DOI: 10.21917/ijct.2021.0378

HOME NETWORK SECURITY INCORPORATING MACHINE LEARNING ALGORITHMS IN INTERNET OF MEDICAL THINGS

Pallavi Arora[1], Baljeet Kaur[2] and Marcio Andrey Teixeira[3]

[1]*Department of Computer Science and Engineering, I.K. Gujral Punjab Technical University, India*
[2] *Department of Electronics and Communication Engineering, Guru Nanak Dev Engineering College, India*
[3]*Department of Informatics, Federal Institute of Education, Science, and Technology of São Paulo, Brazil*

Abstract

The proliferation of chronic disorders such as COVID-19 has recognized the importance of people all over the world having immediate access to healthcare. The recent pandemic has shown deficiencies in the traditional healthcare infrastructure, namely that hospitals and clinics alone are inadequate for grappling with such a disaster. One of the key technologies that favours new healthcare solutions is smart and interconnected wearables. Thanks to developments in the Internet of Things (IoT), these wearables will now collect data on an unprecedented scale. However, as a result of their extensive use, security in these critical systems has become a major concern. This paper presents an intrusion detection mechanism based on Machine Learning Algorithms for healthcare applications used in home network environments. Experiments are carried out on a home network to detect attacks against a health care application. Experiments using the proposed mechanism based on Machine Learning algorithms to detect attacks against a healthcare application are carried out on a home network, and the results show a good performance of the used algorithms.

Keywords:

IoMT, Security, Smart Watch, IDS

1. INTRODUCTION

The Internet of Things (IoT) transpires a concept which illustrates a broad variety composed of interconnected articles and equipment in order to retrieve indication out of its conditions via probes, interpret it, then operate upon real arena via actuators. IoT encompasses a broad range of technologies in diverse fields, including smart electricity grids, industrial management systems, hospitals, transportation, home electronics, and wearables, and is analogous to contemporary cyber-physical systems [1]. Despite the apparent organizational and functional advantages, IoT incorporation has opened up new assault possibilities for remote adversaries [2], [3].

In the healthcare sector, IoT devices that support vital functions of healthcare are referred to as the Internet of Medical Things (IoMT). It exists as a conglomeration composed from medical equipment as well as technology which can use networking tools to communicate to healthcare information systems. Machine-to-machine interaction, which is the foundation of IoMT, is allowed by medical devices configured with Wi-Fi. IoMT devices can be connected to cloud networks like Amazon Web Services, which store and interpret collected data. Healthcare IoT is another acronym for IoMT.

The protection of IoMT components together with medical systems in general (hereafter referred to as IoMT systems) is, indeed, an exigent impediment [4]. The health information collected, transmitted, and stored by IoMT systems should be secured in all stages [5]. Traditional solutions, on the other hand, could fail to provide adequate security guarantees due to their high-power consumption and other device requisites [6], [7]. Currently, researchers have suggested several different approaches that are specifically tailored for IoMT and IoT networks [7], [8].

In this paper, we analyse the performance of an IDS based on Machine Learning (ML) algorithms to detect attacks against the IoMT devices used in environments like a home network [9]. The home networks traditionally do not have any sophisticated security equipment, and most of them have only the security configured in their local Wi-Fi routers. So, we will train a ML framework for understanding its patterns including its home network in identifying attacks. So, we will verify if this kind of solution could be suitable for these environments.

The rest of the paper is structured as follows: Section 2 presents a literature survey regarding the usage of machine learning algorithms for security of IoMT devices. Section 3 represents proposed intrusion detection mechanism and experimental scenario used in home networks. Section 4 represents numerical results and interpretations related to them. Finally, Section 5 gives concluding remarks of this paper.

2. LITERATURE REVIEW

In [10], the Smart Healthcare Systems (SHS) provide patients with enhanced assessment options and medication, but they also pose a number of security issues, as addressed in this paper. The authors also introduced the HealthGuard, a revolutionary ML-based surveillance platform which could represent the actual state of a SHS and assess whether any harmful behavior had occurred with in framework. To detect suspicious behaviors in a SHS, HealthGuard employs four ML-based identification strategies (Artificial Neural Network, Decision Tree, Random Forest, and k-Nearest Neighbor). HealthGuard is an appropriate security system for SHS, according to author's exhaustive assessment, with a 91% accuracy rate and a 90% of F1 score.

The authors of [11] defined IoT threat types including teaching based IoT protection techniques such as IoT verification, user identification, trojan identification, as well as safe unloading, that have already been demonstrated to be potential IoT safety.

