Potential Threats of Man In The Middle Attack and Its Plausible Solutions

Parshva Anish Maniar, Dr. Santhi H (hsanthi@vit.ac.in), Rachit Biswas, Ramanathan Annamalai, Sanchit Anand and Tejas Rahul Rokade, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore-632 014, Tamil Nadu, India

***Abstract­­***— **The Internet of Things (IoT) is expanding as technology progresses. To improve consumer quality of life, many suppliers are developing IoT devices. These Smart grids, smart homes, and smart health care are examples of devices. systems, intelligent transportation systems, and numerous more uses. IoT Devices communicate with each other and the surroundings utilizing actuators and sensors. Nevertheless, the pervasive spread of Cybersecurity risks posed by IoT devices are numerous. The IoT gadgets Interconnection provides access to potential attackers looking to gain Unauthorized use of these gadgets. numerous IT networks, building security and trust while using the gadget challenging. Additionally, devices may leak crucial information, this is a major issue in cybersecurity. Wi-Fi network security can be severely jeopardized by Man in the Middle (MITM) attacks, in which an intruder overhears on and intercepts traffic. This kind attack seeks to collect sensitive information like credit card numbers, logging into accounts and performing other crucial financial operations. In this paper, we performed a survey of 5 special machine studying fashions on a publicly available dataset and assessed many criteria, including precision, accuracy, most crucially, training and testing time, recollection, F1 score because the majority of IoT networks are hosted on limited resources such as Raspberry Pi devices. Also, these ML techniques are used to detect and identify MITM attacks and possible ways to mitigate it. We have divided our paper into sections, namely, Section-I: Introduction, Section-II: Literature Review/Related Works, Section-III: Results & Discussions, Section-IV: Conclusion & Future Work, Section-V: References.**

***Keywords*--- Man in the middle (MITM), Internet of Things (IoT), Cybersecurity, Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), Message Queuing Telemetry Transport (MQTT), Machine Learning (ML).**

# INTRODUCTION

 The Internet of effects (IoT) is a network of real- world objects that are linked together by means of assiduity-standard protocols in order to partake data acquired from their surroundings. The idea of the IoT has been in actuality for times. IoT bias are used by numerous artificial sectors to satisfy consumer prospects. IoT widgets ameliorate the homeowner's quality of life in smart homes, which is profitable. IoT device deployment has soared along with technology's quick development. Wi- Fi, Bluetooth, Z- Wave, and ZigBee are just a many of the communication protocols that IoT bias use to communicate with one another. The IoT Smart Home assiduity is anticipated have a request cap of 397 USD billion soon.

A MITM attack is a laptop- grounded assault in which an stranger impersonates one or each events in a two- manner communication situation, deceiving one group into thinking that it is speak me with the opposite. In those circumstances, the bushwhacker can pay attention in on the exchanges between the 2 ignorant agencies to advantage records. Both connected and unwired companies are vulnerable to these attacks, with the ultimate being more helpless.

One of the major issues we are now passing, especially in the IoT age, is the discovery of MITM assaults on Wi- Fi networks. Due to the significance of data moving over the network, particularly in situations like those involving operations for independent driving, the discovery and forestalment of these assaults is a largely delicate field of exploration.

Internet of effects (IoT) bias need a messaging and communication protocol in order to communicate data from a distance. For this protocol to be simple to apply on colourful low memory and computational power bias, it must have lo power consumption, lo quiescence, lo bandwidth, and a bitsy dataset footmark. These conditions are addressed by the transport protocol MQTT (Communication Queuing Telemetry Transport).

This featherlight dispatches protocol was developed by IBM and first published in 1999. It uses the cantina / sub pattern to understand dispatches between bias, waiters, and operations. We need a broker and guests to use MQTT. The broker oversees routing dispatches between the sender and the approved donors. This armature isolates guests, enabling for a largely scalable system that does not calculate on data directors and data consumers.

