works that use an offline approach, the online model can be more scalable and usable in real-time applications. The execution of the two proposed models in the two different networks presented similar behaviors due to the defined parameters. Through the results, it was possible to analyze the performance of the proposed solution, which, despite having achieved low overall accuracy, detected anomalies with an average accuracy rate of 79%. In addition, the online learning technique solves the problem of retraining and domain change in traditional Machine Learning techniques.

Keywords: IOT; traffic; machine learning algorithms; MQTT protocol; IFASD algorithm.
AMS Subject Classification: 22E46, 53C35, 57S20.

Introduction

Machine learning (ML) is constantly evolving, and supervised learning techniques are one of the most prominent trends in this field.(M. Taye , 91, 2023)[1] In response to these advances and the problems detected in IoT (Internet of Things) network security, a solution has been developed focused on collecting data from accurate MQTT communications over Wi-Fi in IoT networks.(S. Abdulmalek, A. Nasir, W. Jabbar, M. Almuhaya, A. Bairagi, M. Khan and S. Kee , 1993, 2022)[2] MQTT, Message Queuing Telemetry Transport, is a lightweight messaging protocol that uses a publish/subscribe model to transfer information between devices in an IoT deployment. Due to their ubiquitous and diverse nature, such networks are prone to different types of anomalies that make it difficult to transmit information reliably. The proposed approach focuses on detecting, classifying, and handling such anomalies efficiently and accurately. (X. Liu, T. Zhang, N. Hu, P. Zhang and Y. Zhang, 169–176, 2020)[3]

In the first instance, real-time data is collected from various IoT devices that use the MQTT protocol. Promising results are obtained, with positive metrics in classifying and detecting anomalous traffic in MQTT payloads. (S. Abdulmalek, A. Nasir, W. Jabbar, M. Almuhaya, A. Bairagi, M. Khan and S. Kee, 1993, 2022)[4] The models can recognize multiple anomalies, from minor irregularities to severe problems that can affect the functionality of the IoT network. In addition to the above, The research

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

aims to improve the efficiency of the models, both in terms of the computational resources required and the time needed for training and validation. Therefore, given the information presented so far, this work aims to create a solution for the security of IoT networks that use the Message Queuing Telemetry Transport (MQTT) protocol for communication based on machine learning techniques for real-time anomaly detection through data streaming and automatic creation of rules in an intrusion detection system. (A. Al Hanif and M. Ilyas , 1782, 2024)[5] The experiments were conducted on two different networks to validate the functioning of the proposal in various infrastructures and communications. The data from these two networks were saved in datasets, which were necessary to evaluate the results of the experiments. The first was obtained from work on attack detection in MQTT networks(A. Sultan, S. Mehmood and H. Zahid , 118–121, 2022)[6], and a second dataset was collected from our network simulation.(H. Siddharthan, T. Deepa and P. Chandhar , 33095–33110, 2022)[7] .The Isolation Forest and Isolation Forest Algorithm for Streaming Data algorithms were applied and evaluated using evaluation metrics to compare the results of the proposal's application in real-time and offline.Thus, the elaboration and conclusion of this work bring as contributions: (i) The proposal of an anomaly detection system in IoT environments of different infrastructures, but using the MQTT protocol; (ii) The evaluation of two machine learning algorithms, one common in offline training and another adapted and created for data streaming; (iii) The evaluation of the integration between machine learning techniques and signature-based intrusion detection systems.

Statement of the Problem

The growth of cyberspace in recent years, accompanied by the explosion in the use and implementation of devices and mass consumption of the Internet as a result of the COVID-19 pandemic, in the context of IoT(R. De, N. Pandey and A. Pal , 102171, 2020)[8] ensures that by 2023, home automation applications will cover the majority of machine-to-machine (M2M) connections, with an estimated value of 48% of this category. On the other hand, vehicular communications (V2V) have experienced rapid growth that is expected to reach thirty percent in 2023 compared to the

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

document presented by CISCO in 2020. (C. Xu, S. Wang, P. Song, K. Li and T. Song, 6796, 2023)[9]

Additionally, in sensing systems such as those deployed in IoT networks, one of the requirements is the collection of reliable data. This is because these data are used in the systems' decisions provided by the information collection nodes. It is in this framework that a clear solution is consolidated to identify a possible anomaly (data affected by environmental, network, application or information transmission reasons) in the payload sent to the edge or core of the network without substantially degrading the use of bandwidth or processing time. (S. Moore, C. Nugent and S. Zhang, 147–163, 2020)[10]

Background of the Problem

IoT connections are comprised of various technologies for their operation, such as WiFi, Bluetooth, Zigbee, LoRaWAN, and Sigfox. Of the listed technologies, one stands out for its open protocol, wide use in the market and ease of acquiring equipment, such as WiFi.(H. Nejati S. Aldin, M. Ghods, F. Nayebipour and M. Torshiz , 100258, 2024)[11] However, there are still challenges that sensor networks (WSN) must face, such as variable latencies due to physical channel occupancy, bandwidth, connectivity, energy management and security. The latter will be addressed in detail later, paying particular attention to using Machine Learning techniques to detect anomalous behavior within deployments. Within the anomalies, there is a field to explore, which includes anomalies in data collection, also known as sensing failures. (R. Ahmad, R. Wazirali and T. Abu-Ain, 4730, 2022)[12]

MQTT Protocol

MQTT is a communication protocol that can communicate flexibly and quickly on the network it operates on and does not cause conflict on the network. (J. Roldán-Gómez, J. Carrillo-Mondéjar, J. Gómez and S. Ruiz-Villafranca , 10991, 2022)[13] DAD is a labeled dataset consisting of seven days of network activity with anomalous packets spread over five days.(L. Vigoya, D. Fernandez, V. Carneiro and F. Cacheda , 3745, 2020)[14] This dataset has three types of anomalies: duplication, tampering, and modification of the MQTT message.

Clients using the MQTT protocol can be publishers and subscribers. Devices can range from a microcontroller to a server running an MQTT

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

library and connecting to a server/broker on the network. Because MQTT is a lightweight protocol with low implementation costs, it can be implemented on low-resource devices, which makes it very important for IoT networks. (M. Kashyap, V. Sharma and N. Gupta, (M. Kashyap, V. Sharma and N. Gupta, 1611–1618, 2018)[15]

According to Sarker, *et al.* (I. Sarker , 160,2021)[16], one of the areas of application of machine learning algorithms is the problem of finding patterns in data necessary for application in improving companies' marketing, detecting fraud and predicting future issues. In addition, it is important to highlight several types of machine learning, including Supervised and Unsupervised. In supervised learning, the system relies on the data and the result of its interpretation, in other words, labeled data, to predict new values, generating interpretations based on new data based on previous training, such as predicting risks and sales.(J. Afriyie, K. Tawiah, W. Pels, S. Addai-Henne, H. Dwamena, E. Owiredu, S. Ayeh and J. Eshun , 100163, 2023)[17] Examples of supervised learning algorithms are Logistic Regression (LR), Gaussian Naive Bayes (NB), k-nearest, Neighbors (k-NN), Support Vector Machine (SVM), Decision Trees (DT), Random Forests (RF). (I. Sarker , 160,2021)[16] These traditional machine learning methods generally work offline or through batch learning, in which the algorithm is trained on the entire dataset and is immediately deployed to make decisions or predictions without performing any update or retraining, given the costs involved.(S. Uddin, A. Khan, M. Hossain and M. Moni , 1–16, 2019)[18] For this reason, these approaches are not very scalable for real-world applications.

