Journal of Medical Systems (2019) 43: 265
https://doi.org/10.1007/s10916-019-1374-6

SYSTEMS-LEVEL QUALITY IMPROVEMENT

Efficient and Secure Privacy Analysis for Medical Big Data Using TDES and MKSVM with Access Control in Cloud

E. Shanmugapriya[1] · R. Kavitha[2]

Received: 26 March 2019 / Accepted: 5 June 2019 / Published online: 4 July 2019

Abstract

Big Data and cloud computing are two essential issues in the recent years, empowers computing resources to be given as Information Technology services with high efficiency and effectiveness. So as to protect the security of data holders, data are regularly stored in the cloud in an encrypted form. In any case, encrypted data introduce new challenges for cloud data deduplication, which becomes crucial for big data storage and processing in the cloud along with access control. In this paper dissected the medical big data security utilizing encryption with access control process. Big database reduce process Map-Reduce framework with Optimal Fuzzy C means (OFCM) to Clustered data are accumulated in the cloud and furthermore using classification approach to classify sensitive and non-sensitive data in the cloud to encryption. This security process Triple DES (TDES) to encrypted and stored in the cloud and propose practical optimization techniques that further enhance the scheme's performance, at long last authentication phase with attribute-based access control is used to authenticate data in cloud sim. From the proposed method the clustering, classification and encryption results are compared to existing approaches.

Keywords Big data · Cloud · Optimization · Data security · Medical data · Clustering · Classification and encryption

Introduction

In the recent years, people have entered the Big Data generation, which is quick as well as radically changing the way they live, work and thinks [1]. Big Data is a standout amongst the most vital research challenges for the 2020 skyline [2]. Today, an expansive number of big data services are sent or relocate to the cloud for data mining, processing or sharing [3]. Big Data is a data examination philosophy empowered by another technologies and architecture which support high-velocity data capture, storage, and investigation [4]. The current explosion of data that is being created is because of many applications such as mobile sensors and social media services [5]. The volume of data is huge and it is predicted to reach 35 zeta bytes by 2020 [6]. Right when data becomes too big to effectively store and research, an appropriated service, or cloud computing service, becomes a more conceivable choice [7]. Public cloud services can give advantages to enhanced information security, and even enhance privacy practices.

The cloud supplier's users pay for a service and transfer their data to the cloud [8]. Cloud computing alludes to both the applications conveyed as services over the Internet plus the equipment and programming framework [9]. Big data technologies describe new technologies along with its architectures, designed to economically extract an incentive from huge volumes of a wide assortment of data, by empowering high-velocity capture, discovery, and, or examination [10]. Cloud computing has numerous security issues and the data security is imperative and it is to threaten the customer's privacy [11]. An effective technique to enhance the efficiency of training data computation for big data features learning by offloading the costly operations to the cloud [12].

Clustering is an unsupervised technique used to classify huge datasets into correlative gatherings. [13]. Clusters need to incorporate quick interconnection technologies to help high-data transfer capacity and low latency between processor communication among the cluster nodes [14]. An efficient k-means protocol expects to guarantee the privacy protection under a given

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ E. Shanmugapriya
vijaylaya2000@gmail.com

[1] College of Engineering Guindy, Anna University Chennai, Madurai, India

[2] Velammal College of Engineering and Technology, Madurai, India

security without using any cryptographic scheme. [15]. This technique proposes an entropy-based rule for the open data set and introduce a re-identification chance model to prevent the private data leakage from open dataset [16]. The evaluation illustrated that privacy-saving clustering is plausible, and homomorphism-encryption based calculation finished clustering a vast data set. In a distributed situation, it will be greatly scalable and minimal effort by using optimization techniques [17].

Literature review

Wenting Shen et al. 2016 [18] had proposed an example named remote data ownership checked with privacy-protecting authenticators for cloud storage. Both cloud service supplier and the public verifier did not approach the genuine authenticators (marks) for cloud data. The honesty of cloud data is as yet ready to be efficiently checked. To securely protect the privacy of the authenticator, they design another authenticator called Homomorphic Invisible Authenticator (HIA), which protected the privacy of authenticator also supports the block less verification. Based on the HIA method, the paper constructed the principal remote data ownership checking scheme with privacy-saving authenticators for cloud storage. The outcome demonstrated that the proposed technique was secured with more efficient.

Peng Li et al. 2016 [19] had projected a Privacy-Preserving High-Order Neuro-Fuzzy C-Means (PPHOFCM) for clustering heterogeneous data on cloud computing. PPHOFCM clusters the heterogeneous dataset by each heterogeneous data object as a tensor also used the tensor distance to capture the correlations in the high-order tensor space. The cloud computing was used to enhance the clustering efficiency of gigantic heterogeneous data from the Internet of Things (IoT). This scheme used to protect the private data when executing the high-order neuro-fuzzy c-means on cloud computing.

Syam Kumar Pasupuleti et al. 2015 [20] had proposed an efficient and secure privacy-protecting approach for outsourced data of resource-constrained cell phones in the cloud computing. This approach utilized probabilistic public key encryption algorithm for encrypting the data and invoke positioned keyword search over the encrypted data to recover the records from the cloud. This technique was to achieve an efficient framework for data encryption without sacrificing the privacy of data. Thus, data privacy guarantees and computation, through security and performance investigation. This paper demonstrated that this approach was semantically secure and efficient.

Xuefei Cao et al.2016 [21] had proposed to decide the intersection between two data sets while safeguarded the privacy of each set. Malicious participants are considered and the misconduct of participants is avoided. Our strategy depends on a combination of commutative encryption and hash-based commitments. Performance assessment exhibits the effectiveness of the protocol. Security discussion is given demonstrating that the protocol provides data privacy, secure set intersection, and tolerance to participants cheating.

Guiqiang Hu et al. 2015 [22] had proposed a high computational complexity of CS reconstruction process was considered to outsource the cloud data for its plentiful computing and capacity resource. Also how to protect data privacy and at the same time, preserve administration of the image stays challenging. The reconstruction and identification authentication service in cloud incorporated the techniques of signal processing. The cloud decides to supply the reconstruction service was relying upon the personality authentication result. Theoretical examination and empirical assessments demonstrate a satisfactory security performance as well as the low computational complexity of the proposed framework.