On this paper, we supply a hard and fast of machine literacy procedures for detecting and touching on MITM attacks on a wireless communique community. likewise, we assess and validate our method the use of performance measures and examine the consequences to being machine literacy ways.

On an open- source dataset, we also developed five indispensable machine literacy models and examined several characteristics similar as delicacy, perfection, recall, F1 score, training time, and test time.

According to former study, security breaches have grown by 66 in the last seven times, and 94 of HTTPs waiters are vulnerable to MITM attacks.

# Literature Review/Related Works

 A literature evaluates reports on detecting MITM assaults in competition to MQTT-based totally IoT devices carried out using various device studying algorithms. Many techniques have been proposed for identifying man-in-the-middle attacks and we have got conducted studies related to ML algorithms to transform the safety troubles faced by means of IoT-based totally machines.

These results provided by the researchers do not include correct feature selection mechanisms that were not available for a large number of Internet of Things devices. We explored various ways to troubleshoot his IoT security network using machine learning algorithms. We also showed that a synthetic intelligence framework was created to work with distributed computing. He applied ML algorithms such as K Nearest Neighbors, Logistic Regression, Random He Forest to detect faults in IoT networks.

Our research focused only on broad fault detection networks without looking at other types of attacks. This article focused primarily on his MIM attacks and provided a detailed overview of the data collection and preprocessing phases in user networks. Like his previous authors, Watkins talks of similarities between established attack strategies and paid computational difficulties using artificial intelligence frameworks. Researchers have developed a firewall security system to enhance the security of IoT devices.

We developed her MIM detection and location technology for Bluetooth-based network systems that did not show the expected results regarding IoT device threats

Experiments:

(MITM) is the system of obtaining message among end points for a motive of alteration of content; in this illustration acquired dispatches are delivered to the stoner and the mediator and the stoner records is edited. The Kali- Linux division and the Ettercap tool have been exercised to reap an rush to recreate the records set incorporates 110667 frames concerning 3854 raids. There are round ten of them features on which rush vaticination is completed. The crucial bones carry body time, source and the destination IP cope with, supply and the vacation spot harbourage, body quantity and a few other elements.

every other dataset for erecting our version is 194182 compliances, of which,298 compliances had been labelled as usual data and,884 compliances as rush statistics, defining a participation of 47. fifty-two usual and 52.39 rush packets. To make our fashions, we made education and look at information with valuations of 70 and 30. We calculated the translation of our fashions grounded at the distraction matrix, with which we exercised a bracket record to measure appropriate- best prognostications from bracket algorithms. Bracket document lists the number one standard comparable as delicacy, maintain in mind and f1 grievance consistent with magnificence. Metrics are expected the use of genuine and fake fantastic, authentic, and fake inhospitable. Therefore, we're having four methods to test if our prognostications are accurate or incorrect Firstly, authentic inhospitable (TN) is a case in which there was a real marker inhospitable and guessed to be inhospitable Secondly, actual advantageous (TP) the case in which the proper marker changed into high-quality and prognosticated in a fine parkway Thirdly, false inhospitable (FN) Represents the case wherein it's genuine the marker was effective however became expected to be inhospitable fee. Finally, fake high-quality (FP) Represents a case in which there is a true marker changed into inhospitable however prognosticated wonderful. delicacy represents how correct the prognostications are within the model. it is outlined because the fee of actual high-quality to the sum of trues and fake cons for each elegance.

P exactness = TP/ (TP+ FP) --(1)

The don't forget gives the of wonderful cases installation in model. it is mentioned because the price of authentic fantastic valuations to the sum of genuine cons and fake negatives.

Recall = TP/ (TP +FN) --(2)

Delicacy is the number of correct forecasts it consists of each effective and inhospitable casts, separated via the grand variety prognostications made.

Accuracy = (TP +TN)/ (TP +TN+ FP+ FN) --(3)

The F1 score represents the chance of accurate positive prognostications. The F1 grievance is a harmonious- mean of perfection and recall that the stylish grievance is 1.00 and the worst is 0.00. ScoreF1 = 2 ∗ (Recall ∗ Precision) (Recall +Precision P) --(4).