Because of this problem, studies are looking for ways to solve it; one is online learning, a method that trains the algorithm in sequential order for each new data instance. Thus, no additional retraining costs are required, making this approach more scalable for real-world analysis tasks. Several algorithms have been built for this purpose. (S. Ahmed, M. Alam and M. Hassan 13521–13617, 2023,)[19] Other state-of-the-art algorithms are offline or trained in batch and can be used in online learning, such as IFASD, IF adaptation or OCSVM. (I. Sarker , 160,2021)[16,](N. Seliya, A. Zadeh and T. Khoshgoftaar , 1–31, 2021)[20]

With the initial objective of assessing the two types of learning presented in this section (online and offline), algorithms with their versions in both

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

approaches were chosen. Although there are several algorithms for online learning and others for offline learning, there are few that can be used in both. For this reason, this research uses the IF and IFASD algorithms to compare their performances in anomaly detection, one used in offline learning and the other in online education, respectively, but both maintain the same algorithm. (P. Naarayanan, D. Franklin, N. Gouvtham and P. Femi, 1101–1106, 2020)[21]

Isolation Forest (IF)(G. Susto, A. Beghi and S. Mcloone , 89–94, 2017)[22] is an unsupervised learning algorithm for anomaly detection. Its operation is based on isolating anomalies instead of creating standard patterns of instances, as with most common anomaly detection approaches. The proposed method is based on two quantitative properties: (1) There are few anomalous instances in the data read, and (2) These data have values that are very different from standard data. Therefore, it is easier to isolate anomalous data.

The method uses the data to build a forest with isolated trees by randomly dividing the instances, choosing a value between the maximum and minimum of the selected value until all are isolated. (Z. Wang, G. Song and C. Gao, 96319–96333, 2019)[23] Therefore, the anomalies are the data with the shortest paths in each sub-tree. Therefore, given the isolation susceptibility, the anomalies are close to the trees' roots.

The main features of this algorithm are that it allows small samples to be used without losing performance since large trees are not needed to isolate average data, and small samples can reduce the effects of flooding (labeling average data as anomalies) and masking (labeling anomalous data as usual). It also does not use distance or density measures, eliminating computational costs and requiring little time and memory for execution. (H. Xiang, X. Zhang, M. Dras, A. Beheshti, W. Dou, X. Xu, 678–687, 2023)[24]

Isolation Forest Algorithm for Streaming Data (IFASD)

IFASD is an algorithm created to solve common offline learning problems presented here. (M. Kareem and L. Muhammed, 223–228, 2024)[25] Adapted from IF, it aims to detect anomalies by bringing all the capabilities that IF implements and adding the ability to detect them online or in streaming data.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Its operation is based on sliding windows, as shown in Fig. 1. The data is inserted within a sliding window with a pre-defined size. With this data, an initial detector is trained by the IF algorithm. In the testing phase, the detector examines all instances within each sliding window and labels them as anomalies.

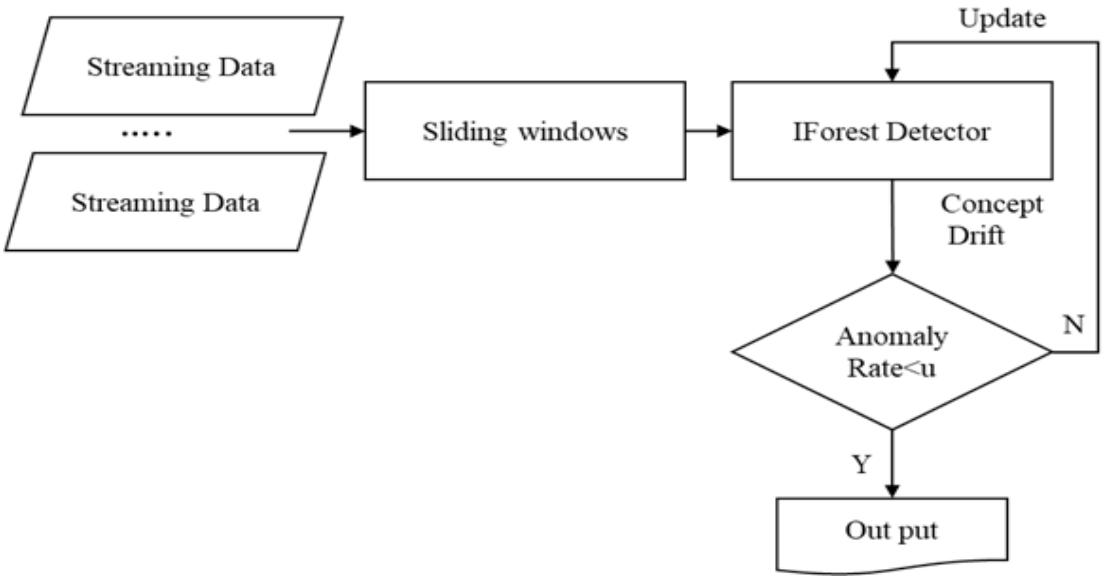

Fig. 1. IFASD Framework. (M. Togbe, Y. Chabchoub, A. Boly, M. Barry, R. Chiky and M. Bahri' 13, 2021)[26]

With the labeled data, the algorithm can continue its operation if the anomaly rate within the sliding window is less than a pre-established limit u. If the anomaly rate exceeds the threshold, the algorithm detects more anomalies than expected, and the domain difference problem may occur. In this case, the algorithm must be discarded with the latest data.(M. Heigl, K. Anand, A. Urmann, D. Fiala, M. Schramm and R. Hable , 1534, 2021)[27]

Review of Literature

Three works were selected that focus mainly on techniques for detecting attacks under the MQTT protocol in IoT networks. Although Castañeda *et al.*,(A. Castañeda, J. Aveleira and H. Aláiz-Moretón , 5281–5296, 2023)[28] Husnain *et al.(* M. Husnain, K. Hayat, E. Cambiaso, U. Fayyaz, M. Mongelli, H. Akram, A. Ghazanfar and G. Shah , 567, 2022)[29] and Alaiz-Tahir, *et al.(* U. Tahir, M. Abid, M. Fuzail, N. Aslam , 01–13,

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

2024)[30] have the same motivation to detect MQTT attacks on brokers in networks, they use different environments, data, attacks, algorithms and evaluation metrics.

Hindy *et al.* (H. Hindy, E. Bayne, M. Bures, R. Atkinson, C. Tachtatzis and X. Bellekens , 73–84,2020)[31] simulate brute force attacks and network scans, using six techniques and five evaluation metrics for each algorithm. Their study uses six techniques and two evaluation metrics for each. Finally, Castañeda *et al..*, *(*A. Castañeda, J. Aveleira and H. Aláiz-Moretón, 5281–5296, 2023)[28] focus on classifying and detecting multiple MQTT-specific attacks using three techniques evaluated by around five metrics.

Hindy *et al.* (H. Hindy, E. Bayne, M. Bures, R. Atkinson, C. Tachtatzis and X. Bellekens , 73–84,2020)[31] study and evaluate the performance of six machine learning techniques on a dataset of attacks based on the MQTT protocol. Initially, the database is created using an environment with several simulated devices, a centralized broker and an attacker. During the simulation, the devices generate average communication data. The attacker performs four attacks: three generic (aggressive scan, UDP scan and brute force SSH) and one specific to MQTT (brute force MQTT). Focusing on the classification of standard data and attacks, the study analyzes six techniques: Logistic Regression (LR), Gaussian Naive Bayes (NB), k-nearest, Neighbors (k-NN), Support Vector Machine (SVM), Decision Trees (DT) and Random Forests (RF). The metrics used to evaluate the results were Overall Accuracy, Precision, Recall and F1-score. The study aimed to capture features based on the MQTT packet, single-stream network and dual-stream network to emphasize then that the data based on the MQTT packet is sufficient to detect the three generic attacks but not to detect the MQTT attack since there is a difficulty in differentiating average MQTT data from MQTT attack data. The importance of single- and dual-stream features for detecting the MQTT attack was also found. Using ML algorithms and techniques, Ciklabakkal *et al.* (E. Ciklabakkal, A. Donmez, M. Erdemir, E. Suren, M. Yilmaz, P. Angin, ARTEMIS , 3690–3692, 2019)[32] propose an intrusion detection system for MQTT attacks in IoT networks. The prototype structure used to collect data consists of sensors connected to a gateway device (Raspberry Pi) that sends the data to a broker. The attack packets

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

were generated by the gateway using the MQTT malaria tool. The broker captures these packets and, through the techniques, trains and predicts whether that packet contains an attack or not. The results obtained are sent to the user through an interactive application. According to the authors, the broker-centric approach minimized computational activities in the gateway (Raspberry Pi). The study by Stoyanova, *et al.* (M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, E. Markakis , 1191–1221,2020)[33] presents a structure in which IoT devices communicate with the central broker to obtain a model with low processing needs required to detect attacks directed at them.

