Monitoring and Detection Time Optimization of Man in the Middle Attacks using Machine Learning

Otily Toutsop
Department of Electrical Engineering
Morgan State University
Baltimore, Maryland, USA
Email: ottou1@morgan.edu

Jean Tshibangu Muabila
Department of Computer Science
Vietnam Nationl University
Hanoi, Building C3, 144 Xuan Thuy 100000
Email: jean.tshibangu-muabila@etu.univ-lyon1.fr

Abstract*—The Internet of Things (IoT) is growing with the advancement of technology. Many vendors are creating IoT devices to leverage the quality of life of consumers. These devices include smart grid, smart homes, smart health care systems, smart transportation, and many other device applications. IoT devices interact with the environment and each other using sensors and actuators. However, the widespread proliferation of IoT devices poses many cybersecurity threats. The IoT devices' interconnection opens the door to attackers who try to gain unauthorized access to the devices. For many IT networks, establishing trust and security during the devices' regular operation is challenging. Further, devices may leak vital information, which is a huge concern in cybersecurity. Prior research has shown that security breaches have increased by 67% in the past five years, and 95% of HTTPs servers are vulnerable to Man in the Middle (MIM) attacks. This paper explores the new attacks dataset from the HCRL (Hacking and Countermeasure Research Lab) collected from real-life internet of things devices that incorporated smart cameras, laptops, and smartphones [1]. We then apply three machine learning algorithms that include the Random Forest, Logistic Regression, and Decision Tree to evaluate our model's performance. Our results show that the overall detection accuracy is 98-100%, which more promising than traditional Intrusion Detection System (IDS).

***Keywords*—** Internet of Things (IoT), cybersecurity, Intrusion Detection System (IDS), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR),Man In The Middle (MIM) attacks.

I. Introduction

The Internet of Things can be viewed as the network of physical devices connected using some set of protocols and exchange information based on the environment. The concept of the Internet of Things has been around for decades. Many industrial sectors are adopting the IoT devices within their organization to satisfy the vast demand of the customers [2], [3]. IoT has also attracted many home users, and its benefit has tremendously affected users' quality of life [4]. With the rapid growth of technology, people tend to get more IoT devices connected to their home networks to facilitate their daily activities [5]. For instance, a smart garage door opener will automatically open the door to its intended users without requiring any physical contact with the door. IoT devices talk to each other using several communication protocols such as Wi-Fi, Bluetooth, Z-Wave, and ZigBee [6], [7]. As shown in Figure 1, the IoT Smart Home market growth will reach up to 397 USD billion in the next future.

The giant IoT ecosystem is the interconnection of multiple devices, ranging from the smart health care system, smart grid, smart home, smart agriculture system, and smart cities [8]. For example, the smart grid system might have more than five million nodes interconnected that exchanged confidential information like the client power consumption, which could be an attack surface for any third party. People do not need to access their devices manually; everything is done via smartphones in just one click [9]. Users' can connect to their home network from anywhere and access the devices. For example, in smart healthcare, a doctor in the hospital can diagnose and prescribe medication to patients remotely without their physical presence [10], [11]. It is almost impossible to live without smart devices, which leads to the critical need for IoT in our life. Those devices must be tied to a network with Internet connectivity to operate correctly. However, with the diverse types of communication networks, an unauthorized user can easily access the user's network and perform malicious activities [12].

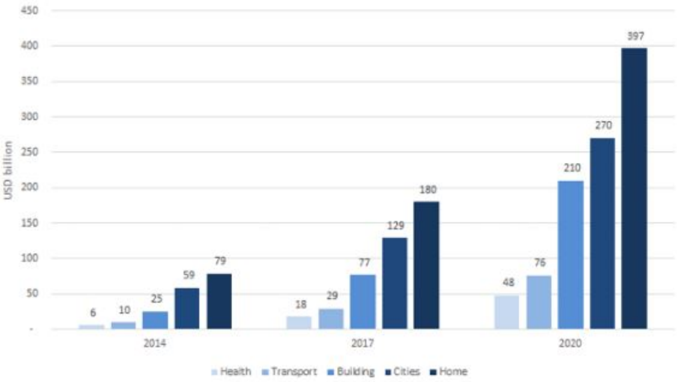


Fig. 1. Global IoT Market Growth in Different Sectors [13]

The intruders are always seeking techniques to penetrate the network and comprised the system. The consequences are tremendous and can put any host device connected to the network at risk and private information such as passwords, bank credentials, social security number, and patient records. Moreover, giving the fact that most of the network today has many devices that can be controlled remotely, monitoring the network traffic from anywhere is a huge concern in the cybersecurity area [14]. Usually, the consumer will rely on the service provider's network to assure the security of their devices [15]. Unfortunately, most of the gateways available in the market, like routers, do not have an intrusion detection system built-in to protect the potential users' against any cyber threat [16]. The network system in place does not have an efficient mechanism to differentiate between Man in the Middle attack type with other cyber threats [17]. Therefore, industry, government, and researchers should work closely together to address the security vulnerabilities related to internet usage of things devices and networks. Although the Intrusion Detection Systems (IDS) are widely used to respond to network attacks, it is still not sophisticated enough to handle attacks from the vast network. Machine learning is a subset of artificial intelligence that utilizes algorithms to recognize patterns from data [18]. The algorithms provide the ability to extract meaningful features from the data for further analysis. The result of the analysis will then be used for future prediction [19]. With a massive amount of data generated by the internet of things devices, it is significant to use machine learning to learn the network traffic behavior, which could be relevant to detect an attack in the system [20]. Vendors could then utilize the result of the prediction and analysis to build more secure systems. On the other side, researchers could exploit that to further their work. The main contributions of this paper are as follows:

