International Journal of Parallel Programming (2020) 48:515–533
https://doi.org/10.1007/s10766-018-0592-8

Triple DES: Privacy Preserving in Big Data Healthcare

R. Ramya Devi[1] · **V. Vijaya Chamundeeswari**[1]

Received: 20 May 2018 / Accepted: 30 July 2018 / Published online: 16 August 2018

Abstract
Big data stand as a technique to retrieve, collect, manage and also analyze a vast quantity of structured and also unstructured data which are tough to process utilizing the traditional database that involves new technologies to examine them. With the expanding success of the big data usage, loads of challenges emerged. Timeless, scalability and privacy are the chief problems that researchers endeavor to work out. Privacy-preserving is at present a highly active domain of research. To guarantee a safe and trustworthy big data atmosphere, it is imperative to pinpoint the drawbacks of the existing solutions furthermore conceive directions for future study. In the given paper, the security and also the privacy-preserving on big data is proposed concerning the healthcare industry and to beat security issues in existing approach. Mainly anonymizations along with Triple DES techniques aimed at security purpose are incorporated. Triple DES offers a fairly simple technique of increasing the key size of DES to shield against such attacks, devoid of necessitates to design an entirely new block cipher algorithm. Data anonymization work as an information sanitizer whose target is to defend the data privacy. It encrypts or takes away the personally recognizable data as of the data sets in order that the persons about whom the data designate remain anonymous. In this work, a combination of anonymization and Triple DES are utilized that are shortly called as the A3DES algorithm. Experimental outcome reveals that the approach performed well when contrasted with all other related approaches.

Keywords Big data · Privacy-preserving · Anonymization and Triple Data Encryption Standard (Triple DES)

✉ R. Ramya Devi
ramyakathir@gmail.com

V. Vijaya Chamundeeswari
vijaychamu@gmail.com

[1] Department of Computer Science and Engineering, Velammal Engineering College, Chennai, India

1 Introduction

1.1 Background

The word "data mining" is recognized as the basic processes in knowledge discovery of databases (KDD) [1]. Data mining research handles the extraction of worthwhile information from a huge mass of data with application areas like market-basket analysis, customer relationship management, etc. [2]. The individual's complete data generally embraces some sensitive information. Dispensing such data directly disrupts individuals' privacy [3]. In recent epoch, Privacy-preserving data mining method is highly essential on account of their increasing capacity to store users' private data. The holding of such private information is the complex process in data mining algorithm [4]. The main target of this technique is to device data mining approaches without maximizing the risk of mishandling of data which are utilized to create those methodologies [5]. To get the privacy preservation, most techniques apply some alterations on the real data [6]. This altered dataset is utilized for mining and it must meet privacy desires without losing the mining benefit [7].

1.2 State-of-the-Art Techniques

In recent epoch, methods like classification, k-anonymity [8], L-diversity [9] clustering, and also association rule mining are recommended to employ privacy-preserving data mining technique [8].

This technique is partition into 2 levels

- Level I of PPDM focuses on shielding the data that are sensitive, like, name, id, address, etc.
- Level II of PPDM focuses on protecting the sensitive knowledge which is exposed by data mining [9].

Map Reduce framework centered big data privacy preservation in Cloud environment were also used in previous techniques [12].

1.3 Problem Statement

A vital portion of Information technology research hard work goes into analyzing and also monitoring data concerning events on the servers, networks and other connected devices. Big data is a quite new concept in modern technological world. Lately, there is an increasing usage of big data, as the problem of security has become very important [13]. To deal with these challenges, service providers gradually more are adopting new techniques, counting the utilization of a third-party auditor to confirm the integrity of data saved in the cloud, and access control centered on data attributes along with semantics, etc. [14]. Numerous privacy-preserving procedures utilize certain sort of alterations to accomplish security [10]. Privacy-preserving is essentially centered on data twisting, data remaking, and also data encryption innovation [11]. Big data fundamentally changed the manner wherein the associations oversee, investigate and use

data in any industry [12]. Security and also protection are imperative issues in enormous data. Security is regularly characterized to ensure delicate data by identifiable social insurance data [13]. It centers about the utilization together with administration of person's near to home data like making arrangements and building up approval necessities to guarantee that patients' near to home data is being gathered, shared and used in right ways. Whilst security is regularly characterized as the insurance against unapproved access, with some including unequivocal say of availability and integrity [14]. It centers on shielding data from malevolent attacks and taking data for the benefit [15]. Despite that the security is fundamental for ensuring data; however, it's deficient for tending to protection.

