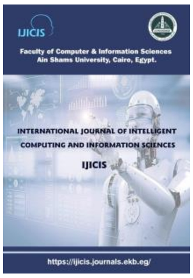
IJICIS, Vol.22, No.1, 16-31
DOI: 10.21608/ijicis.2021.79289.1100

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/


IJICIS
Faculty of Computer & Information Sciences Ain Shams University, Cairo, Egypt.
INTERNATIONAL JOURNAL OF INTELLIGENT COMPUTING AND INFORMATION SCIENCES
IJICIS
https://ijicis.journals.ekb.eg/

A Threshold-based Technique to Cluster Ransomware Infected Medical Records on the Internet of Medical Things

Randa ElGawish*
Bioinformatics, Computer and Information Sciences, Ain Shams University, Cairo, Egypt
randa.mahmoud@cis.asu.edu.eg

Mohamed Hashem
Information System, Computer and Information Sciences, Ain Shams University, Cairo, Egypt
mohamed.hashem@cis.asu.edu.eg

Rania ELGohary
Information System, Computer and Information Sciences, Ain Shams University, Cairo, Egypt
rania.elgohary@cis.asu.edu.eg

Mohamed Abu-Rizka
Computer Science, Computing and Information Technology, Arab Academy for Science and Technology, Cairo, Egypt
m.aborizka@aast.edu.eg

Received 2021-6-5; Revised 2021-9-27; Accepted 2021-10-7

Abstract: *Ransomware attacks have led many healthcare hospitals to migrate back to their traditional methods of monitoring patients using pen and paper instead of using implantable medical devices remotely. Studying the behaviour of payload ransomware on an approved actual healthcare dataset obtained from ICU and correctly clustering them into normal and malicious records after manifestation is the primary focus of this study. The features decided were upon the possibility of being captured remotely and their frequency of occurrences. Data transformation was included, to handle the encrypted values and perform data normalization, prior to the clustering process.*

Unsupervised machine learning gained a lot of attention in the cybersecurity domain for its efficiency and capability of clustering tuples into malicious and benign categories. However, on the internet of medical things (IoMT), due to the constraints of the interconnected nodes, clustering of malicious activities became highly challenging and demanded to secure the infrastructure. This work used unsupervised machine learning techniques of k-means, DBscan, and mean shift compared to a threshold-based method which outperformed them with a precision of 100%. The performance metrics used in this work are; precision, recall and f1 score.

Keywords: *Machine learning, Internet of Medical Things, data science, cybercrime, Internet of Things*

* Corresponding author: Randa ElGawish
Bioinformatics, Computer and Information Sciences, Ain ShamsUniversity, Cairo, Egypt
E-mail addressranda.mahmoud@cis.asu.edu.eg

1. Introduction

Internet of things (IoT), the recent revolution of technology, became the most attractive topic for researchers due to its benefits of implementation in various applications. It comprises small devices known as sensors that capture valuable and highly confidential readings of the surroundings. Implementation of IoT in healthcare systems or (IoMT) provides innovative services in remote monitoring across healthcare systems using a set of interconnected networks and devices or nodes. IoMT made a revolution in technology, especially during the pandemic of the coronavirus, as it enables an infection-controlled environment. However, due to its nature to sensitive and confidential information of patients, ensuring security is a critical issue in the development of an IoT-based healthcare system.

Highly qualified cybercriminals build nowadays intelligent, dynamic and untraceable algorithms that infect the system. These algorithms can cause an overall failure in extreme conditions. IoMT shares the same architecture of IoT concerning the data transmitted within the system. In addition to the equivalence in the security requirements, the malware targeting IoT technologies differ from one domain to another. For instance, in 2016, the collecting value of personal health information worth $50 compared with the prices of credit card information which is $1.50 and $3 for social security numbers [4]; these values are far more nowadays.

For example, patients' lives with critical cases depend on the medical information captured by IoT implantable medical devices. Any losses, errors or modifications within this medical information in certain situations could lead to loss of human lives or severely affecting them directly. Therefore, the development of self–intelligent, autonomous security algorithms is highly demanded in cybersecurity. A 2017 report from Cybersecurity Ventures [14], anticipated ransomware harms would cost the world $5 billion out of 2017, up from $325 million out of 2015 — a 15X expansion in only two years. The harms for 2018 were anticipated to reach $8 billion, and for 2019 the figure is $11.5 billion.

The most recent expectation is that worldwide ransomware harm costs will reach $20 billion by 2021 – which is 57X more than it was in 2015. This makes ransomware the quickest developing sort of cybercrime [1]. On October 29th,2020, American emergency clinics are being targeted with a flood of ransomware assaults as coronavirus in the U.S. break records, pushing the nation's wellbeing framework as far as possible [2].

The Federal Bureau of Investigation, the Cybersecurity and Infrastructure Security Agency, tweeted a dramatic warning to the health care providers against an imminent ransomware attack. The agencies said hackers are using Ryuk ransomware, malicious software used to encrypt and keep information locked, and the infected computer Trickbot network to steal information, interrupt health care services and extort money from health care facilities.