1) We explore the new attacks datasets from the HCRL (Hacking and Countermeasure Research Lab) collected from real-life internet devices that incorporated smart cameras, laptops, and smartphones.
2) We perform an in deep study of the detection of MIM attacks in the network that has different Internet of Things devices connected.
3) We build our model using three machine learning algorithms that include Random Forest, Decision Tree, and Logistic Regression to optimize the network's detection time and have an overall accuracy that ranges between 98-100%

The rest of this paper is organized as follows. An overview of the motivation is presented in section II, a literature review of the related works and gaps is presented in section III, while section III presents the methodology utilized to optimize the detection time of the attacks. Section VI shows the results and analysis. Lastly, Session VII elaborates on the future work and conclusion.

II. MOTIVATION

In his article titled ”80 Eye-Opening Cyber Security Statistics for 2019”, Hashedout published a disturbing report summarizing cybersecurity statistics for 2019, which shows that security breaches have increased by 67% past five years .Among their 80 statistics list, there are MIM attacks, which, according to IBM's X-Force Threat Intelligence Index 2018, more than a third of the inadvertent exploitation of weaknesses involved MIM attacks. Also, 95% of HTTPS servers are vulnerable to MIM, according to NetCraft. MIM attacks consist of intruders intercepting traffic between two parties, the source, and destination, posing as a legitimate source at the destination and masquerading as the source's legitimate destination [21].

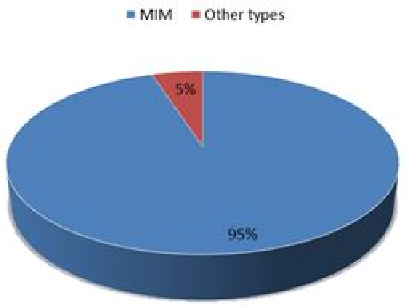


Fig. 2. MIM Attacks Statistics

The MIM attack is made without both the attacker source and the victim destination suspecting the communication channel's corruption. The Internet of Things devices, with their low computational power, low memory, can barely handle high-level encryption algorithms, thus the exposure to critical cyber-attacks due to the presence of backdoors [22]. Therefore, protecting the network system with adequate security prevention mechanisms such as IDS (Intrusion Detection System), firewalls do not guarantee the prevention or mitigation of cyber threats [23].

III. RELATED WORK

The Internet of Things has significantly leveraged the consumer's quality life. Besides, with the vast amount of benefits that range from interconnecting the devices in any given network, being able to monitor the devices without any physical presence, and the capability of the user's to integrate end devices in their network setup independently of the manufacturers [24]. However, the majority of those IoT devices' present some limitations in terms of security features which often open backdoors to hackers who can infiltrate into the network using the devices. Man in the Middle attack is known as one the most common attack used by hackers to harm the system. Many researchers have proposed some methods to detect the MIM attack in the network; however, their solution is not always suitable for constraint devices with very low computational power [25], [26], [27], [28].

Dong et al. [20] developed a technique to detect and localize the MIM in the wireless sensor network based system which is does not shown promising results when applying to the attacks coming from Internet of Things devices. Liu et al. [29], conducted their experiment using Internet of Things devices data and applied some machine learning algorithms such as KNN, Logistic Regression (LR), and Random Forest to detect anomaly in the IoT Network; however, in their work the only focused on broad range of anomaly detection in the network without deep dive into any particular type of attacks. The authors implemented the SVM (Support Vector Machine) algorithm which gave them the overall accuracy of 70%, which is not promising. In addition, the Logistic Regression method used in their work gave an accuracy of 86%. This paper focused on MIM attacks particularly and presents more details about the data acquisition and pre-processing phase which is more sophisticated for the detection of this particular type of attack in the user's network.

The previous approach is different from the one discuss in this paper which main focus is to optimize the detecting time of MIM attack. Bagaa et al. [30] have designed a machine learning security framework to leverage cyber threat in Software Define Networking system which again barely detect attacks from Internet of Things devices in the network. Hussain et al. [27]have investigated on different possibilities to apply machine learning technique to solve IoT security network which target technologies such as Software Defined Network, Cloud Computing, and Fog computing.

Santhosh et al. [31] has carried out some research related to the usage of machine learning technique to solve the security issue faced by Internet of Things devices. Their experimental results do not incorporated appropriate features selection mechanism which could not be applied to a large scale network of Internet of Things devices. Gupta et al.[26] have created a firewall based system to improve the security of internet of things devices. The firewall is built using a Raspberry as a gateway to monitor the network traffic and secure the communications patterns through the cloud database. The proposed firewall based system presents some security limitations because any third party can eavesdrop into the communication, then compromise the entire cloud database. Another limitation of this work is related to the scalability due to the fact a Raspberry Pi is not powerful enough to handle complex encryption algorithm.