Jesu VedhaNayahi et al. 2016 [23] had proposed privacy-protecting data mining based on secured high communication as well as computational cost. Data anonymization is a promising technique in the field of privacy-saving data mining used to protect the data. Recently, data anonymization using data mining techniques had demonstrated a significant change in data utility. All things considered, the current techniques lack the effective treatment of attacks. The anonymized data was distributed on Hadoop Distributed File System. In this work, the data utility was measured as far as accuracy with respect to various classifiers. The outcome demonstrated that the accuracy and the execution time of the classification algorithms on the privacy-safeguarded data set to be secured.

Chih-Fong Tsai et al. 2015 [24] had proposed a big data which depends on isolating a substantial issue into littler ones and each of them is carried out by single processor independently. There are two common procedures used to tackle the big data issue. First one is the circulated procedure based on the data parallelism, where a given big dataset can be physically separated into n subsets, and algorithms are respectively executed for the corresponding n subsets. The last outcome can be attained from a combination of the outputs produced by the n calculations. The second one is the Map Reduce based procedure under the cloud computing stage. In this paper, the technique was to compare the performance differences between the distributed and Map Reduce systems over vast scale datasets in terms of mining accuracy and efficiency. Furthermore, the Map-Reduce procedure required the minimum computational cost to process these big data sets.

Ayong Ye et al. 2015 [25] had proposed location obscurity scheme based on the false inquiries in continuous location-based services. To avoid attackers from tracing a mobile user by continuous inquiries, some false questions will be arbitrarily injected by a Trusted Third Party. These

fake questions can avoid the invert mapping from user identity to query content. Also, a partitioning and a clustering algorithm were introduced to reduce the computation and communication working cost further. Security examination demonstrated that the scheme can oppose the continuous questions attack and the trajectory attack. The brief theoretical examinations and experimental outcomes demonstrated the effectiveness and practicality of the proposed scheme.

Qingchen Zhang et al. 2017 [26] had proposed a high-order PCM algorithm (HOPCM) for big data clustering by optimizing the objective function in the tensor space. Further, this paper designed a conveyed HOPCM strategy based on MapReduce for a lot of heterogeneous data. At long last, they devised a privacy-preserving HOPCM calculation (PPHOPCM) to protect the private data on the cloud by applying the BGV encryption scheme to HOPCM. Results indicated that PPHOPCM can effectively cluster an extensive number of heterogeneous data using cloud computing without disclosure of private data.

Problem identification

From the examination of several existing papers about privacy preserving over big data in cloud some threats are noted and illustrated as below:

- Probabilistic C means (PCM) algorithm cannot be applied to big data clustering directly since it is initially designed for the small structured dataset. Especially, it cannot capture the complex correlation over multiple modalities of the heterogeneous data object [26].
- High order clustering algorithm can concatenate the features from different modalities linearly and ignore the complex correlations hidden in the heterogeneous data sets, so they are not able to produce desired results.
- It has high time complexity, making them only applicable to small data sets. Thus, they cannot cluster large amounts of heterogeneous data efficiently [16].
- Cloud Service Provider (CSP) can act naturally intrigued, untrusted and conceivably malicious. It endeavors to conceal data loss incident because of management mistakes, Byzantine failures et cetera [8–10].
- One of the encryption is homomorphic encryption which is higher operational cost; also it has less security and privacy. Performance of the completely homomorphic is completely infeasible even from a worst, and amusingly terrible, best case scenario.
- All these limitations motivated us to move on to the proposed method which could rectify all those issues.

Methodology

Enhance the privacy-preserving big data in the cloud the proposed technique makes to exhibit the innovative and efficient model (Fig. 1). Here the big dataset which is taken as a medical dataset. This database contains a lot of data as lakhs as or more than lakhs. Since the volume of big data is high it is basic to reduce its size by using “MAP REDUCE” technique without the loss of data. Here, the input dataset is at first mapped into several groups in order to reduce the volume of big data to maintain a strategic distance from scalability issues. In this examination OFCM clustering algorithm is used to map the data, it decreases the execution span. This centroid optimization is achieved by using inspired optimization techniques like Krill Herd Optimization (KHO). At that point, the data is fed into reducer, the reducer classifies the data as sensitive and non- sensitive medical datasets based on the score value. MKSVM is used to classify the data. Hence, to enhance the privacy safeguarding of sensitive data, it is secured in the cloud with the assistance of Triple Data Encryption Standard (TDES) model. Encrypted data can be securely accessed because authorized data holders can acquire the symmetric keys used for data decryption. In authentication phase, attribute-based access control framework held to encrypted information, this procedure classification methodology used to assemble the characteristics of the attribute. At the point, when a user is denied from the summary of the authorized users, different users who have the common property with the repudiated user, necessary to refresh their private keys. Information Owner (producer of record) first scrambles the document and then stores into the cloud.

Clustering based map reduce model

Map reduce is a structure for efficiently processing the analysis of big data on an extensive number of servers. At the point, when the information dataset is given to a MapReduce data, it is part of autonomous data chunks which are, at that point processed by the map tasks in parallel. Outputs of the map assignments are then given to the reduce tasks as input. The map function executes the procedure of allocating each specimen to the closest center while the reduce function performs the procedure of updating the new centers. For each mapping, OFCM clustering is used to construct mappers in big medical data.

Mapper model

In the mapper model, the dataset is split and internationally broadcast to all mappers. OFCM construct all global variant centers which is an array containing the information about centers of the clusters and subsequently the distance computations are executed. That is, a mapper can calculate the closest