Online networks are also compromised by hijacking data, causing many to pay millions of dollars to restore their services [3]. It can be transmitted via attachments, links in phishing emails, malicious websites, downloaded drives or infected USB sticks to computers and smartphones. The malware prevents access to the device until a machine or network becomes infected with ransomware to encrypt data on that system. Cybercriminals demand victims to pay a ransom to regain access to their computers or restore their data. Ransomware exists in two types crypto and locker ransomware. Crypto ransomware encrypts valuable files on a computer or can encrypt data within files. When data or, as in this case study, medical records become infected, we refer to this malware as payload ransomware. This malware encrypts the physiological measurement tests obtained through patients. Moreover, locker ransomware does not encrypt files but locks all access means into the targeted computer or network.

The purpose of this research is to study how payload ransomware affects medical data captured within the intensive care unit of Beth Israel Medical Deaconess Medical Center, Boston, U.S. [6]. The study includes implementing three unsupervised machine learning techniques and a threshold-based clustering technique to cluster medical records into normal and malicious. The evaluation of the results was verified using precision, recall or true positive rate, and f1 score.

This paper is divided into four sections. The first section is the introduction, then section 2, the experimental methodology that compromises a detailed description of the dataset, feature selection, transformation, manifestation processes. The methodology is followed by section 3, which implements the unsupervised machine learning and the threshold-based clustering techniques. Lastly, section 4, which contains the evaluation of results and conclusion.

2. Related Work

Plenty of research work utilizes ML to cluster, classify, and monitor payload transfer among IoT devices and even expose their security vulnerabilities during manufacturing processes. Moreover, in critical domains such as the healthcare system, availability and reliability are major security requirements.

P. Kitsos et al., in [7] showed that IoT environment operating from normal to extreme circumstances has a significant impact on reliability. Nodes were evaluated under a range of threats using the k-means clustering algorithm for malicious detection, according to a study in [8]. With supervised training, the findings demonstrated successful detection.

In [9], G. Guofei et al. implemented an unsupervised clustering technique to distinguish between botnets and benign traffic. The results showed real-world botnets such as IRC-based, HTTP-based and P2P botnets could be detected with a low false-positive rate.

In another study [10], S. Zhao et al. demonstrated that unsupervised algorithms could suffer performance degradation due to the high dimensionality of data. As a result, principal component analysis (PCA) was incorporated to increase detection performance.

In another study in [11], P. Gogoi et al. illustrated that the utilization of supervised and unsupervised techniques to investigate existing ML approaches for detecting attacks in network traffic was carried out. The results showed that unsupervised detection was better, but it also revealed a sizeable false-positive rate.

Further study in [12] was carried out by C.C. Aggarwal et al.; clustering was accomplished using k-means and BIRCH in a distributed IoT site. Clusters that are within three times the standard deviation gap are grouped into one class. Results revealed an accuracy of 96.5 % with a 0.2 % false-positive rate.

It is required to study the effect of crypto-ransomware on medical data since it can directly affect patients' medical conditions and disables the caregivers to provide exemplary service at the right time. Therefore, after providing fulfilling the Health Insurance Portability and Accountability Act (HIPAA) standards and signing a data user agreement (DUA), the MIMIC III dataset became accessible and was obtained to fulfil our research purpose.

3. Experimental Procedure

Once the dataset was obtained, a number of features were selected and transformed to be manifested by the payload ransomware. Before the clustering process, similar preprocessing steps had to be implemented to prepare the data for the final phase of clustering. Each step presented in the flow chart above is described in detail in the following sections.

3.1 Dataset

MIMIC III is a vast and freely accessible medical database consisting of several medical measurements, tests, procedures and reports made at the bedside through two different clinical information systems CareVue and MetaVision. The dataset as a whole consists of 29 different tables; however, the table used in this research is the "CHARTEVENTS", in which all the charted data for all patients stay in the intensive care unit stored in a comma-separated (CSV) file format.

A number of the tests were recorded. However, below other names consistent with their associated clinical information system. Once inspecting the dataset, it absolutely was found that some tests were monitored by one amongst the clinical information systems, whereas the opposite did not. Therefore, it was necessary to manually undergo the dataset to settle on the features, particularly those under shorthand terms.

The table stores different types of examinations and their values conducted on patients and some administrative data such as identification numbers, etc. The total number of different tests is 353 tests. Considering all of them in this study wouldn't aid in the investigation process. Therefore, only the essential features were selected out of the whole set described in the next section.

3.3 Feature Selection

Each examination in this dataset had a separate number of occurrences (i.e. how many times it was conducted). Therefore, we have picked the most frequent tests in addition to static information about the patient to identify for who these tests were carried out and their caregivers. The frequency of different examination tests ranges from 1 to 8263. Therefore, all tests that were carried out more than 5000 times in addition to their date and time of conduction, patient's intensive care identification number and caregiver identification number were considered as our features. The most frequent medical tests or features within the dataset are shown in the table below:

Table 1 Features Selected

Test	Frequency

Heart Rate	8263
Respiratory Rate	8213
O2 Saturation Pulseoximetry	8148
Non-invasive blood pressure systolic	5526
Non-invasive blood pressure diastolic	5524
Non-invasive blood pressure means	5559

3.4. Data Transformation

Currently, the dataset consists of nine features; date, time, patient's intensive care unit identification number, heart rate, non-invasive blood pressure systolic, non-invasive blood pressure diastolic, non- invasive blood pressure means, respiratory rate and O_2 saturation pulseoximetry. The choice of these tests doesn't depend entirely just on their frequency; it has to have the capability to be captured by devices remotely to stimulate the environment of IoMT.