Alsoufi, *et al.(* M. Alsoufi, S. Razak, M. Siraj, A. Ali, M. Nasser, S. Abdo , 659–675, 2020)[34] proposes a system for detecting anomalies or unknown attacks that uses machine learning techniques to learn the normal behavior of the monitored environment and alert any different behavior. The system was developed to be installed on IoT devices. Four attacks targeting the network and devices were evaluated. Two algorithms were tested and assessed using about three metrics. Hasan *et al.* (2019) simulate an environment generating average data and data from seven different attacks and use them to evaluate five algorithms with about five metrics.

Hasan *et al.* (H. Mahmudul, M. Islam, M. Zarif and M. Hashem, 100059,2019)[35] propose a model for detecting attacks and anomalies in an IoT network infrastructure. The dataset was obtained through an infrastructure simulated by an intelligent space orchestration system (DS2OS) (PAHL; AUBET, 2018). It contains average communication data and denial of service (DoS) attacks, Data Type Probing, Malicious Control, Malicious Operation, Scan, Spying and Wrong Setup. The algorithms evaluated were Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) and Artificial Neural Network (ANN). The evaluation metrics used were accuracy, precision, recall, F1 score and Receiver Operating Characteristic Curve. By running the model, the main result obtained by the study showed that the Random Forest (RF) algorithm had greater accuracy than the other models, in addition to having predicted more DoS attack samples.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Methodological procedures

Initially, a literature review was carried out on the functioning of Internet of Things networks and their protocols, specifically MQTT, which this research focused on, defence mechanisms used in these networks, such as intrusion detection and prevention, and ways to improve the security provided by these mechanisms through machine learning techniques, to automate defences and recognise new attacks. Thus, the knowledge acquired was necessary to develop a proposal for an architecture to improve security in IoT networks that use the MQTT protocol in their communication.

The procedures of this work are Data Collection, Data Preprocessing and Selection, Model Definition and Training, SDI Configuration and Rule Creation, and Validation and Analysis of Results.

Data Collection

To evaluate the proposal in different networks, data collection initially used a pre-existing dataset containing several classes of MQTT attacks. These data were generated in a network simulation in the study. (A. Singh, A. Kumar, V. Kumar, 209–213, 2022)[36] Communication data from a simulated network was also collected. The network information was captured by a dedicated collector agent programmed in Python.

Data Preprocessing and Selection

During the network data collection, the agent captures all communication from the application, transport, and network layers. For this reason, preprocessing of the data is necessary to adapt it to the input of the proposed model. In this procedure, missing values and textual values were replaced by unique numerical identifiers, given that the algorithm receives numerical values as input, explaining the procedures previously performed in executing the processing and selection steps.

In addition to preprocessing, with knowledge of the domain and focus of this proposal on MQTT networks, only the values related to the MQTT application and communication layer were selected from the data to demonstrate that it is possible to detect attacks using only the information from the MQTT protocol.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Model Definition and Training

The algorithms used were IF for offline training and IFASD for online training. The IFASD model is implemented through the PySAD framework built-in Python.(N. Jeffrey, Q. Tan and J. Villar , 3283, 2023)[37] IF is implemented using the sklearn framework. For both algorithms, the datasets used are divided into 70% of the data for training and 30% of the data for testing. Furthermore, since this is anomaly detection, the algorithms are initially trained with all the normal data from the training data set without labels. A contamination parameter is required for both models, so it is necessary to analyse the dataset beforehand to know how much anomalous data it contains, which can be used in defining the model.

SDI Configuration and Rule Creation

A proprietary agent was used with SDI Suricata to detect threats, an open-source, network-based, and signature-based tool. In other words, it analyses network traffic and uses the rules defined in its configuration to detect attacks.

The agent creates the rules using the model outputs and transforms them into the pattern expected by Suricata. In this way, new rules will be generated for new attacks so that SDI can identify them and alert the user or administrator. In general, the operation of the SDI together with the model aims to save the rules created and detect attacks that occurred in the past, considering that the ML algorithm can change its concept; for this reason, it may no longer detect attacks previously detected, these are detected from the rules saved in the SDI.

Validation and Analysis of Results

For validation and analysis of the results of the experiments, the datasets obtained were used to execute the proposed models. The model was evaluated using five metrics: Precision, Recall, F1-score, Accuracy and AUC ROC.

Each metric uses indicators in its calculation: False Positive, False Negative, True Positive and True Negative. These indicators define the class of data predicted by the algorithm, allowing the evaluation of its correct and incorrect predictions.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Since this research aims to predict anomalous data, it is essential to emphasise that anomalies are treated as the positive class and normal data as the negative class. Therefore, the indicators for understanding can be defined as:

- True Positive (TP): Predicted anomaly, and the data refer to an anomaly.
- False Positive (FP): Predicted anomaly, but the data refer to standard data.
- True Negative (TN): Normal data is predicted and refers to standard data.
- False Negative (FN): Normal data predicted, but the data refers to an anomaly.

Therefore, the metrics are described as follows:

- Precision (P): Defines the number of correct predictions for a given class based on the predictions made. Eq. (1) defines, given the predicted anomalies, the percentage of anomalies correctly predicted. Eq. (2) defines, given the predicted average data, the percentage of average data correctly predicted.

$$Precision\ (P) = \frac{TP}{TP+FP}, \quad (1)$$

$$Precision\ (P) = \frac{TN}{TN+FN}. \quad (2)$$

- Recall (R): Defines the number of correct predictions for a given class. Eq. (3) defines the percentage of correct anomaly predictions given the number of anomalies in the instances. Eq. (4) defines the percentage of correct normal data predictions given the total number of normal data contained in the instances.

$$\text{Recall (R)} = \frac{TP}{TP+FN}, \quad (3)$$

$$\text{Recall (R)} = \frac{TN}{TN+FP}. \quad (4)$$

- F1-Score (F1): Defines a combination of the Precision and Recall metrics to be observed together through a single number. Defined by Eq. (5).

$$\text{F1} - \text{Score (F1)} = 2.\frac{P \times R}{P+R}. \quad (5)$$

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

• AUC ROC: The area under the ROC evaluates the algorithm's performance based on the limit of the relationship between true positives and true negatives. With values from 0 to 1, being:
• algorithm predicts anomalies as normal data and normal data as anomalies;
• (0.5) algorithm cannot distinguish between the two classes;
• (>0.5) performance rate of the prediction of the two classes by the algorithm (NARKHEDE, 2018).

In the execution of the proposal in real-time, the model does not have information on the assertiveness of the prediction, so the analysis of its functioning was given through the visualization of the alerts generated by the SDI in moments of simulations of attacks on the network. Therefore, knowledge of the domain is necessary to evaluate its functioning.

Proposals

The proposed solution consists of an architecture, shown in Fig. 2, composed of a collector agent, an ML agent, and an SDI agent. In addition, the visualization of data, network devices and alerts generated by the SDI and ML agents is inserted through a Web Page to clarify the network's current behavior.