A standout among the encouraging arenas, where huge data able to be get connected to make certain change is "medicinal services" [16]. Big health insurance data can (i) possibly enhance persistent results, (ii) anticipate outbreaks of plagues, (iii) increase significant bits of knowledge, (iv) maintain a strategic distance from preventable infections, (v) decrease the human services conveyance cost and (vi) ameliorate the personal satisfaction as a rule [17]. But, deciding on the allowable utilization of data whilst sustaining security and patient's privacy right is a challenging mission [18]. Even though big data stands useful in the improvement of medical science and vital for the success of the entire healthcare organizations, it is utilized if privacy and security problems are completely addressed. To assure a trustworthy and secure big data situation, it is compulsory to recognize the boundaries of prevailing solutions and predict directions for upcoming research.

Anonymization along with triple DES is proposed here. In this case, it usually alludes to hiding identifier attributes (attributes that exclusively identify individuals) like full name, voter id, license number, etc. The chief issue with data anonymization is that data might look anonymous but re-identification can be made effortlessly by linking it to other external data.

2 Related Work

Xu et al. [19] suggested privacy-preserving data integrity verification model via utilizing Fully Homomorphic encryption-centered Merkle Tree (FHMT) technique of streaming authenticated data structures aimed at Health-CPS. The architecture, design idea, formal definition, and security definition are clarified in detail. The key structure of this model includes data appending, initialization, scale expansion, verification and data query.

Kaur et al. [20] suggested PPCF structure on ADD centered on a multiparty random mask and polynomial aggregation techniques. Here, 2 phases were regarded namely: online prediction generation and off-line model generation. For the process of privacy-preservation 3 protocols are utilized and also these protocols get analyzed separately. For computing the vector length securely, Paillierhomomorphic encryption structure was used.

Yang et al. [21] recommended a secure system to develop a twofold access control approach. This methodology was self-adaptive in the normal situations and also in the emergency situations. In the former situations, the healthcare staff with secret keys containing appropriate attributes has the rights to access the data. In later application,

patient's historical medical data was retrieved by utilizing a password centered break-glass access methodology. A secured de-duplication method is designated to remove the duplicate medical files with matching data, which may get encrypted with diverse access strategies. Such de-duplication method overcomes the storage overheads of the big-data storage scheme. This method allows all the authenticated users to access the medical file after the de-duplication by the diversified original access strategies.

Lu et al. [22] presented a method, called Lightweight Privacy-preserving Data Aggregation (LPDA), intended for executing fog computing-ameliorate IoT which supports aggregation of data. The suggested LPDA was categorized by implementing Chinese Remainder Theorem, Paillierhomomorphic encryption practices, etc. to aggregate hybrid IoT devices information into one and also to formerly filter injected incorrect data in the network edge. Comprehensive security analysis displays LPDA is certainly secure and privacy was enhanced with diversified privacy methods. In addition, extended performance evaluations were accompanied, and the outcomes specify LPDA is certainly lightweight in fog computing-enhanced IoT.

Ara et al. [23] suggested secure privacy-preserving data aggregation (SPPDA) structure centered on bilinear pairing to enhance data privacy and data aggregation efficacy of isolated health monitoring schemes. Bilinear ElGamal cryptosystem technique's homomorphic property is utilized by SPPDA to execute secure computation of privacy-preserving. Then associate it with the total signature system, which assists data authenticity/integrity on the WBAN. Under the decisional bilinear Diffie–Hellman assumption, the suggested SPPDA was ascertained as acceptably secure. Security analysis determines that the suggested structure preserves data privacy, confidentiality, and authenticity; it also withstands passive intrusion and playback attacks.

