Network Forensics and Intrusion Detection in MQTT-Based Smart Homes

Lama AlNabulsi [1], Sireen AlGhamdi [2], Ghala AlMuhawis[3], Ghada AlSaif[4], Fouz AlKhaldi[5], Maryam AlDossary[6], Hussian AlAttas [7] and Abdullah AlMuhaideb [8]

2180005711@iau.edu.sa 2180002698@iau.edu.sa 2180003634@iau.edu.sa 2180006143@iau.edu.sa
2180003624@iau.edu.sa mmsaldossary@iau.edu.sa htalattas@iau.edu.sa amalmuhaideb@iau.edu.sa

University of Imam Abdulrahman bin Faisal, College of Computer Science and Information Technology, KSA

Summary

The emergence of Internet of Things (IoT) into our daily lives has grown rapidly. It's been integrated to our homes, cars, and cities, increasing the intelligence of devices involved in communications. Enormous amount of data is exchanged over smart devices through the internet, which raises security concerns in regards of privacy evasion. This paper is focused on the forensics and intrusion detection on one of the most common protocols in IoT environments, especially smart home environments, which is the Message Queuing Telemetry Transport (MQTT) protocol. The paper covers general IoT infrastructure, MQTT protocol and attacks conducted on it, and multiple network forensics frameworks in smart homes. Furthermore, a machine learning model is developed and tested to detect several types of attacks in an IoT network. A forensics tool (MQTTracker) is proposed to contribute to the investigation of MQTT protocol in order to provide a safer technological future in the warmth of people's homes. The MQTT-IOT-IDS2020 dataset is used to train the machine learning model. In addition, different attack detection algorithms are compared to ensure the suitable algorithm is chosen to perform accurate classification of attacks within MQTT traffic.

Keywords:

Internet of things (IoT), Message Queuing Telemetry Transport (MQTT), machine learning, anomaly detection, smart home.

1. Introduction

Imagine the possibility of waking up to the smell of freshly made coffee in the morning, to the sunlight from your window without having to draw back your curtain, not having to worry whether you locked the door on your way out to work, or whether your kids have made it home safely after school. Luckily, with the rapid emergence of new technologies, our day-to-day lifestyles are becoming easier and simpler with the integration of Internet of Things (IoT). IoT is a framework that permits devices to be connected and remotely observed across the Internet.

Recently, the IoT field has had strong development, being right now utilized in different domains like smart homes. A smart home is a piece of the IoT paradigm and means to incorporate home automation, permitting devices and items in a home to be connected through the Internet, enabling customers to observe and control them remotely [1]. The security issues, threats, and attacks related to IoT have been declared as a demanding and promising region of exploration [2].

2. Background

Smart homes are spreading and have succeeded in acquiring people's attention in recent years. But with the spread of it, new security challenges and targets for attackers are raised [3]. Although there is much work done in digital forensics, the amount of work done in IoT forensics is very limited. This section demonstrates the IoT infrastructure, IoT network protocols, and the attacks against Message Queuing Telemetry Transport (MQTT) protocol.

2.1 IoT Infrastructure

IoT architecture is an arrangement of the underlying systems, which deliver services using IoT. Smart homes, digital factories, automated warehouse, etc., depend on the underlying infrastructure to convey their IoT capabilities [4]. IoT architecture varies incredibly based on the implementation; thus, it must be open and adaptable enough for open protocols to deal with network applications [5]. Despite the fact that there is no single IoT architecture that is globally agreed upon, the most basic and broadly accepted format is the three-layer architecture which consists of: perception, network, and application layers [6] as shown in Figure 1.

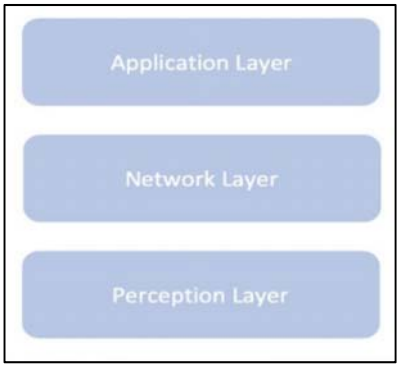

Fig. 1 IoT Infrastructure [6].

