PJAEE, 17 (9) (2020)

PalArch's Journal of Archaeology of Egypt / Egyptology

Optimal attribute based encryption for secure data storage on Social Distributed Environmen

Lekshmy P L[1], Dr.M. Abdul Rahiman[2]*

[1]Assistant Professor, Computer Science and Engineering, L B S Institute of Technology for Women, Trivandrum, University of Kerala,
[2]Director, LBS Centre for Science and Technology, Trivandrum, Kerala
Email: [1]lekshmyvinod@gmail.com

Lekshmy P L[1*], Dr.M. Abdul Rahiman[2,] Optimal attribute based encryption for secure data storage on Social Distributed Environmen - Palarch's Journal Of Archaeology Of Egypt/Egyptology 17(9). ISSN 1567-214x, Keywords- encryption, dataset, Swarm Optimization algorithm, optimal attribute based encryption.

Abstract

Preserving privacy and security in cloud computing is a big challenge. There are a lot of systems that can provide cloud computing services. Cloud Computing is a service that is quickly increasing its development in the IT industry in recent years. In a public cloud environment, users transfer their data to a public cloud server and cannot control its remote data. Thus, information security is one of the significant problem in public cloud storage, such as high computation cost, confidentiality of the data, integrity, availability and authenticity. To resolve this, in this paper, optimal attribute based encryption (OABE) is proposed. ABE aims to strengthen the sensitive data confidentiality in public cloud storage. To enhance the ABE algorithm, the key values are optimally selected with the help of Beetle Swarm Optimization algorithm (BSO). BSO is a new meta-heuristic algorithm, which is based on the foraging principle of the beetle.

I. INTRODUCTION

Nowadays, social media is expanded by increasing the number of users across the world, and industries also started advertising through social media to improve and broaden their businesses [1]. A virtual community, in which people shared their interest such as a specific activity, can interact and socialize among themselves with the help of social media [2]. A social community is a platform to build social relations among people who share interests, activities, backgrounds, or real-life connections [3]. People were using internet applications to share their personal and private data's in their groups. They expect shared details to be secured. But the major problem faced by most of the users is information leakage to the service provider. A way to protect the privacy-sensitive data of the user from the service provider is having a trusted third party that keeps the data and runs the algorithm instead of

service provider [8]. By using advanced data storage capabilities of the computer, varieties of data mining algorithms were developed [4].

Sharing and storing data in the cloud environment raises the serious problems of individual privacy when computing information including storage which is provided by third-party service providers for the adoption of cloud computing technologies [5]. When processing and sharing data is performed in a distributed environment, data privacy is a stringent need. More computation cost and high communication requires for secure multiparty computation in collaborative privacy-preserving data mining. In recent years, many encryption algorithms have been used to encrypted data. Here attribute based encryption (ABE) have used to strengthen the data and the beetle swarm optimization (BSO) algorithms are used to select the optimal key [6].

In Cryptography-based secure data storage and sharing using HEVC and public clouds developed by Usman, M., *et al* [7]. Their objectives were to support real-time processing with power saving constraints in mind. Advanced Encryption Standard (AES) was used as the basic encryption technique by their program. Simulation results clearly show that this program surpasses AES-256 by reducing processing time to 4.76% and increasing data volume by approximately 0.72%. Moreover, Jiang, T., *et al*., [8] have developed secure and reliable cloud storage data. In this paper, a probabilistic challenge response scheme was to prove that users' files were available and stored in a specified cloud server. Security and performance analysis demonstrates that economically rational cloud service providers offer strong benefits from re-outsourcing its customer data to certain other cloud providers. Additionally, Mohanram, K., *et al*, [9] have developed a secured data storage and retrieval techniques for transport data for securely storing traffic data, including vehicles, records, and payments. The techniques have been tested using data from the Tamil Nadu Department of Transport in India and have been found to be highly efficient in terms of increasing time, memory, and security. The main advantages of the system are increased data availability from anywhere at any time and enabling online payment.