Rahman et al. [24] presented PriSens-HSAC structure that offers maximal privacy to RFID centered healthcare schemes. The PriSens constituent delivers enhanced privacy contrasted to the authentication protocols called RFID whilst considering an RFID tag in the healthcare setting. By utilizing P-RBAC strategy, the HSAC restricts unauthenticated access to patient's private information. Though the chief motivation was to augment the users' privacy in an RFID centered healthcare scheme, their recommended PriSens-HSAC structure also addresses all the security desires.

Abouelmehdi et al. [18] surveyed the top-notch security together with privacy difficulties on big data as implemented to healthcare field, evaluated how security together with privacy issues occurs in case of big healthcare data and discussed ways wherein they might be addressed. They mainly concentrated on the lately suggested methods centered on anonymization along with encryption, compared their strengths and limits, furthermore envisioned upcoming research directions.

3 Flowchart For Proposed System

In the given paper, a technique is proposed intended for the preservation of privacy in big data. Security method, say, data encryption and also anonymization technique is utilized. The encryption technique utilized here is Triple DES, which has the benefit of reliability, longer key-length, eliminates numerous attacks and diminishes the quantity of time. The workflow architecture intended for the system which is proposed is displayed in Fig. 1.

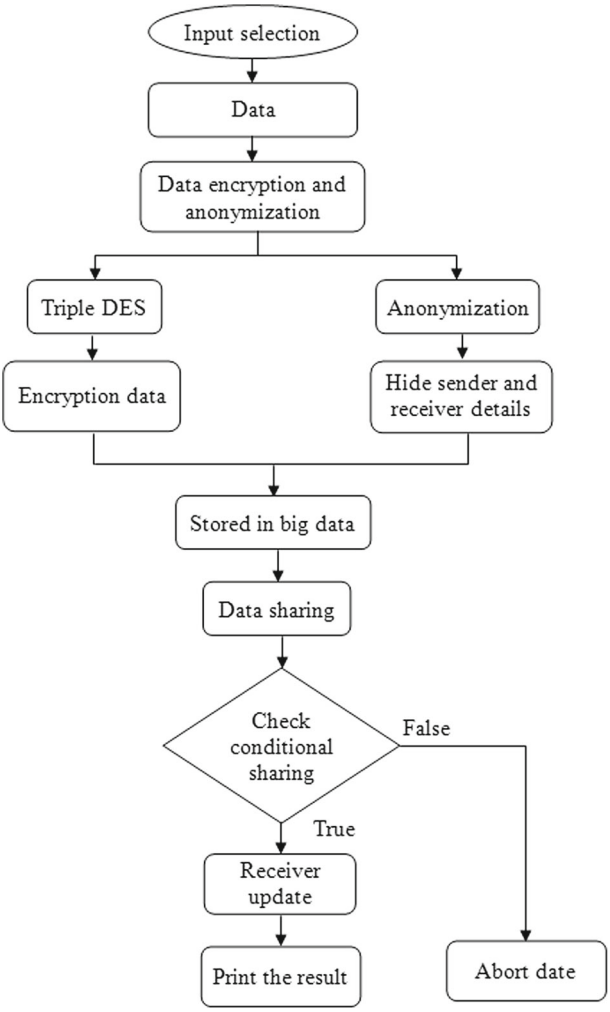

Fig. 1 Architecture diagram for proposed system

3.1 Input Selection

This is also labeled as data selection. The patient dataset is taken as the input since the healthcare field is considered. This dataset holds some attributes like name, age, gender, month, locations and symptoms of the patients. From this attributes, necessary data's are selected and inputted to the encryption along with anonymization techniques. Data stands as any assortment of characters which is collected and interpreted mostly for analysis purpose. It might be any character, counting text, numbers, pictures, sound or video. The inputted data can well be of a word or number, explicitly the patient name, gender, age, birth date.