Manuscript received April 5, 2023
Manuscript revised April 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.4.12

The perception layer represents the physical level which has sensors for gathering and processing information. In this level, there are objects that have the ability to interact with the outside world and are equipped with computing abilities, therefore having intelligent and smart sense. The network layer is responsible for transferring the data given from the perception layer to the application layer. It incorporates all the technologies and protocols that make this interconnection occur. The application layer incorporates all the software needed to offer services. Data is collected, stored, and filtered in this layer and databases, analysis software, etc., are used [7].

2.2 IoT Application Layer Protocols

IoT protocols are methods of correspondence to ensure protection and security of information transmitted between smart devices [8]. There are a few normalized protocols for IoT application layer to lead Machine-to-Machine (M2M) communication. This section discusses the two most well-known IoT protocols utilized for information transmission in smart homes: Constrained Application Protocol (CoAP) and MQTT [9].

2.2.1 CoAP

This protocol was initially delivered for IoT frameworks dependent on Hypertext Transfer Protocol (HTTP). CoAP is a simple and lightweight request–response protocol that is based on an asynchronous trade of data running over User Datagram Protocol (UDP) connection, which does not provide any dependability. CoAP was created to be utilized in the middle of devices on the network. It may be run in two distinct modes, confirmed and non-confirmed delivery modes. For sensors dealing with relaxed reliability as a necessity, the non-confirmed mode is sufficient. Confirmations may be requested uniquely for certain packets [8][9][10].

2.2.2 MQTT

MQTT protocol uses publish/subscribe model and is explicitly intended for M2M correspondence. It joins PCs and networks into software and middleware and runs over TCP/IP, which gives Quality level of Service (QoS). This level ensures the reliability of message transmission [8][11]. The structure of MQTT consists of a broker - server- and clients (Figure 2). A client can be both a publisher and a subscriber. When a client device subscribes to a topic, the broker delivers the published messages on that topic to the device [12][13].

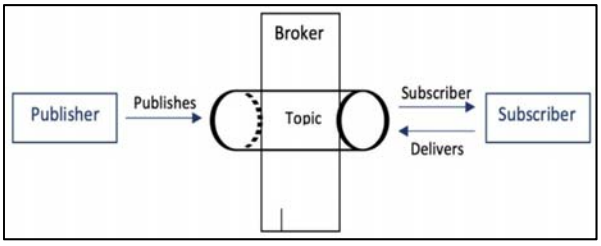

Fig. 2 MQTT Communication Process [13].

MQTT has a data packet size with low overhead, at least [> 2 bytes]. This protocol is a data-agnostic protocol that can communicate information in different structures like Extensible Markup Language, JavaScript Object Notation, text and binary files [4].

Since MQTT is responsible for the communication between devices, IoT environments face several attacks conducted by attackers in order to gain network access and disrupt or stop the broker during the subscriber's and publisher's communication. The most common attacks carried out by attackers against this protocol are Denial of service (DoS), Man in the Middle (MITM), and Botnet over MQTT.

2.2.2.1 Denial of service attack (DoS)

A DoS can be conducted on the broker by sending a vast amount of connection requests as often as possible, thus making the broker occupied as in flooding attacks. In the event that numerous connection requests reach simultaneously, the buffer will be depleted, and the broker won't be ready to deal with all the new approaching requests. During the DoS attack, there is a fast pace of expansion in the quantity of request packets, which prompts to stop the brokers functions and restricts the working of the IoT environment network as shown in Figure 3 [14].

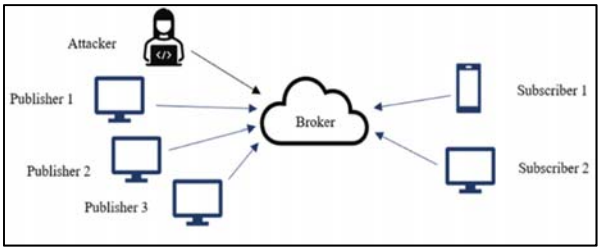

Fig. 3 DoS Attack on MQTT [14].

2.2.2.2 Man in the Middle attack (MITM)

MITM intrudes on the messages transmitted between two devices, which are the broker and sensor, to adjust the content. Despite the fact that security was not in the design

of this protocol, it has a few security precautions. MQTT provides a two-way handshake by permitting authentication to the client. On the off chance that SSL/TLS is implemented on the obliged asset devices, then this system takes into consideration the encryption of information in the message. If SSL/TLS is not implemented, the client's username and password that validate the client are in plain text. Therefore, this two-way handshake is not sufficient against MITM attacks, thus both authentication and encryption are required to prevent MITM attacks [14].