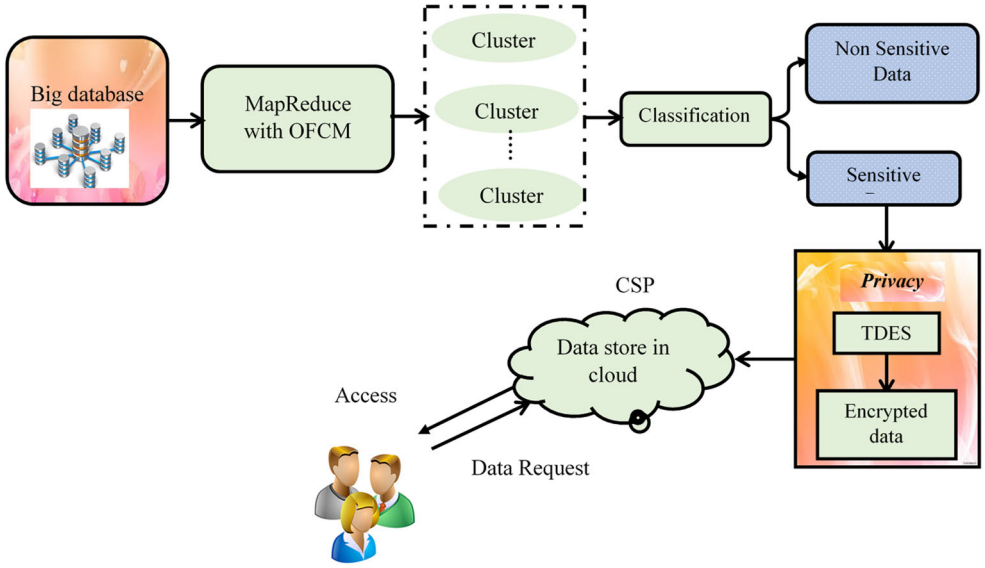

Fig. 1 Schematic model for proposed big data security

center point for each data. This optimal cluster design for mapper function appears in Fig. 2.

Fuzzy C means clustering

Clustering algorithms i.e. fuzzy clustering is based on the optimization of a c-means objective function. The fuzziness of the consequent segment controls the parameter "m" and it is used in this study. The objective function is minimized when pixels near the focuses of their clusters are relegated high membership values, and low membership values are allotted to pixels with far from the focuses of their clusters.

Optimal FCM FOR MR process

Here, the optimal FCM clustering based on the MapReduce framework is proposed for the clustering of big medical data. The authors used a KHO approach to decompose the big data into a few data segments to be used as a part of clustering. To enhance centriole clustering and lead to the robotization of the semantic clustering, apply MapReduce system to the KHO and it will enhance the data analysis task [27]. This optimal FCM discussed in below section.

Step 1: Let choose a fuzzy weighted index, the cluster number of categories, the number of data samples $b = \{b_1, b_2, ...b_n\}$ and cluster center $c = \{c_1, c_2, ...c_n\}$. For select the cluster center or centroid in the mapper model using KHO.

Step 2: Compute the objective function of fuzzy process to choose the centroid

$$O_m = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \left\| b_i - c_j \right\|^2 \quad (1)$$

Above equation calculate the fuzzy membership function u_{ij} using

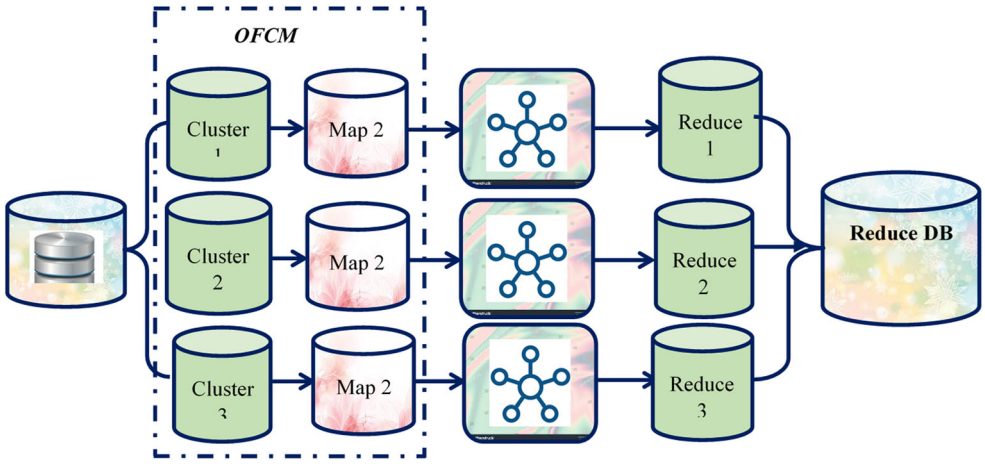

Fig. 2 Fuzzy-based MapReduce model

$$u_{ij} = \frac{1}{\sum \left(d_{ij}/d_{ik}\right)^{(2(m-1))}} \tag{2}$$

Step 3: Compute the fuzzy centers c_j

$$c_j = \left(\frac{\left(\sum_{i=1}^{n} \left(u_{ij}\right)^m c_i \right)}{\left(\sum_{i=1}^{n} \left(u_{ij}\right)^m \right)} \right) \quad \forall_j = 1, 2, ...c \tag{3}$$

In order to attain optimal centroid of FCM, consider the mapping of the FCM to the Map as well as Reduce primitives, it is necessary for FCM to be partitioned into two MapReduce jobs since just a single would not be sufficient.

Centroid optimization model

The centroid optimization process in FCM is done by using KHO technique. This KHO procedure considers two principle objectives, for example, increasing krill thickness and Reaching sustenance, so the crowding behavior of expanding thickness and discovering the food. The main elements of KHO are Movement induced by other krill individuals, foraging activity, Random diffusion, crossover and mutation process.

Objective Function for OFCM The accuracy of data mapping model is attained by evaluating the fitness function, where the whole data is divided into various clusters. The objective function i.e. centroid optimization is achieved by the following equation

$$F_i = MAX\,(Accuracy) \tag{4}$$

New Centroid Updating Procedure It is realized that an optimization algorithm should be capable of searching spaces of subjective dimensionality. In this manner, the accompanying Lagrangian is summed up in an n-dimensional decision space.

$$\frac{dK_i}{dt} = I_i + G_i + P_i \tag{5}$$

Here $I_i\rightarrow$ denotes motion induced by other krill individuals; $G_i\rightarrow$ foraging motion, and $P_i\rightarrow$ physical diffusion of the *ith*krill individuals.

Movement Induced by Other Krill Individuals In the improvement, the course of movement of a krill individual is determined both by the neighborhood swarm thickness (local impact), an objective swarm thickness (target impact) and a shocking swarm thickness (repulsive effect) [28]. The krill optimization can be characterized as

$$I_i^{new} = I^{\max}\gamma_i + \omega_n I_i^{old} \tag{6}$$

In above conditions, the documentation is clarified as just like$I_{\max}$ the greatest incited rate, ω_n is the optimal weight of the movement instigated to the extent [0, 1],

Foraging Motion The scavenging movement is thought as far as two fundamental successful parameters. The first is the sustenance area and the second one is the past experience about the nourishment area. This movement can be communicated to the i^{th}krill individual as takes after:

$$G_i = F_m\delta_i + \omega_m G_i^{old} \tag{7}$$

Here F_m is the scavenging velocity, ω_m is the dormancy weight of the searching movement in the reach [0, 1], is the last scrounging movement, δ_i^{food} is the nourishment appealing and δ_i^{best} is the impact of the best fitness of the krill as such. As indicated by the estimations of the searching rate it is taken 0.02 (ms-1).