The input data expression is discussed below:

$$S_i = S_1, S_2, S_3, \ldots, S_n \tag{1}$$

where S_i signifies the number of input selection and S_n is the n number of inputs in the selection field.

3.2 Data Encryption and Anonymization

After the data are collected, it must be encrypted. Encryption algorithm which is utilized here is Triple Data Encryption Standard (Triple DES) algorithm. Triple DES stands as an open encryption standard. It proffers 112-bit and also 168-bit strength encryption. It employs a symmetric key algorithm wherein the data get read, write or transmitted. In this proposed system, encryption together with anonymization methods is utilized to obtain better security. Triple DES algorithm is the best one, as it employs symmetric key generation and offers better security. After encryption, the data's are saved in the big data.

Triple DES technique is typically defined by,

$$E^1 = E^3 = E, E^2 = D \tag{2}$$

where E signifies the (single) DES encryption function, in addition D, signifies its decryption counterpart.

Anonymization technique is utilized for hiding the receiver's information. Data anonymization is defined as the technology which converts the clear data into an unreadable and also irreversible form, counting pre-image resistant hashes as well as the encryption methods where the decryption key is jettison. It enables the transfer of information across a boundary, for instance, betwixt two departments within an agency or betwixt two agencies, whilst diminishing the hazard of unintentional exposure. Additionally, in a certain specified situation, it enables the evaluation along with analytics post-anonymization.

The general form of anonymization is,

$$(personid, \{item_1, item_2, \ldots, item_n\}) \tag{3}$$

where $\{item_1, item_2, \ldots, item_n\}$ signifies the collection of items with person id.

Generally, in the medical data milieu, anonymized data signifies to data as of that the patient can't be identified by the one who receives the information. The name, address, and full postcode should be detached, accompanied by any other information that is in tandem with other data that could identify the patient if grasped by or revealed to the recipient.

De-anonymization stands as the procedure of cross-referencing the anonymous data with all the data sources with the intention of re-identifying its source i.e., reverse process. The two prominent anonymization approaches aimed at relational data are (i) Generalization, (ii) perturbation.

3.2.1 Triple DES

This is beneficial since it has an extensively sized key length that stands longer than the majority key lengths associated with all other encryption modes. The Advanced Encryption standard replaced the DES algorithm. Therefore DES is at the present