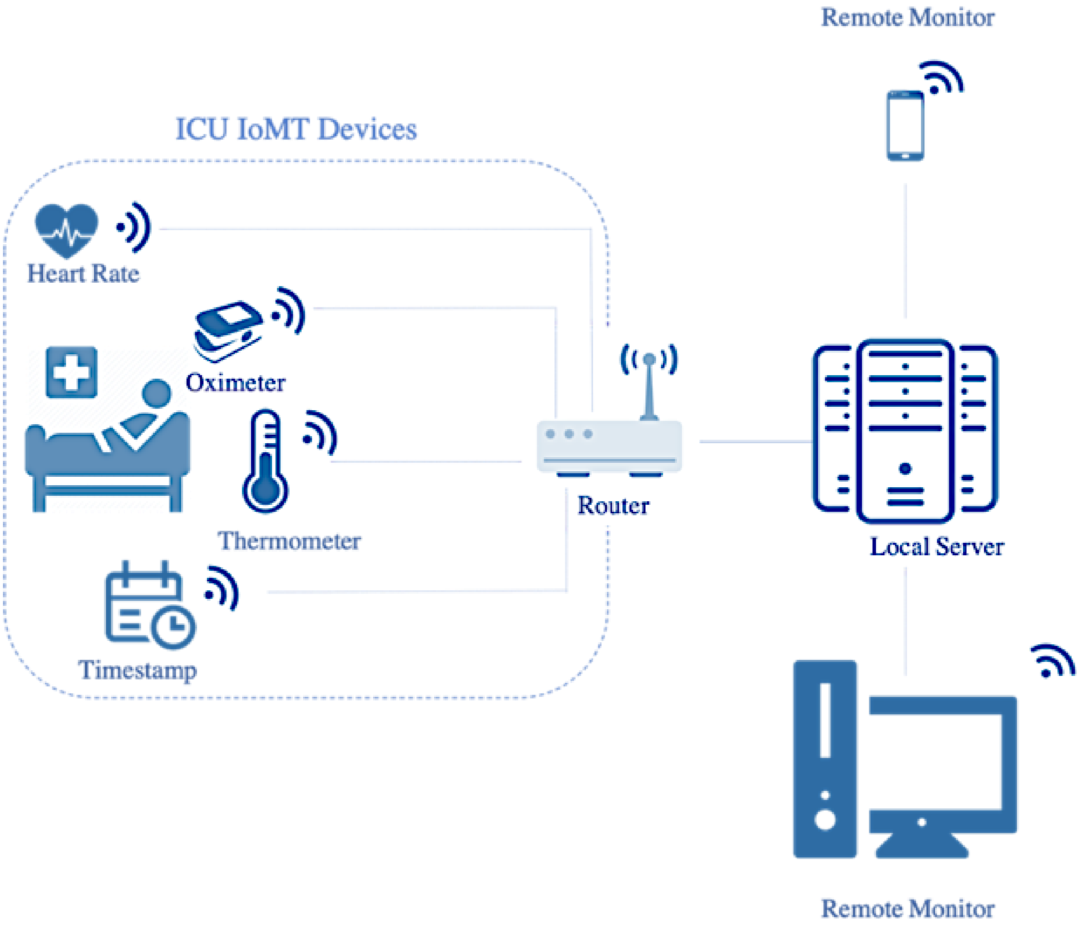


Figure. 2 : Data Capturing Mechanism

The medical records in the raw dataset were stored as the features string name and its value next to it inthe same field; therefore, it must be transformed into a transpose tabular format.

$$M = M^{T} \qquad (1)$$

This step was time-consuming because all of the patients were tested for the same physiological test. Some had only one feature tested; others had only three, and others had all of the features tested. For those who had certain features tested innumerable times, the means of these values were computed forevery day instead.