Similarly, Manjula, S., *et al*, [10] have presents a cloud data environment for secure data storage. The hacker sees the stored file; it would not find any part of the file, because it was encrypted. Secure distribution of files was possible with proper verification. If any cloud servers fail due to duplication, the system must be able to restore data from the remaining cloud server. Confirms file encryption and file split data privacy using distributed algorithm. Reliable and secure data storage and retrieval in a cloud analyzed by Atukuri, V. R. R., *et al* [11]. Here, they guarantee input/output privacy and authenticity/sound. So the customer's data was secure in the cloud server, and the customer data was safe in the cloud and may not be corrupted by the personal data server. Here, they also tested the cloud's communication delay to monitor the system's performance. Moreover, Kangavalli, R., *et al*, [12] have developed a homomorphic encryption scheme for secure data storage in cloud. Providing protection for data stored in the cloud using homomorphic encryption schemes is a new way of cloud data storage. They present the feedback they received during their study of FHE projects. In this study, the byte-level auto-morphism system was proposed.

II. PROPOSED METHOD

Our aim is to protect the private data from service provider. For that, we present cryptographic protocol that clusters user in social network using optimal attribute based

encryption. The data of users must be transferred through the service provider. It is an unavoidable process. By PFCM clustering algorithm the users are separated into similar groups by the service provider.

There are huge numbers of users $A=(a_1,a_2,a_3,.....,a_n)$ in a social network connected with the service provider. The number of groups $P=(p_1,p_2,p_3,.....,p_n)$ depends on the number of users in a system. Inside the group, the users are clustered based on their similarities which means according to their properties. The cluster is denoted by (R). Our aim is to hide user's privacy data from the service provider i.e.) the data that is shared by the user must be secured and send it to service provider which is in cloud. So the service provider chooses one helper user (H) from each group. Helper user that is chosen by service provider is fully trusted by both the user and service provider. When the user sends the message the *H* receives the message and then sends the message to the sanitization process with the optimal key. By using optimal key the message is sanitized using sanitization process. Then the sanitized data is send to the service provider. The overall proposed method is given below.

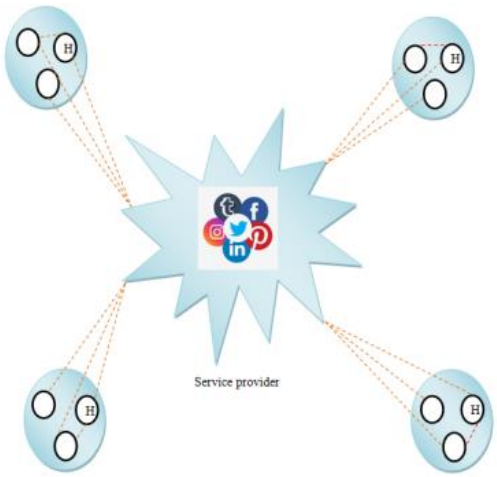


Figure 1: The overall proposed method

2.1 Clustering using PFCM

The background of the PFCM technique development is effectively presented. It is accompanied by the presentation of the data through the innovative schemes. In this regard, a possibility-fuzzy c-means (PFCM) model characterizes a clustering technique proficiently employed to determine the centroid. The innovative approach amazingly performs the activity of concurrently generating the memberships and possibilities, together with the usual point prototypes or cluster centers for each cluster, which represents a hybridization of the Possibility c-means (PCM) and fuzzy c- means (FCM) which was able to consistently avert several challenges faced by the PCM, FCM and the FPCM. In addition, the novel PFCM method incredibly resolves the noise sensitivity deficiency of the FCM significantly overwhelms the coincident clusters challenge of the PCM and eradicates the row sum parameters of the FPCM. In thispaper, for the purpose of clustering the data the PFCM is effectively employed and it represents a fusion of FCM and PCM and utilizes the membership and typicality facets from both the clustering techniques. It is competent to proficiently overwhelm the thorny issues of the FCM and PCM techniques and elegantly

carries out the clustering procedure. The objective function of PFCM is furnished by means of Equation 9 shown below.

$$J_{PFCM}(U,T,V;X) = \sum_{k=1}^{n}\sum_{i=1}^{c}(aU_{ik}^{m}+bT_{ik}^{\eta})\times\|x_k - v_i\|_A^2 + \sum_{i=1}^{c}\gamma_i\sum_{k=1}^{n}(1-T_{ik})^{\eta} \quad (1)$$