# Results and Discussions

 MITM attacks are labeled into two kinds; ordinary information does no longer contain any assaults. This statistic is labelled as every day, Attack1, and Attack2. To prepare the statistics for the education segment, the labels are transformed into numerical values which includes everyday=zero and assault 1, assault 2=1. The data in the selected data set is depicted in Figure 1 below.

 Number of Attack1-Attack 2-Normal Packets



Fig. 1. Dataset Distribution

The attack is based on a deep learning approach called LSTM within the Detection of guy within the center attack making use of ML techniques to begin with. The model is trained using a previous dataset. Following that, the training is completed. Figure 2 depicts the Confusion Metrics for the aforementioned Model.

 

Fig. 2. CM for LSTM

Many unique Attributes are calculated to affirm the performance of the LSTM version. The accuracy of the model is calculated to be 0.92

* TP = 447
* FP = 423
* FN = 98
* TN = 5136
* Rec =0.827
* Prec =0. 525
* F1-Score =0.637

TP=True Positive

FP=False Positive

FN=False Negative

TN=True Negative

Now, for MITM attack detection, the Random Forest Technique is utilised. Random wooded area (Random decision forest) is an ensemble studying approach for type, regression, different tasks that works by way of producing awesome choice timber in the course of education. As shown in Figure 3, this method comprises of many decision trees that are trained and integrated to produce accurate, exact, and consistent results.

 

 Fig. 3. Random Forest Algorithm

After applying the Random Forest Algorithm for the attack detection, the accuracy is found to be 0.94 for which a confusion matrix is presented in Figure 4.

 

Fig. 4. CM for RF algorithm

The other performance metrics are:

* TP=611
* FP=262
* FN=95
* TN=5138
* Rec =0.876
* Prec =0.701
* F1-Score =0.777

TP=True Positive

FP=False Positive

FN=False Negative

TN=True Negative

Now the performance is measured using the SVM algorithm, in device gaining knowledge of, a aid vector system is a supervised mastering model with an associated learning algorithm that analyses information for classification and regression analysis. Figure 5 depicts the two-class classification by a hyperplane.



Fig. 5. Hyperplane by SVM

With the creation of the confusion matrix for SVM, we got the accuracy for the model to be 0.857, during the conversion of normal data to attacks the accuracy was lost. Figure 6 depicts the illustrated results of the confusion matrix which applies to the SVM.

 

 Fig. 6. CM for SVM

The other values are:

* TP= 0.0100
* FP=872
* FN=0.0100
* TN=5237
* Rec =0.0010
* Prec =0.0010
* F1-Score =0.0001

TP=True Positive

FP=False Positive

FN=False Negative

` TN=True Negative

The comparative graph for the above 3 algorithms is shown below in Figure 7 which is used to select the best algorithm:

 

 Fig. 7. Accuracy of the Used techniques

Table 1: Comparison of LSTM with the other algorithms:



Table 1 depicts the comparison of deep learning with other machine earning algorithms.

With a different dataset, many different algorithms are used to test the Attack detection.

The algorithms employed have been *ANN* (synthetic Neural network), that is a computational model that simulates how nerve cells within the human brain behave (or neural network). artificial neural networks (ANNs) use gaining knowledge of strategies that allow them to alter their responses autonomously in reaction to new enter.

In fact, *Naive* Bayes classifiers are a own family of uncomplicated "probabilistic classifiers" which are primarily based on the application of Bayes' theorem with strong independence assumptions between the capabilities. notwithstanding being one of the only Bayesian network models, they will gain high levels of accuracy when mixed with kernel density estimation.

*Gradient Boosting* at Extremes, Gradient boosting is a machine learning set of rules this is used for classification and regression applications, among different things. It offers a prediction version in the form of a selection tree-like ensemble of vulnerable prediction fashions.

