Computer Science and Information Technologies
Vol. 5, No. 2, July 2024, pp. 176~185
ISSN: 2722-3221, DOI: 10.11591/csit.v5i2.p176-185

Clustering man in the middle attack on chain and graph-based blockchain in internet of things network using k-means

Sari Nuzulastri[1], Deris Stiawan[1], Hadipurnawan Satria[2], Rahmat Budiarto[3]
[1]Department of Computer Engineering, University of Sriwijaya, Palembang, Indonesia
[2]Department of Informatics Engineering, University of Sriwijaya, Palembang, Indonesia
[3]Department of Computer Science, College of Computing and Information, Al-Baha University, Al Bahah, Saudi Arabia

Article Info

Article history:

Received Dec 6, 2023
Revised May 27, 2024
Accepted Jun 4, 2024

Keywords:

Blockchain
Internet of things
K-means
Man in the middle
Network security

ABSTRACT

Network security on internet of things (IoT) devices in the IoT development process may open rooms for hackers and other problems if not properly protected, particularly in the addition of internet connectivity to computing device systems that are interrelated in transferring data automatically over the network. This study implements network detection on IoT network security resembles security systems from man in the middle (MITM) attacks on blockchains. Security systems that exist on blockchains are decentralized and have peer to peer characteristics which are categorized into several parts based on the type of architecture that suits their use cases such as blockchain chain based and graph based. This study uses the principal component analysis (PCA) to extract features from the transaction data processing on the blockchain process and produces 9 features before the k-means algorithm with the elbow technique was used for classifying the types of MITM attacks on IoT networks and comparing the types of blockchain chain-based and graph-based architectures in the form of visualizations as well. Experimental results show 97.16% of normal data and 2.84% of MITM attack data were observed.

Corresponding Author:

Deris Stiawan
Department of Computer Engineering, University of Sriwijaya
Indralaya, Ogan Ilir-30662, Palembang, Indonesia
Email: deris@unsri.ac.id

1. INTRODUCTION

The development of internet of things (IoT) as a smart device in several technologies [1]–[4] that changes the world with the development of internet networks [5]–[8] are seen in collecting data, and controlling tools to do certain things through the internet network. Self organization and communication using the cloud as a data storage medium are vulnerable to attacks because many devices are connected to the internet [9], [10]. Network security in IoT devices is used to protect data during the data transmission process to keep them safe because devices connected to IoT devices can open gaps for hackers and other problems [11]. Mallik *et al.* [12] and Nayak and Samaddar [13] explain about the type of man in the middle (MITM) attack that aims to retrieve information in a network protocol or secure sockets layer and transport layer security (SSL/TLS) MITM attack and the domain name system (DNS) spoofing attack that provides different data (data falsification) [14]. Choi, *et al.* [15] explain the blockchain-based MITM security system that detects MITM attacks by filtering, detecting, and comparing networks implemented on a network security system on the blockchain in the IoT.

Singh *et al.* [16], Li and Kassem [17] describe the distributed ledger technology (DLT) which is part of the blockchain that provides a decentralized data management system in storing and sharing data on every network transaction. Ferraro *et al.* [18] explain the directed acyclic graphs (DAGs) in the blockchain

architecture for the DLT can make transactions easier and more linear because the network is peer to peer [19]–[21]. It provides a detailed analysis of security attack patterns applied to IoT devices. The security system that exists on a decentralized blockchain that stores and shares data is a decentralized data management system [22] and peer to peer characteristics can hinder the improvement of blockchain technology in several aspects of life.

Blockchain technology is categorized into several parts based on the type of architecture that suits its use case. In the context of blockchain chain based and graph based, there are two types of data structures used by blockchain to store transaction data and build evidence of consensus [23]. The chain-based blockchain has a data structure in each block forming a chain and it will continue to grow. In contrast, graph-based uses a random graph-shaped data structure and each transaction can be directly connected to several other transactions in the network whose use depends on the purpose of the blockchain being used [24].

The use of the k-means algorithm in the IoT network for grouping data according to their characteristics has been implemented such as in [25], [26] and show the accuracy in the clustering process of 99.94% with confusion matrix accuracy in the true negative section of 98.62%, true positive of 100%, false negative of 0.00% and false positive of 1.38%. Related research on DLT in IoT networks that had been carried out previously discussed the benefits of the data transmission transaction process [27], [28]. These studies explain the stochastic mechanism in the transaction process that existed in the blockchain architecture for DLT to make transactions faster and more stable using the Markov chain Monte Carlo (MCMC) algorithm, which was proven by a numerical balance of 25% on each transaction sent through the protocol.

In general, in security system processes of the IoT networks, it is very important to have an immutable transaction record to analyze a parasitic chain attack, which aims to see the resilience and security by using the MCMC algorithm in reducing parasitic chain attacks [28]. As for some research, it was found that the improvement process that focuses on the number of transactions called Tangle [29], [30] have proven that by using the tip selection algorithm (TSA) method, the level of confidence and sustainability were getting better along with the increase in the number of transactions. On the other hand, research in 2021 [31]–[33] explains that attacks on IoT networks have increased by up to 20% for the security level of the identification process in IoT networks integrated with blockchain technology. The use of the elliptic curve cryptography (ECC)-based algorithm was needed because of the privacy of the security protocol [10]. The development of IoT aims to connect data through the internet network in the issue of identity security (data privacy) [34] from various attacks such as MITM attacks that steal passwords, and personal identification numbers [35]. It generally estimates the theoretical complexity of attacks that allow for multiple combinations of increased MITM attacks [15], [36].