Subject to the parameters $\sum_{i=1}^{c} U_{ik} = 1 \ \forall k$ and $0 \le U_{ik}$, $T_{ik} \le 1$. Here $a > 0$, $b > 0$, $m > 1$ and $\eta > 1$. In (1) $\gamma_i > 0$ is the user specified constant. The constant $a \ and \ b$ defined the comparative significance of the fuzzy membership and typicality values in the objective function. In addition, in equation (1) U_{ik} value is identical to that of the FCM membership function is given as,

$$U_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|}\right)^{\frac{2}{m-1}}} \quad (2)$$

$$v_i = \frac{\sum_{i=1}^{n}(U_{ik})^m x_k}{\sum_{i=1}^{n} U_{ik}} \quad (3)$$

Subject to;

$$\sum_{i=1}^{c} U_{ik} = 1, \ j = 1,2,\ldots,n$$
$$1 \ge U_{ik} \ge 0, i = 1,2,\ldots,c, \ j = 1,2,\ldots,n$$
$$n > U_{ik} > 0, \ i = 1,2,\ldots,c \quad (4)$$

It is found that the appraisal of the FCM, incredible and elongated time duration involves together with extreme sensitivity to the initial guess and vulnerability to noise. Likewise, T_{ik} represents an interpretation similar to that of typically as in PCM is given as,

$$T_{ik} = \frac{1}{1+\left[\frac{D^2(x_k,v_i)}{\eta_i}\right]^{1/(m-1)}} \quad (5)$$

The parameter η_i is assessed for each and every cluster independently.

$$\eta_i = A\sum_{k=1}^{n} T_{ik}^m \ D^2(x_k,v_i) \Big/ \sum_{k=1}^{n} T_{ik}^m \quad 6)$$

As illustrated in (6) η_i is in direct proportion to the average fuzzy intra cluster distance of cluster v_i. Habitually A is selected as 1.

2.2 Data encrypted using Attributed based encryption

Encryption refers to the task of updating basic text in the hard cipher text. Before storing the data it can be encrypted using the OABE algorithm. The step by step procedure is of encryption is given by,

Setup

It does not take any inputs other than security parameters, the random prime numbers are referred to as PK and MK is the master key.The security parameters of the ABE encryption is represented as β and universal description can be denoted as U. We define a group of prime order is P_{g1} and then s and t is represented as bilinear map e: P_{g1}*P_{g1}-> P_{g2} and p can be denoted as a generation of P_{g1}. By defining h_1, h_2….h_U P_{g1} randomly which are associated with universal attributes U and then it, chooses random exponents are s, t Z_g. The public key is shown below,

(7)

Here h can be represented as hash function.

Key generation

This algorithm takes as input a set of attributes associated with the user and the master key. Its outputs random prime keys that enable the user to decrypt a message encrypted under an access tree structure.This method takes the set of prime numbers as input. This method is used to select the optimal key from this set of prime numbers.M is the prime set of matrix with size l*n, ρ is an injective function, which the row of M. For a short description, we suppose want the messages are the same. Then the random vector $v = (y_1, y_2, \ldots \ldots, y_n) \in$ which is based on shared exponents A. It calculates. $\gamma_i = v.M; i = 1,2, \ldots$ is represents the total number of prime numbers. The user output key is given by,

$SK = p^i$ i=1, 2, 3…

(8)

The BSO calculation is a streamlining calculation that joins the insect rummaging instrument and PSO algorithm. Similar to the PSO algorithm, the beetle also share information but the distance and direction of the beetles are determined their speed and the intensity of the information to be detected by their long antennae. Studies have exhibited that the two antennae of beetles are utilized to investigate encompassing locales. At the point when a reception apparatus at one side distinguishes highly thought food smells, the creepy-crawly will go to the antennae on the current side. As per this basic rule, beetles can viably discover food. The step by step procedure of BSO is given by,

Step 1: Here, the solution is considered as the population beetles. The position of beetles and the speed of the beetles are initialized in the three environmental models. The populations of the beetles are represented by,