2.2.2.3 Botnet over MQTT

Botnet is a network of many bots -type of malware installed on a compromised computer- controlled by BotMaster. The BotMaster uses a specific broker to control numerous IoT devices with one distributed massage in a particular topic as visualized in Figure 4. Also, the BotMaster can receive victim status and subscribe to the status of every IoT device (botnet). This attack is extremely proficient, particularly if we assume that BotMaster provides a single command to all botnets at the same time (e.g., spamming and DoS attack) [15].

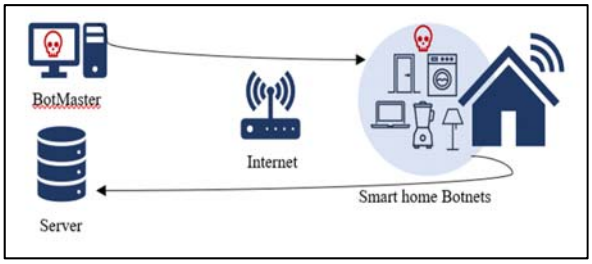

Fig. 4 Botnet Attack over MQTT [16].

3. Literature Review

3.1 IoT in Smart Homes

One of the most fundamental trademark provisions of smart devices in an organization dependent on IoT is to accumulate a more extensive set of information that has been made, and afterwards move the assembled information to the sender/receiver server through the web. IoT gives consistent client experiences which altogether improve individuals' daily lives and is shown by how conspicuous such devices are today. Nonetheless, the expansion of smart devices isn't just inside the homegrown environment, it is likewise the main thrust behind the improvement of an interconnected information-based world; our economies and societies, more explicitly, Container Network Interface (CNI) ideas, like smart homes, smart transportations, smart cities and so forth, are subject to smart technologies and IoT [17].

Kim et al. [3] paper demonstrates that smart homes nowadays are highly demanded, because they provide many automation services with the ability to sense, collect and distribute sounds, images, and videos. Smart devices can be remotely controlled through controllers such as smartphones. Moreover, smart home IoT devices gather and process data related to motion, temperature, lighting control, and the complex data are stored within a different device. As a result, it will be helpful for forensic investigations.

3.2 Smart Homes Security Challenges

Although smart devices support the undertakings of regular daily existence, Anthi et al. [18] highlighted in their paper that their reliance on Information Communication Technology (ICT) and IoT devices accompany gigantic security hazards. In this manner, IoT-based organizations are generally powerless against straightforward or complex attacks that should be distinguished in the early phase of information transmission for saving the organization from these malevolent attacks.

Dorobantu et al. [19] proposed some security solutions to IoT oriented attacks. They also conveyed that when attempting to dispatch a new IoT device as fast as possible, IoT originators regularly disregard security aspects. Need of solid security in IoT is the main interest of the day due to the expanding number of IoT devices and cyber-attacks. Many organizations have examined the smart frameworks as of now accessible and inferred that their security is totally unacceptable. Notwithstanding their weaknesses, these devices don't have programmed schedules for automatic updates and use decoded or poorly encoded communication channels. Hence, more and more smart devices (phones, network gadgets, CCTV cameras etc.) are associated with largescale cyber-attacks. The main cyber threats to IoT are data breaches, botnets and malware attacks, and DDoS attacks. IoT challenges the current security procedures, and it associates the internet with the real world. This has genuine security implications, as the danger of attacks transits from the digital to the actual world. Thus, the attack region is definitely extended from known dangers and known devices to extra security dangers of new devices, conventions and work processes. Numerous working frameworks are moving from shut frameworks to IP-based frameworks that further grow the attack surface.

From the standpoint of law enforcement and digital forensic investigators, Perumal et al. [20] said that IoT presents a variety of issues in smart homes. Cyber

criminals can utilize routers, televisions, refrigerators, and other internet connected devices to launch broad and spread attacks. Conti et al. [21] paper's contribution is in smart home forensics. Their paper highlighted the challenge of how to extract useful information and detect if there is any malicious activity. Furthermore, it's hard to acquire evidence because many IoT nodes gather and handle private information, resulting in the difficulty of identifying meaningful evidence.

3.3 Signature-based Detection

The authors of [22] implemented the Myers algorithm under the MAPCG framework to match Snort signatures against packet traces. The Myers algorithm is a fuzzy search algorithm that matches strings by calculating the edit distance between them. The edit distance is the number of operations needed to convert a string s1 into string s2 or vice versa. A match is found when the edit distance between s1 and s2 is less than or equal to a predefined threshold. The paper demonstrates how the implementation of the Myers algorithm increases the speed of MAPCG, along with its effect on other frameworks.