The solution was developed in modules to allow each agent to be installed on different devices or all on the same device if necessary. In addition, since the SDI Agent and Broker used to exist tools, replacing them with others that meet the requirements for the different networks is possible. The following sections will present the operation of each of the proposed modules in detail.

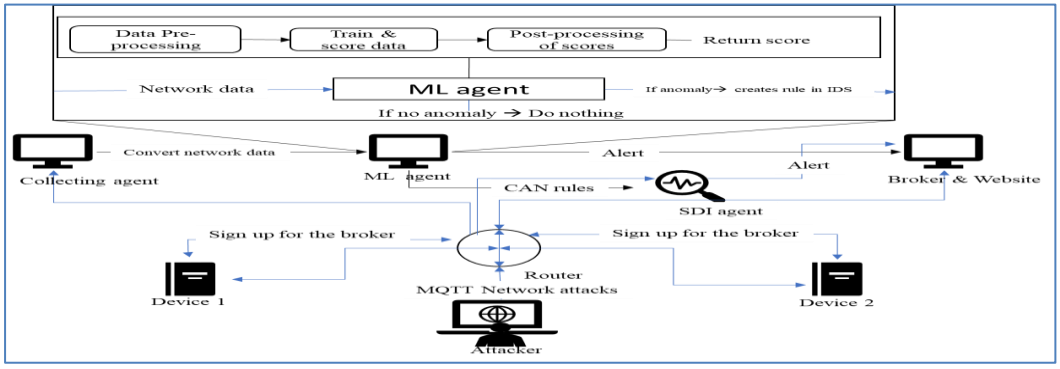

Fig. 2. Architecture of the proposed solution.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Collector Agent

The Collector Agent is responsible for reading and processing network traffic, then inserting it into a dataset, sending it to the ML Agent, or both. It was programmed in Python and the main libraries used by it are: *csv*, used to generate the dataset; socket: used to open sockets and allow reading of network traffic; *pahomqtt*: used to transform the agent into an MQTT client to send data to the ML Agent. The parameters *args* to communicate with the ML Agent via MQTT *client/broker connection.host, args.port, args.topic* are required, meaning the ML Agent identifier, port (default: 1883) and topic (default: network/data) respectively.

The packets generated in the communication are extracted to obtain the TCP and IP headers as well as the payload and MQTT header. Only packets with source/destination port 1883 or 8883 (MQTT standards) are used.

Machine Learning (ML) Agent

The ML Agent is responsible for receiving the data, pre-processing the categorical features, transforming them into unique numeric values using a word dictionary and using them in the algorithm's training and scoring. The score is post-processed by a calibrator and . classified as an anomaly If so, the model forwards the data from that packet to the SDI Agent to create a threat detection rule. The data is also sent to the Broker to view the alert since the SDI Agent maintains alerts to detect attacks that are the same as those previously detected.

It was programmed in Python and its communication with the Collector Agent is done through the MQTT protocol. The agent was created as an MQTT2 broker and subscribes to the user-defined topic (default: network/data), through which it receives data from the Collector Agent. It also has an MQTT client that connects to the SDI Agent using the required port, host and topic (default: ml/rules) to send the model results.

The algorithm used by the Agent is IFASD, implemented in the PySAD framework. The parameters used in the algorithm were the default ones, so the adopted sliding window size was 2048 data instances, as explained in the algorithm operation. The default operation of the Agent is to collect data for a pre-defined initial time to train the network's normal behavior.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

At this point, a thread is executed parallel to count the time while traffic is received and saved in a list in memory. After the time is up, the data is trained in the model, which then scores all subsequent data.

Each scored data is transformed into a probability employing a conformal calibrator, as shown in Source Code 4, also implemented in the PySAD framework. It receives the following parameters: windowed True and window_size 2048, the same number of instances used in the model definition. In this way, each data will likely be an anomaly. If it exceeds the pre-defined limit, it is characterized as an anomaly, and the model sends the data to create the rule for the SDI Agent.

Intrusion Detection Agent (IDA)

The Intrusion Detection Agent is responsible for receiving data from the ML Agent indicating an anomaly transforming them into detection rules for a rule-based intrusion detection system. It is also responsible for reading the alerts generated by the IDA and sending them to the broker.

It was programmed in Python, and its communication with the ML Agent is done through the MQTT protocol. The agent was created as an MQTT broker using the Mosquitto tool and subscribes to the user-defined topic (default: ml/rules), through which it receives data from the ML Agent. It also has an MQTT client that connects to the broker using the port, host and topic required to send alerts.

The intrusion detection system used was Suricata. Its installation is very simple, and it can be used once installed. To create custom rules, insert them into a new file with the format accepted by SDI3 and restart the service for detection.

On Linux systems, the path to the Suricata configuration file is /etc/suricata/suricata.yaml, The path to the custom rules file must be added to the custom rules file using the rule-files option contained therein. After the rules file is populated, Suricata has the command suricata -c ruleset-reload-nonblocking to update the rules non-blocking (in the background) without having to restart the Suricata service completely. Alerts are read and sent to the broker using the file /var/log/suricata/eve.json, which generates alerts in JSON format.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Broker and Web Page

Since the proposed network has an MQTT communication environment, a broker must receive information from the devices. In this work, Dojo T4 an open-source middleware and broker platform was chosen that facilitates the development of solutions in the IoT ecosystem.

Its own micro-services architecture allows the management of the life cycle of devices, the construction of flows and data for real-time processing and data persistence, and an interface for accessing data in real-time.

This work used it to manage the broad visualization of devices and alerts generated in the SDI Agent. Its architecture has an internal broker that receives messages via MQTT and inserts them into devices created for visualization.

Devices and Attacker

The devices were simulated using two MQTT clients programmed in Python using the paho – mqtt library. The first device sends random words in a JSON. The second sends a location simulation in JSON format.

The attacker was also simulated as a client using the Malaria5 tool, generating fuzzy payloads and anomalous connection data.

Experiments and Results

This chapter will present and discuss the results obtained in the experiments performed. Initially, the procedures performed in both experiments are described. The following sections will present the configurations adopted for each experiment and their executions and results.

The first experiment was performed to compare the two proposed algorithms and the algorithm's functioning in the ML Agent, considering that the data had already been collected from another network. The second experiment was performed based on a simulation in the proposed network, thus comparing the two algorithms and evaluating the functioning of the Collector Agent, ML Agent, SDI Agent, and the visualization of the alerts generated.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

All experiments reported here were performed on a VMWARE virtual machine with 12GB of RAM, Lubuntu 20.04 LTS operating system, 4 logical processors, and 2 cores per Intel Core i7 virtualized processor.

Initial Procedures

Pre-processing was necessary to adapt the datasets for the proposed models In both cases, all categorical values were transformed into unique numeric values using a label encoding preprocessor. Of the features, the following were transformed:

mqtt. conack. flags, mqtt. conflags, mqtt. hdrflags, mqtt. protoname, mqtt. msg

In addition, to maintain all instances related to the network domain, a selection was made of features related only to the application layer data and communication via the MQTT protocol. For this reason, of the features contained in the datasets, the following were removed: tcp. flags, tcp. time_delta, tcp. len.

MQTTset Network Experiment

This experiment used the dataset collected in the study, which contains data from 10 sensors connected to a broker and data from an attacker performing approximately 6 MQTT attacks on the network (Flooding DoS, MQTT Publish flood, SlowITe, Malformed data, Brute Force).

The model provided the results obtained in Figs. 3 and 4 with the IF algorithm using the following parameters:

- max_samples: 0.2, maximum number of samples that can be used in each estimator, with 100 being the default number of estimators.
- random_state: 110, number chosen to define the pseudo-randomization the algorithm performs when choosing each feature.
- contamination: 0.3, contamination rate of the dataset.
- n_jobs: -1, number of jobs used in parallel executing the algorithm. -1 means that all jobs were used.