$P = (P_1, P_2, \ldots, P_n)$, n=1, 2, 3……. (9)

The position of the beetles are represented by,

$$P_j = (P_{j1}, P_{j2}, \ldots, P_{js})^T \quad (10)$$

Where S represents the search space with j^{th} beetle and also represents the potential solution. The speed of the beetles are given by,

$$X_j = (X_{j1}, X_{j2}, \ldots, X_{jS})^T \quad (11)$$

Step 2: After finishing initialization process, the fitness functions are calculated. Here the maximum key breaking time can be considered as the objective function of fitness calculation. Fitness function is given by,

$$Fitness = Max(Key\ breaking\ time) \quad (12)$$

Step 3: The individual beetles are represented as $B_i = (B_{i1}, B_{i2}, \ldots, B_{iS})^T$, and the group extreme of the beetles are denoted as $B_g = (B_{g1}, B_{g2}, \ldots, B_{gS})^T$. It is represents the updation of attribute of beetles.

Step 4: Iterations are performed until the end of optimization. The mathematical model for simulating behavior is given by,

$$P_{jS}^{k+1} = P_{jS}^{k} + \delta X_{jS}^{k} + (1-\delta)\chi_{jS}^{k} \quad (13)$$

Where S=1, 2, ..., n; k is considered as the current number of iteration. X_{jS} is denoted as the speed of the beetles and χ_{jS}^{k} represents the increase in the beetle position movement. δ is considered as positive constants. The speed formula is given by,

$$X_{jS}^{k+1} = \gamma X_{jS}^{k} + a_1 r_1 \left(B_{iS}^{K} - P_{jS}^{k}\right) + a_2 r_2 \left(B_{gS}^{k} - P_{gS}^{k}\right) \quad (14)$$

Where a1 and a2 are two constants, and the r1 and r2 denotes the random function in the range [0, 1]. γ represents the inertia weight. Decreasing the weight of inertia is given by,

$$\gamma = \gamma_{\max} - \frac{\gamma_{\max} - \gamma_{\min}}{I} * k \quad (15)$$

Where $\gamma_{\max}$ and $\gamma_{\min}$ represents the minimum and maximum values of γ . K and k are the maximum number of iteration and the current number of iteration. The maximum and the minimum iteration is set to be 0.4 and 0.9 respectively. So that the algorithm can search a larger range at the beginning of evaluation and find the optimal solution area as quickly as possible. As γ gradually decrease, the beetles speed decreases and then enters local search. The γ function which defines the incremental function is calculated as,

$$\gamma_{jS}^{k+1} = \lambda^{k} * X_{js}^{k} * sign\left(g\left(P_{rS}^{k}\right) - g\left(P_{rS}^{k}\right)\right) \quad (16)$$

In this step, we extend the updation to a high dimension. λ indicates step size. The search behavior of the right antenna and the left antenna are represented as,

$$P_{rS}^{k+1} = P_{rS}^{k} + X_{is}^{k} * \frac{d}{2}$$

$$P_{lS}^{k1} = P_{lS}^{k} - X_{iS}^{k} * \frac{d}{2} \quad (17,18)$$

Step 5: Termination: The optimization process terminates when is achieves theminimum error value and the maximum number of iteration. Once we get the best fitness solution, the iteration stops. The optimal solution is reached.

Encryption

It takes the input random prime numbers. The output of the encryption is ciphertext. This algorithm is mainly used to encrypt the owner's data. Before encryption the data d can be divided into d_1, d_2,……, d_n according to their different privacy levels and every part of the corresponding attributes level is A. Then A_1, A_2… A_n is represented as an attribute set. It chooses random A_1, A_2, ……,A_n Z_g. The ciphertext is given by,

T_1 (19)

T_2 : (20)

T_n = (21)

Decryption

The output of the ciphertext is encrypted under an attribute set A, public key PK and secrete k SK of an access structure. If attribute set A∈ access structure, the output will be m (i.e.) mass otherwise false symbol.