Chang et al. [23] propose a bidirectional and parallel processing string-matching algorithm to improve the performance of Aho-Corasick algorithm. The bidirectional algorithm works in both positive and reverse directions. In their implementation, a string is matched by both directions concurrently, increasing the speed of matching. According to the authors, the bidirectional string-matching appears to be two times better than the AC string-matching algorithm when the input string is large.

In their work, Pal et al. [24] divide a detection scheme into three phases; preparing a database of virus signatures for training and computing their hashes, using the Mid-Square method for computing the hashes of received data, reducing false positives and false negatives by applying a deterministic finite automation for the filtered signatures. For detection, hashes of the received data are compared to virus signature hashes. If no match was found, data can pass as clean and benign. Otherwise, data is delivered to the finite state modules which perform further investigations to check for intrusions.

3.4 Anomaly-based Detection

Yassin et al. [25] proposed a host-based packet header anomaly detection model that statistically analyses the behaviour of packet headers to identify suspicious activity. The performance of their proposed model was tested on ISCX 2012 Intrusion Detection Evaluation and DARPA 1999 benchmark datasets and achieved successful detection of above 90% suspicious packets.

In their paper, Ahmad et al. [26] introduced a novel anomaly detection algorithm that meets the constraints of applying unsupervised anomaly detection. The proposed system is based on Hierarchical Temporal Memory (HTM) framework. One of HTM's constraints mentioned was that it lacks the ability to model anomalies scores. The authors addressed the issue and as a result, an anomaly is detected when the likelihood measure passes a certain threshold.

Hosseinzadeh et al. [27] discusses and compares a variety of Support Vector Machine (SVM) types that were used in anomaly detection schemes (ADS). SVM is a supervised learning algorithm that is often applied on classification problems. According to their study, an SVM-based ADS is commonly evaluated using the following metrics: Detection rate, accuracy, training time, testing time, false positive rate, classification rate, false negative rate, false acceptance rate, recall, and correlation coefficient.

Decision Tree (DT) is a machine learning classification algorithm that consists of multi-layered nodes where each parent node performs a test on the input. Based on the test result, the algorithm then branches to the suitable child node and repeats the process until a leaf node is reached eventually. Each leaf node represents a specific classification such that an input variable is classified according to the leaf node it stops at [28].

The production of few decision trees is the concept applied by the random forest algorithm. After the running of the produced decision trees, autonomic results will be delivered. The predicted outcome by most decision trees is then selected by random forest. Dilli [29] implemented and applied different machine learning algorithms in the anomaly detection in Domain Name System (DNS) query data. The overall accuracy of nearly all algorithms is over 85%. He concluded that the best result was the random forest with 90.80% accuracy in the anomaly detection of Domain Generation Algorithm and Fast Flux botnets. Therefore, he proposed the selection of the random forest algorithm as a detection model. Additionally, Padmanabhan et al. [30] implemented the random forest algorithm to enhance anomaly detection performance of Network Intrusion Detection System (NIDS) in active routers. Using the algorithm, the ability to identify the active routers, attacks, and packet corruptions was accomplished.

XGboost is a machine learning library that is extensible, distributed, Gradient Boosting Decision Tree (GBDT). It is the leading machine learning library in problem ranking, regression, classification and supplies parallel tree boosting. Henriques et al. [31] proposed a

framework that combines K-means and XGboost in anomaly detection in enormous log files. The approach achieved the highlight and identification of anomalies in log data. Results showed that the method proposed is applicable in both fields, forensics and auditing, compliance. Wang and Lu [32] proposed an anomaly-based intrusion detection system (IDS) architecture for IoT devices. The implementation of two machine learning methods to collect the sequence of the systems' calls as the dataset was done through XGboost and LSTM in order to build a stacking-based model to distinguish anomaly from normal behavior. Tests done show the framework has greater classification performance and is valid.

Authors of [33] have tested multiple machine learning algorithms for classification on the MQTT-IOT-IDS2020 dataset, including SVM, Random Forests, Decision Trees, and Logistic Regression. According to their experiments, the Decision Tree algorithm has achieved the highest detection accuracy with an average of 96.15%.