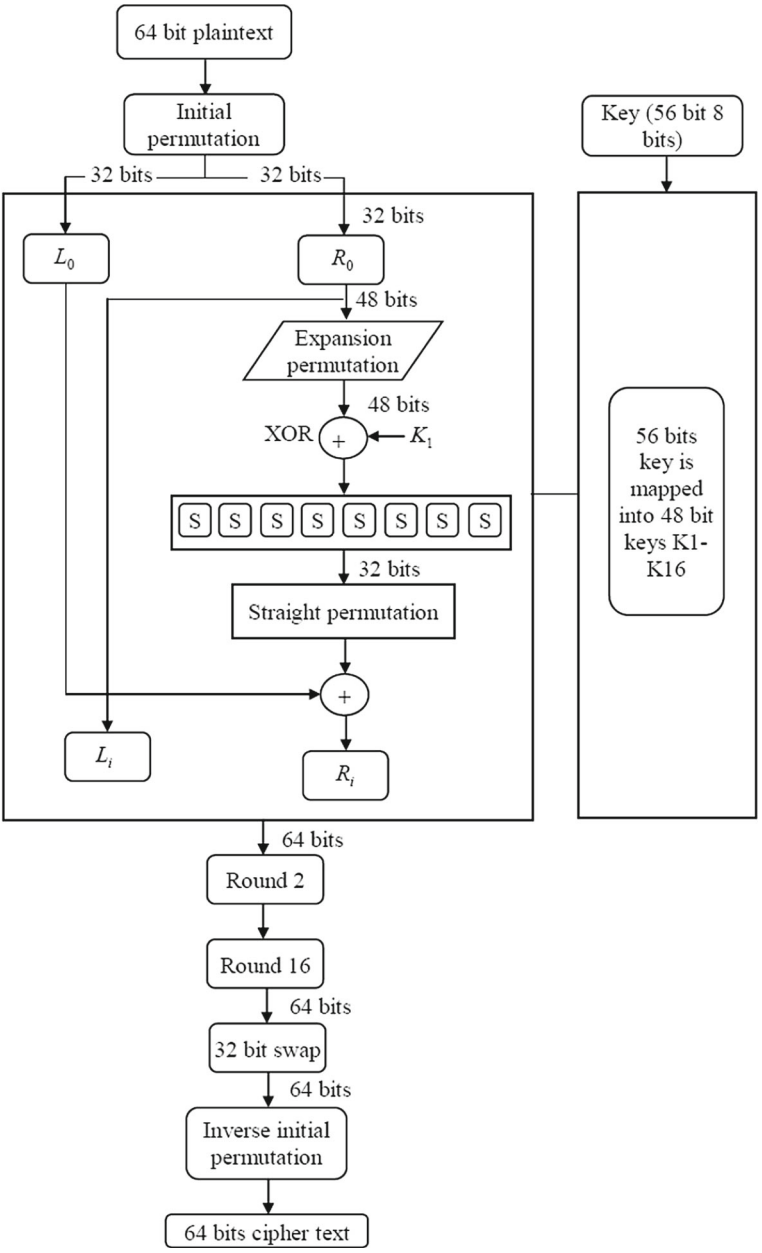

Fig. 2 General depiction of DES

regarded to be outdated. It is obtained by utilizing the single DES thrice and engages 3 subkeys and key padding when needed. Keys must be augmented to 64 bits in length. Recognized for its flexibility and compatibility it can effortlessly be converted aimed at Triple DES inclusion (Fig. 2).

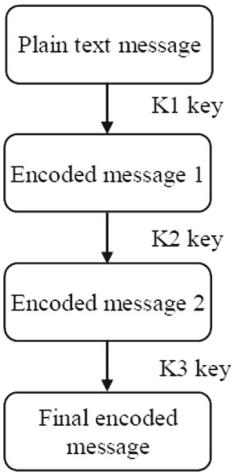

Fig. 3 Block diagram for 3DES

Several sorts of triple DES encryption are normally recognized:

- DES-EEE3: 3 DES encryptions using 3 diverse keys;
- DES-EDE3: a diverse key for each of the 3 DES operations (encryption, decryption, encryption);
- DES-EEE2 and DES-EDE2: a diverse key for the secondary operation (decryption) (Fig. 3).

Let $E_K(I)$ and $D_K(I)$ signify the DES encryption as well as decryption of I utilizing DES key K correspondingly. Every TDEA encryption/decryption operation stands as a compound operation of DES encryption as well as decryption operations. The successive operations are utilized:

1. TDEA encryption operation: a block I of 64-bit is transformed to a block O of 64-bit which is characterized as follows:

$$O = E_{K3}(D_{K2}(E_{K1}(I))) \tag{4}$$

2. TDEA decryption operation: a block I of 64-bit is transformed to a block O of 64-bit which is characterized as follows:

$$O = D_{K1}(E_{K2}(D_{K3}(I))) \tag{5}$$

The standard denotes the subsequent keying options for bundle $(K1, K2, K3)$

a. Keying Option 1: $K1$, $K2$ and $K3$—independent keys;
b. Keying Option 2: $K1$ and $K2$—independent keys and $K3 = K1$;
c. Keying Option 3: $K1 = K2 = K3$.

A TDEA means of operation stands backward consistent with its single DES counterpart if, with well-matched keying options aimed at TDEA operation,

1. An encrypted plaintext calculated using a single DES means of operation is decrypted properly by an equivalent TDEA means of operation;

Begin

1. Option 1, the preferred option, employs three mutually independent keys $(K1 \neq K2 \neq K3 \neq K1)$. It gives key space of $3 \times 56 = 168$ bits.
2. Option 2, employs two mutually independent keys and a third key that is the same as the first key $(K1 \neq K2 and K3 = K1)$. This gives key space of $2 \times 56 = 112$ bits.
3. Option 3, is a key bundle of three identical keys $(K1 = K2 = K3)$. This option is equivalent to DES Algorithm.