This algorithm will help users to decrypt the ciphertext with the prime numbers SK. SK is assoc with an access structure (M, ρ) and T_1 is associated with the attribute set A_1. Then we defin $\{i: \rho(i) \epsilon A_1\} C \{1,2,3, \ldots, l\}$, there will be asset of constant $\{x_i \in z_g\}_{i \in F}$ which satisfies $\sum_{i \in F} x_i \gamma_i$ = The message d is the ratio of the optimum key to the ciphertext is given by,

$$d = \frac{T \prod_{i \in F} \left((e(T_1', T_i) e(D_{x'}, G_i)^{\hat{i}} \right)}{e(T_1', G)} \quad (22)$$

$$d = \frac{d e(p,p)^A}{e(p,p)^A} (23)$$

III. RESULT AND DISCUSSION

This section is described a result and discussion based on the proposed method such as optimized ABE algorithm for privacy preserving in social environment. Here the proposed methods are executed by python and the proposed method. the detailed description of movie lens dataset is given as,

MovieLens Dataset- 1: This dataset contains 100,000 integer ratings in the range of (0.5) for 1682 movies by 943 users. It contains the age, gender, occupation and rating of the user. The dataset is in the form of matrix.

MovieLens Dataset- 2: Stable benchmark dataset which contains 20 million ratings and 465,000 tag applications applied to 27,000 movies by 138,000 users. Includes tag genome data with 12 million relevance scores across 1,100 tags.

3.1 Performance matrices

$$Clustering\ Accuracy = \frac{C_D}{C_S} \times 100 \quad (24)$$

$$\text{Processing Time} = \sum (C_T + O_T + T_T) \quad (25)$$

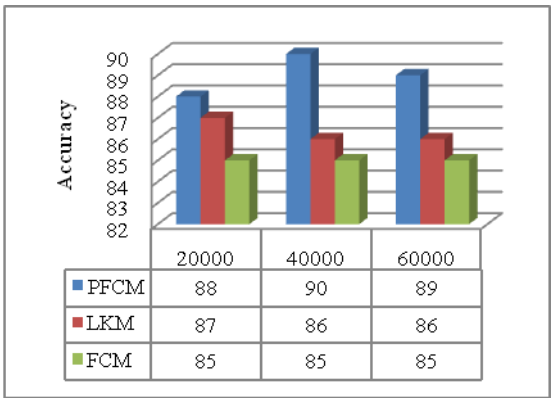
Where C_T represents the clustering time, O_T denoted as optimization time and T_T represents the transmission time. C_D represents the clustering data and C_S represents the size of the clustering data.

Key breaking time: The key breaking time is an essential one to ensure the duration taken for the hackers to hack the key and get access to the secured data. When the duration for key breaking is increased then the data hacking time is also increased.

Data transmission time: Data transmission time represents thetime taken to complete the transmission of the data from helper user to the service provider.

3.2 Comparative analysis

In this section experimental results of quality metrics have been analyzed based on our proposed privacy preserving optimization process in distributed clustering network.



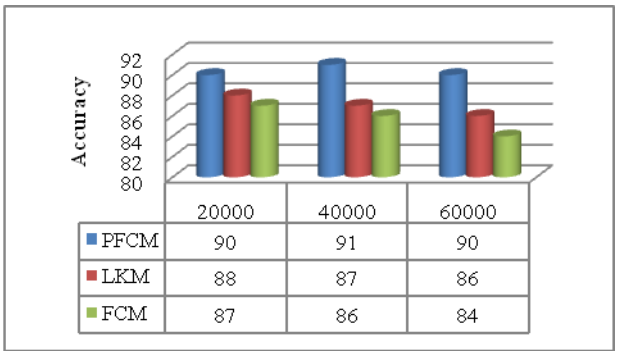


Figure2: Comparative analysis of clustering accuracy dataset 1 and dataset 2

Figure (2) represents the comparative analysis of the proposed and the existing method against the accuracy. The accuracy value of proposed method in data set 1 is 88% in the cluster is 2 and the accuracy value of data set 2 is 90%. According to the comparative analysis the accuracy of existing method is smaller than the proposed method. Our approach effectively groups the relevant users that lead to maximum clustering accuracy. Therefore by observing these values we get effective clustering accuracy.