*Decision Tree* are a sort of supervised gadget getting to know wherein the schooling statistics is continuously segmented depending on a sure parameter, with you specifying the enter and the corresponding output. the two additives that may be utilised to illustrate the tree are choice nodes and leaves.

*K-Nearest Neighbour* (KNN) it's far a supervised gadget getting to know algorithm. The approach may be used to cope with problem statements which includes class and regression. The letter "k" represents the range of nearest neighbours to a brand-new unknown variable that should be classified Table 2 shows a comparison of the aforementioned algorithms.

Table 2-Comparison of the above-stated Algorithms:



The outcomes of the other studies now include an exam of guy inside the center assault the use of device studying Algorithms. First, the Logistic Regression technique is used to are expecting the explicit structured variable the usage of a distinctive set of independent variables. The output of a specific structured variable is expected using logistic regression. Figure 8 depicts the Logistic Regression confusion matrix.

 

 Fig. 8. CM for Logistic Regression

The selection tree approach is now used to discover the records’ final results. A choice tree is a selection help tool that makes use of a tree-like model of choices and their capacity outcomes, consisting of hazard event outcomes, useful resource fees, and software. that is one method for demonstrating an algorithm that most effective employs conditional manages statements. figure 9 depicts the decision Tree confusion matrix.

 

Fig. 9. CM for Decision tree

Following the LR and DT approaches, the RF strategy is now used to extract the needed data from the dataset. A RF is a ML technique for dealing with classification and regression problems. It employs ensemble learning, a method for resolving complex problems by integrating numerous classifiers. A random forest method employs many DT’s. The confusion matrix shown in Figure 10 depicts the results of Logistic Regression.

 

 Fig. 10. CM for RF

Comparing the results,

Table 3: Evaluation of Logistic regression (LR), choice tree (DT) and Random Forest (RF)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ALGO** | **ACC** | **PREC** | **RECALL** | **F1- SCORE** |
| **LR** | 0.998 | 0.999 | 0.995 | 0.993 |
| **DT** | 1.000 | 1.000 | 1.000 | 1.000 |
| **RF** | 1.000 | 1.000 | 1.000 | 1.000 |

# Limitations:

Despite the fact that the Extreme gradient boosting technique demonstrated 99.7% accuracy, its training period is significantly longer than that of other algorithms.

# Conclusion and future work

 Man in the Middle attacks are extremely risky due to the fact they now not only harm the transmission service however additionally have an effect on in the theft of fragile and touchy data and information exchange between the bushwhacker and legal druggies. As technology progresses, bushwhackers will decreasingly use IoT bias to gain unwanted network access. Traditional ways of detecting intrusions have limitations, challenging the operation of our machine learning technology. In this composition, we observed that all the models were further than 85 accurate. The extreme grade boosting system was 99.6 accurate. The logistic retrogression approach was 99.7 accurate, the arbitrary timber was 100 accurate, and the decision tree was 100 accurate with a veritably high F1 score. In addition, two styles, SVM- Random Forest, we have used to compare the issues of the deep literacy network. We hope to induce our own network dataset in the future by automating assault scripts of multitudinous cyber-attacks. We will also use Deep literacy to make a model that will learn from a particular dataset in order to help this type of attack from spreading.