End

Fig. 4 Algorithm for Three Keying Options

2. An encrypted plaintext calculated utilizing a TDEA means of operation is decrypted properly by an equivalent single DES means of operation.

When employing Keying Option 3, TECB, TCBC, TCFB and also TOFB mode are backward consistent with single DES means of operation ECB, CBC, CFB, OFB correspondingly (Fig. 4).

3.2.2 Modified Anonymization

Let $\Phi : DB \rightarrow DB$ be a database transformation function, furthermore $\Delta \in DB$ be a database. The transformed database $\Phi(\Delta)$ is anonymized if

$$\forall\Psi\forall\Delta 1 \ldots \forall\Delta N \Pr(\Psi(\Phi(\Delta), \Delta 1, \ldots, \Delta N) * \Delta) \approx 0 \tag{6}$$

where

$$\Psi : DB^{N+1} \rightarrow DB(N = 0, \ldots, \infty) \tag{7}$$

is an arbitrary function, and

$$\Delta_i, i = 1, \ldots, N \tag{8}$$

are database representations of the entire available background knowledge If $\Phi(\Delta)$ is anonymized intended for any $\Delta \in DB$ then Φ is a data anonymization function.

Data anonymization is viewed as a process of facade or taking away the sensitive data as of a document whilst preserving its novel format. This process is vital for sharing data without exposing any sensitive information to the third parties that were present in databases or documents.

Anonymization of medical report is critical for publishing along with utilizing the clinical data for research reasons without the hazard of illegal access to the patients' identification. Most of those anonymization systems are applied to English texts. Though few of those systems are multilingual, there is not any specific anonymization system for Portuguese texts.

Anonymization is a cognitive process; practitioners ought to comprehend what could induce the identification of a person besides the apparent; direct data access physically or logically, carelessness, etc.

Every person encompasses some natural identifiers, i.e. data that typify an individual's; name, age, month, passport number, birth date, social security number, cellular phone number etc. some of it might not identify a person exclusively, but in proper contexts, they shall be presumed to be unique identifiers. In this paper, the term direct identifiers are used for these forms of data. There is a set of natural identifiers called indirect identifiers which together provide a unique identification, e.g. birth date, mother's name, address. One must notice that the data that are personal and data which enable identification of a human being are not necessarily different things. For example, thoughts, types of expression, activities, friends, medical case history, etc., may as well identify people; called as unintentional identifiers.

To protect privacy one must understand the potential threats, i.e. possible re-identification strategies:

- Direct re-identification: data themselves without any further action reveal the data subject identity.
- Re-identification via linking: sometimes, data set is believed to be de-identified while using publicly available or legally accessible databases that enable the re-identification of data subjects. For example, Netflix prize award dataset contains pseudonymized user ids and movie ratings. Netflix ratings were easily correlated to IMDB ratings where user ids are often personal names; re-identification was made possible through linking the preferences.
- Publishing anonymization algorithms or settings aimed at predictive algorithms: publishing always makes ways to de-construct or to invert applied functions using direct re-mapping, guessing, etc. If looking for information on an individual, one may deduce the future medical condition using medical predictive function centered on their observed symptoms.
- Re-identification via extremities: outlier values, rare or very unusual behavior are specific by definition to an extremely restricted number of persons that might lead to re-identification. For example, if inhabitants of a little town suffer from the same malady, it is easy to deduce the medical condition of an individual from that town.
- Background knowledge-based re-identification: sometimes, not structured, not stored, or single facts or knowledge known or accessed by a limited persons are applied to retrieve someone's identity, For example, custom habits of neighbors when and how they leave, or activities and photos published on a social network portal. Background knowledge is the utmost probable attacking strategy in the social networks era.
- Re-identification through event sequencing: frequencies or the ordering of data items also may exclusively identify certain individuals. A company based on a sick-leave registry may easily reveal employee medical condition if published, the health care database contains only dates and medical conditions.