IV. METHODOLOGY

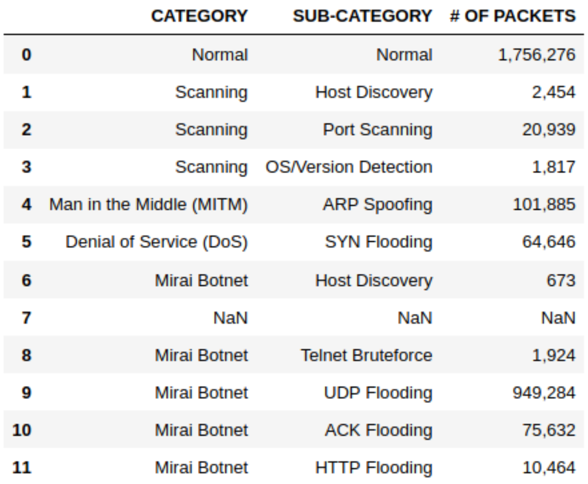
The paradigm of the Intrusion Detection System (IDS) is not explicitly and formally defined in the literature. Besides, traditional tools such as firewalls and anti-virus are not designed to detect MIM coming from IoT devices. Consumers usually rely on routers to protect their home network against cyber threats and do not necessarily consider IoT devices a second entry point for hackers [32]. In this work, we develop a model that focused on anomaly detection, particularly the Man-in-the-Middle attack, given that in the last five years, statistics show that 95% of HTTP servers are vulnerable to these attacks, hence our motivation. A benefit with this perspective is that it manages to optimize the Man-In-The-Middle attack's monitoring and detection time.To experiment, we use three Machine Learning algorithms in the data set for accurate real-time anomaly detection. The approaches include Random Forest (RF), Logistic Regression (LR), and Decision Tree. We then apply binary classification by classifying Man-In-The-Middle attacks as one (1) and zero (0) as the normal packet.

A. Data Acquisition Phase

In this work, we are developing and experimenting with our models on the intrusion dataset on the loT network, created by a group of researchers from the Hacking and Countermeasure Research Lab [1]. This dataset is publicly available and developed explicitly for academic use only. The dataset employs two typical smart home devices - SKT NUGU (NU 100) and the EZVIZ Wi-Fi camera (C2C Mini O Plus 1080P). Both devices were connected via Wi-Fi with other smartphones and laptops to create the data. All devices, including some laptops and smartphones, were connected to the same wireless network. Packet files were captured using the monitor mode of the wireless network card. To maintain the privacy of the users, Aircrack-ng was utilized to remove wireless headers. ll attacks except for the Mirai Botnet category are packets captured when simulating attacks using tools such as Nmap. In the case of the Mirai Botnet category, the attack packets were generated on a laptop and then manipulated to appear as if they came from the IoT device. The dataset consists of 42 raw network packet (pcap) files at different times; including six pcap files containing Man in the Middle attack [1].

Table I provides an overview of the dataset set. Each pcap file contains 7 features. The first feature is the sequence number of the packet; the second feature is the time, which represents the transmission time of the packet; the third feature is the source IP address of the packet; the fourth feature is the destination IP address of the packet; the fifth feature is the protocol in which the packet is transmitted; the sixth feature is the length of the packet, measured in bytes; the seventh feature is Info, which contains additional details corresponding to the packet. The dataset used containing the Man in the Middle attack has 194184 observations.

TABLE I
Table of the dataset



	CATEGORY	SUB-CATEGORY	# OF PACKETS
0	Normal	Normal	1,756,276
1	Scanning	Host Discovery	2,454
2	Scanning	Port Scanning	20,939
3	Scanning	OS/Version Detection	1,817
4	Man in the Middle (MITM)	ARP Spoofing	101,885
5	Denial of Service (DoS)	SYN Flooding	64,646
6	Mirai Botnet	Host Discovery	673
7	NaN	NaN	NaN
8	Mirai Botnet	Telnet Bruteforce	1,924
9	Mirai Botnet	UDP Flooding	949,284
10	Mirai Botnet	ACK Flooding	75,632
11	Mirai Botnet	HTTP Flooding	10,464

The dataset contains 42 raw network packet files recorded at various time intervals. Each pcap file for the Man in the Middle attack contains up to 65768 packets. In this dataset, a description file is provided specifying the filtering rules for separating attack packets from benign packets. In addition, the dataset includes 5 main categories and 11 subcategories. The 5 categories include one category of normal transmissions and 4 groups of cyber-attacks. The 4 groups of attacks cover the most common attacks for loT networks, i.e. scan attacks, Man in the Middle (MIM) attacks, Denial of Service (DoS) attacks and Mirai Botnet attacks. In addition, 11 subcategories are derived from the 5 main categories, including 1 subcategory for non-normal transmission and 10 unique attack types. In this paper, we have exploited only Man-in-the-Middle attacks due its critical presence in the consumer's network. Figure 3 is showing the proporsion of MIM in the network of Internet of Things devices extracted from the attacks the original datasets.