Physical Diffusion The physical dissemination of the krill people is thought to be an irregular procedure. This movement can be express as far as a most extreme dispersion speed and an irregular directional vector. It can be defined as takes after:

$$P_i = P^{\max}\lambda \tag{8}$$

Here $D^{\max}$ is the maximum diffusion speed, and d is the random directional vector and its arrays are random values between −1 and 1.

Crossover and Mutation The crossover operator is initially used in GA as a viable procedure for worldwide enhancement. The hybrid rate computation as takes after. The mutation is controlled by a mutation probability (Mr) using eq. (9).

$$cr = 0.2F_i \quad and \quad Mr = 0.5/F_{i\,best} \tag{9}$$

Using this new mutation probability, the mutation probability for the global best is equal to zero and it increases with increasing the fitness value.

Reducer model

Based on this optimal clustering process the input big data reduce function is the data obtained from the combine function of each host. As explained in the combine function, the data includes the partial sum of the examples in a similar cluster and the specimen number. In reduce function, whole every one of the examples and compute the total number of tests doled out to a similar cluster.

Sensitive data classification model

Cloud data security model, training classifier algorithm utilized to classify sensitive and non-sensitive data from reducer model. This approach data classification based on the score values, here MKSVM used. Classification of data helps in characterizing the baseline security controls for protecting the data and classified based on its level of sensitivity and the effectiveness of the association if the data are disclosed, modified or destroyed without authorization.

Multi-Kernel SVM

Multi-kernel SVM to classify the data's without the use of hyper planes, the kernel functions such as linear and quadratic function. Training vectors xi is mapped into a higher dimensional space by the function Φ [29]. At that point SVM finds a linear separating hyperplane with the maximal edge in this higher dimensional space, its appeared in Fig. 3. Its function appeared in underneath condition.

$$\text{Linear Kernel} : Lin_k\left(b_i, b_j\right) = b_i^T b_j + C \tag{10}$$

$$\text{Quadratic Kernel} : quad_k\left(b_i, b_j\right) = 1 - \frac{\left\|b_i - b_j\right\|^2}{\left\|b_i - b_j\right\|^2 + C} \tag{11}$$

In the innovative technique, linear and quadratic kernel functions were combined and then the average of two functions is used here to separate the data of the same classification. The innovative work employs the integration of two kernel functions given as

$$MKSVM = \frac{\left(Lin_k\left(b_i, b_j\right) + Quad_k\left(b_i, b_j\right)\right)}{2} \tag{12}$$

Where *MKSVM* in the above equation symbolizes the integrated kernel function representation of the linear and quadratic kernels.

Then apply this parameter set to the training dataset and then to the classifier. From the classification model classify sensitive and non-sensitive data and also classifier to classify the testing dataset to achieve the generalization accuracy.

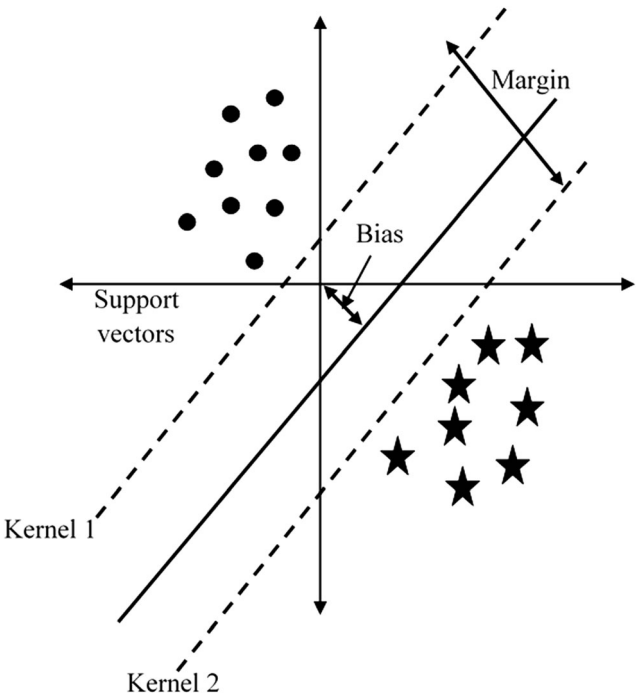

Fig. 3 MKSVM model

Encryption model for big data privacy

In Big data privacy process, encryption is considered for the privacy of data in cloud computing. The results to decide the confidence data the certification procedure will be used to encrypt the data, TDES process. Once these stages are done, the last stage is the transmission of data. For the transmission of data, it is necessary to select particular nodes, with the help of selected nodes that appropriate transmission of data can be finished. After this process, access control process is used as a part of authentication stage.

Triple data encryption standard (TDES)

In cryptography, TDES is a most common encryption model. In TDES algorithm, block cipher algorithms are applied to each data block three times and size of the key is increased to ensure extra security through encryption.

In TDES technique three keys are selected i.e. Key 1, Key 2 and Key 3 (K1, K2&K3). Triple DES algorithm utilizes three cycles of common DES cipher. The key process appeared in underneath.

- Encrypted the plain data withK_1 that is EK1
- Decrypted cipher data K_1 that is EK2
- Again encrypted cipher data with K_3

TDES runs three times slower than DES; however is more secure if it is used legitimately. The procedure for decrypting data is similar to the procedure for encryption, except its execution. For enhancing the performance of TDES process, enhance the keys in the framework [30].

Steps for TDES Triple DES is useful because it has a significantly estimated key length, which is longer than most key lengths partnered with other encryption modes. The accompanying Fig. 4 is the block diagram of TDES as appeared.