3.5 Comparison of Existing Network Forensics Frameworks

Researchers have proposed many frameworks in the field of IoT network forensics. In [34], the authors developed the Particle Deep Framework which defines the process of network flow analysis to detect and trace attack behaviors in an IoT network. They use a Particle Swarm Optimization algorithm to base the developed deep neural network on. The proposed framework succeeded in achieving high accuracy rate, which is 98%, in detecting abnormal activities.

In [35], the paper proposes a system that uses Raspberry Pi to investigate and collect data from layer 5 (i.e., session layer) on smart home devices by using access points to connect IoT devices and the Raspberry Pi. The system can sniff, connect to access points, obtain PCAP files, perform analysis and more.

In [36], The authors have proposed a digital forensics framework that uses multiple low-cost blockchain networks as temporary storage before passing the data to Etherscan. They evaluated the approach on popular blockchains such as EOS, Stellar, and Ethereum by performing a cost analysis.

In paper [37], authors have proposed a forensic investigation framework for smart home environments. Three different case studies were presented to demonstrate the performance, usability, and accuracy of the framework. In case study 1, Arduino IDE simulator and ZigBee technology were used to capture files and filter them. In the second case study, they identify digital evidence in smart home devices by using local LAN router. Finally, the last case study was by obtaining and examining artifacts through a third-party service provider. They found that a range of artifacts of forensic relevance could be recovered. In another study [38], authors presented a methodology that can extract digital traces from IoT devices with wider range of device categories in both hardware and software. The extraction was possible from difference locations: from the network, directly from the memory of the IoT and from the smartphone application.

Based on the reviewed literature, Table 1 summarises and demonstrates five existing forensics frameworks used in smart home investigations; all the frameworks have the same aim but with different methods applied. The first two frameworks, [34] and [35], both have the same source of data collection (PCAP files). [34] Focused on detecting malicious activities by analysing network traffic, while [35] concentrated on the forensics of unauthorized access to access points in the IoT environment.

Whereas in [36], incident data are stored mainly in EOS and Stellar, and only a daily summary of all transactions is written to Ethereum. In [37], the data is collected from volatile memory, logs, or sniffed traffic, and analysed using a variety of tools and techniques. Finally, the authors in [38] were using device-specific approaches to analyse data that was collected from network, smart phone applications, and devices' memory.

Table 1: Requirements

			Requirements							
				Security				Quality Attributes		
		References	Performance	Confidentiality	Integrity	Availability	Non-repudiation	Reliability	Usability	Accuracy
Type	Network forensics	[34]	✓	x	✓	x	x	x	x	✓
		[35]	✓	✓	✓	x	x	x	x	x
	Digital forensics	[36]	✓	x	✓	✓	x	✓	x	x
	Network and digital forensics	[37]	✓	x	✓	✓	x	✓	✓	✓
		[38]	✓	✓	x	x	✓	x	x	x

Due to the absence of an MQTT-specific network forensics tool, this paper proposes a tool to automate the process of extracting digital evidence from MQTT traffic and detecting intrusions.

4. Proposed Solution

With every new technology a new door opens for security issues and cyberattacks. Attackers always find new ways to violate policies and mechanisms to achieve

their goals. When the MQTT was built, security was not an aspect of the design. Security aspects apply only when MQTT is combined with another protocol. The investigation of MQTT protocol will contribute to building the knowledge needed for manufacturers, protocol developers and law enforcement. It will provide them with the ability to identify malicious activities, know and study the environment in order to employ security measures in upcoming products in case an incident occurs. MQTTracker is proposed to analyse and investigate MQTT network traffic passing through IoT devices, providing the user with information about connected devices and their communication.

4.1 Design

MQTTracker is designed to receive a network capture file and extract all useful information from the MQTT traffic. It uses deep packet inspection to process and examine the traffic to find digital evidence. After processing the uploaded file, extracted information will be displayed to the users enabling them to view the analysis of the PCAP file that was done by MQTTracker. The displayed information contains all devices' details, including their IP addresses, MAC addresses, the topics that they subscribed to, and the exchanged messages between them. The tool also displays the topic hierarchy, and the attacks that were detected in the traffic with the help of a trained machine learning model. Additional functionalities incorporate searching and filtering data and eventually generating a report with either all acquired evidence or the tagged information only. Figure 5 visualizes the flow of data in the proposed solution.