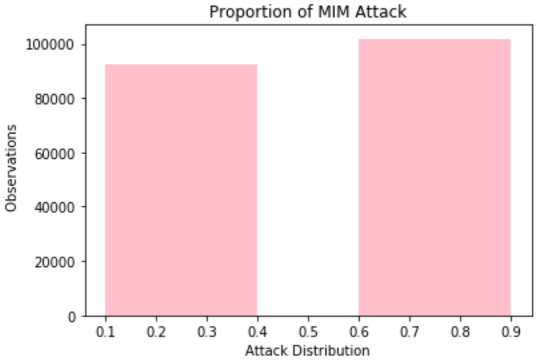


Fig. 3. Proposion of MIM Attacks Distribution

B. Approach for Data Preprocessing

The data acquisition stage mentioned previously is essential to understand the features use in the raw data packets. The second step is called Pre-processing which is very important because it helps us to analyze and transform the data for prediction. In the case of this experiment, we need to pre-process the raw data files. Since the filtering rules have been given, we proceed as follows to process the data: In the first phase, we run the shell scripts applying the filtering rules in order to get two csv files, abnormal csv file (Man in the Middle attack) and normal file. Then we move on to label some features, and add a column named Target, which qualifies the attack flow or not, and we do the same for the normal data. The Target column is labeled 1 for the presence of the attack and 0 for normal. Then we merge the two csv files and remove the redundant data. We perform this mechanism for all 6 pcap files containing the Man-in-the-Middle attack. In the end, we concatenate all the files into one file.

This work focuses on the detection of the Man in the Middle Attack and the figure below presents the screenshot of the attack packets.

C. Description of the Features

The table below describes some of the important features extracted during the pre-processing phase:

The final output after the pre-processing operation is showed in this table 3:

V. Implementation and Evaluation

The data set to build our model is 194184 observations, of which 92299 observations were labeled as normal data, and 101885 observations as attack data, which represents a participation of 47.53% normal packets and 52.47% attack packets. To build our models, we dissected the training and test data with 70% and 30%

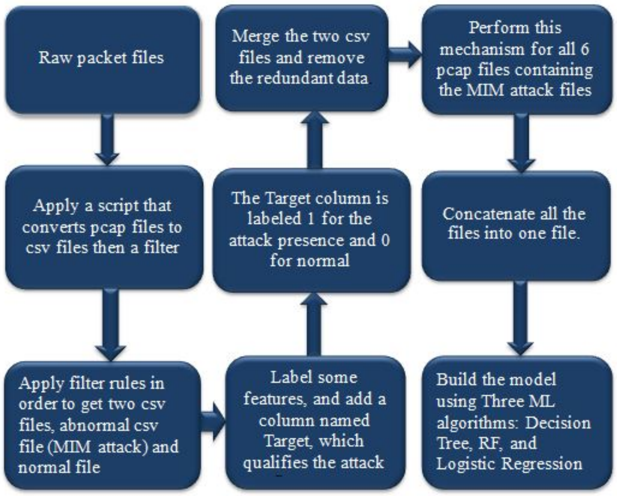
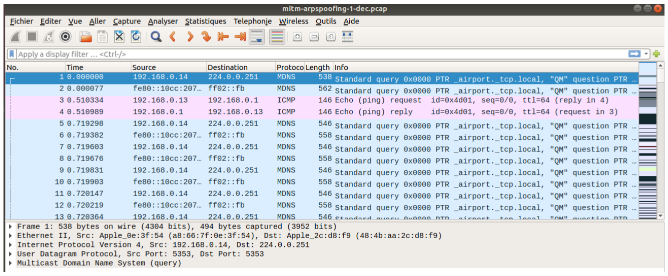

Raw packet files
Merge the two csv files and remove the redundant data
Perform this mechanism for all 6 pcap files containing the MIM attack files
Apply a script that converts pcap files to csv files then a filter
The Target column is labeled 1 for the attack presence and 0 for normal
Concatenate all the files into one file.
Apply filter rules in order to get two csv files, abnormal csv file (MIM attack) and normal file
Label some features, and add a column named Target, which qualifies the attack
Build the model using Three ML algorithms: Decision Tree, RF, and Logistic Regression