REFERENCES

1. BAKSHI, Aman; DUJODWALA, Yogesh B. Securing cloud from ddos attacks using intrusion detection system in virtual machine. In: 2010 Second International Conference on Communication Software and Networks. IEEE, 2010. p. 260-264.
2. JEYANTHI, N.; IYENGAR, N. Ch SN. Packet resonance strategy: a spoof attack detection and prevention mechanism in cloud computing environment. International Journal of Communication Networks and Information Security, 2012, 4.3: 163.
3. 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.
4. 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
5. ZHU, Shasha; GONG, Guang. Fuzzy authorization for cloud storage. IEEE Transactions on Cloud Computing, 2014, 2.4: 422-435.
6. 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.
7. VISCONTI, Andrea; TAHAYORI, Hooman. A biologically-inspired type-2 fuzzy set-based algorithm for detecting misbehaving nodes in ad-hoc wireless networks. International journal for infonomics, 2010, 3: 373-82.
8. WATKINS, Damian. Tactical manet attack detection based on fuzzy sets using agent communication. MORGAN STATE UNIV BALTIMORE MD, 2004.
9. 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.
10. 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.
11. Saed, Muhanna, and Ahamed Aljuhani. “Detection of Man in the Middle Attack Using Machine Learning.” *Detection of Man in the Middle Attack Using Machine Learning*, 27 Jan. 2022, ieeexplore.ieee.org/abstract/document/9711555.
12. Sultan, Ali Bin Mazhar, et al. “Man in the Middle Attack Detection for MQTT Based IoT Devices Using Different Machine Learning Algorithms.” *Man in the Middle Attack Detection for MQTT Based IoT Devices Using Different Machine Learning Algorithms*, 17 May 2022, ieeexplore.ieee.org/document/9773590.
13. Toutsop, Otily, et al. “Monitoring and Detection Time Optimization of Man in the Middle Attacks Using Machine Learning.” *Monitoring and Detection Time Optimization of Man in the Middle Attacks Using Machine Learning*, 10 May 2021, ieeexplore.ieee.org/document/9425304.
14. Anand, Gokul, et al. “Detection of Man In The Middle Attacks in Wi-Fi Networks by IP Spoofing.” Detection of Man In The Middle Attacks in Wi-Fi Networks by IP Spoofing, 13 Dec. 2019, ieeexplore.ieee.org/abstract/document/8939063.
15. Ekparinya, Parinya, et al. “Impact of Man-In-The-Middle Attacks on Ethereum.” Impact of Man-In-The-Middle Attacks on Ethereum, 17 Jan. 2019, ieeexplore.ieee.org/abstract/document/8613949.
16. Sun, Da-Zhi, et al. “Man-in-the-Middle Attacks on Secure Simple Pairing in Bluetooth Standard V5.0 and Its Countermeasure - Personal and Ubiquitous Computing.” SpringerLink, 26 Sept. 2017, link.springer.com/article/10.1007/s00779-017-1081-6.
17. P. Sampath et al., “Iot based health—related topic recognition from emerging online health community (med help) using machine learning technique,” Electronics, vol. 9, no. 9, 2020. [Online]. Available: <https://www.mdpi.com/2079-9292/9/9/1469>
18. A. N. Alvi et al., “Ogmad: Optimal gts-allocation mechanism for adaptive data requirements in ieee 802.15.4 based internet of things,” IEEE Access, vol. 7, pp. 170 629–170 639, 2019.
19. T. H. Team, “Introducing the mqtt protocol - mqtt essentials: Part 1,” 2015. [Online]. Available: <https://www.hivemq.com/blog/mqttessentials-part-1-introducing-mqtt/>
20. H. C. Hwang, J. Park, and J. G. Shon, “Design and implementation of a reliable message transmission system based on mqtt protocol in iot,” Wireless Personal Communications, vol. 91, no. 4, pp. 1765–1777, 2016.
21. S. Shin et al., “A security framework for mqtt,” in 2016 IEEE Conference on Communications and Network Security (CNS). IEEE, 2016, pp. 432–436.
22. J. Aveleira-Mata, “Network data set from an iot system with mqtt protocol attacks.” [Online]. Available: <https://joseaveleira.es/dataset>
23. N. F. Syed et al., “Denial of service attack detection through machine learning for the iot,” Journal of Information and Telecommunication, vol. 4, no. 4, pp. 482–503, 2020.
24. H. Hindy et al., “Machine learning based iot intrusion detection system: an mqtt case study (mqtt-iot-ids2020 dataset),” in International Networking Conference. Springer, 2020, pp. 73–84.
25. I. Vaccari et al., “Mqttset, a new dataset for machine learning techniques on mqtt,” Sensors, vol. 20, no. 22, p. 6578, 2020.