3.5 Data Manifestation

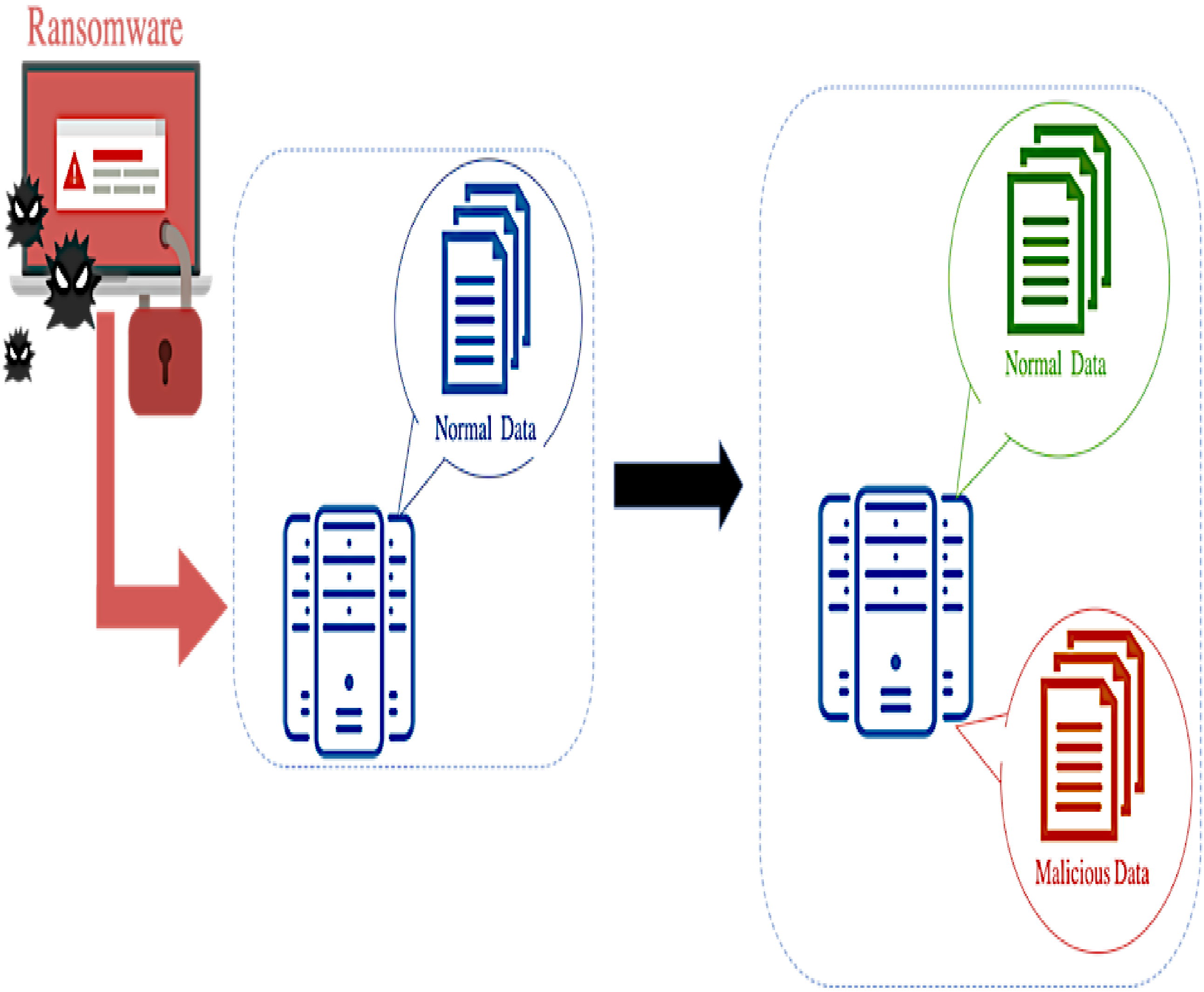


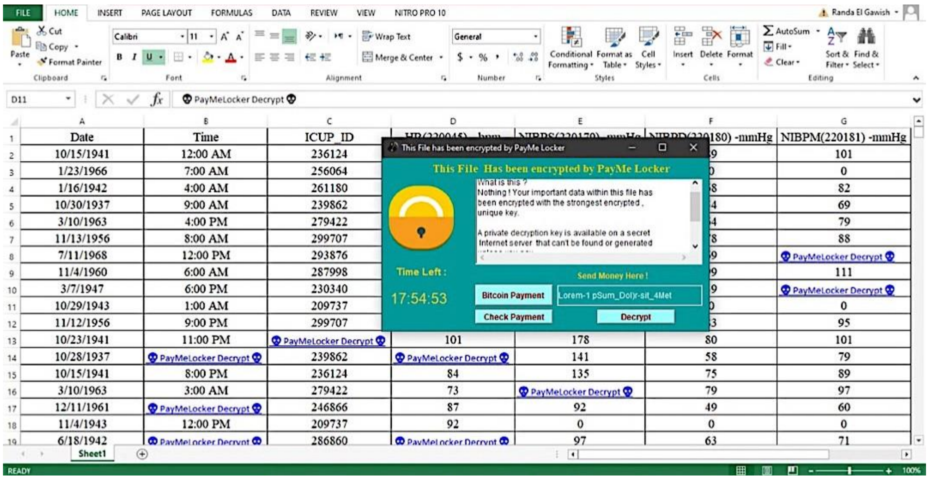
Figure. 3 : Data Manifestation Overview

After converting the dataset into transpose, it was split into two portions randomly. 51% of the dataset proceeded with the manifestation process, while 49% were considered normal records. The larger dataset is fed into the payload ransomware manifestation function using the following algorithm to encrypt the stored feature value.

Algorithm 1: Payload Ransomware Manifestation

```
Require: Data in Transpose;
function INFECT (file)
    Data ← file
    for each row € Data do
        for each feature € Data do
            Generate state      /* 0 or 1 state
            if state = 0 then
                feature++
            else
                Data [row][feature] = Encrypt the value stored in the feature
            end if
        end for
    end for
end function
```

The dataset after manifestation rejoins the normal dataset and the tuples become randomly distributed.A tuple is considered a rectangular block of the previously mentioned features.