Fig. 4. Procedure for data preprocessing


mitm-arpspoofing-1-dec.pcap
Fichier Editer Vue Aller Capture Analyser Statistiques Telephonie Wireless Outils Aide
Apply a display filter ... <Ctrl-/>

```
No.   Time      Source             Destination   Protoco Length Info
    1 0.000000  192.168.0.14       224.0.0.251   MDNS    538 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
    2 0.000077  fe80::10cc:207…    ff02::fb      MDNS    562 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
    3 0.510334  192.168.0.13       192.168.0.1   ICMP    146 Echo (ping) request  id=0x4d01, seq=0/0, ttl=64 (reply in 4)
    4 0.510989  192.168.0.1        192.168.0.13  ICMP    146 Echo (ping) reply    id=0x4d01, seq=0/0, ttl=64 (request in 3)
    5 0.719298  192.168.0.14       224.0.0.251   MDNS    546 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
    6 0.719382  fe80::10cc:207…    ff02::fb      MDNS    558 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
    7 0.719603  192.168.0.14       224.0.0.251   MDNS    546 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
    8 0.719676  fe80::10cc:207…    ff02::fb      MDNS    558 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
    9 0.719831  192.168.0.14       224.0.0.251   MDNS    546 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
   10 0.719903  fe80::10cc:207…    ff02::fb      MDNS    558 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
   11 0.720147  192.168.0.14       224.0.0.251   MDNS    546 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
   12 0.720219  fe80::10cc:207…    ff02::fb      MDNS    558 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR …
   13 0.720364  192.168.0.14       224.0.0.251   MDNS    546 Standard query 0x0000 PTR _airport._tcp.local, "QM" question PTR
▸ Frame 1: 538 bytes on wire (4304 bits), 494 bytes captured (3952 bits)
▸ Ethernet II, Src: Apple_0e:3f:54 (a8:66:7f:0e:3f:54), Dst: Apple_2c:d8:f9 (48:4b:aa:2c:d8:f9)
▸ Internet Protocol Version 4, Src: 192.168.0.14, Dst: 224.0.0.251
▸ User Datagram Protocol, Src Port: 5353, Dst Port: 5353
▸ Multicast Domain Name System (query)
```

Fig. 5. mitm-arpspoofing

values, respectively. We evaluated our models' performance based on the confusion matrix, with which we used a classification report to measure the quality of the predictions from the classification algorithms. The classification report presents the main classification metrics, such as accuracy, recall, and f1 score per class. The metrics are calculated using true and false positives, true and false negatives. Hence, there are four ways to check whether our predictions are right or wrong:

- True Negative(TN): is a case in which the actual label was negative and predicted negative
- True Positive(TP): a case in which the actual label was positive and predicted positive
- False Negative (FN): Represents a case in which the actual label was positive but predicted to be negative.
- False Positive (FP): Represents a case in which the actual label was negative but predicted positive

The precision represents how accurate the predictions are in the model. It is defined as the ratio of true positives to the sum of true and false positives for each class.

$$Precision = TP/(TP + FP) \tag{1}$$

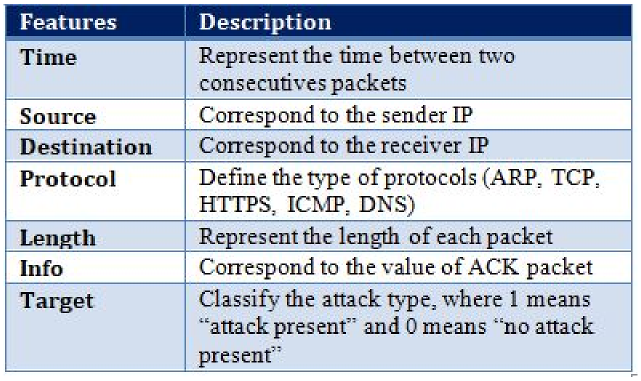
The recall represents the percentage of positive cases detected in the model. It is defined as the ratio of true positives to the sum of true positives and false negatives.

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

The accuracy is the number of correct predictions, which includes both positive and negative predictions, divided by the total number of predictions made.

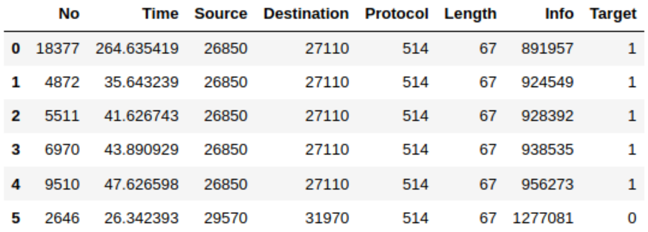
$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{3}$$

TABLE II
DESCRIPTION OF THE FEATURES



Features	Description
Time	Represent the time between two consecutives packets
Source	Correspond to the sender IP
Destination	Correspond to the receiver IP
Protocol	Define the type of protocols (ARP, TCP, HTTPS, ICMP, DNS)
Length	Represent the length of each packet
Info	Correspond to the value of ACK packet
Target	Classify the attack type, where 1 means "attack present" and 0 means "no attack present"

TABLE III
DATASET AFTER PREPROCESSING



	No	Time	Source	Destination	Protocol	Length	Info	Target
0	18377	264.635419	26850	27110	514	67	891957	1
1	4872	35.643239	26850	27110	514	67	924549	1
2	5511	41.626743	26850	27110	514	67	928392	1
3	6970	43.890929	26850	27110	514	67	938535	1
4	9510	47.626598	26850	27110	514	67	956273	1
5	2646	26.342393	29570	31970	514	67	1277081	0

Score F1 represents the percentage of correct positive predictions. Score F1 is a weighted harmonic average of precision and recall such that the best score is 1.0, and the worst is 0.0.

$$ScoreF1 = 2 * (Recall * Precision)/(Recall + Precision) \tag{4}$$

VI. RESULTS

A. Logistic Regression Approach

In this section, we present the different results of our experiments. The experiments were carried out with 10 trials for each approach. Tables V, VI, and **??** show the results. As a result, Table V illustrates the logistic regression approach, which has a precision of 99%, a recall rate of 99%, and an f1 score of 99%. The performance is virtually inferior to the other perspectives. Figure 6 illustrates the confusion matrix of the logistic regression. It shows that the number of errors in class 1 is seven times greater than the number of errors in class 0, resulting in an error percentage of 1.21% for class 1 and 0.17% for class 0. The logistic regression model gives an accuracy of 98.6%.