Therefore, it is necessary to analyze the improvement in the detection of attacks in producing a lower rate of misclassification of attacks so that the process of sending data in transmission is safe and integrated using the k-means method. This research discusses the comparison of the performance of blockchain chain-based and graph-based transactions on data of MITM attack on IoT networks where the traffic features are extracted using principal component analysis (PCA) and clustered using the k-means method. The results then were displayed in the form of visualizations. The discussion in this research was as follows: section 2 discusses the proposed method in determining the data to be clustered. Section 3 provides the results of clustering data of the MITM attacks and section 4 provides conclusions and hopes for future research.

2. METHOD

In general, the steps in the research methodology used to assist in the preparation of this research required a clear framework in its stages. The research framework is shown in Figure 1, which consists of a literature review by reviewing research in recent years, followed by data preparation using a dataset of 550,000 data samples. Next is data preprocessing by performing feature extraction followed by testing, analyzing the results and drawing conclusions.

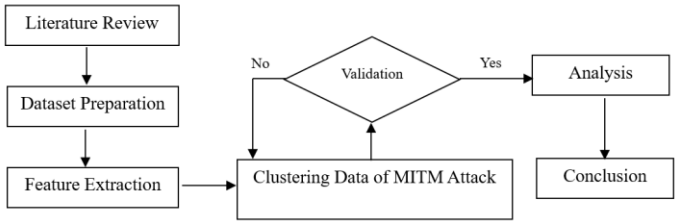

Figure 1. Research methodology

2.1. Feature extraction

Feature extraction is an important part of this process. In this study, datasets were taken from a journal data [29], which were then preprocessed using the PCA method to reduce the dimensions of the data without significantly reducing the characteristics of the data. The flow of data preprocessing was depicted in Figure 2. This preprocessing stage can be divided into two, i.e.: feature extraction process and feature selection process using PCA method that can reduce the dimensionality of data without significantly reducing the characteristics of the data [37], [38]. In this process the data was made using simpler features so that it could be analyzed and interpreted properly in order to produce accurate and reliable data using several techniques including data cleaning, data transformation and data reduction. The processed data was saved in csv format. Figure 3 shows an example of a dataset that had been saved in csv format.

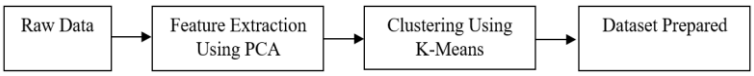

Figure 2. Dataset preprocessing flow

sender	timestamp	status	height	
io1jafqlvntcxgyp6e0uxctt3tljzc3vyv5hg4ukh,	2019-04-22T13:18:10Z,	1	4001	ae54fd89ee871ecca6dce
io159fv8mu9d5djk8u2t0flgw4yqmt6fg98uqjka8,	2019-04-22T13:18:20Z,	1	4002	346a29a7d4a5715406664
io1ar5l5s268rtgzshltnqv88mua06ucm58dx678y,	2019-04-22T13:18:30Z,	1	4003	f5d24ce169074b18d7b67
io1z7mjef7w528nasnsafan0rp6yuvkvq405l6r8j,	2019-04-22T13:18:40Z,	1	4004	cfa5c4f8e8103781bd227c
io1cup9k8hl8fp40vrj29ex8djc346780dk223end,	2019-04-22T13:18:50Z,	1	4005	c1481a15bcab567411c78(
io1wv5m0xyermvr2n0wjx2cjsqwyk863drdl5qfyn,	2019-04-22T13:19:00Z,	1	4006	54c12d42e42d0367981f5
io1u5xy0ecnrjrdkzyctfqh37lgr5pcfzphgqrdwt,	2019-04-22T13:19:10Z,	1	4007	3868673975feb76c6fac9b
io1nz40npqa3yvek4zdasmqaetl2j4h6urejfkera,	2019-04-22T10:31:30Z,	1	3001	4ea009983c9c70e852416
io12yxdwewry70gr9fs6fphyfaky9c7gurmzk8f4f,	2019-04-22T10:31:40Z,	1	3002	827d1c5f0c9369fa097076
io1u5xy0ecnrjrdkzyctfqh37lgr5pcfzphgqrdwt,	2019-04-22T10:31:50Z,	1	3003	44a0b8adb9637fad0f3ce
io14vmhs9c75r2ptxdaqrtk0dz7skct30pxmt69d9,	2019-04-22T13:19:20Z,	1	4008	dc8bf71fd45f55b2d5e8cf
io1gf08snppu2a2wfd50pjas2j6q2kcxjzqph3pep,	2019-04-22T10:32:00Z,	1	3004	a0e96aaa4e2238b1c63b8
io10kyvvzu08pjeylymq4umknjal25ea3ptfknrpf,	2019-04-22T13:19:30Z,	1	4009	c44760ddae88ed5f563e1
io159fv8mu9d5djk8u2t0flgw4yqmt6fg98uqjka8,	2019-04-22T10:32:10Z,	1	3005	a32d4d2d1ed1241cc465a
io12yxdwewry70gr9fs6fphyfaky9c7gurmzk8f4f,	2019-04-22T13:19:40Z,	1	4010	ede2a4856485c202f8c22
io14vmhs9c75r2ptxdaqrtk0dz7skct30pxmt69d9,	2019-04-22T10:32:20Z,	1	3006	a46ee2b1391eeae74d1f
io1gh7xfrsnj6p5uqgjpk9xq6jg9na28aewgp7a9v,	2019-04-22T13:19:50Z,	1	4011	f78185e3a739dc92e80b0
io1ar5l5s268rtgzshltnqv88mua06ucm58dx678y,	2019-04-22T10:32:30Z,	1	3007	543d1ca83a2fb7f62a185c
io1xsx5n94kg2zv64r7tm8vyz9mh86amfak9ka9xx,	2019-04-22T13:20:00Z,	1	4012	41c8b0aa29a6b35b85da7
io1qs785af9k9xf3xgd6vut7um9zcthtrvsn2xap2,	2019-04-22T10:32:40Z,	1	3008	bd91dc9b40a074f83ff93b