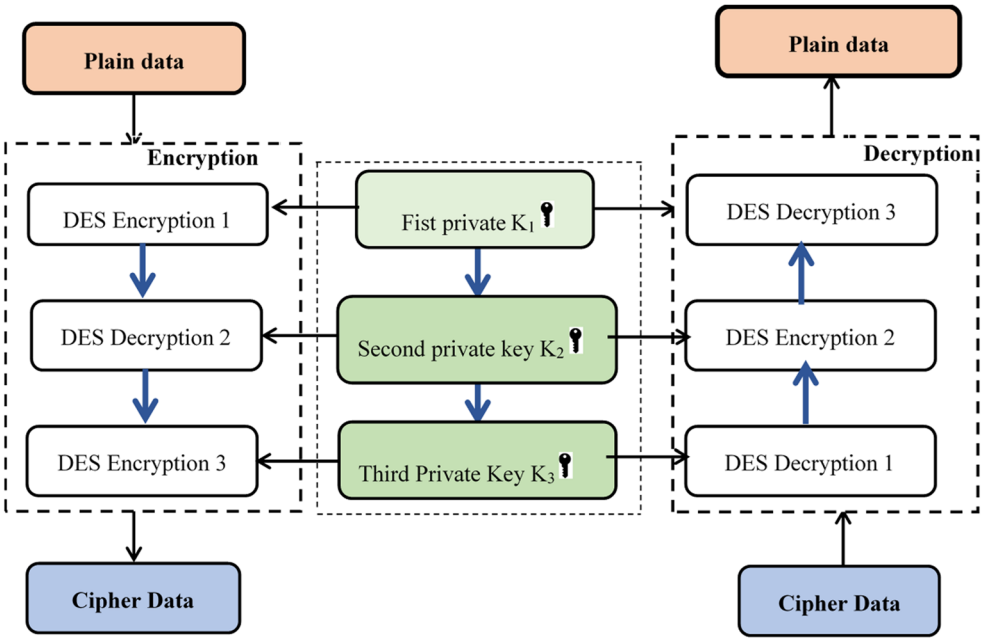

Fig. 4 Triple-DES process

Step 1: Encrypt the data using DES Algorithm with the help of the first keyK_1.
Step 2: Using decryption in the second step during encryption provides backward compatibility with common DES algorithm. In these case first and second secret keys or second and third secret keys are the same whichever key.
Step 3: With the use of second key K_2, the first step output is decrypted using DES algorithm.
Step 4: Finally, encrypt the output of the second step using DES Algorithm with the help of the third key K_3

$$C = E_3(D_1(E_1(d))) = E_3(s) \quad (13)$$

$$C = E_3(D_3(E_1(d))) = E_1(s) \quad (14)$$

Here $C \rightarrow$Cipher data, $D_1 \rightarrow$Decryption withK_1, $E_1 \rightarrow$ Encryption with $K_1$and $s \rightarrow$Secure Data.

It is conceivable to use 3DES cipher with a secret 112-bit key. The first and third secret keys are similar in this case.

It obtains from single DES, however, the technique is used as a part of triplicate and includes three subkeys and key

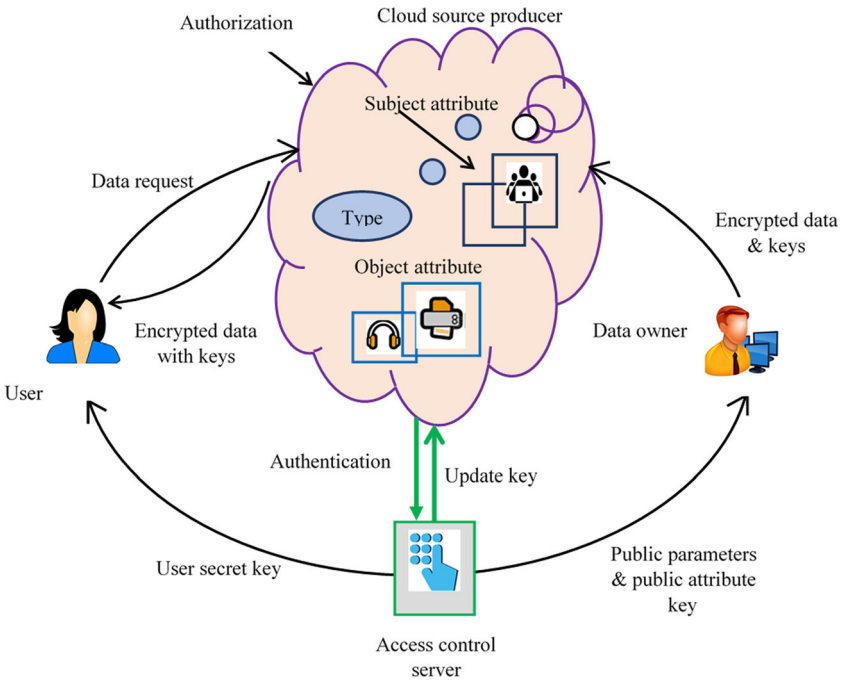

Fig. 5 Model for ABAC

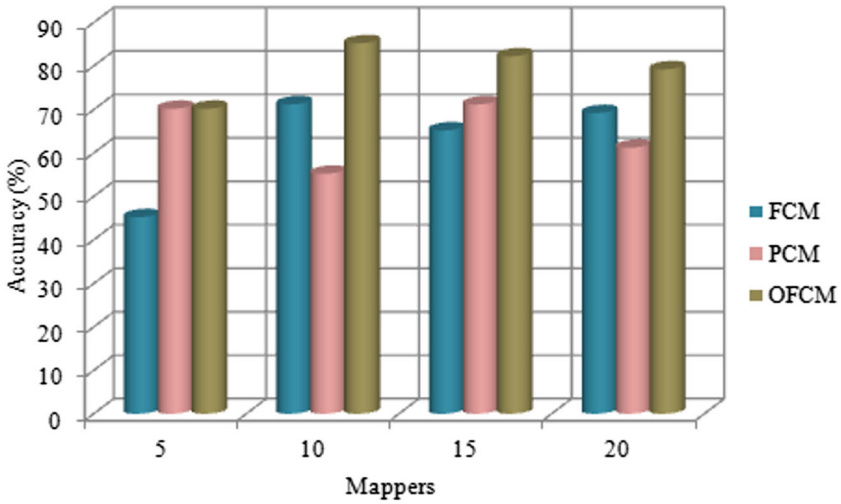

Fig. 6 Clustering accuracy analysis

padding when necessary. Keys must be increased to 64 bits long Known for its compatibility and adaptability can be converted for Triple DES inclusion. For authentication purpose, use access control based on encrypted data.