The parameters were chosen based on a series of model runs to evaluate the best parameters. Since the IF algorithm works best with few instances and a low contamination rate, the lowest number of samples and lowest contamination rate were chosen for the two proposed models, considering the performance of the algorithms with the given parameters.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

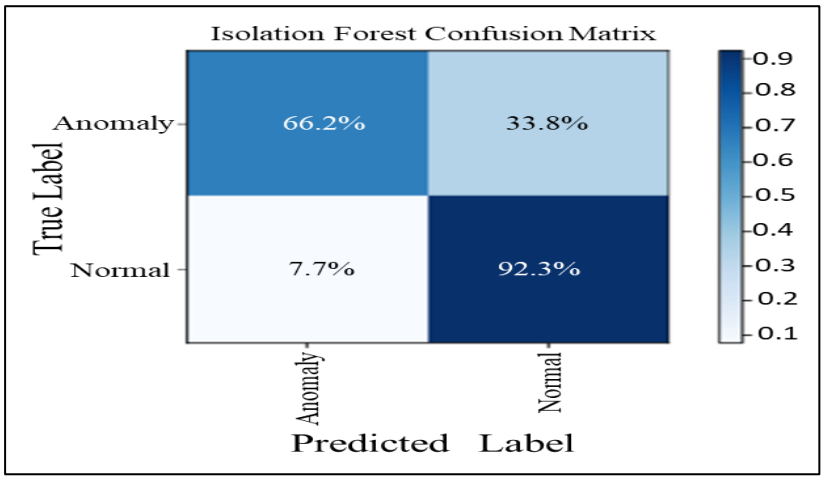

Fig. 3. IF Confusion Matrix - MQTTset Experiment.

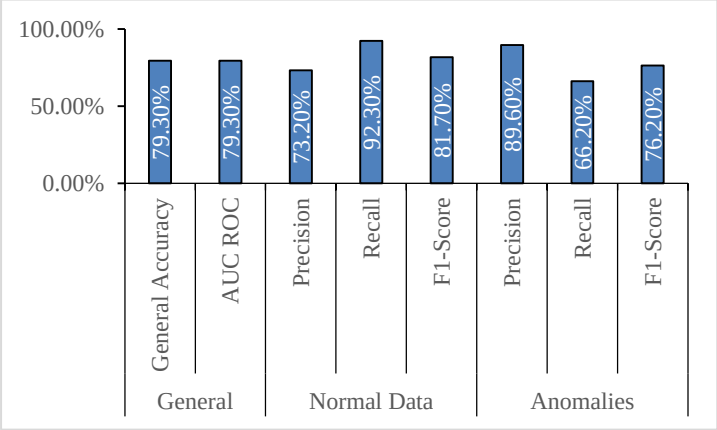

Fig.4. IF Evaluation Metrics - MQTTset Experiment.

Using the same contamination rate of 0.3 and the sliding window size of 2048 instances, the IFASD algorithm obtained the results shown in Fig. (5) and Fig. (6).

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

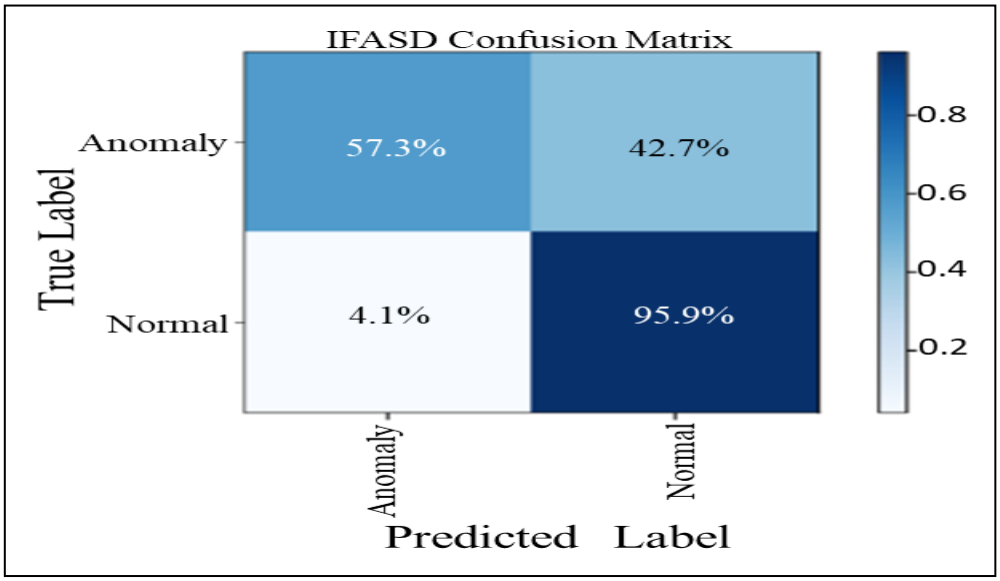

Fig. 5. IFASD Confusion Matrix - MQTTset Experiment.

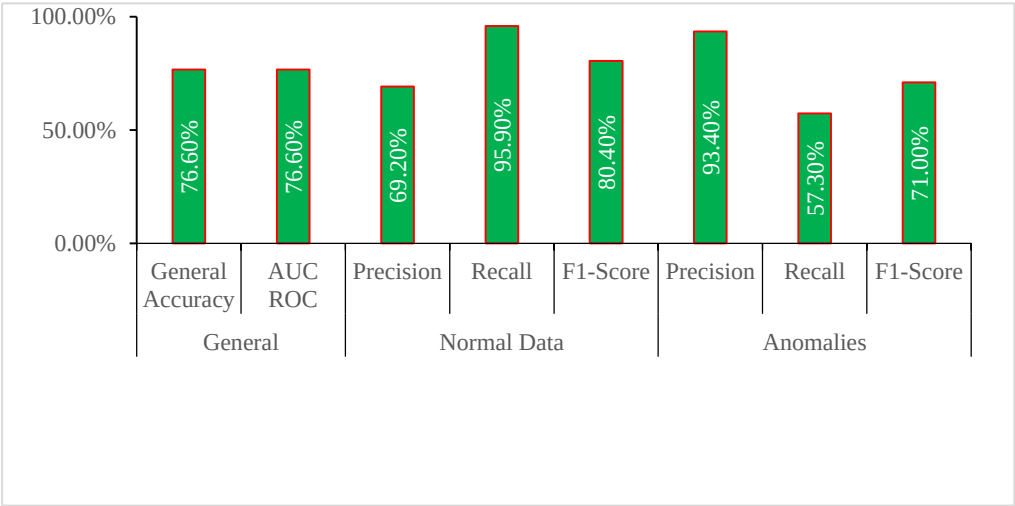

Fig. 6. IFASD Evaluation Metrics - MQTTset Experiment.

Proposed Network Experiment

In this experiment, a simulation was performed on the proposed network, containing 2 sensors connected to a broker and an attacker performing attacks. A dataset with the features obtained from the simulation will be used to evaluate the proposed models offline. It was also possible to analyze the functioning of the SDI Agent in creating the rules through the alerts generated and viewed on the Web Page.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

The model provided the results obtained in Figs 7 and 8 with the IF algorithm using the following parameters:

- max_samples: 0.2, the maximum number of samples that can be used in each estimator, with 100 being the default number of estimators.
- random_state: 110 is a number chosen to define the pseudo-randomization the algorithm performs when choosing each feature.
- contamination: 0.3, contamination rate of the dataset.
- n_jobs: -1, number of jobs used in parallel in the algorithm's execution. -1 means all jobs were used.