Figure 3. Research dataset in CSV format

2.2. Clustering with k-means

Stages of clustering with the k-means method is a grouping with a specified number of clusters using different cluster shapes [39], [40]. The MITM types are grouped in the form of sample data that has a lot in common with each other. The flow chart of the working system can be seen in Figure 4.

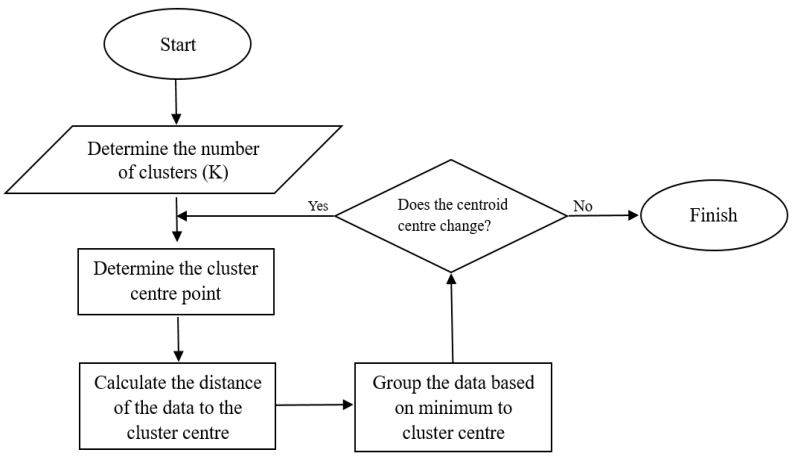

Figure 4. K-Means clustering flowchart

Determining the number of clusters at each center point (centroid) by presenting the cluster, the centroid value can be found using the formula in (1).

$$c = \sum_{i=1}^{n} \frac{x_i}{n} \quad (1)$$

Where, c is centroid value, x_i is point value/the i-th object, n is number of objects. The formula in (1) can be rewritten as (2).

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

where, μ_k is centroid of the k-th cluster, x_q is the q-th object from the k-th cluster, and N_k is number of data (samples) from the k-th cluster.

2.3. Confusion matrix calculation

The proposed method's performances are measured, in terms of accuracy, sensitivity, precision, and F1 score using a confusion matrix. Confusion matrix has four values, i.e.: True positive (TP), false positive (FP), true negative (TN), and false negative (FN). Accuracy describes how accurate the model is in classifying correctly. it can be calculated by dividing the number of correct predictions by the total number of predictions made, the accuracy calculation uses (3).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

The correctly predicted precision can be calculated by dividing the number of positive prediction results by the number of positive predictions using (4).

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Sensitivity measures how good the model is at identifying positive classes by dividing the number of positive predictions by the total number of positive cases as in (5).

$$Sensitivity = \frac{TP}{TP+FN} \quad (5)$$

F1 score provides a balanced average value between sensitivity and precision and expressed as (6).

$$F1\ score = \frac{2*(precision*sensitivity)}{precision+sensitivity} \quad (6)$$

3. RESULTS AND DISCUSSION

This section presents the results of the experiments of the MITM attack on the blockchain of the IoT network. Results of the feature data extraction process used in the clustering process is discussed first, followed by the clustering result itself. Since the clustering is done with k-means, it is silhouette score is also analyzed to determine the quality of the clusters. After that, the evaluation result is discussed.