It is investigated in [12] whether Internet of things information is gathered for ML approaches, as well as the present hurdles in promoting creative IoT technologies. They also proposed a methodology for IoT apps to study from some of the similar IoT apps dynamically, as well as a situational analysis as to how the methodology might be expanded to field experiments. Finally, they go through the most important considerations that will influence potential intelligent IoT implementations.

In [1], the authors categorize IoT communication protocols based on their usage in IoMT. The key characteristics of IoT network topologies employed at consumer product sensing, network, and access layer are then outlined. The intrinsic security features and shortcomings of IoMT-specific communication protocols are investigated. Based on realistic risks, they outline acceptable remediation procedures that can be implemented to protect IoMT interactions, and also current testing as well as deployment constraints.

Machine Learning (ML) has seen a significant technical breakthrough, opening up several new testing avenues to solve current and potential IoT problems [13]. To detect threats and detect suspicious activities in smart devices and networks, machine learning is being used as an important technology. Following a detailed literature review on ML approaches to IoT defense in terms of various types of potential threats, the design of IoT is explored in [13]. Furthermore, ML-based possible solutions for IoT protection have been described, along with potential risks.

Detecting threats and flaws in Internet of Things (IoT) technology has now become a major challenge throughout the IoT industry. Threats and assaults on IoT infrastructure are increasing in tandem with the expanded usage of IoT infrastructure in every industry. Denial-of-service, Numeric Values Invasion, Fraudulent Controlling, Disruptive Procedure, Scanning, and Incorrect Configuration are all examples of assaults and anomalies that might lead an IoT device to fail.

In In [14] the efficiency of many machine learning models for predicting attacks and abnormalities on IoT systems is compared. Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN) are among the machine learning (ML) algorithms that have been included in this study. The measurement parameters that were used to compare results were correctness, clarity, recollection, f1 measure, and arc under Roc Curve. For Decision Tree, Random Forest, and ANN, the device achieved a test accuracy of 99.4%. Despite having the similar correctness, additional measurements suggest that Random Forest outperforms them.

Authors in [15] use IDS to address the security requirements of IoT systems in healthcare. The purpose of this research is to assert that combining systems and biological measurements as characteristics surpasses utilizing just one of the two. They presented an integrated Enhanced Healthcare Monitoring System (EHMS) test platform which collects network traffic measures as well as patient biometrics. The obtained information is routed to a cloud database for additional diagnosis as well as treatment possibilities. Man-in-the-middle cyber-attacks were deployed, and a database of over 16 thousand entries of normal and malicious medical records was created. After that, the algorithm uses a variety of machine learning techniques to train and validate the dataset against these attacks. The results reveal that in some situations, efficiency has grown from 7% to 25%, indicating the suggested program's robustness in delivering accurate detection mechanism.

3. PROPOSED INTRUSION DETECTION

This section describes the proposed intrusion detection mechanism based on ML algorithms. The main idea of our intrusion detection mechanism is training ML models to detect attacks in a home network. So, we can divide our development into two steps: (a) obtain a representative dataset to train and test the ML models; (b) Training and test the ML models using a dataset and evaluate their performance in a home network.

3.1 DATASET DEVELOPMENT

The main component used to train and test a ML algorithm is the dataset because the ML algorithms learn from the dataset. So, the choice of a representative dataset is a very important step in the development of an ML project. In our case, we are creating a ML model to detect attacks against an IoT medical device at the home network. This necessitates the usage of a dataset including network traffic traces, and this dataset needs to have normal traffic and attacked traffic.

The home network traffic varies from home to home. In this case, we have created our dataset based on the typical devices and applications used on a home network. The Fig.1 illustrates our home network testbed.

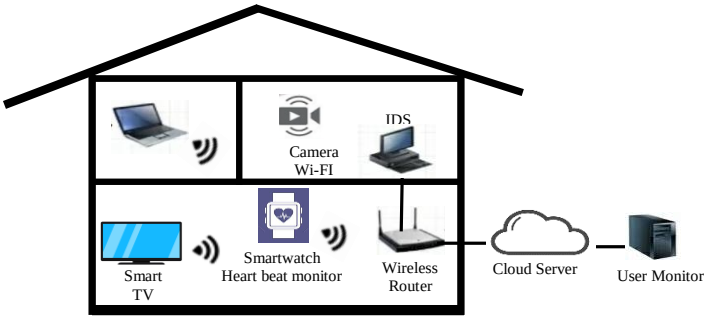

Fig.1. Experiment scenario

Our home network testbed is composed of traditional devices used at home networks like laptops, smart TVs, and security cameras. All devices are connected to the Internet by a wireless router generating heterogeneous network traffic. In our scenario, a user is being monitored by a smartwatch device. The smartwatch reads the beat heart signals from the user and sends this information through the Internet to a user monitor. So, this user can be monitored 24 hours a day.