D11 ☠ PayMeLocker Decrypt ☠

	A	B	C	D	E	F	G
1	Date	Time	ICUP_ID	HR(220045) - bpm	NIBPS(220179) -mmHg	NIBPD(220180) -mmHg	NIBPM(220181) -mmHg
2	10/15/1941	12:00 AM	236124			9	101
3	1/23/1966	7:00 AM	256064			0	0
4	1/16/1942	4:00 AM	261180			8	82
5	10/30/1937	9:00 AM	239862			4	69
6	3/10/1963	4:00 PM	279422			4	79
7	11/13/1956	8:00 AM	299707			8	88
8	7/11/1968	12:00 PM	293876			9	☠ PayMeLocker Decrypt ☠
9	11/4/1960	6:00 AM	287998			9	111
10	3/7/1947	6:00 PM	230340			9	☠ PayMeLocker Decrypt ☠
11	10/29/1943	1:00 AM	209737			0	0
12	11/12/1956	9:00 PM	299707			3	95
13	10/23/1941	11:00 PM	☠ PayMeLocker Decrypt ☠	101	178	80	101
14	10/28/1937	☠ PayMeLocker Decrypt ☠	239862	☠ PayMeLocker Decrypt ☠	141	58	79
15	10/15/1941	8:00 PM	236124	84	135	75	89
16	3/10/1963	3:00 AM	279422	73	☠ PayMeLocker Decrypt ☠	79	97
17	12/11/1961	☠ PayMeLocker Decrypt ☠	246866	87	92	49	60
18	11/4/1943	12:00 PM	209737	92	0	0	0
19	6/18/1942	☠ PayMeLocker Decrypt ☠	286860	☠ PayMeLocker Decrypt ☠	97	63	71

This File has been encrypted by PayMe Locker

This File Has been encrypted by PayMe Locker

What is this ?
Nothing ! Your important data within this file has been encrypted with the strongest encrypted , unique key.

A private decryption key is available on a secret Internet server that can't be found or generated

Time Left :
17:54:53

Send Money Here !
Bitcoin Payment
Lorem-1 pSum_Dol)r-sit_4Met
Check Payment
Decrypt

Figure. 4 : Dataset After Payload Ransomware Manifestation

3.6 Data Preprocessing

In any machine learning process, data preprocessing is that step in which data gets transformed, or encoded, to bring it to such a state that now the machine can quickly parse it. In other words, the algorithm can now easily interpret the data's features. The preprocessing phase consists of two procedures; handling encrypted values and data scaling.

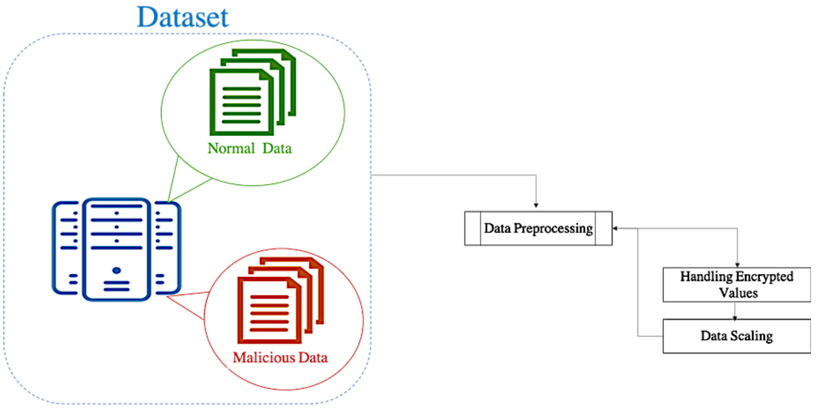


Figure. 5 : Data Preprocessing Module

3.6.1 Handling Encrypted value

The values were overridden using the “□ 💀PayMeLocker Decrypt □”💀signature, hiding its actual values. Which is in critical condition could lead to catastrophic results if not treated on time. The first step here was to replace the encrypted values with distinct large values within each column. This value is twice the maximum value of each feature. These values are represented in the table below:

Table 2 Features Maximum and Distinct Values

Feature n	Maximum Value	Distinct Value
Feature 1	26758	53516
Feature 2	23.5	47
Feature 3	299707	599414
Feature 4	200	400
Feature 5	265	530
Feature 6	193	386
Feature 7	197	394
Feature 8	100	200
Feature 9	114	228

3.6.2 Data Scaling

The next step in this phase is data scaling. Data scaling is a technique that normalizes the data values of features within a given dataset into a particular range. For example, after the encrypted values were replaced with numerical values, the data within each column were normalized to floating numbers ranging from 0 to 1. This step was implemented using the library of preprocessing and MinMaxScaler().

4. Clustering Implementation

In this work, three unsupervised machine learning techniques were used to compute the clustering process in addition to the threshold-based technique. The methods used were k-means, DBscan, mean shift clustering techniques show an outstanding performance whenever the features are dependent on each other; however, in this case, the features were independent. This section describes the methods used. Therefore, the threshold-based technique outperformed the unsupervised learning technique; the result of each method is discussed in detail in the section of evaluation and results.

4.1. Unsupervised Machine Learning Techniques

K-means is computationally fast, robust, easy to understand and produce a tighter cluster. However, k-means may not be suitable whenever data scales up and inefficient in non-linear data. In addition to its lack of consistency whenever there are overlapping data samples. The following figure shows the resulting clusters; the blue cluster indicates the number of normal records while the red ones are malicious. The normal ones are essentially identified as records that haven't been modified or overridden in any way. However, k-means assumed many infected records as normal ones and another large number of the normal ones as malicious. This is discussed in the evaluation and results section.