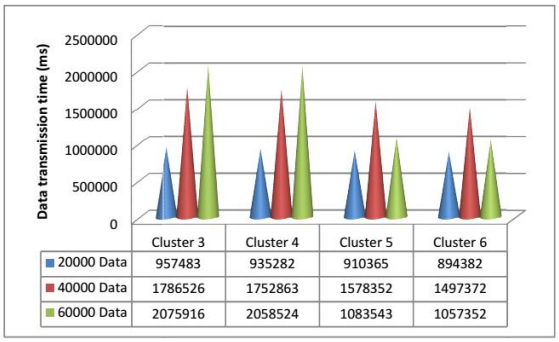


Figure3: Comparative analysis of Data transmission in data set 2

Figure (3) represents the comparative analysis of the proposed and the existing method in data set 1 and data set 2. When the data size of the proposed method is 20000, the transmission time of data set 1 is 935282ms and the data set 2 is 901545ms. When comparing the data transmission time in the proposed and the existing method, the proposed method is very much better than the existing method.

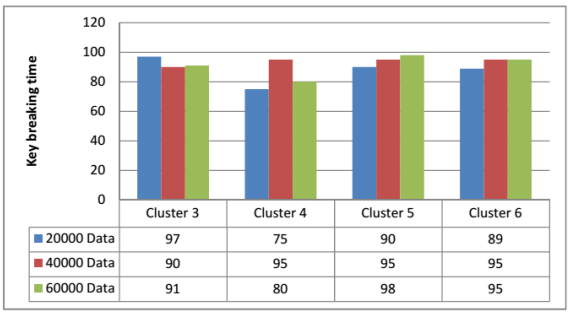


Figure 4: Key breaking time in data set 1and data set 2

The key breaking time of our proposed method is calculated experimentally in dataset 1 and dataset2 shown in Figure (4). The key breaking time must be higher for more privacy purpose. The key breaking time is calculated in percentage.By using 20000 data the key breaking time of the proposed data set 1is 91 and dataset 2 is 90. If the key breaking time is increased the hackers couldn't access our data easily. Hence the key breaking time of our proposed method is analyzed and the value is higher.

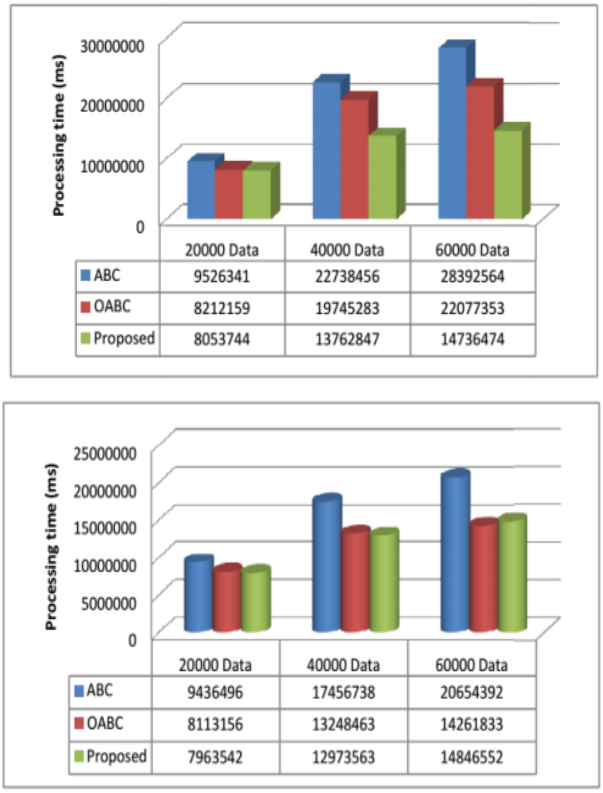


Figure 5: Comparative analysis of processing time in data set 1 and 2

The above figure 6 shows the performance time of different method by varying number of data against the dataset 1 and dataset 2. When analyzing figure 6, our proposed method takes minimum time to processing the clustering compare to other method. This because of prototype based hybrid is utilized in leader based clustering algorithm which is speedup the clustering process.