All network traffic was stored for 4 hours. During this time, we carried out a reconnaissance attack against the smartwatch. The network traffic was monitored by a computer that has an IDS as shown in Fig.1. We used the Argus program to monitor and store all home network traffic flow.

An IP data stream would be a series comprising messages sent from a source (any equipment connected to an IP network) to a receiver. A system movement is made up of all the packets in a particular transit link or medium flows. The packets associated with the flow possess a set of common characteristics, such as ports, protocols, etc. The home network traffic flow is monitored and stored by Argus [16]. Argus is a system monitoring application used to monitor the status of the network services. It consists of two components: a main program (also called daemon) used to detect the network flows and log them to a file, and the clients that connect to the main program to have access to the logs.

Once collected all network flows from the home network, the data is classified into normal traffic and attacked traffic. This classification is made using the IP number of the computer attack.

Our goal in this step is to create our dataset for training and testing the machine learning algorithms to detect the attack against the home network. Table 1 shows statistical information about our dataset, and Fig.2 shows the information about the features belonging to each network flow.

Table.1. Statistical information on the traffic during the reconnaissance and command injection attacks

Measurement	Value
Duration of capture	4 Hours
Number of observations	6,257
Normal traffic (%)	99.05%
Attacked traffic (%)	0.95 %

The selection of good features is an important step to build a ML model, because how the higher the number of features used to build a ML model, the more complex the model will be. In our case, verified the correlation of the features belonging to the network flow using the Pearson correlation coefficient [17]. The Fig.3 illustrates our dataset utilized to develop and evaluate the Machine Learning programs, as well as the features used after the Pearson correlation coefficient was applied.

Features	Type	Description
Source Port (Sport)	Integer	Source port number
Destination Port (Dport)	Integer	Destination port number
Source Packets (SrcPkts)	Integer	Source/Destination packet count
Destination Packets (DstPkts)	Integer	Destination/Source packet count
Total Packets (Tpkts)	Integer	Total transaction packet count
Source Bytes (Sbytes)	Integer	Source/Destination bytes count
Total Bytes (TBytes)	Integer	Total transaction bytes count
Source Load (Sload)	Float	Source bits per second
Destination Load (Dload)	Float	Destination bits per second
Total Load (Toad)	Float	Total bits per second
Source Rate (Srate)	Float	Source packets per second
Destination Rate (Drate)	Float	Destination packets per second
Total Rate (Trate)	Float	Total packets per second

Fig.2. Features of the network flow captured by Argus

SrcPkts	DstPkts	TotPkts	TotBytes	Traffic
2	0	2	286	Normal
1	0	1	187	Normal
4	3	7	925	Normal
1	1	2	128	Attack
1	0	1	154	Normal
1	0	2	128	Attack
1	1	2	90	Attack
21	0	21	7328	Normal
1	0	1	154	Normal
1	1	2	128	Attack

Fig.3. Dataset used for training and testing the machine learning algorithms

The created data is utilized to test and evaluate the ML programs. The dataset is split into two parts, where 80% is used to train the model and 20% is used to test the model. The results of the attack detection are shown in the next section.

3.2 PERFORMANCE METRICS

The metrics obtained from the confusion matrix have been intended to evaluate the effectiveness of the machine learning programs. The Table.2 depicts the confusion matrix. According to the confusion matrix, the variables used during the research to test the application of the ML procedures are as follows:

Table.2. Confusion Matrix

	Predict Class	
Data Class	**Classified as Normal**	**Classified as Abnormal**
Normal	True Negative (TN)	False Negative (FN)
Abnormal	False Positive (FP)	True Positive (TP)

- **Accuracy**: When the whole number of forecasts is taken into consideration, it is the proportion (%) of accurately forecast samples.

$$\text{Accuracy \%} = (TP+TN)/(TP+TN+FP+FN) \times 100 \quad (1)$$

- ***False Alarm Rate (FAR)***: The proportion of real information falsely labeled as an anomaly (attack) by the model.

$$\text{FAR \%} = FP/(FP+TN) \times 100 \quad (2)$$

- ***Training time***: It is time taken by ML model to learn patterns and classify data into normal and abnormal.

4. SIMULATION RESULT

Here the performance of the ML model developed in this paper is shown in this section. The Fig.4 shows the accuracy metric of all ML models used in this work.