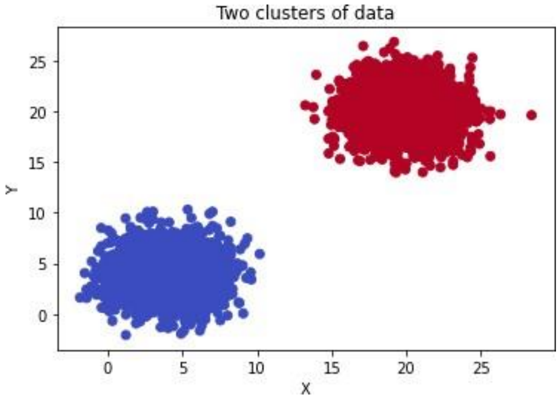


Figure. 6 : Clusters using K-Means

DBScan is excellent at outlier or noise detection and could be able to find a new cluster that is surrounded by distinct clusters. However, in IoMT, DBscan is not suitable for multiprocessor systems and with altering datasets, its clustering performance is not promising. Significantly if the density constantly varies and is too sparse [13]. Moreover, according to figure 7, DBscan has shown that the dataset contains noises or outliers, although it has not. In addition to half of the medical records were incorrectly labelled, although the signature is quite different from the measurements data type. The results of this algorithm are described in the evaluation and results section.

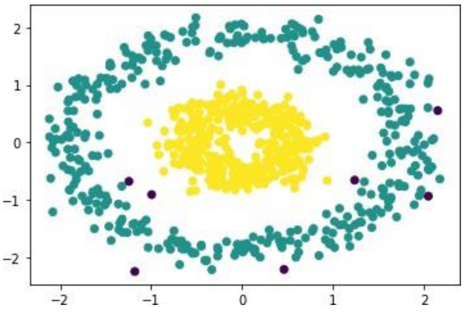


Figure. 7 : Outlier Detection using DBScan

The third ML technique used was the mean shift algorithm. Mean shift has shown a slightly better performance than DBscan algorithms but equivalent to k-means. Figure 7 shows the green cluster as the normal ones while the blue ones as the infected ones. However, as previously mentioned, there was a lotof incorrect labelling of the medical records. Mean shift is not the ultimate choice for data clustering with numerous features, so utilizing it with all of the 354 features won't enable correct labelling that could aid in further processes of classification or detection.

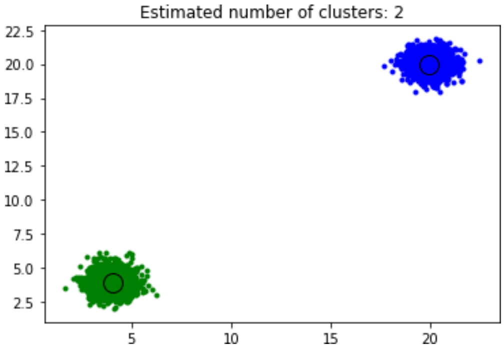


Figure. 8 : Clusters Using Mean Shift

4.2. Threshold-based technique

The research-based clustering technique proved excellent performance compared to the previous methods. The clustering technique that is used should be based on the problem it solves. The methods base its clustering decision depending on a threshold within each feature which was 1, except for the 5thfeature, which had a threshold of 0.43.

These values were obtained after the data scaling phase. The maximum values were all mapped to 1 except for the 5th feature, which had a maximum floating number of 0.43. These values are way more extensive than actual physiological measurements; therefore, they are easy to distinguish from the normal range. Each data sample that has a value equal to the threshold throughout its features is considered infected. Conversely, the data sample is deemed to be normal if all values were less than thethreshold stated within every feature.

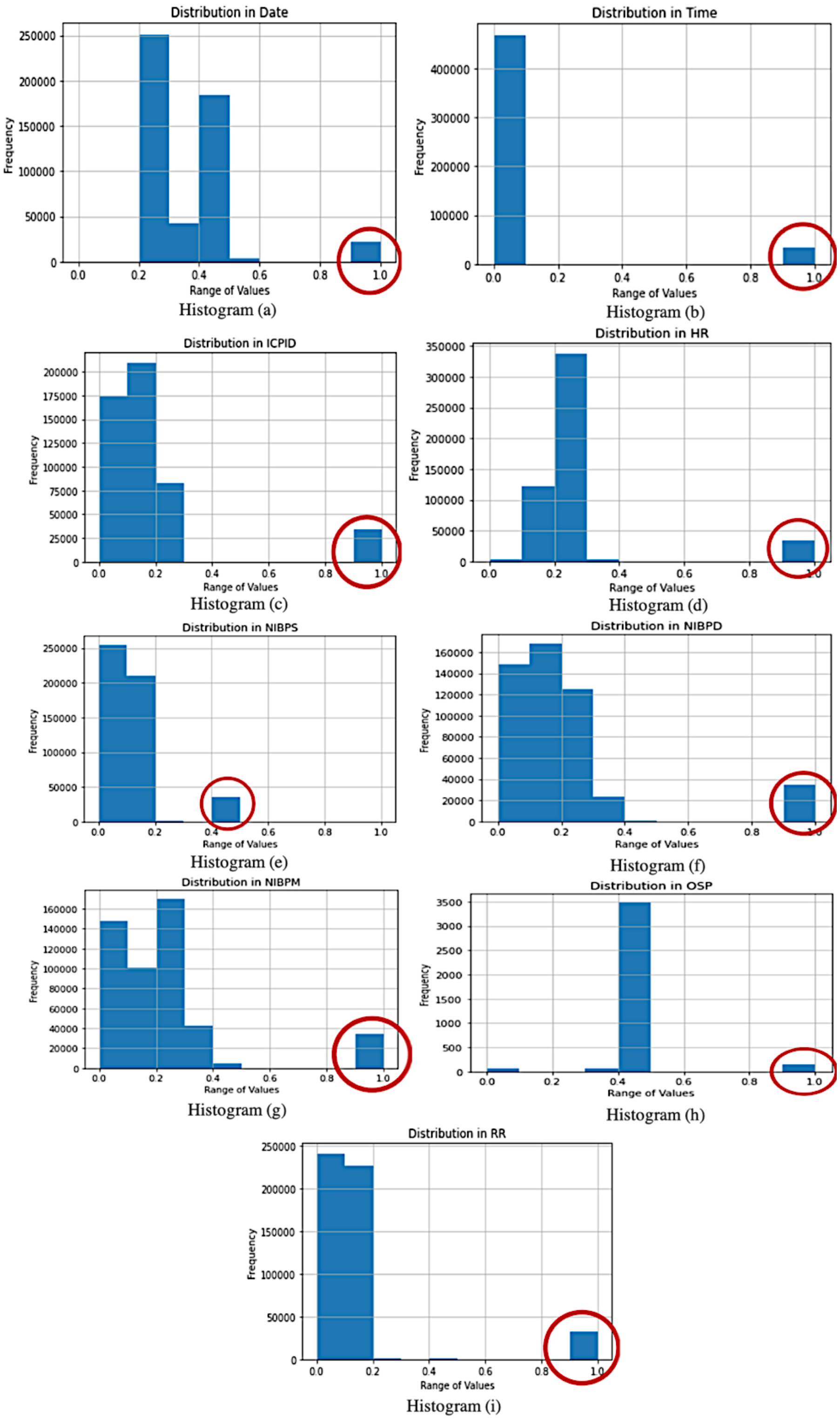
Algorithm 2: Label Dataset

```
Require: Dataset after normalization
function Label (Normalized Dataset)
    $X \leftarrow$ Normalized Dataset
    for each row € X do
        for each feature € X do
            if feature = 5 then
                if X[row][ feature] =0.43 then
                    Create Label
            else
                if X[row][feature] = 1 then
                    Create Label
                end if
            end if
        end if
    end if
    end for
end for
end function
```