IV. PRIVACY PERFORMANCE

4.1. Various Security Attacks

The suggested method employs different security attacks for the privacy purpose in a data transmission such as Denial of service (DOS) attack and Man in Middle (MIM) Attack. These attacks are applied by malicious to collapse the original data. To prove the security of the proposed method, the DoS and MIM attacks are applied.

DOS Attack

DOS attacks have turn out to be a most significant hazard to current computer networks. DOS or distributed denial of service (DDoS) attack is an effort to make a machine or network resource connected to its future users. DOS attacks were initiated from distributed attacking hosts. In our work DoS attack is applied in data to check the hacking percentage.

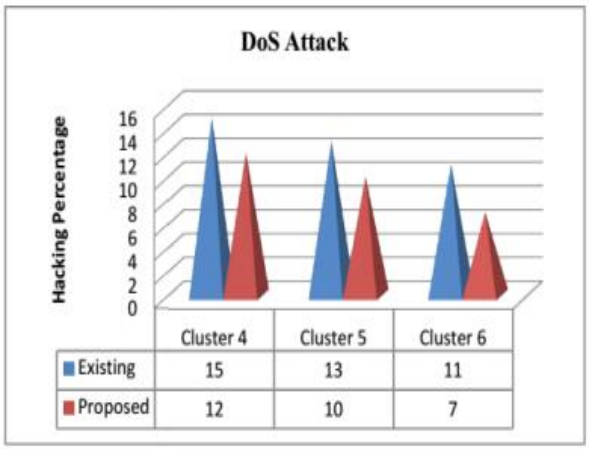


Figure 6: DoS attack in data set 1

Here DOS attack percentage is compared with existing hybrid speedup k-clustering method as shown in Figure 7. When executing hacking percentage with 4 clusters, in existing method there is 15% of possibilities and for proposed ABE process there is only 13% of chance. It is a difficult task to hack the user's data because the optimal key value cannot be predicted.

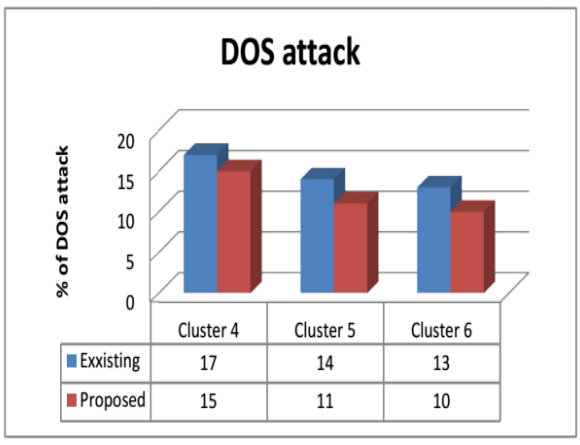


Figure 7: Dos attack in data set 2

In Figure 8, DOS attack percentage is compared with existing hybrid kernel k-means clustering using homomorphic method. When executing hacking percentage with 4 clusters, in existing method there is 17% of possibilities and for proposed ABE process there is only 15% of chance. The order of values is similar for 5 and 6 clusters. The hacking probability is lower for the proposed method than existing model.

MIM Attack

In cryptography the attacker secretly relays and possibly alters the communication between two parties, in MIM attack who believes they are directly communicating with each other. MIM attack is also applied in sanitized data to check the hacking percentage.