3.1. Feature extraction

The PCA was used to extract and compress the dataset. This stage was carried out to select features that were used for clustering. Initially the raw data consists of 16 blockchain features. Some of the features were dropped because they are deemed unsuitable to be used with k-means. Also, some of the blockchain features must be first transformed into numerical forms. This leaves the number of features down to seven. Then all the data are normalized before being fed to the PCA which reduces the number of features to three. With the dimension reduced to three, the dataset can be easily visualized with 3D graphs. Figure 5 shows the features that were used before and after the PCA process. PCA generates new features that are a linear combination of actual features, as such the resulting features have no associated meaning with the actual blockchain.

	nonce	gas_limit	gas_price	sender	timestamp	heigh	gas_consumed
0	0	0	0.0	-4316393735033259258	1.555939e+09	4001	0
1	0	0	0.0	996546070771703395	1.555939e+09	4002	0
2	0	0	0.0	2620931106237483442	1.555939e+09	4003	0
3	0	0	0.0	-4834417782576484420	1.555939e+09	4004	0
4	0	0	0.0	7707969396558710493	1.555939e+09	4005	0

	PCA1	PCA2	PCA3
0	-1.363132	2.008574	-0.600256
1	-1.430676	1.971390	0.312096
2	-1.451324	1.960015	0.591040
3	-1.356532	2.012176	-0.689213
4	-1.515992	1.924404	1.464599

Figure 5. Blockchain features used for clustering

3.2. Clustering results with k-means

To measure the quality of the similarities within clusters and the differences between clusters as the result of clustering using k-means method its silhouette score is calculated. This score uses a measurement range of [-1,1] which means the higher the score of the silhouette, the more optimal number of clusters. The result of the quality test measurement with silhouette score with six clusters was 0.417. This score indicates that the K- Means was able to create distinct enough clusters while perhaps not the best possible.

Figure 6 shows the silhouettes of each cluster with the vertical line marking the average silhouette score. The cluster heights in the figure denote the variation of the nodes within each cluster. Most clusters have consistent height, except cluster 4. The maximum scores of each cluster are in fact close to 0.7, which is considered strong. But some clusters have negative scores, notably cluster 1, 2 4 and 5. These negative scores indicate outliers but their existences are still minimal. Nevertheless, they are responsible for reducing the average down to 0.4 even though all clusters have maximum scores above 0.6.

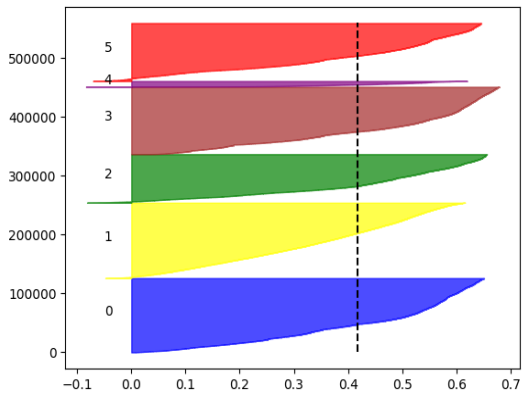

Figure 6. Silhouette score of the clustering

The clustering with k-means method produced a total of six clusters, shown in Figure 7 where nodes belonging to cluster 0, cluster 1, cluster 2, cluster 3, cluster 4 and cluster 5 are marked with color blue, yellow, green, brown and red respectively. As can be seen in the figure, most clusters have notably clear boundaries and contain nodes that are all close together with only a few outlier nodes. Cluster 4 (purple) though, has more spread-out nodes. Compared to other clusters, cluster 4 also has the least amount of nodes. This is consistent with the previously discussed silhouette plot where it was the only cluster with low height or score variations.

To help understanding the clustering result, Figure 8 shows the parallel coordinates plot of each cluster against all of the features. Using this plot, the relation of each feature of data nodes and the cluster they belong to can be analyzed. Note that the values in y axis are normalized, hence only their relative values are meaningful for the analysis. Also note that all six subfigures are scaled differently, their maximum values in the y axis are different and must be considered when comparing one cluster to the others.

In the Figure 8, cluster 0 (blue) and cluster 2 (green) appear to be very similar, only differing at sender, where cluster 0 has values around 0 and cluster has values around 1. Furthermore, values of feature gas_price on both clusters gather in two groups, one group near zero and another group near six. These two clusters are the only one exhibiting this trait. Compared to other clusters, the other distinct traits are the very low values of timestamp, height and gas_consumed.