Authentication phase

Access control is basic when an unauthorized user tries to access the data from the cloud so that only authorized users can access the data. Our proposed authentication process Attribute Based Access Control (ABAC) isn't to characterize consents directly amongst subjects and objects, yet rather to utilize their attributes as the reason for approvals [31]. A basic ABAC for big data privacy appeared in Fig. 5 and this ABAC consists of four principal components

Data owner: Attribute-based access policy, separating his own particular data into various parts, encrypting each part by using symmetric encryption strategies under various content keys.
User: The individual or endeavor that has an arrangement of attributes relying on its parts in the framework. The user is approved to access the data if the access policy associated with the ciphertext which was characterized by data owner is fulfilled by data attribute.
Cloud service provider: The CSP consists of data servers to control data access, and a data service director to deal with the attributes of users.
Central authority: This is a completely trusted gathering that is in charge of entitling, denying, and updating the attributes of users. It produces public and private parameters for the frameworks and stipends the diverse access to users based on their attributes.

This model considers a few attributes from the stored cloud data to the authentication process, it uses the database for maintaining information about registered groups and users, stored keys of users. Groups, message authentication code of the records and the usernames and group names with their distinctive benefits. Encrypting your data previously it is sent to the service provider ensures that if the supplier's security measures are breached, data is as yet secure.

Result and discussion

This proposed could big data high security with encryption with access control model is implemented in Java programming language with JDK 1.7.0 in a windows machine

Table 1 Time and memory analysis for proposed technique

File Size(MB)	Execution Time(ms)	Encryption Time(ms)	Decryption Time(ms)	Memory(bits)
1	48,489	56,471	41,214	1,121,441
2	45,215	55,878	40,154	1,211,152
3	51,232	53,145	39,854	1,284,568
4	58,652	52,477	38,445	1,322,475
5	60,125	53,452	39,445	1,322,457

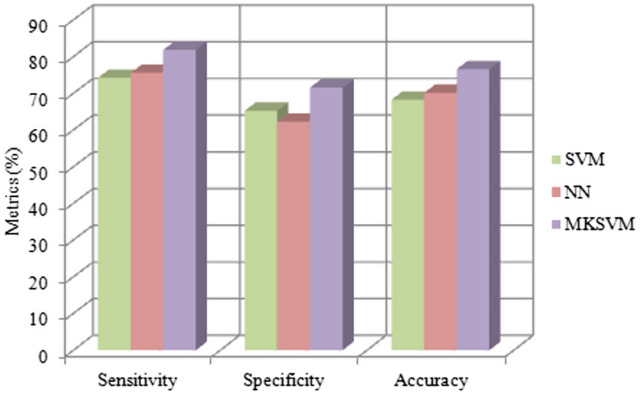

Fig. 7 Classification comparative analysis

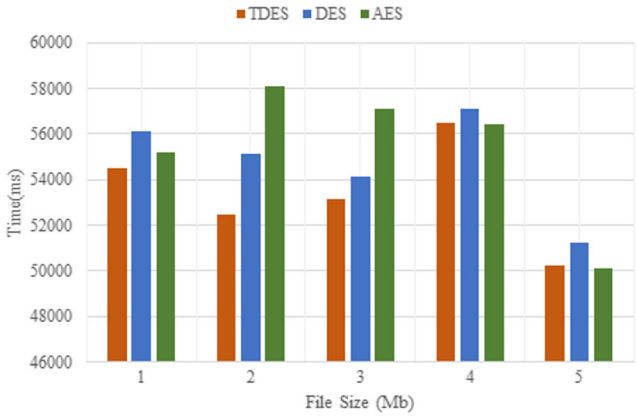

Fig. 8 Comparative analysis for encryption techniques

containing configurations Intel (R) Core i5 processor, 1.6 GHz, 4 GB RAM and the operating system platform. Big Data through Map-Reduce Framework in Cloud environment with the information set of different medical data. Comparative analysis and database description are discussed below.

Database description

This big data security process has combined medical database from UCI machine learning storehouse. Here considered Switzerland, heart disease and breast cancer databases, absolutely the most extreme the size of databases as 1,000,000 data's. Switzerland database: his database contains 76 attributes, yet all distributed examinations allude to utilizing a subset of 14 of them. In particular, the database is the special case that has been used by ML researchers to this date. The "objective" field alludes to the presence of heart disease in the patient. Breast cancer Database: Samples arrive periodically his clinical cases. The database in this manner reflects this chronological gathering of the data. This gathering information shows up quickly beneath, having been expelled from the data itself. Heart database contains 76 crude attributes, just 14 of them are actually utilized. Along these lines I've risked making 2 copies of each database: one with the entire attribute and 1 with the 14 attributes actually consider for investigation reason.

Figure 6 explains the accuracy analysis of clustering with the clustering approaches like FCM, PCM, and OFCM. Big data clustering using FCM technique achieves 42% for mapper 5, 69.42% for mapper 10, 62.12% for mapper 15 and 67% for mapper 20. Compared to FCM and PCM clustering, the proposed OFCM attains a maximum accuracy of 82.21%.

Table 1 describes the time and memory analysis of the proposed technique for the different file size. The size of the file is denoted in terms of Mega Bytes (MB). For 1 MB file size, the time taken for executing the file is 48,489 ms, 56,471 ms is for encrypting the big data and for decryption at the end of the process, it takes 41,214 ms. The memory allocation of 1 MB file size is 1,121,441 bits. Similarly, the analysis of execution time, encryption and decryption time also its memory allocation is illustrated in Table 1 for the 2, 3, 4 and 5 MB file size.

Figure 7 shows the comparative analysis of classification approaches like Support Vector Machine (SVM), Neural Network (NN) and the proposed Multi Kernel Support Vector Machine (MKSVM). The measures like sensitivity, specificity, and accuracy are compared for analyzing the system efficiency. From the graphical analysis, the proposed MKSVM attains more sensitivity, specificity, and accuracy in the classification of big data. It concludes that the MKSVM classifier performs better when compared to other two approaches.