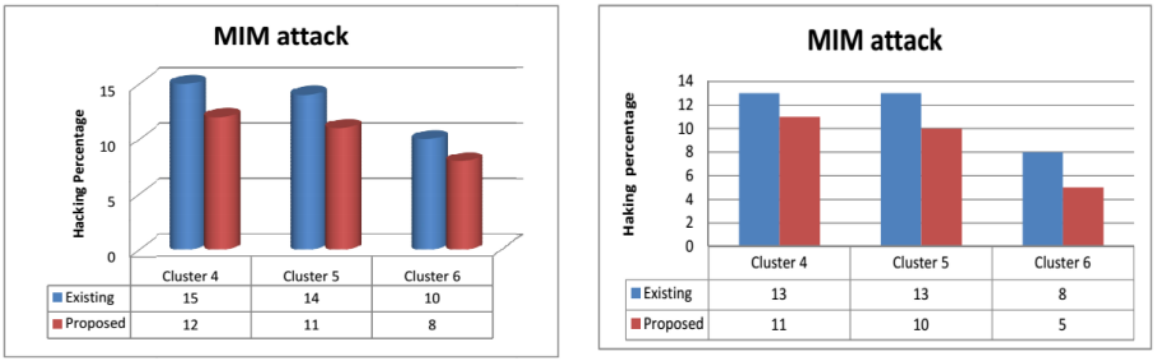


Figure 8: MIM attack in data set 2

Here MIM attack percentage is compared with existing hybrid kernel k-means clustering method as shown in Figure 9. When executing hacking percentage with 4 clusters; in existing method there are 15% of possibilities and for proposed ABE process there is only 12% of chance in data set 1. When executing hacking percentage with 4 clusters; in existing method there are 13% of possibilities and for proposed ABE process there is only 11% of chance in data set 2. The result is similar for 5 and 6 clusters. Therefore, the hacking probability is lower for the proposed method than existing model.

V. CONCLUSION

In this paper, secure data storage on the social environment has been clarified. Here, the movie lens data has been scrambled utilizing the OABE algorithm. The BSO has been improved by methods for the OABE algorithm. The scientific articulation of both the algorithm has been explained. The performance of the proposed system has been analyzed as far as various measurements to be specific accuracy, information loss, processing time, DoS attack, MIM attack and key breaking time. Our recommended method has less encryption and decryption time than the present system, as appeared by exploratory results. Consequently, our proposed strategy is profoundly ideal than current strategies. Finally, we achieve good accuracy and less information loss.

REFERENCE

[1] S. Baur, H. Boche, “Robust secure storage of data sources with perfect secrecy,” IEEE Workshop on Information Forensics and Security (WIFS), 2017.

[2] Potey, M. Manish C. A. Dhote, and Deepak H. Sharma. & quot; “Homomorphic Encryption for Security of Cloud Data,” ." Procedia Computer Science 79 (2016): 175-181.

[3] Y. Ren, Leng. Y, Y. Cheng, and J.Wang, “Secure data storage based on blockchain and coding in edge computing,” Math. Biosci. Eng, 16(4), pp.1874-1892, 2019.

[4] R. Di Pietro, M. Scarpa, M. Giacobbe, and A. Puliafito, “Secure storage as a service in multi-cloud environment,” In International Conference on Ad-Hoc Networks and Wireless (pp. 328-341). Springer, Cham. 2017, September.

[5] Zhihong Tian, Wei Shi, Yuhang Wang, Chunsheng Zhu, Xiaojiang Du, Shen Su, Yanbin Sun and NadraGuizani, "-Real Time Lateral Movement Detection based on Evidence Reasoning Network for Edge Computing Environment", IEEE Transactions on Industrial Informatics. 2019.

[6] Zhihong Tian, Shen Su, Wei Shi, Xiaojiang Du, Mohsen Guizani and Xiang Yu, "A data-driven method for future Internet route decision modeling", Future Generation Computer Systems, vol.95, pp. 212-220, 2019.

[7] M. Usman, M. Ahmad Jan, X. He, Cryptography-based secure data storage and sharing using HEVC and public clouds. Information Sciences, 387, 90–102, 2017.

[8] T. Jiang, X. Chen, J. Li, D. S, Wong, J. Ma, J. K. Liu, "Towards secure and reliable cloud storage against data re-outsourcing," Future Generation Computer Systems, 52, 86–94, 2015.

[9] K.. Mohanram, T. T. Mirnalinee, Secured Data Storage and Retrieval Techniques for Effective Handling of Transport Data. 2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), 2017.

[10] S. Manjula, M. Indra, & R. Swathiya, "Division of data in cloud environment for secure data storage," International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16). 2016.

[11] V. R. R. Atukuri, & R. S. R. Prasad, "A novel approach: Reliable and secure data storage and retrieval in a cloud," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017.

[12] R. Kangavalli, & S. Vagdevi, "A mixed homomorphic encryption scheme for secure data storage in cloud," 2015 IEEE International Advance Computing Conference (IACC), 2015.