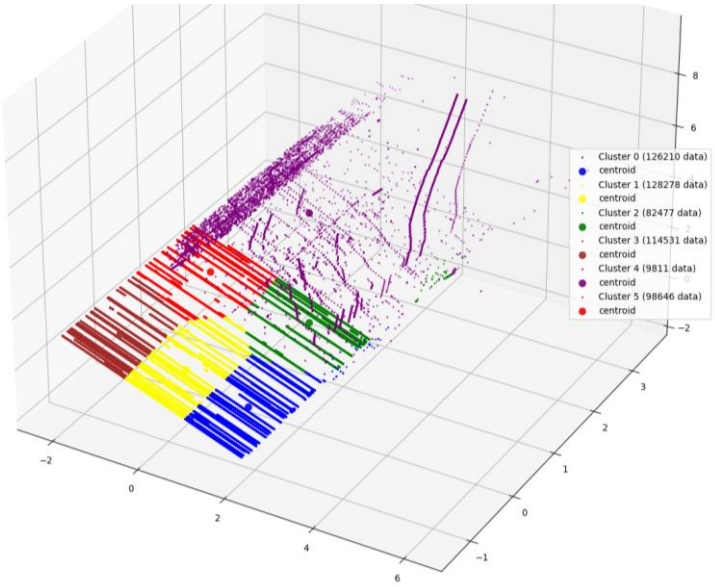

Figure 7. Clustering result

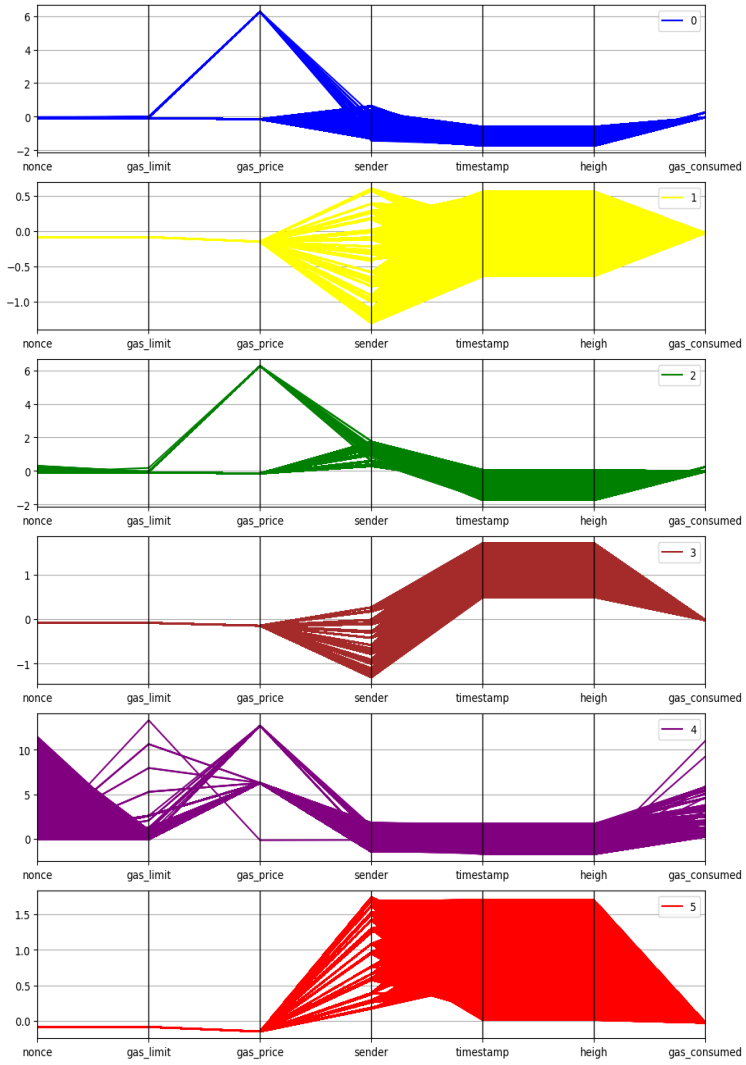

Figure 8. Mapping clusters to each feature

Cluster 1 (green) in particular is more compressed than other clusters, as signified by it is figure's y axis max value of 0.5, while other clusters max values are as high as 12. Generally speaking, all features of this cluster are near zero. Feature nonce, gas_limit, gas_price and gas_consumed are the most compressed, with all nodes having values nearing 0. feature sender has the most spread-out values, ranging from -1.5 to 0.5 while timestamp and height are somewhere in between.

Cluster 3 (brown) and cluster 5 (red) are actually quite similar even though their general shapes appear different. Just like most clusters the nodes in these clusters have nonce, gas_limit, gas_price and gas_consumed values near zero. The rest of the features though are notably different. The sender values of cluster 3 are more compressed than those of cluster 5. On the contrary, the values of timestamp and height of cluster 5 are more compressed and on the higher side in comparison to cluster 3.

Cluster 4 is the most distinctive among the six clusters. It is sender, timestamp and height values are quite similar to other clusters, in that all the values are near zero. But it has multiple groups of values for gas_price, gas_consumed and gas_limit. Also, its nonce values are the most spread out, ranging from zero to around 11. Another notable distinction is the multiple appearance of solitary values of the gas_limit feature which may indicate outlier nodes within the cluster. Determination of the cluster class based on the similarity of features in the clustering process on the blockchain includes several aspects such as transaction time which is a significant feature because it can identify at a certain time, transaction size where data grouping is based on the size of the data to be transferred, transaction security in identifying groups based on security characteristics (transaction security attributes such as digital signatures).

3.3. Validation result

The following are the results of simulation experiments from scenarios focused on MITM attacks, as for attacks carried out by changing the value in packet data. Based on the number of validation results in the training and testing phases of the dataset with 550,000 data samples. The data visualization in Figure 9 shows that the normal data (represented by blue) is 97.16% and MITM attacks are 2.84% (represented by orange), which means that normal data amounted to 534,380 data and as much as 15,620 data were MITM attacks.

Confusion matrix usually uses training data to train the proposed model and measure the performance of the clustering algorithm on the testing data. The following parameters were used to measure the performance, namely TP, FP, TN and FN. Then, the results of the Confusion Matrix calculation can measure how accurate the results of the Man in the Middle attack detection. Figure 10 displays the confusion matrix observations.