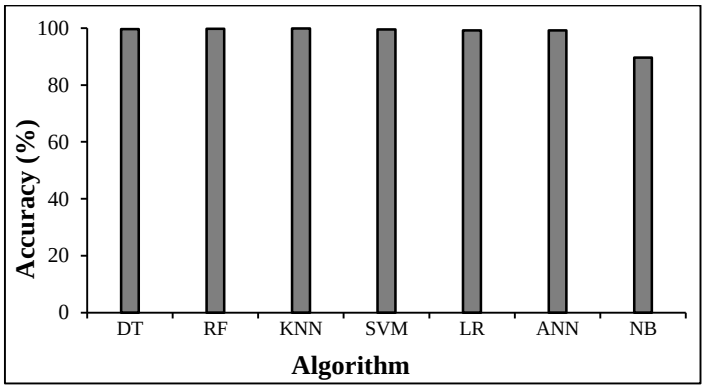

Fig.4. Accuracy

As shown in Fig.4, the KNN algorithm has better performance than the other algorithms. However, the difference of the KNN accuracy result comparing to Decision Tree, Random Forest, SVM, Logistic Regression and ANN algorithms is low. The algorithm Naïve Bayes has the worst performance. Fig.5 presents the False Alarm Rate (FAR) metric.

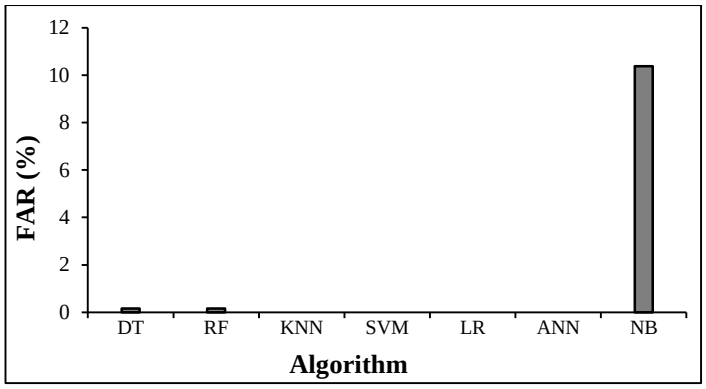

Fig.5. False alarm Rate (FAR) results

The FAR measures represent the normal traffic that has been erroneously classified as abnormal traffic by the model. So, these metrics represent the percentage of false alarms, and in this case, the lower FAR value is considered better. As shown in Fig.5, the algorithm Naïve Bayes has the higher FAR value, having the worst performance considering this metric. The Fig.6 shows the time wasted to train the ML algorithm, also called training time.

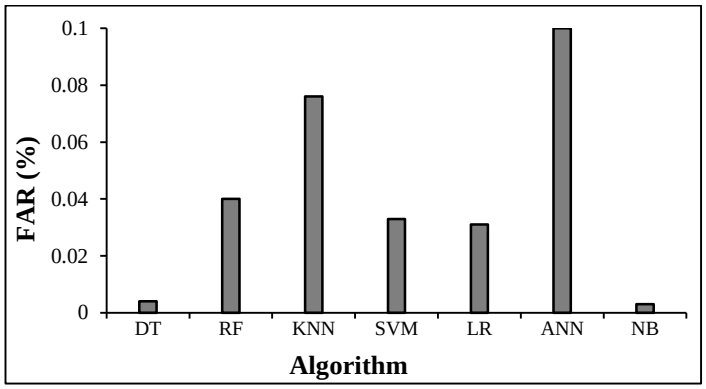

Fig.6. Training time

As can be seen in Figure 6, the ANN waste more time to train the model comparing to other algorithms. The ANN algorithm is more complex to train comparing, for example, to Logistical Regression algorithm. On the other hand, Decision Tree and Naive Bayes take less time to train the model.

If we have a situation where we need to rebuild the ML model in accordance with the characteristics of the traffic in a dynamic way, the training time is an important metric. However, we need to consider all metrics to define what ML algorithm is better to identify attacks against a home network. So, it is possible to see in our results that the Decision Tree is had the better performance to be used in our model. KNN algorithm had better accuracy, but it takes more time to train the model comparing to Decision Tree algorithm. The accuracy results of the KNN and Decision Tree algorithms is tiny.

The ML models were built for our home network. However, to be used in a different network, the models need to be rebuilt in accordance with the traffic of the network. As future work, we suggest the development of a new mechanism that can automatically classify the normal traffic of the network. So, the ML models can be built to detect attacks according to the network traffic of the different home networks.