1) Analysis of the confusion matrix for the logistic regression: We would analysis the confusion matrix based on the metrics values that include the True Positive, True Negative, False Positive, and False Negative.For a better performance, we would like to have our True Positive and True Negative metrics high. For this Logistic Regression case, the average f1 score is 99%.

B. Decision Tree Approach

The Decision Tree is another machine learning uses to build model. The main idea behind the decision tree is to have multiple choices when it comes to making decision about what to do next.This method operates by choosing a criteria to classify attacks in the network. The number of observations is obviously the same for the Logistic Regression case. Subsequently, Table V illustrates the decision tree approach, which shows a 100% accuracy, a 100% recall rate, and

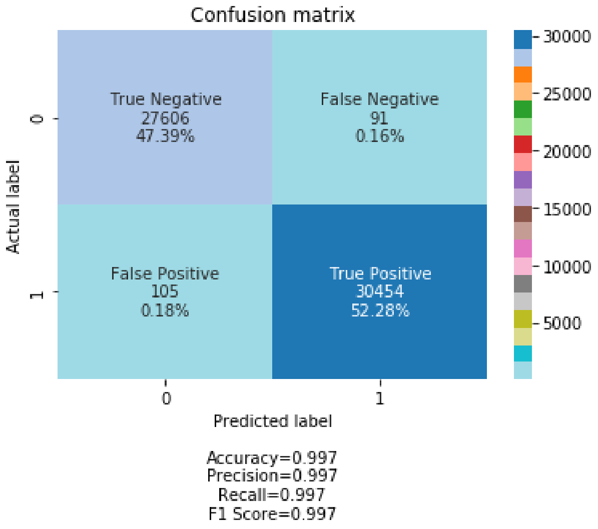


Fig. 6. Confusion Matrix Logistic Regression

TABLE IV
TABLE LOGISTIC REGRESSION

Classe	Precision	Recall	f1-score	Support
Classe 0	0.98	1.00	0.99	27697
Classe 1	1.00	0.98	0.99	30559
avg / total	0.99	0.99	0.99	58256

an F1 score of 100%, only with the default hyperparameters. The performance is almost superior to logistic regression and equal to Random Forest model. Figure 7 illustrates the confusion matrix of the decision tree. It shows that our model has no prediction errors; therefore, we have a perfect model. The RF model gives an accuracy of 100%. Similarly, the max_depth (maximum depth) has not been limited for the decision tree to let the forest in the RF classifier increase the level as much as possible when needed.

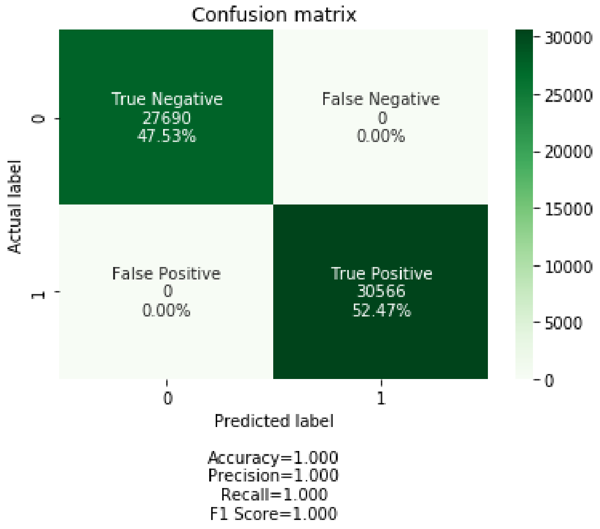


Fig. 7. Confusion Matrix Decision Tree

TABLE V
TABLE LOGISTIC REGRESSION

Classe	Precision	Recall	f1-score	Support
Classe 0	1.0	1.0	1.0	27690
Classe 1	1.0	1.0	1.0	30566
avg / total	1.0	1.0	1.0	58256

C. Random Approach

Next, Table VI illustrates the Random Forest approach, which has 100% accuracy, a 100% recall rate, and an F1 score of 100%, only with the default hyper-parameters. The performance is almost superior to logistic regression and equal to the decision tree. Figure 8 illustrates the confusion matrix of the logistic regression. The True Negative and True Positive values show that our model has no prediction any errors; therefore, we have a perfect model. The RF model gives an accuracy of 100%. The max_depth (maximum depth) has not been limited to let the forest in the RF classifier increase the level as much as possible when needed. The performance of the Random Forest model is evaluated using the confusion matrix hyper-parameters such as the precision, recall, f1 score, and the number of observations.