For validation purposes, training data of 80% and the testing data of 20% are used, and obtained an accuracy value of 99.78%. Table 1 shows the confusion matrix using 80% off the testing data. The use confusion matrix in the use of K-Mean's method is to show the level of accuracy of the prediction results that have been done in seeing the accuracy value of the data labeling that has been done.

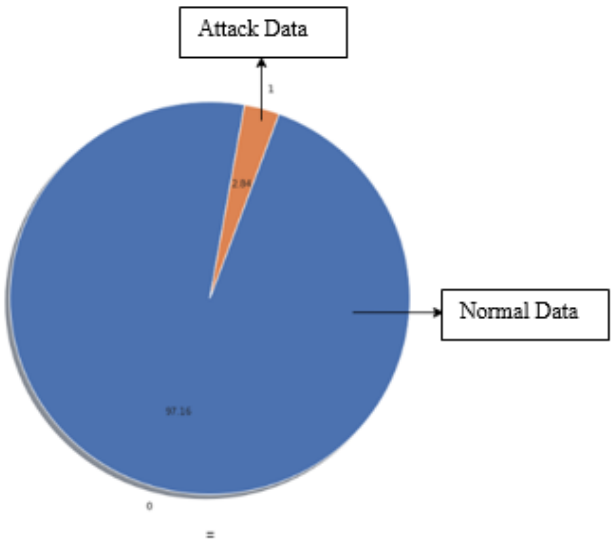

Figure 9. Visualization data transaction

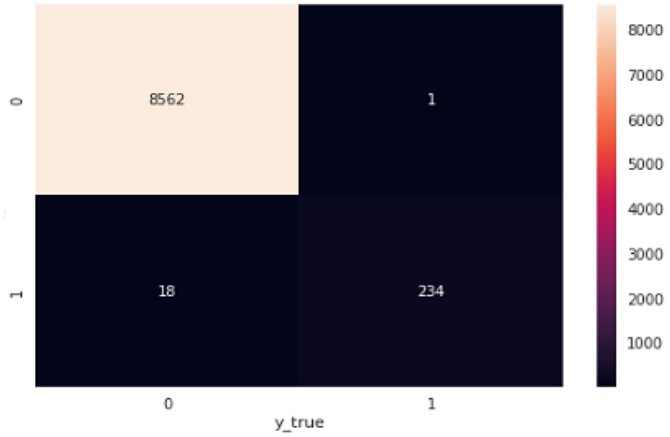

Figure 10. Confusion matrix display

Table 1. Confusion matrix using 80% testing data

Measurement	Value
True Positive (TP)	8,562
True Negative (TN)	234
False Positive (FP)	1
False Negative (FN)	18

Based on the results obtained using the k-means method which shows the advantages in identifying patterns and finding data for those tested. This is in accordance with the advantages of k-means, namely simplicity and efficiency. In addition, k-means is easily applied to large data and has better data computation time efficiency than other methods, while the disadvantage is that it must determine the initial number of clusters (k value). In this study, the determination of the initial cluster value (k) uses the silhouette score technique in clustering.

4. CONCLUSION

The PCA method used in feature extraction from incoming transaction data on the IoT network, reduces the number of features from 16 to 3 features to build a classification model in the clustering process. The clustering process with the k-means method implemented on the IoT network was carried out by performing an extraction process on the MITM attack data types. The results of the clustering analysis using the k-means method with 6 clusters in the transaction process with a silhouette score were 0.417. The detected Normal data was 97.16%, while the MITM attacks data was 2.84%. In the future, it is hoped that newly available datasets on the blockchain can be applied to get different features and characteristics using the implementation of the GMM clustering method and spherical k-means clustering to see better results and visualization. Other clustering methods can also be explored, especially methods that are derived from k-means but with more suitable characteristics to be used with the blockchain dataset.

REFERENCES

[1]. S. Kumar, P. Tiwari, and M. Zymbler, "Internet of things is a revolutionary approach for future technology enhancement: a review," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0268-2.
[2]. "Number of internet of things (IoT) connected devices worldwide from 2019 to 2030, by vertical," Statista. [Online]. Available: https://www.statista.com/statistics/1194682/iot-connected-devices-vertically/
[3]. "Internet of things (IoT) annual revenue from 2020 to 2030, by region," Statista. [Online]. Available: https://www.statista.com/statistics/1194715/iot-annual-revenue-regionally/
[4]. S. Sinha, "State of IoT 2023: Number of connected IoT devices growing 16% to 16.7 billion globally," IoT Analitics. [Online]. Available: https://iot-analytics.com/number-connected-iot-devices/
[5]. I. Mistry, S. Tanwar, S. Tyagi, and N. Kumar, "Blockchain for 5G-enabled IoT for industrial automation: A systematic review, solutions, and challenges," *Mechanical Systems and Signal Processing*, vol. 135, Jan. 2020, doi: 10.1016/j.ymssp.2019.106382.
[6]. A. Vaghani, K. Sood, and S. Yu, "Security and QoS issues in blockchain enabled next-generation smart logistic networks: A tutorial," *Blockchain: Research and Applications*, vol. 3, no. 3, Sep. 2022, doi: 10.1016/j.bcra.2022.100082.
[7]. T. Rathod *et al.*, "Blockchain for future wireless networks: a decade survey," *Sensors*, vol. 22, no. 11, May 2022, doi: 10.3390/s22114182.