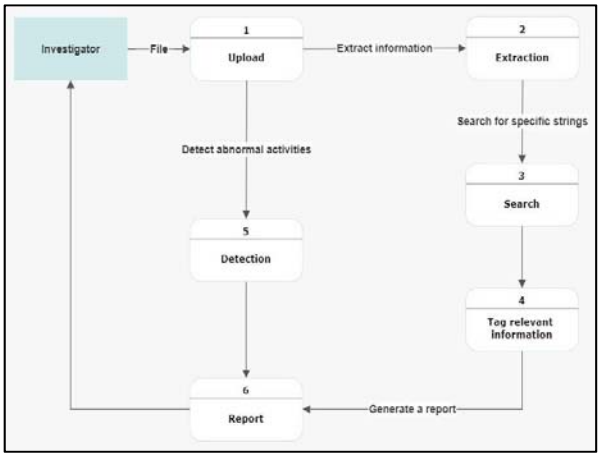

Fig. 5 Data flow diagram (Level 0).

4.2 Classification Model

For intrusion detection, a classification model must be trained to detect several types of attacks in the network traffic. This section presents the used dataset, the feature selection process, and the testing of multiple machine learning algorithms.

4.2.1 Data Description

The dataset used in the anomaly detection part of this paper is MQTT-IOT-IDS2020. The dataset chosen simulates a pragmatic MQTT IoT network traffic captured using tcpdump. Network packets were recorded and collected through ethernet traffic and are afterwards exported to PCAP files. The dataset is open source and is published in [33]. It consists of the recording of five different scenarios within the IoT environment, each of which is recorded separately. As listed below, the scenarios incorporate normal day to day operations and four different types of attacks:

- Normal operations
- MQTT brute-force attack
- Sparta SSH brute-force
- Aggressive scan
- UDP scan

The dataset chosen and used in the project has four-way significance. First, it is one of the first datasets that provide MQTT traffic attacks and scenarios. Second, it conveys a realistic MQTT IoT environment. Third, the dataset was made so researchers can use and benefit from it. Finally, IoT intrusion detection systems can be developed and tested on this dataset [33].

4.2.2 Feature Selection

Alongside the PCAP files, the dataset provided processed features. The features constitute: packet, unidirectional and bidirectional flow features. The features list is taken from [33]. In Table [33, 1], columns 1,2,3 describe the features of the dataset, while the fourth column expresses the extracted features form the basic packet. As for columns 5,6 they reflect the feature list for unidirectional and bidirectional features. Noting that for bidirectional flow data flows forward and backwards. Therefore, some of its features are denoted as (*) pointing out it has two values.

4.2.3 Algorithms Testing

Four classification algorithms are tested against the bidirectional flow features from the MQTT-IOT-IDS2020 dataset in this section: XGBoost (XGB), Decision Tree (DT), K-Nearest Neighbor (KNN), and Random Forest (RF). According to the authors of [33], source and

destination IP addresses, protocol, and MQTT flags features had to be dropped to avoid their influence. The data was split as follows: 75% for training and 25% for testing. Table 2 shows the accuracy of each tested algorithm. As seen in the table, XGB and RF achieve the highest score of accuracy. However, XGB was slower than RF during testing, leading RF to be the best option of the four algorithms to perform the classification in the proposed solution.

Table 2: Accuracy score of the tested algorithms.

XGB	DT	RF	KNN
99.9383%	99.9275%	99.9383%	99.8241%

5. Conclusion and Future Work

The emergence and integration of IoT into our daily lives raised security challenges and thus the idea of MQTTracker came to life. Due to the lack of forensics tools in the IoT field, MQTTracker will facilitate the investigation of MQTT traffic. This paper covers a background on IoT, its infrastructure, and two of the most used IoT application layer protocols. A review of literature was conducted on smart homes and their security challenges, detection types, and a comparison between existing network forensics frameworks. In the proposed solution section, the tool design is demonstrated along with the machine learning model and results of the tested algorithms.

Future enhancements to MQTTracker can be applied to expand its scope to support more IoT protocols in different environments other than smart homes, along with a real-time packet sniffer linked to an IDS. A voting classifier can be added to the classification part to let multiple classification algorithms decide on the right class (i.e., the attack type). Due to the lack of IoT, specifically MQTT, datasets, the detection was limited to four types of attacks. In the future, detection should include a wider range of attacks for the proposed tool to be useful.

References

[1] C. Stolojescu-Crisan, C. Crisan and B. Butunoi, "An IoT-Based Smart Home Automation System," Sensors (Basel, Switzerland), vol. 21, (11), pp. 3784, 2021.