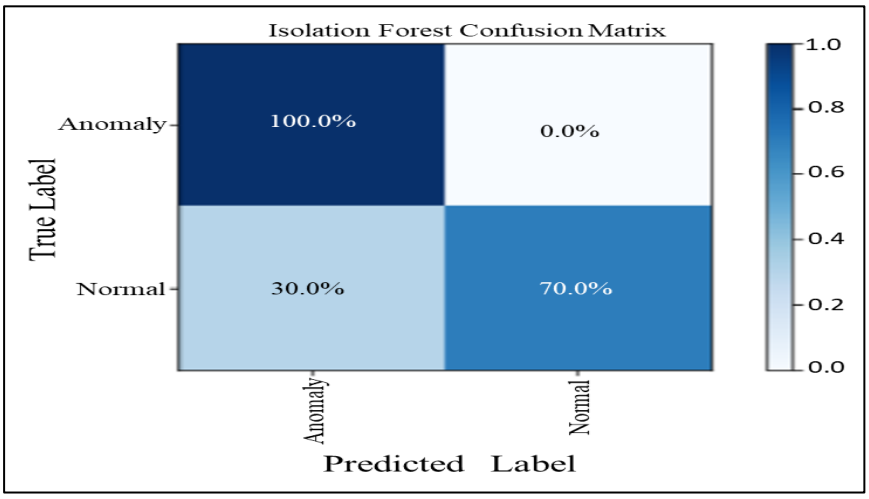

Fig. 7. IF Confusion Matrix - Proposed Experiment.

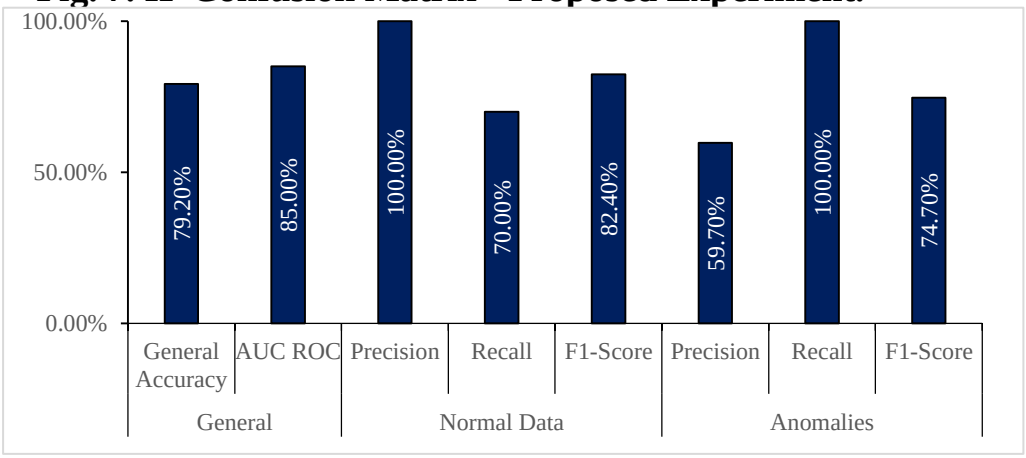

Fig. 8. IF Evaluation Metrics - Proposed Experiment.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Using the same contamination rate of 0.3 and the sliding window size of 2048 instances, the IFASD algorithm obtained the results shown in Figs. 9 and 10.

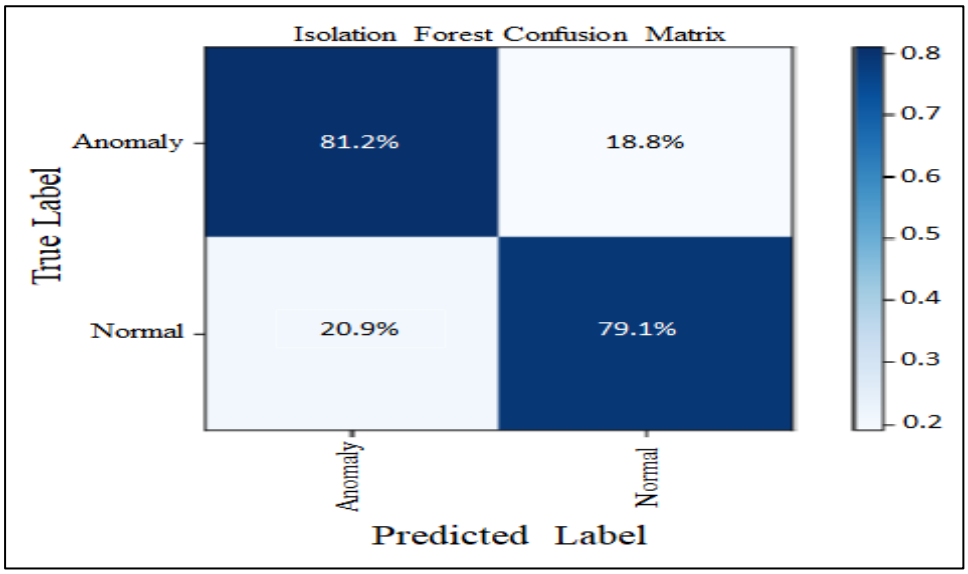

Fig. 9. IFASD Confusion Matrix - Proposed Experiment.

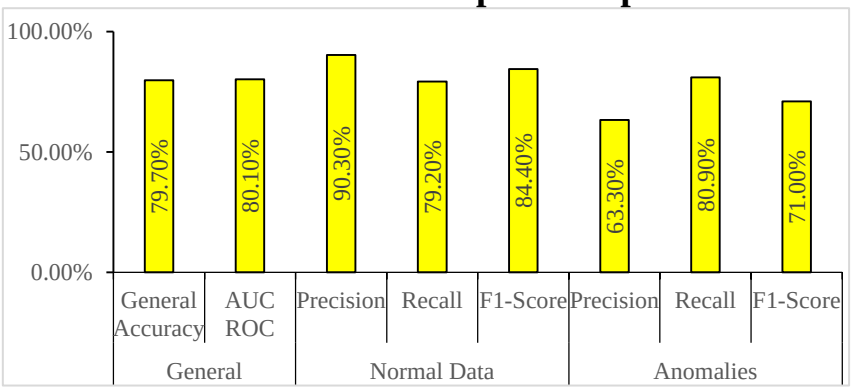

Fig. 10. IFASD Evaluation Metrics - Proposed Experiment.

When viewing the operation of the SDI agent, it is possible to observe the alerts generated on the Web Page, in addition to the data sent by the devices.

Discussion

Despite achieving a low accuracy rate compared to related works that used different features and algorithms, the applied models obtained a higher recall in normal data, which becomes interesting in defense in IoT networks because several SDIs suffer from a high number of false

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

positives. In addition, the accuracy did not differ much between the two approaches (offline and online). This indicates that the online model can be more scalable and usable in real-time applications, unlike related works that use an offline approach.

In general, the accuracy of the applied models maintained an average between 75% and 82%, and standard data recall was obtained above 80% and attacks above 60% using the online approach.

In the experiment of the proposed network, a fixed anomaly rate was used, one of the reasons for the decrease in the model's performance, since in real-time applications, it is not yet possible to define the number of attacks that are being carried out due to the total amount of data in the communication. This phenomenon can be seen in Fig. 11.

Finally, executing the two proposed models in the two networks presented similar behaviors due to the defined parameters. The IFASD algorithm becomes scalable and usable in the proposed network, considering the real-time communication of the devices and the implementation and operation of the collector, ML and SDI agents.