Table 2 Classification performance measures for proposed security model

Mappers	5				10				15			
Training Data (%)	50	60	70	80	50	60	70	80	50	60	70	80
Sensitivity	79.38	80.06	81.16	82.47	80.31	82.11	83.64	85.07	75.22	80.56	80.85	76.23
Specificity	69.38	70.19	71.28	72.36	70.16	71.68	72.46	74.83	65.22	72.12	70.1	68.45
Accuracy	74.62	75.18	76.28	77.18	75.38	77.19	78.61	80.39	78.2	81.2	75.4	82.22
FPR	30.62	29.81	28.72	27.64	29.84	28.32	27.54	25.17	31.21	30.2	29.45	22.19
FNR	20.62	19.94	18.84	17.53	19.69	17.89	16.36	14.93	15.22	21.22	17.52	14.5

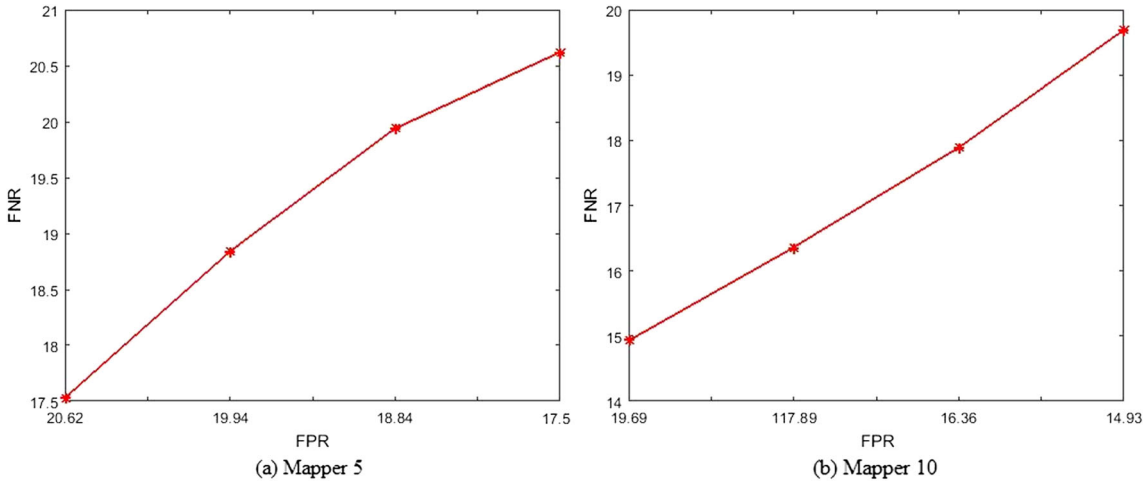

Fig. 9 ROC curve for ABAC performance

Table 2 describes the performance measures of the proposed security model for the mappers 5, 10 and 15. For the 80% training data, sensitivity attains 82.47%, specificity achieves 72.36%, classification accuracy accomplishes 77.18%, False Positive Ratio (FPR) attains 27.64 and False Negative Ratio (FNR) achieves 17.53. Likewise, 50%, 60%, 70% training data classification performance are described in Table 2.

,

Figure 8 represents the comparative analysis of three encryption techniques such as TDES, DES and AES. The overall analysis shows that the encryption time varies from 50,000 to 58,000. In the medical big data privacy analysis, Triple DES (TDES) achieves most secure data and stored in the cloud. Compared to DES and AES, the proposed encryption method requires minimum time to encrypt the medical data.

Figure 9 demonstrates the ROC curve for ABAC performance for mapper 5 and mapper 10. The curve is drawn between FNR (varies from 17.5 to 21) and FPR (varies from 20.62 to 17.5) ratio. The function of the mapper is to generate an arbitrary number of intermediate key-value pairs. For each mapper, the false negative and the false positive ratio are evaluated to increase the efficiency of big data privacy analysis.

Figure 10 represents the average time analysis of data security based on a number of user attributes. The authentication process Attribute Based Access Control (ABAC) is not to define permissions directly between subjects and objects, but instead to use their attributes as the basis for authorizations. Such observation indicates that adding cloud servers can increase the privacy efficiency further.

Conclusion

The proposed medical big data security in cloud computing was analyzed by data encryption with access control process. With the intention of attaining efficient privacy analysis, Map Reduce framework with Optimal Fuzzy C means (OFCM)

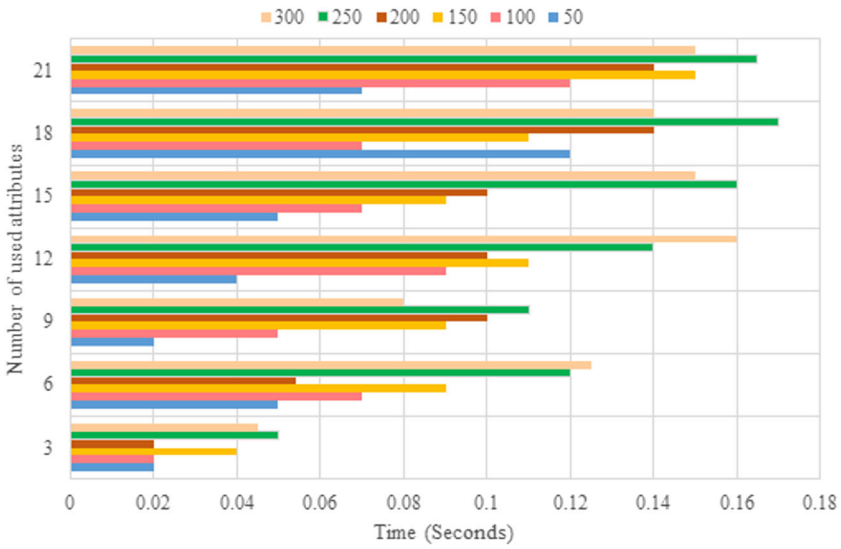

Fig. 10 Average time for ABAC for data security

was proposed to map the data, it decreases the execution period. And also the centroid optimization was done by KHO algorithm. For data classification, MKSVM classifier was utilized to classify sensitive and non-sensitive data in the cloud to encryption model. For enhancing the privacy of data encryption, Triple DES (TDES) was proposed to improve the scheme's performance; finally, authentication phase using attribute based access control was utilized to authenticate data in cloud sim. From the proposed model the clustering, classification and encryption results are compared to existing approaches. The comparative analysis concluded that the proposed TDES encryption technique accomplished minimum execution time with highly secured data compared to existing methods.