[2] S. Sathwara, N. Dutta and E. Pricop, "IoT forensic A digital investigation framework for IoT systems," in 2018. DOI: 10.1109/ECAI.2018.8679017.

[3] S. Kim et al, "Smart Home Forensics—Data Analysis of IoT Devices," Electronics, vol. 9, (8), pp. 1215, 2020. Available: https://library.iau.edu.sa/scholarly-journals/smart-home-forensicsdata-analysis-iot-devices/docview/2429598317/se-2?accountid=136546. DOI: http://dx.doi.org.library.iau.edu.sa/10.3390/electronics9081215.

[4] A simple guide to IOT architecture. Total Phase Blog. (2019, October 22). Retrieved October 20, 2021, from https://www.totalphase.com/blog/2019/10/simple-guide-iot-architecture/.

[5] 3-layer IOT architecture. GeeksforGeeks. (2021, April 13). Retrieved October 20, 2021, from https://www.geeksforgeeks.org/3-layer-iot-architecture/.

[6] Reynolds, I. (2021, May 31). IOT architecture: 3 layers, 4 stages explained. Custom Software Development Insights | Zibtek Blog. Retrieved October 20, 2021, from https://www.zibtek.com/blog/iot-architecture/.

[7] M. Lombardi, F. Pascale and D. Santaniello, "Internet of Things: A General Overview between Architectures, Protocols and Applications," Information, vol. 12, (2), pp. 87, 2021. Available:https://library.iau.edu.sa/scholarly-journals/internet-things-general-overview-between/docview/2535221489/se-2. DOI: http://dx.doi.org/10.3390/info12020087.

[8] F. Alsuhaym, T. Al-Hadhrami, F. Saeed and K. Awuson-David, "Toward Home Automation: An IoT Based Home Automation System Control and Security," *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, 2021, pp. 1-11, doi: 10.1109/ICOTEN52080.2021.9493464.

[9] T. Song, R. Li, B. Mei, J. Yu, X. Xing and X. Cheng, "A Privacy Preserving Communication Protocol for IoT Applications in Smart Homes," in IEEE Internet of Things Journal, vol. 4, no. 6, pp. 1844-1852, Dec. 2017, doi: 10.1109/JIOT.2017.2707489.

[10] A. Larmo, A. Ratilainen and J. Saarinen, "Impact of CoAP and MQTT on NB-IoT System Performance," Sensors, vol. 19, (1), 2019. Available: https://library.iau.edu.sa/scholarly-journals/impact-coap-mqtt-on-nb-iot-system-performance/docview/2301580589/se-2. DOI: http://dx.doi.org/10.3390/s19010007.

[11] Atmoko, Rachmad & Riantini, Rona & Hasin, M. (2017). IoT real time data acquisition using MQTT protocol. Journal of Physics: Conference Series. 853. 012003. 10.1088/1742-6596/853/1/012003.

[12] A. Cornel - Cristian et al, "Smart home automation with MQTT," in 2019, . DOI: 10.1109/UPEC.2019.8893617.

[13] A. van den Bossche et al, "Specifying an MQTT tree for a connected smart home," in Anonymous Cham: Springer International Publishing, 2018, pp. 236-246.

[14] Bhanujyothi, H. C., Vidya, J., TJ, S. J., & Sahana, D. S. Diverse Malicious Attacks and security Analysis on MQTT protocol in IoT. Available at: http://www.paideumajournal.com/gallery/6-april2020.pdf [Accessed 20 October 2021].

[15] S. Andy, B. Rahardjo and B. Hanindhito, "Attack scenarios and security analysis of MQTT communication protocol in IoT system," in 2017, . DOI: 10.1109/EECSI.2017.8239179.

[16] 2021. [online] Available at: <http://biometricnews.blog/how-your-own-home-can-be-prone-to-a-massive-cyber-attack-must-read/> [Accessed 21 October 2021].

[17] G. Kalnoor and S. Gowrishankar, "IoT-based smart environment using intelligent intrusion detection system," Soft Computing (Berlin, Germany), vol. 25, (17), pp. 11573, 2021.

[18] E. Anthi, L. Williams, M. Słowińska, G. Theodorakopoulos and P. Burnap, "A Supervised Intrusion Detection System for Smart Home IoT Devices," in IEEE Internet of Things Journal, vol. 6, no. 5, pp. 9042-9053, Oct. 2019, doi: 10.1109/JIOT.2019.2926365.