This algorithm had yielded nine clustering results. This could be seen as a truth table of zeros and 1s forall of these features for each record. The final clustering result, which is the label, is done by applying OR logic operator between the clustering results of all the features. This step produced accurate labelling of the dataset.

$$\boldsymbol{Label_i = f_0 + f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 + f_9} \qquad (2)$$

The following histograms show the distribution of the normal records and the infected ones. For instance, the first histogram shows the distribution between the tuples in the first feature within thenormal normalized values, while the malicious ones have the highest value of 1.



Histogram (a)

Histogram (b)

Histogram (c)

Histogram (d)

Histogram (e)

Histogram (f)

Histogram (g)

Histogram (h)

Histogram (i)

Figure. 9 : Distribution of Normal and Infected Tuples for the 9 Features

5. Evaluation and Results

5.1. Evaluation Metrics

The clustering system was evaluated by measuring the performance metrics: recall, precision and F1 score. Recall (R) and precision (p) are highly vivid performance metrics to evaluate the clustering performance when imbalanced clustering occurs. In addition to a precision and recall curve that highlights the values of these two metrics for each clustering technique.

Precision is the positive predictive value and outlines how good a model predicts the positive (anomaly) label. To calculate precision, we use the following formula:

$$\boldsymbol{Precision} = \frac{\boldsymbol{TruePositives}}{\boldsymbol{TruePositive?FalsePositives}} \quad (3)$$

The recall is the ratio between the number of true positive labels divided by the sum of the true positive values and the false negative values. To calculate recall, we use the following formula:

$$\boldsymbol{Recall} = \frac{\boldsymbol{TruePositives}}{\boldsymbol{TruePositive?FalseNegatives}} \quad (4)$$

A true positive is an outcome where the model *correctly* predicts the *positive* class. Similarly, a true negative is an outcome where the model *correctly* predicts the *negative* class. A false positive is an outcome where the model *incorrectly* predicts the *positive* class. And a false negative is an outcome where the model *incorrectly* predicts the *negative* class. F1 score is a measure of a test's accuracy. It depends on the values of precision and recall. To calculate the F1 score, we use the formula below:

$$\boldsymbol{F1} = \boldsymbol{2*precision*recall}\underline{\boldsymbol{(precision?recall)}} \quad (5)$$

Table 3 TP, TN, FP and FN Values of the clustering techniques implemented

Clustering Techniques	TP	TN	FP	FN
k-means	38658	36428	36018	38238
DBScan	38428	36208	36248	38458
Mean Shift	38628	36408	36048	38268
Proposed Clustering approach	76896	72465	0	0

Table 3 shows the number of TP values in k-means, DBScan and Mean Shift and the threshold-based technique. The clustering approach based on the threshold of all features correctly labels the medical records with zero errors. Table 4 maps the precision, recall and f1 score upon the obtained values from table 3.

Table 4 Evaluation Metrics Results

Clustering Techniques	Precision	Recall	F1 Score
k-means	51.70 %	50.30 %	50.90 %
DBScan	51.50 %	49.98 %	50.60 %
Mean Shift	51.70 %	50.2 %	50.90 %
Proposed Clustering approach	100 %	100 %	100 %

The following P-R curve shows the clustering performance of each of the previously implemented techniques upon the payload ransomware signature. As obviously seen , the threshold-based approach offers accurate labelling performance as it has considered all of the features state within each medical tuple.