5. CONCLUSION AND FUTURE SCOPE

Because of the significant increase in utilization of IoMT detectors in order to reduce medical expenses and deliver quality looking after of the clients, securing such instruments is becoming crucial. IoMT devices, on the other hand, offer scarce capabilities while ones which have been presently installed require the usage of external equipment to protect them. In this paper, we have proposed an intrusion detection mechanism for home networks using machine learning approaches. The results showed that it is feasible to use mechanisms based on ML algorithms to detect attacks against the healthcare application in a home network. However, the ML algorithms need to be trained in all home networks to learn the patterns of the home network, once each home network has it owns characteristics. As a future work, we suggest the development of a mechanism that can be used to identify the best features of the home network traffic in an automatic way. Then, ML models can be adjusted according to the home network traffic.

REFERENCES

[1] D. Koutras, G. Stergiopoulos, T. Dasaklis, P. Kotzanikolaou, D. Glynos and C. Douligeris, "Security in IoMT Communications: A Survey", *Sensors*, Vol. 20, No. 17, pp. 4828-4848, 2020.

[2] F. Alsubaei, A. Abuhussein and S. Shiva, "Security and Privacy in the Internet of Medical Things: Taxonomy and Risk Assessment", *Proceedings of IEEE 42nd International Conference on Local Computer Networks*, pp. 112-120, 2017.

[3] S. S. Hameed, "A Systematic Review of Security and Privacy Issues in the Internet of Medical Things; the Role of Machine Learning Approaches", *Computer Science*, Vol. 7, No. 4, pp. 44-56, 2021.

[4] M. Papaioannou "A Survey on Security Threats and Countermeasures in Internet of Medical Things (IoMT)", *Emerging Telecommunications Technologies*, Vol. 23, pp. 1-12, 2020.

[5] A. Hockey, "Uncovering the Cyber Security Challenges in Healthcare", *Network Security*, Vol. 2020, No. 4, pp. 18-19, 2020.

[6] G. Thamilarasu, A. Odesile and A. Hoang, "An Intrusion Detection System for Internet of Medical Things", *IEEE Access*, Vol. 8, pp. 181560-181576, 2020.

[7] G. Hatzivasilis, O. Soultatos, C. Verikoukis, G. Demetriou, C.I. Tsatsoulis and N. Systems, "Review of Security and Privacy for the Internet of Medical Things (IoMT)", *Proceedings of IEEE 42nd International Conference on Distributed Computing in Sensor Systems*, pp. 1-8, 2019.

[8] J.J. Hathaliya and S. Tanwar, "An Exhaustive Survey on Security and Privacy Issues in Healthcare 4.0," *Computer Communications*, Vol. 153, pp. 311-335, 2020.

[9] H.K. Bharadwaj, A. Agarwal, V. Chamola, V. Hassija, M. Guizani and B. Sikdar, "A Review on the Role of Machine Learning in Enabling IoT Based Healthcare Applications", *Proceedings of IEEE International Conference on AI and IoT for Smart Health*, pp. 32-45, 2021.

[10] A.I. Newaz, A.K. Sikder, M.A. Rahman and A.S. Uluagac, "HealthGuard: A Machine Learning-Based Security Framework for Smart Healthcare Systems", *Proceedings of IEEE International Conference on Social Networks Analysis, Management and Security*, pp. 389-396, 2019.

[11] L. Xiao, X. Wan, X. Lu, Y. Zhang and D. Wu, "IoT Security Techniques Based on Machine Learning: How Do IoT Devices Use AI to Enhance Security?", *IEEE Signal Processing Magazine*, Vol. 35, No. 5, pp. 41-49, 2018.

[12] E. Adi, A. Anwar, Z. Baig and S. Zeadally, "Machine Learning and Data Analytics for the IoT", *Neural Computing and Applications*, Vol. 32, No. 20, pp. 16205-16233, 2020.

[13] S.M. Tahsien, H. Karimipour and P. Spachos, "Machine Learning based Solutions for Security of Internet of Things (IoT): A Survey", *Journal of Network and Computer Applications*, Vol. 161, pp. 1-13, 2020.

[14] M. Hasan, M.M. Islam, M.I.I. Zarif and M.M.A. Hashem, "Attack and Anomaly Detection in IoT Sensors in IoT Sites using Machine Learning Approaches", *Internet of Things*, Vol. 7, pp. 1-19, 2019.

[15] A.A. Hady, A. Ghubaish, T. Salman, D. Unal and R. Jain, "Intrusion Detection System for Healthcare Systems Using Medical and Network Data: A Comparison Study", *IEEE Access*, Vol. 8, pp. 106576-106584, 2020.

[16] Openargus-Home, Available at https://openargus.org/, Accessed at 2021.

[17] Pearson Correlation an Overview, Available at https://www.sciencedirect.com/topics/computer-science/pearson-correlation, Accessed at 2021.