[19] O. Georgiana Dorobantu and S. Halunga, "Security threats in IoT," 2020 International Symposium on Electronics and Telecommunications (ISETC), 2020, pp. 1-4, doi: 10.1109/ISETC50328.2020.9301127.

[20] S. Perumal, N. M. Norwawi, and V. Raman, “Internet of Things(IoT) digital forensic investigation model: Top-down forensic approach methodology,” in 2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC), 2015, pp. 19–23.

[21] M. Conti, A. Dehghantanha, K. Franke, and S. Watson, “Internet of Things security and forensics: Challenges and opportunities,” Future Gener. Comput. Syst., vol. 78, pp. 544–546, 2018.

[22] M. Aldwairi, A. M. Abu-Dalo and M. Jarrah, "Pattern matching of signature-based IDS using Myers algorithm under MapReduce framework," EURASIP Journal on Information Security, vol. 2017, (1), pp. 1-11, 2017.

[23] G. Chang, "An Efficient String-Matching Algorithm Using Bidirectional and Parallel Processing Structure for Intrusion Detection System," KSII Transactions on Internet and Information Systems, vol. 4, (5), pp. 956-967, 2010.

[24] U. Dixit, S. Gupta and O. Pal, "Speedy Signature Based Intrusion Detection System Using Finite State Machine and Hashing Techniques," International Journal of Computer Science Issues, vol. 9, (5), pp. 387, 2012.

[25] W. Yassin et al, "Packet header anomaly detection using statistical analysis," in Anonymous Cham: Springer International Publishing, pp. 473-482.

[26] S. Ahmad et al, "Unsupervised real-time anomaly detection for streaming data," Neurocomputing (Amsterdam), vol. 262, pp. 134-147, 2017.

[27] M. Hosseinzadeh et al, "Improving security using SVM-based anomaly detection: issues and challenges," Soft Computing (Berlin, Germany), vol. 25, (4), pp. 3195-3223, 2020; 2021.

[28] S. Uddin et al, "Comparing different supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, vol. 19, (1), pp. 281-281, 2019.

[29] Dilli, R., 2022. *Anomaly detection based on machine learning techniques*. [online] Krex.k-state.edu. Available at: <https://krex.k-state.edu/dspace/handle/2097/40286>.

[30] Prashanth, G. & Prashanth, V. & Padmanabhan, Jayashree & Srinivasan, N.. (2008). Using Random Forests for Network-based Anomaly detection at Active routers. 93 - 96. 10.1109/ICSCN.2008.4447167.

[31] Henriques, João & Caldeira, Filipe & Cruz, Tiago & Simoes, Paulo. (2020). Combining K-Means and XGBoost Models for Anomaly Detection Using Log Datasets. Electronics. 9. 10.3390/electronics9071164.

[32] Xiali Wang, Xiang Lu, "A Host-Based Anomaly Detection Framework Using XGBoost and LSTM for IoT Devices", Wireless Communications and Mobile Computing, vol. 2020, Article ID 8838571, 13 pages, 2020. https://doi.org/10.1155/2020/8838571

[33] H. Hindy et al, "Machine Learning Based IoT Intrusion Detection System: An MQTT Case Study (MQTT-IoT-IDS2020 Dataset)," 2020.

[34] N. Koroniotis, N. Moustafa and E. Sitnikova, "A new network forensic framework based on deep learning for Internet of Things networks: A particle deep framework," Future Generation Computer Systems, vol. 110, pp. 91-106, 2020.

[35] N. Widiyasono et al, "IoT forensic: Optimizing Raspberry Pi for investigation on the smart home network," IOP Conference Series. Materials Science and Engineering, vol. 550, (1), pp. 12019, 2019.

[36] S. Mercan, M. Cebe, E. Tekiner, K. Akkaya, M. Chang, and S. Uluagac, “A cost-efficient IoT forensics framework with blockchain,” in *2020 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 2020, pp. 1–5.

[37] A. Goudbeek, K. R. Choo and N. Le-Khac, "A forensic investigation framework for smart home environment," in 2018, . DOI: 10.1109/TrustCom/BigDataSE.2018.00201.

[38] F. Servida and E. Casey, "IoT forensic challenges and opportunities for digital traces," Digital Investigation, vol. 28, pp. S22-S29, 2019.