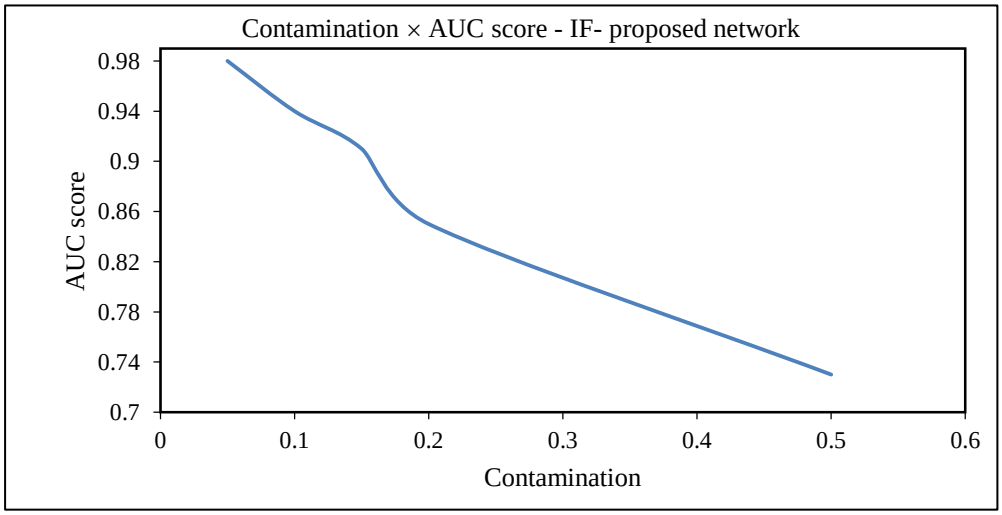

Fig. 11. Contamination and AUC ROC relationship

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

Conclusions and Future Work

With the increase in IoT devices and the parallel emergence of several security flaws, accompanied by the lack of defense capacity on the part of the devices due to their low resources, it has become necessary to build a security solution that detects not only known attacks but also anomalous attacks, known as zero-days.

Therefore, this work proposed a modular solution based on data analysis from IoT networks operating with the MQTT protocol through intrusion detection agents linked to Machine Learning techniques for data streams. Through the results, it was possible to analyze the performance of the proposed solution, which, despite having achieved a low overall accuracy, detected anomalies with an average accuracy rate of 60%. In addition, the online learning technique solves the problem of retraining and domain change in traditional Machine Learning techniques. As future work, to build an even more accurate solution, other algorithms based on online learning can be used together, obtaining the results of several algorithms and generating a more accurate result from them. In addition, better data extraction for differentiation in the algorithm can be done, and techniques for extracting context from features can be useful, given the functioning of anomaly detection in data that is very different from normal data.

In the context of application in IoT networks, scalability assessments of the algorithms and agents used can provide information used to improve the performance of the solution, such as the assessment of the training and detection time of the models, the number of resources spent on using the agents, and others. Furthermore, due to the heterogeneity property of these networks, the solution requires additional evaluations in networks with different protocols and new devices.

In addition, when using the IFASD algorithm, it is necessary to pre-define the network anomaly rate. In real environments, detecting the rate of contained attacks is impossible, so research is needed to change the anomaly rate automatically to avoid reducing the performance of the algorithms with different numbers of attacks.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

References

[1]. M. Taye, Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions, *Computers 12*(5) (2023) 91.

[2]. S. Abdulmalek, A. Nasir, W. Jabbar, M. Almuhaya, A. Bairagi, M. Khan and S. Kee, IoT-Based Healthcare-Monitoring System towards Improving Quality of Life: A Review, *Healthcare (Basel)* 10(10) (2022) 1993.

[3]. X. Liu, T. Zhang, N. Hu, P. Zhang and Y. Zhang, The method of Internet of Things access and network communication based on MQTT, *Comput. Commun.* 153 (2020) 169–176.

[4]. S. Abdulmalek, A. Nasir, W. Jabbar, M. Almuhaya, A. Bairagi, M. Khan and S. Kee, IoT-Based Healthcare-Monitoring System towards Improving Quality of Life: A Review. *Healthcare (Basel)* 10(10) (2022) 1993.

[5]. A. Al Hanif and M. Ilyas, Effective Feature Engineering Framework for Securing MQTT Protocol in IoT Environments, *Sensors (Basel)* 24(6) (2024) 1782.

[6]. A. Sultan, S. Mehmood and H. Zahid, Man in the Middle Attack Detection for MQTT based IoT devices using different Machine Learning Algorithms, in *Int. Conf. Artif. Intell.* (2022) 118–121.

[7]. H. Siddharthan, T. Deepa and P. Chandhar, SENMQTT-SET: An Intelligent Intrusion Detection in IoT-MQTT Networks Using Ensemble Multi Cascade Features, *IEEE Access* 10 (2022) 33095–33110.

[8]. R. De, N. Pandey and A. Pal, Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice, *Int. J. Inf. Manage*. 55 (2020) 102171.

[9]. C. Xu, S. Wang, P. Song, K. Li and T. Song, Intelligent Resource Allocation for V2V Communication with Spectrum-Energy Efficiency Maximization, *Sensors (Basel)* 23(15) (2023) 6796.

[10]. S. Moore, C. Nugent and S. Zhang, IoT reliability: a review leading to 5 key research directions, *CCF Trans. Pervasive Comp. Interact.* 2 (2020) 147–163.

[11]. H. Nejati S. Aldin, M. Ghods, F. Nayebipour and M. Torshiz, A comprehensive review of energy harvesting and routing strategies for IoT

sensors sustainability and communication technology, *Sens. Int.* 5 (2024) 100258.
[12]. R. Ahmad, R. Wazirali and T. Abu-Ain, Machine Learning for Wireless Sensor Networks Security: An Overview of Challenges and Issues, *Sensors* 22(13) (2022) 4730.
[13]. J. Roldán-Gómez, J. Carrillo-Mondéjar, J. Gómez and S. Ruiz-Villafranca, Security Analysis of the MQTT-SN Protocol for the Internet of Things, *Appl. Sci.* 12(21) (2022) 10991.
[14]. L. Vigoya, D. Fernandez, V. Carneiro and F. Cacheda, Annotated Dataset for Anomaly Detection in a Data Center with IoT Sensors, *Sensors (Basel)* 20(13) (2020) 3745.
[15]. M. Kashyap, V. Sharma and N. Gupta, Taking MQTT and NodeMcu to IOT: Communication in Internet of Things, *Procedia Comput. Sci.* 132 (2018) 1611–1618.
[16]. I. Sarker, Machine Learning: Algorithms, Real-World Applications and Research Directions, *SN Comput. Sci*. 2(3) (2021) 160.
[17]. J. Afriyie, K. Tawiah, W. Pels, S. Addai-Henne, H. Dwamena, E. Owiredu, S. Ayeh and J. Eshun, A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions, *Decis. Anal. J.*, 6 (2023) 100163.
[18]. S. Uddin, A. Khan, M. Hossain and M. Moni, Comparing different supervised machine learning algorithms for disease prediction, *BMC Med. Inform. Decis. Mak.* 19(1) (2019) 1–16.
[19]. S. Ahmed, M. Alam and M. Hassan, Deep learning modelling techniques: current progress, applications, advantages, and challenges, *Artif. Intell. Rev.* 56(11) (2023) 13521–13617.
[20]. N. Seliya, A. Zadeh and T. Khoshgoftaar, A literature review on one-class classification and its potential applications in big data, *J. Big Data* 8 (2021) 1–31.
[21]. P. Naarayanan, D. Franklin, N. Gouvtham and P. Femi, Comparing the Performance of Anomaly Detection Algorithms, *Int. J. Eng. Res. Technol.* 9(7) (2020) 1101–1106.
[22]. G. Susto, A. Beghi and S. Mcloone, Anomaly Detection through on-line Isolation Forest: An application to plasma etching, in *Annu. SEMI Adv. Semicond. Manuf. Conf.* (2017) 89–94.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