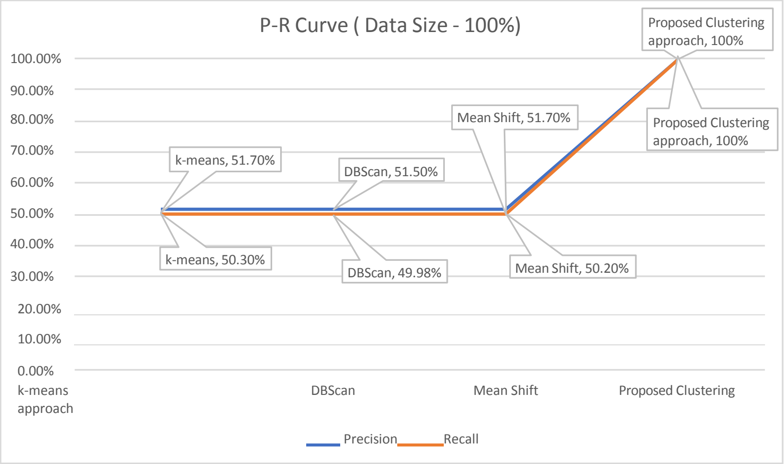


Figure. 10 : Precision-Recall Curve

In this work, we have clustered infected medical records using unsupervised machine learning techniques of k-means, DBScan and mean shift in contrast with the threshold-based approach. The payload ransomware affects the medical dataset with a clear evident signature overriding its true values under their associated features. Results have shown that the proposed clustering approach has successfully labelled the medical tuples with a 100 % precision score.

6. Conclusion

IoT or IoMT share the same paradigm and security vulnerabilities; however, the type of data transmittedwithin the application draws attention to specific attacks. For example, medical data has been targetedby ransomware extensively, especially during the pandemic of covid-19. This has caused thousands of hospitals to be affected in the U.S., causing them to lose track of the physiological levels of students, which in serious situations could lead to death.

According to the previously shown results in the clustering process of payload ransomware, the threshold-based method has shown the best accuracy of score 100 % compared to the unsupervised machine learning techniques of k-means, DBScan and mean shift. However, the scale of the dataset definitely affects the performance of the clustering process; therefore, it's highly vivid to consider the type of application and its corresponding data when it comes to choosing the suitable method for clustering the normal tuples from the malicious ones.

References

1. Cybersecurity Ventures, Global Ransomware Damage Costs Predicted To Reach $20 Billion (USD) By 2021. https://cybersecurityventures.com/global-ransomware-damage-costs-predicted-to-reach-20-billion-usd-by-2021/, 2019 (accessed 21 October, 2019).
2. MIT Technology Review, A wave of ransomware hits U.S. hospitals as coronavirus spikes. https://www.technologyreview.com/2020/10/29/1011436/a-wave-of-ransomware-hits-us-hospitals- as-coronavirus-spikes/, 2020 (accessed 29 October 2020).
3. National Public Radio, U.S. Hospitals Targeted In Rising Wave Of Ransomware Attacks Federal Agencies Say, https://www.npr.org/2020/10/29/928979988/u-s-hospitals-targeted-in-rising-wave-of-ransomware-attacks-federal-agencies-say, 2020 (accessed 29 October 2020).
4. R. Altawy , A. M. Youssef, Security tradeoffs in cyber physical systems: A case study survey onimplantable medical devices , IEEE Access 4(1) (2016) 959-979.
5. A. Johnson T. Shen , L. Lehman , M. Feng, M. Ghassemi , B. Moody , P. Szolovits , L. Celi , R.Mark , MIMIC-III, a freely accessible critical care database, Sci Data 3(160035) (2016) , 1-9.
6. PhysioNet , MIMIC-III Clinical Database Demo (version 1.4), https://doi.org/10.13026/C2HM2Q ,2019 (accessed 24 April 2019).
7. P. Kitsos, N. Sklavos, and A. G. Voyiatzis, Ring oscillators and hardware trojan detection in *Hardware Security and Trust*, Springer (2017) 169–187.
8. S. Papafotikas and A. Kakarountas, A Machine-Learning Clustering Approach for Intrusion Detection to IoT Devices, In: IEEE International Conference on South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference, 2019,p.1-6.
9. G. Guofei, R. Perdisci, J.zhang, W. Lee, BotMiner: Clustering Analysis of Network Traffic forProtocol- and Structure-Independent Botnet Detection, CCS'08 (2008) 139-154.
10. S. Zhao, W. Li, T. Zla, A. Zomaya, A Dimension Reduction Model and Classifier for Anomaly-Based Intrusion Detection in Internet of Things, In: IEEE International Conference on PervasiveIntelligence and Computing (PICom 2017), 2017, p.836-843.
11. P. Gogoi, B. Borah, D.K. Bhattacharyya, *Anomaly detection analysis of intrusion data using supervised & unsupervised approach,* Journal of Convergence Information Technology 5(1)

(2010)95-110.

12. C.C. Aggarwal, J. Han, J. Wang, P. S. Yu, A Framework for Clustering Evolving Data Streams,In: Proceedings of the 2003 VLDB Conference, 2003 p.81-92.
13. S. Dang, Performance Evaluation of Clustering Algorithm Using Different Datasets, IJARCSMS3(1) (2016) (167-173).
14. Ransomware Damage Report, Cybersecurity Ventures. https://cybersecurityventures.com/ransomware-damage-report-2017-5-billion/, 2017(accessed 18 May 2017).