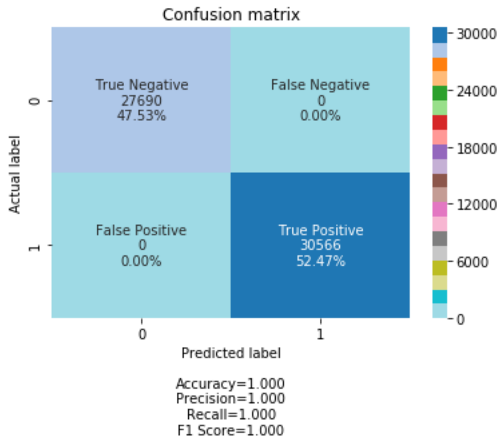


Fig. 8. Confusion Matrix Random Forest

TABLE VI
TABLE LOGISTIC REGRESSION

Classe	Precision	Recall	f1-score	Support
Classe 0	1.0	1.0	1.0	27690
Classe 1	1.0	1.0	1.0	30566
avg / total	1.0	1.0	1.0	58256

D. Comparison of our Results WITH Previous Work

In this paper, we study the MIM attacks behavior in deep to understand how we can easily detect its presence in any given system. The reason why we focused on the MIM attacks is because the article [21] has stated that MIM is one of the most critical attack today. Compare to the most recent work that has been done on the detection of anomaly in IoT network intrusion using machine Learning [29], our approach presents better performance with an enhancement in terms of the metrics parameters as well.

E. Proposed Internet of Things Testbed for MIM Attack Data Collection

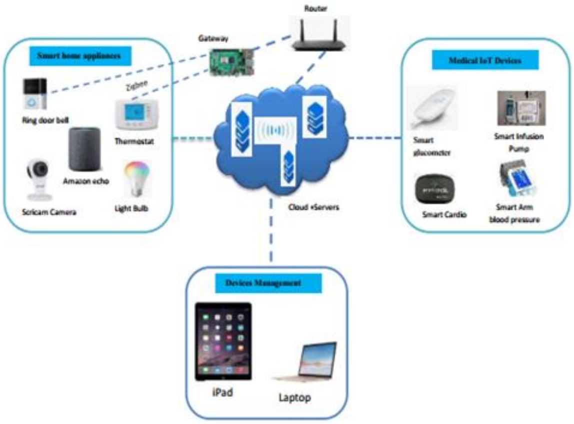


Fig. 9. IoT Testbed Architecture

VII. Conclusion and Future Work

IoT devices are gaining popularity as technology improves, which is why attackers use them to gain illegitimate access to the network. Traditional tools have some limitations in detecting an intrusion, hence the need for our machine learning approach. In this paper, we proposed a technique based on three machine learning algorithms to optimize the detection time of the Man in the Middle attack in the network. With these three algorithms, the logistic regression approach showed an accuracy of 98.6%, the random forest showed an accuracy of 100%, and the decision tree showed an accuracy of 100% with a very high F1 score. Our results showed that the technique implemented in this paper will improve the state of the art of traditional IDS. Future work will use a Deep Learning technique to build model that will learn from data to stop propagating this attack in the network.

References

[1] H. Kang, D. H. Ahn, G. M. Lee, J. D. Yoo, K. H. Park, and H. K. Kim, "Iot network intrusion dataset," *IEEE Dataport*, 2019, doi:10.21227/q70p-q449.

[2] R. Sivapriyan, K. M. Rao, and M. Harijyothi, "Literature review of iot based home automation system," in *2020 Fourth International Conference on Inventive Systems and Control (ICISC)*, 2020, pp. 101–105.

[3] A. Gai, S. Azam, B. Shanmugam, M. Jonkman, and F. De Boer, "Categorisation of security threats for smart home appliances," in *2018 International Conference on Computer Communication and Informatics (ICCCI)*, 2018, pp. 1–5.

[4] J. Qiu, Z. Tian, C. Du, Q. Zuo, S. Su, and B. Fang, "A survey on access control in the age of internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4682–4696, 2020, doi10.1109/JIOT.2020.2969326.

[5] K. Chopra, K. Gupta, and A. Lambora, "Future internet: The internet of things-a literature review," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 135–139, doi:10.1109/COMITCon.2019.8862269.

[6] J. C. Talwana and H. J. Hua, "Smart world of internet of things (iot) and its security concerns," in *2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2016, pp. 240–245, doi:10.1109/iThings-GreenCom-CPSCom-SmartData.2016.64.

[7] Y.-K. Chen, "Challenges and opportunities of internet of things," in *17th Asia and South Pacific design automation conference*. IEEE, 2012, pp. 383–388, doi:10.1145/2628071.2635931.

[8] J. Voas, "Building blocks of the internet of things," in *2016 IEEE Symposium on Service-Oriented System Engineering (SOSE)*, 2016, pp. 1–2, doi:10.1109/SOSE.2016.36.

[9] A. J. Brush, J. Albrecht, and R. Miller, "Smart homes," *IEEE Pervasive Computing*, vol. 19, no. 2, pp. 69–73, 2020, doi:10.1109/MPRV.2020.2977739.

[10] I. Chiuchisan, H. Costin, and O. Geman, "Adopting the internet of things technologies in health care systems," in *2014 International Conference and Exposition on Electrical and Power Engineering (EPE)*, 2014, pp. 532–535, doi:10.1109/ICEPE.2014.6969965.