[23]. Z. Wang, G. Song and C. Gao, An isolation-based distributed outlier detection framework using nearest neighbor ensembles for wireless sensor networks, *IEEE Access* 7 (2019) 96319–96333.
[24]. H. Xiang, X. Zhang, M. Dras, A. Beheshti, W. Dou, X. Xu, Deep Optimal Isolation Forest with Genetic Algorithm for Anomaly Detection, in *IEEE Int. Conf. Data Min*. (2023) 678–687.
[25]. M. Kareem and L. Muhammed, Anomaly Detection in Streaming Data using Isolation Forest, in *Int. Women Data Sci. Conf. Prince Sultan Univ*. (2024) 223–228.
[26]. M. Togbe, Y. Chabchoub, A. Boly, M. Barry, R. Chiky and M. Bahri, Anomalies Detection Using Isolation in Concept-Drifting Data Streams, *Computers* 10(1) (2021) 13.
[27]. M. Heigl, K. Anand, A. Urmann, D. Fiala, M. Schramm and R. Hable, On the Improvement of the Isolation Forest Algorithm for Outlier Detection with Streaming Data, *Electronics* 10(13) (2021) 1534.
[28]. A. Castañeda, J. Aveleira and H. Aláiz-Moretón, Characterization of threats in IoT from an MQTT protocol-oriented dataset, *Complex Intell. Syst.* 9(5) (2023) 5281–5296.
[29]. M. Husnain, K. Hayat, E. Cambiaso, U. Fayyaz, M. Mongelli, H. Akram, A. Ghazanfar and G. Shah, Preventing MQTT Vulnerabilities Using IoT-Enabled Intrusion Detection System, *Sensors* 22(2) (2022) 567.
[30]. U. Tahir, M. Abid, M. Fuzail, N. Aslam, Enhancing IoT Security through Machine Learning-Driven Anomaly Detection, *VFAST Trans. Softw. Eng*. 12(2) (2024) 01–13.
[31]. H. Hindy, E. Bayne, M. Bures, R. Atkinson, C. Tachtatzis and X. Bellekens, Machine Learning Based IoT Intrusion Detection System: An MQTT Case Study, in *Int. Netw. Conf.* (2020) 73–84.
[32]. E. Ciklabakkal, A. Donmez, M. Erdemir, E. Suren, M. Yilmaz, P. Angin, ARTEMIS: An Intrusion Detection System for MQTT Attacks in Internet of Things, *Symp. Reliab. Distrib. Syst.* (2019) 3690–3692.
[33]. M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, E. Markakis, A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches and Open Issues, *IEEE Commun. Surv. Tutor*. 22(2) (2020) 1191–1221.

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

[34]. M. Alsoufi, S. Razak, M. Siraj, A. Ali, M. Nasser, S. Abdo, Anomaly Intrusion Detection Systems in IoT Using Deep Learning Techniques: A Survey, in *Int. Conf. Reliab. Inf. Commun. Technol.* (2020) 659–675.
[35]. H. Mahmudul, M. Islam, M. Zarif and M. Hashem, Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches, *Internet of Things* 7 (2019) 100059.
[36]. A. Singh, A. Kumar, V. Kumar, A Study on MQTT Protocol and its Cyber Attacks, *Int. Adv. Res. J. Sci. Eng. Technol.* 9 (2022) 209–213.
[37]. N. Jeffrey, Q. Tan and J. Villar, A Review of Anomaly Detection Strategies to Detect Threats to Cyber-Physical Systems, *Electronics* 12(15) (2023) 3283.

تطبيق تقنيات تعلم الآلة لكشف شذوذ حركة مرور البيانات في بيئات إنترنت الأشياء

اسامة وليد عبد القهار[1]
متوسطة حلب للبنين
مديرية تربية نينوى
وزارة التربية

جلال قيس جميل[2]
اسماء وليد عبد القهار[3]
لمياء حسن رهيف[4]
كلية الطب /فرع الفسلجة،
الجامعة المستنصرية،ةبغداد،العراق

[2]jalalalqaisy1@uomustansiriyah.edu.iq
[3]physiology_aw@uomustansiriyah.edu.iq oms-wa82@nan.epedu.gov.iq
[4]physiology_l@uomustansiriyah.edu.iq

مستخلص البحث:
يقدم هذا العمل حلاً مجزأً (Modular) يعتمد على تحليل البيانات المتدفقة من شبكات إنترنت الأشياء (IoT) العاملة ببروتوكول"MQTT" ، وذلك عبر وكلاء لكشف التسلل (Intrusion Detection Agents)مرتبطين بتقنيات تعلم الآلة المخصصة لتدفقات البيانات (Data Streams).تكمن مساهمة هذا البحث في اقتراح نظام لكشف الشذوذ (Anomaly Detection System)في بيئات إنترنت الأشياء ذات البنى التحتية المتنوعة التي تستخدم بروتوكول MQTT، وذلك بالاعتماد على خوارزميتين من خوارزميات تعلم الآلة. كما يُقيّم البحث مدى التكامل بين تقنيات تعلم الآلة وأنظمة كشف التسلل القائمة على التواقيع (Signature-based IDS). ويتميز هذا الحل بقدرته على رصد الهجمات غير المعروفة (Unknown Attacks)من خلال هذا التكامل الهجين. ولتقييم المقترح عبر شبكات مختلفة،

مجلة كلية التربية الاساسية
كلية التربية الاساسية – الجامعة المستنصرية
P- ISSN: 3007-3650 E- ISSN: 2706-8536
Journal of the College of Basic Education Vol.32 (No.139) 2026, pp.800-827

اعتمد جمع البيانات في البداية على مجموعة بيانات جاهزة (Pre-existing Dataset) تتضمن فئات متنوعة من هجمات MQTT التي تم توليدها عبر محاكاة للشبكة في هذه الدراسة، كما تم جمع بيانات الاتصال من شبكة محاكاة أخرى؛ حيث التُقطت معلومات الشبكة بواسطة وكيل جمع مخصص (Collector Agent) تم برمجته بلغة بايثون.(Python)
تضمّنت عملية تقييم النموذج خمسة مقاييس أساسية هي: الدقة (Precision) ، والاستدعاء (Recall)، ومقياس(F1-score) F1، والصحة العامةAccuracy) ، والمساحة تحت منحنى خصائص التشغيل للمستقبل ROC، AUC وقد حققت النماذج المطبقة معدل استدعاء أعلى في البيانات القياسية(Standard Data) ، وهو أمر بالغ الأهمية في الدفاع عن شبكات إنترنت الأشياء نظراً لأن العديد من أنظمة كشف التسلل التقليدية تعاني من كثرة الإنذارات الكاذبة.(False Positives) حافظت النماذج المطبقة على متوسط صحة عامة يتراوح بين 75% و82%، في حين تجاوز معدل استدعاء البيانات القياسية 80%، وتخطى 60% للهجمات باستخدام النهج الفوري/المباشر .(Online Approach) وخلافاً للأعمال السابقة التي تعتمد على النهج غير المباشر (Offline Approach) ، فإن النموذج الفوري (Online Model) يتميز بقابلية أعلى للتوسع وعملية أكبر في تطبيقات الوقت الفعلي (Real-tim Applications).وقد أظهر تشغيل النموذجين المقترحين في شبكتين مختلفتين سلوكيات متطابقة نتيجة للمحددات (Parameters) التي تم ضبطها. ومن خلال النتائج، أتيحت إمكانية تحليل أداء الحل المقترح، والذي نجح في رصد الشذوذ بمتوسط معدل صحة بلغ 79% على الرغم من تسجيله دقة عامة منخفضة نسبيًا. علاوة على ذلك، تغلبت تقنية التعلم الفوري (Online Learning)على مشكلة إعادة التدريب (Retraining) وتغير النطاق (Domain Change) التي تواجه تقنيات تعلم الآلة التقليدية.

الكلمات المفتاحية: إنترنت الأشياء(IoT) ، حركة مرور البيانات(Traffic) ، خوارزميات تعلم الآلة، بروتوكولMQTT ، خوارزمية.IFASD