[8]. E. Borgia, "The internet of things vision: key features, applications and open issues," *Computer Communications*, vol. 54, pp. 1–31, Dec. 2014, doi: 10.1016/j.comcom.2014.09.008.
[9]. J. Zhou, Z. Cao, X. Dong, and A. V. Vasilakos, "Security and privacy for cloud-based iot: challenges, countermeasures, and future directions," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 26–33, 2017.
[10]. S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Computer Networks*, vol. 76, pp. 146–164, Jan. 2015, doi: 10.1016/j.comnet.2014.11.008.
[11]. O. Eigner, P. Kreimel, and P. Tavolato, "Detection of man-in-the-middle attacks on industrial control networks," in *2016 International Conference on Software Security and Assurance (ICSSA)*, IEEE, Aug. 2016, pp. 64–69. doi: 10.1109/ICSSA.2016.19.
[12]. A. Mallik, A. Ahsan, M. M. Z. Shahadat, and J.-C. Tsou, "Man-in-the-middle-attack: Understanding in simple words," *International Journal of Data and Network Science*, pp. 77–92, 2019, doi: 10.5267/j.ijdns.2019.1.001.
[13]. G. Nath Nayak and S. Ghosh Samaddar, "Different flavours of man-in-the-middle attack, consequences and feasible solutions," in *2010 3rd International Conference on Computer Science and Information Technology*, IEEE, Jul. 2010, pp. 491–495. doi: 10.1109/ICCSIT.2010.5563900.
[14]. B. Bhushan, G. Sahoo, and A. K. Rai, "Man-in-the-middle attack in wireless and computer networking-A review," in *2017 3rd International Conference on Advances in Computing, Communication and Automation (ICACCA) (Fall)*, IEEE, Sep. 2017, pp. 1–6. doi: 10.1109/ICACCAF.2017.8344724.
[15]. J. Choi, B. Ahn, G. Bere, S. Ahmad, H. A. Mantooth, and T. Kim, "Blockchain-based man-in-the-middle (MITM) attack detection for photovoltaic systems," in *2021 IEEE Design Methodologies Conference (DMC)*, IEEE, Jul. 2021, pp. 1–6. doi: 10.1109/DMC51747.2021.9529949.
[16]. S. Singh, A. S. M. S. Hosen, and B. Yoon, "blockchain security attacks, challenges, and solutions for the future distributed IoT network," *IEEE Access*, vol. 9, pp. 13938–13959, 2021, doi: 10.1109/ACCESS.2021.3051602.
[17]. J. Li and M. Kassem, "Applications of distributed ledger technology (DLT) and Blockchain-enabled smart contracts in construction," *Automation in Construction*, vol. 132, Dec. 2021, doi: 10.1016/j.autcon.2021.103955.
[18]. P. Ferraro, C. King, and R. Shorten, "On the stability of unverified transactions in a DAG-based distributed ledger," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3772–3783, Sep. 2020, doi: 10.1109/TAC.2019.2950873.
[19]. J. Sedlmeir, H. U. Buhl, G. Fridgen, and R. Keller, "The energy consumption of blockchain technology: beyond Myth," *Business and Information Systems Engineering*, vol. 62, no. 6, pp. 599–608, Dec. 2020, doi: 10.1007/s12599-020-00656-x.
[20]. S. Kably, M. Arioua, and N. Alaoui, "Lightweight direct acyclic graph blockchain for enhancing resource-constrained IoT environment," *Computers, Materials and Continua*, vol. 71, no. 3, pp. 5271–5291, 2022, doi: 10.32604/cmc.2022.020833.
[21]. R. Paulavičius, S. Grigaitis, A. Igumenov, and E. Filatovas, "A decade of blockchain: review of the current status, challenges, and future directions," *Informatica*, vol. 30, no. 4, pp. 729–748, Jan. 2019, doi: 10.15388/Informatica.2019.227.
[22]. A. Abdelmaboud *et al.*, "Blockchain for IoT applications: taxonomy, platforms, recent advances, challenges and future research directions," *Electronics*, vol. 11, no. 4, Feb. 2022, doi: 10.3390/electronics11040630.
[23]. Q. Zhu, J. Pei, X. Liu, and Z. Zhou, "Analyzing commercial aircraft fuel consumption during descent: A case study using an improved K-means clustering algorithm," *Journal of Cleaner Production*, vol. 223, pp. 869–882, Jun. 2019, doi: 10.1016/j.jclepro.2019.02.235.
[24]. H. Y. Wu, X. Yang, C. Yue, H.-Y. Paik, and S. S. Kanhere, "Chain or DAG? Underlying data structures, architectures, topologies and consensus in distributed ledger technology: A review, taxonomy and research issues," *Journal of Systems Architecture*, vol. 131, Oct. 2022, doi: 10.1016/j.sysarc.2022.102720.
[25]. D. Stiawan *et al.*, "Ping flood attack pattern recognition using a k-means algorithm in an internet of things (IoT) network," *IEEE Access*, vol. 9, pp. 116475–116484, 2021, doi: 10.1109/ACCESS.2021.3105517.
[26]. M. J. Brusco, E. Shireman, and D. Steinley, "A comparison of latent class, K-means, and K-median methods for clustering dichotomous data.," *Psychological Methods*, vol. 22, no. 3, pp. 563–580, Sep. 2017, doi: 10.1037/met0000095.
[27]. S. Popov, O. Saa, and P. Finardi, "Equilibria in the tangle," *Computers and Industrial Engineering*, vol. 136, pp. 160–172, Oct. 2019, doi: 10.1016/j.cie.2019.07.025.
[28]. A. Cullen, P. Ferraro, C. King, and R. Shorten, "On the resilience of DAG-based distributed ledgers in IoT applications," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7112–7122, Aug. 2020, doi: 10.1109/JIOT.2020.2983401.
[29]. F. Guo, X. Xiao, A. Hecker, and S. Dustdar, "Characterizing IOTA tangle with empirical data," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/GLOBECOM42002.2020.9322220.
[30]. P. Kumar, R. Kumar, G. P. Gupta, R. Tripathi, A. Jolfaei, and A. K. M. Najmul Islam, "A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system," *Journal of Parallel and Distributed Computing*, vol. 172, pp. 69–83, Feb. 2023, doi: 10.1016/j.jpdc.2022.10.002.
[31]. B. K. Mohanta, D. Jena, S. Ramasubbareddy, M. Daneshmand, and A. H. Gandomi, "Addressing security and privacy issues of iot using blockchain technology," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 881–888, Jan. 2021, doi: 10.1109/JIOT.2020.3008906.
[32]. H. H. A. Emira, "Authenticating IoT devices issues based on blockchain," *Journal of Cybersecurity and Information Management*, pp. 35–40, 2020, doi: 10.54216/JCIM.010202.
[33]. Q. Fan, J. Chen, L. J. Deborah, and M. Luo, "A secure and efficient authentication and data sharing scheme for Internet of Things based on blockchain," *Journal of Systems Architecture*, vol. 117, Aug. 2021, doi: 10.1016/j.sysarc.2021.102112.
[34]. C. Fan, S. Ghaemi, H. Khazaei, and P. Musilek, "Performance evaluation of blockchain systems: a systematic survey," *IEEE Access*, vol. 8, pp. 126927–126950, 2020, doi: 10.1109/ACCESS.2020.3006078.
[35]. N. Sivasankari and S. Kamalakkannan, "Detection and prevention of man-in-the-middle attack in iot network using regression modeling," *Advances in Engineering Software*, vol. 169, Jul. 2022, doi: 10.1016/j.advengsoft.2022.103126.
[36]. A. Canteaut, M. Naya-Plasencia, and B. Vayssière, "Sieve-in-the-middle: improved MITM attacks," in *Annual Cryptology Conference*, 2013, pp. 222–240. doi: 10.1007/978-3-642-40041-4_13.
[37]. L. Smith, "A tutorial on PCSA," *Department of Computer Science, University of Otago.*, pp. 12–28, 2006.
[38]. H. Choi, H. Lee, and H. Kim, "Fast detection and visualization of network attacks on parallel coordinates," *Computers and Security*, vol. 28, no. 5, pp. 276–288, Jul. 2009, doi: 10.1016/j.cose.2008.12.003.
[39]. M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: a comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, Aug. 2020, doi: 10.3390/electronics9081295.
[40]. H. Qabbaah, G. Sammour, and K. Vanhoof, "Using k-means clustering and data visualization for monetizing logistics data," in *2019 2nd International Conference on New Trends in Computing Sciences, ICTCS 2019-Proceedings*, 2019. doi: 10.1109/ICTCS.2019.8923108.