Compliance with Ethical Standards

Conflict of Interest This paper has not communicated anywhere till this moment, now only it is communicated to your esteemed journal for the publication with the knowledge of all co-authors.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Mathew, P. A., Dunn, L. N., Sohn, M. D., Mercado, A., Custudio, C., and Walter, T., Big-data for building energy performance: Lessons from assembling a very large national database of building energy use. *J. Appl. Energ.* 140:85–93, 2015.
2. Cecchinel, C., Jimenez, M., Mosser, S., and Riveill, M., An architecture to support the collection of big data in the internet of things. Proceedings of 2014 IEEE 10th World Congress on Services. 442–449, 2014.
3. Vennila, S., and Priyadarshini, J. J., Scalable privacy preservation in big data a survey. *Proc. 2nd Int. Conf. Big Data Cloud Comput.* 50: 369–373, 2015.
4. Bernice M. Purcell, "Big data using cloud computing", Journal of Technology Research, pp.1–8, 2014
5. Fernández, A., del Río, S., López, V., Bawakid, A., del Jesus, M. J., Benítez, J. M., and Herrera, F., Big data with cloud computing: An insight on the computing environment, MapReduce, and programming frameworks. *Article: Wiley Interdiscip. Rev.: Data Mining Knowl. Discov.* 4(5):380–409, 2014.
6. Goli-Malekabadi, Z., Sargolzaei-Javan, M., and Akbari, M. K., An effective model for the store and retrieve big health data in cloud computing. *J. Comput. Methods Programs Biomed.* 132:75–82, 2016.
7. Litchfield, A. T., and Althouse, J., A systematic review of cloud computing, big data, and databases on the cloud. *Proc. Conf. Inform. Syst.* 1–19, 2014.
8. Neves, P. C., Schmerl, B., Bernardino, J. and Cámara, J., Big data in cloud computing: Features and issues: 1–8, 2013.
9. Sadashiv, N. and DilipKumar, S. M., Cluster, grid and cloud computing: a detailed comparison. Proceedings of 6th International Conference on Computer Science & Education. 447–482. IEEE 2011.
10. Kchaou1, H., Kechaou, Z., and Alimi, A. M., Towards an offloading framework based on big data analytics in mobile cloud computing environments. *Proc. INNS Conf. Big Data* 53:292–297.
11. Shen, J., Liu, H., Shen, J., Tan, H. and He, D., Privacy preserving search schemes over encrypted cloud data: A comparative survey. Proceedings of first international conference on computational intelligence theory, systems and applications. 197–202, 2015.
12. Zhang, Q., Yang, L. T. and Chen, Z., Privacy preserving deep computation model on cloud for big data feature learning. *Trans. Comput.* 1351–1362, 2016.
13. Sanse, K., and Sharma, M., Clustering methods for big data analysis. *Proc. Int. J. Adv. Res. Comput. Eng. Technol.* 4:642–648, 2015.
14. Yeo, C. S., Buyya, R., Pourreza, H., Eskicioglu, R., Graham P. and Sommers, F., Cluster computing: high-performance, high-availability, and high-throughput processing on a network of computers.521–551, 2016.
15. Gheid, Z. and Challal, Y., Efficient and privacy-preserving k-means clustering for big data mining. TrustCom-BigData SE-ISPA: 791–798, 2016.
16. Kim, S.-H., Jung C. and Lee, Y.-J., An entropy-based analytic model for the privacy-preserving in open data. Proceedings of international conference on Big Data. 3676–3684, 2016.
17. Jha, S., Kruger, L. and McDaniel, P., Privacy preserving clustering. ESORICS: 397–417, 2017.
18. Shen, W., Yang, G., Yu, J., Zhang, H., Kong, F., and Hao, R., Remote data possession checking with privacy-preserving authenticators for cloud storage. Proceedings of Future Generation Computer Systems. 1–46, 2017.
19. Li, P., Chen, Z., Yang, L. T., Zhao, L. and Zhang, Q., A Privacy-preserving High-order Neuro-Fuzzy c-Means Algorithm with Cloud Computing. Proceedings of Neurocomputing. 1–15, 2016.
20. Pasupuleti, S. K., Ramalingam, S., and Buyya, R., An efficient and secure privacy-preserving approach for outsourced data of resource constrained mobile devices in cloud computing. *Proc. J. Netw. Comput. Applic.* 64:12–22, 2015.
21. Cao, X., Li, H., Dang, L. J., and Lin, Y., A two-party privacy preserving set intersection protocol against malicious users in cloud computing. *Proc. Comput. Standards Interf.* 54:41–45, 2016.
22. Hu, G., Di Xiao, T. X., Bai, S., and Zhang, Y., A compressive sensing based privacy preserving outsourcing of image storage and identity authentication service in cloud. *Proc. Inform. Sci.* 387:132–145, 2017.
23. Jesu, J., Nayahi, V., and Kavitha, V., Privacy and utility preserving data clustering for data anonymization and distribution on Hadoop. *Proc. Future Gen. Comput. Syst.* 74:393–408, 2016.
24. Tsai, C.-F., Lin, W.-C., and Ke, S.-w., Big data mining with parallel computing: A comparison of distributed and MapReduce methodologies. *Proc. J. Syst. Softw.* 122:83–92, 2015.
25. Ye, A., Yacheng Li, and Li, Xu, Anovellocation privacy-preserving scheme based on l-queries for continuous LBS. Proc. Comput. Commun. 1–10, 2016.
26. Zhang, Q., Yang, L. T., Chen, Z., and Li, P., PPHOPCM: Privacy-preserving high-order Possibilistic c-means algorithm for big data clustering with cloud computing. *IEEE Trans. Big Data*. 1–11, 2017.
27. Yang, M.-S., and Nataliani, Y., Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. *J. Pattern Recogn.* 71:45–59, 2017.
28. Abualigah, L. M., Khader, A. T., Al-Betar, M. A., and Gandomic, A. H., A novel hybridization strategy for krill

herd algorithm applied to clustering techniques. *J. Appl. Soft Comput.* 60:423–435, 2017.

29. Feng, C., and Liao, S., Scalable Gaussian kernel support vector machines with sublinear training time complexity. *J. Inform. Sci.* 418:480–494, 2017.
30. Karthik and Muruganandam, Data encryption and decryption by using triple DES and performance analysis of crypto system. *J. Sci. Eng. Res.* 2(1):24–31, 2014.
31. Sookhak, M., Yu, R., KhurramKhan, M., Xiang, Y., and Buyya, R., Attribute-based data access control in mobile cloud computing: Taxonomy and open issues. *J. Fut. Gen. Comput. Syst.* 72:273–287, 2017.