[11] F. Ahamed and F. Farid, "Applying internet of things and machine-learning for personalized healthcare: Issues and challenges," in *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, 2018, pp. 19–21, doi:10.1109/iCMLDE.2018.00014.

[12] K. N. Tongay, "Sensor data computing as a service in internet of things," in *2016 Symposium on Colossal Data Analysis and Networking (CDAN)*, 2016, pp. 1–4, doi:10.1109/CDAN.2016.7570963.

[13] Online. Iot market growth. [Online]. Available: https://asia.ub-speeda.com/en/the-internet-of-things//-taking-the-technology-and//-communication-space-by-storm/

[14] N. Kumar, J. Madhuri, and M. ChanneGowda, "Review on security and privacy concerns in internet of things," in *2017 International Conference on IoT and Application (ICIOT)*, 2017, pp. 1–5, doi:10.1109/CSNET.2017.8242006.

[15] R. Kurte, Z. Salcic, and K. Wang, "A distributed service framework for the internet of things," *IEEE Transactions on Industrial Informatics*, vol. 16, pp. 4166–4176, 2020.

[16] A. Jurcut, T. Niculcea, P. Ranaweera, and N.-A. Le-Khac, "Security considerations for internet of things: A survey," *SN Comput. Sci.*, vol. 1, p. 193, 2020.

[17] M. Husamuddin and M. Qayyum, "Internet of things: A study on security and privacy threats," in *2017 2nd International Conference on Anti-Cyber Crimes (ICACC)*, 2017, pp. 93–97, doi:10.1109/Anti-Cybercrime.2017.7905270.

[18] S. Earley, "Analytics, machine learning, and the internet of things," *IT Professional*, vol. 17, no. 1, pp. 10–13, 2015, doi:10.1109/MITP.2015.3.

[19] K. Sharma and R. Nandal, "A literature study on machine learning fusion with iot," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 1440–1445, doi:10.1109/ICOEI.2019.8862656.

[20] Z. C. Dong, R. Espejo, Y. Wan, and W. Zhuang, "Detecting and locating man-in-the-middle attacks in fixed wireless networks," *Journal of computing and information technology*, vol. 23, no. 4, pp. 283–293, 2015, doi:10.2498/cit.1002530.

[21] HashedOut. 80 eye-opening cyber security statistics for 2019. [Online]. Available: https://www.thesslstore.com/blog/80-eye-opening-cyber-security-statistics-for-2019/

[22] M. Mamdouh, M. A. I. Elrukhsi, and A. Khattab, "Securing the internet of things and wireless sensor networks via machine learning: A survey," in *2018 International Conference on Computer and Applications (ICCA)*, 2018, pp. 215–218, doi:10.1109/COMAPP.2018.8460440.

[23] P. Laplante and S. Applebaum, "Nist's 18 internet of things trust concerns," *Computer*, vol. 52, no. 6, pp. 73–76, 2019, doi:10.1109/MC.2019.2908544.

[24] I. Psychoula, L. Chen, and O. Amft, "Privacy risk awareness in wearables and the internet of things," *IEEE Pervasive Computing*, vol. 19, no. 3, pp. 60–66, 2020, doi:10.1109/MPRV.2020.2997616.

[25] H. Hsu, G. Jong, J. Chen, and C. Jhe, "Improve iot security system of smart-home by using support vector machine," in *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, 2019, pp. 674–677, doi:10.1109/CCOMS.2019.8821678.

[26] N. Gupta, V. Naik, and S. Sengupta, "A firewall for internet of things," in *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*, 2017, pp. 411–412, doi:10.1109/COMSNETS.2017.7945418.

[27] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, "Machine learning in iot security: current solutions and future challenges," *IEEE Communications Surveys & Tutorials*, 2020, doi:10.1109/COMST.2020.2986444.

[28] P. Kaur, V. Bharti, and S. Maji, "Comparison of machine learning approach in smart wearables," in *2019 Women Institute of Technology*

Conference on Electrical and Computer Engineering (WITCON ECE), 2019, pp. 135–138, doi:10.1109/WITCONECE48374.2019.9092921.

[29] Z. Liu, N. Thapa, A. Shaver, K. Roy, X. Yuan, and S. Khorsandroo, "Anomaly detection on iot network intrusion using machine learning," in *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*. IEEE, 2020, pp. 1–5, doi:10.1109/icABCD49160.2020.9183842.

[30] M. Bagaa, T. Taleb, J. B. Bernabe, and A. Skarmeta, "A machine learning security framework for iot systems," *IEEE Access*, 2020, doi:10.1109/ACCESS.2020.2996214.

[31] N. N. Santhosh, "Future black board using internet of things with cognitive computing: Machine learning aspects," in *2016 International Conference on Communication and Electronics Systems (ICCES)*, 2016, pp. 1–4, doi:10.1109/CESYS.2016.788987 .

[32] M. El-hajj, M. Chamoun, A. Fadlallah, and A. Serhrouchni, "Analysis of authentication techniques in internet of things (iot)," in *2017 1st Cyber Security in Networking Conference (CSNet)*, 2017, pp. 1–3, doi:10.1109/CSNET.2017.8242006.