BIOGRAPHIES OF AUTHORS

Sari Nuzulastri currently a Master student in Universitas Sriwijaya. She received her undergraduate degree in the same university, majoring in informatics. Her areas of interest include internet of things, machine learning, and cyber security. She can be contacted at email: sari.anhar88@gmail.com.

Deris Stiawan received his Ph.D. degree in Computer Engineering from Universiti Teknologi Malaysia, Malaysia. He is currently a Professor at Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya. His research interests include computer networks, intrusion detection/prevention system, and heterogeneous networks. He can be contacted at email: deris@unsri.ac.id.

Hadipurnawan Satria received his Ph.D. degree in Computer Science from Sun Moon University, South Korea. He is currently a Lecturer at the Department of Informatics Engineering, Faculty of Computer Science, Universitas Sriwijaya. His research interests include platform-based development, embedded systems, and software engineering. He can be contacted at email: hadi@ilkom.unsri.ac.id.

Rahmat Budiarto received his Doctor of Engineering in Computer Science from Nagoya Institute of Technology, Japan in 1998. Currently, he is a full professor at Department of Computer Science, College of Computing and Information, Albaha University, Saudi Arabia. His research interests include intelligent systems, brain modeling, IPv6, network security, wireless sensor networks, and MANETs. He can be contacted at email: rahmat@bu.edu.sa.