Data anonymization as a cipher is also discussed in many preposition methods that are discussed below:

- For a data $Set^3 X_{(n,p)}$ with n records and p attributes $\left(X_1, \ldots, X_p\right)$, its anonymized version $Y_{(n,p)}$ can at all times be written, despite the anonymization methods utilized, as:

$$Y_{(n,p)} = \left(P_1 X_1, \ldots, P_p X_p\right)_{(n,p)} + E_{(n,p)} \tag{9}$$

where $P_1, \ldots, P_p$ is a collection of p permutation matrices and $E_{(n,p)}$ is a matrix of small noises.

- For a data set $X_{(n,p)}$ with n records and p attributes $\left(X_1, \ldots, X_p\right)$, its anonymized version $Y_{(n,p)}$ can at all times be written, despite the anonymization methods utilized, as:

$$Y_{(n,p)} = \left(A_1^T D_1 A_1 X_1, \ldots . A_p^T D_p A_p X_p\right)_{(n,p)} + E_{(n,p)} \tag{10}$$

where $E_{(n,p)}$ is a matrix of small noises, $Y_{(n,p)}$ anonymized version of $X_{(n,p)}$

- The three-tuple $\Gamma = (P, K, E)$ with the following conditions satisfied:

1. P is a finite set of possible original and anonymized data sets of $n \geq 2$ records and $p \geq 1$ attributes.
2. K is the key space, a finite set of possible key groups k, each containing p permutation-based keys.
3. For every key groups $k \in K$ there presents a collection of p permutation-based encryption rules $\varepsilon k \in E$, where each group

$$\varepsilon k : P \rightarrow P \tag{11}$$

is a function such that,

$$\varepsilon k(x) = y \quad for\ \forall x, y \in P \tag{12}$$

Is a cipher for data anonymization. Where εk is the permutation based encryption.

In the proposed method, noise anonymization method is utilized; the noise is added to the sensitive data. The attacker can't be certain concerning the real data value in any but all particular record. This solution inhibits exploring sensitive data, thus linking external data sources provide no further information. The proposed method is described as displayed in Tables 1 and 2.

Table 1 Patient's original records

ID	Date of Birth	Gender	Location
Annie	21-01-1988	Female	U.S
Bill	24-03-1986	Male	U.S
Dennis	27-02-1981	Female	U.K
Elise	21-01-1991	Male	Australia
Fred	24-03-1976	Male	U.S

Table 2 Anonymized patient's record

ID	Date of Birth	Gender	Location
Annie	21-1-1988	Female	U.S
Bill	24-1-1986	Male	U.S.
Dennis	27-2-1981	Female	U.K
Elise	21-3-1991	Male	Australia
Fred	24-3-1976	Male	U.S

Above table clearly explained the detail about anonymization technique, added the noise to the original records, that is to say, unwanted characters are incorporated in the original records. Thus, it camouflages the sender–receiver information.

In noise anonymization technique, particle swarm optimization (PSO) algorithm is utilized for selecting the string position. Since, this technique reduces the selection time and it as well gives the speed of the system which is proposed. Some information concerning the PSO algorithm is discussed below:

3.3 Big Data Storage

This chiefly supports storage and also input/output operations on storage with a numerous data files and objects. A typical big data storage structural design is composed of a redundant along with a scalable supply of direct attached storage (DAS) pools, scale-out or clustered network attached storage (NAS) or else an infrastructure centered on object storage format. The storage infrastructure is linked to computing server nodes which allow quick processing and also retrieval of big measures of data. In this work, big data storage for healthcare is utilized.

3.4 Data Sharing

It implies that the data are stored in big data additionally conditional sharing method is used to prevent the same dataset from being changed by two people simultaneously. Data sharing is a prime attribute of the database management system (DBMS).

3.5 Conditional Sharing

In this work, conditional sharing has 3 stages that are,

- Registration
- Login
- Authentication

These steps are discussed below,

3.5.1 Registration

In Registration phase, the user requires registering at the server by providing appropriate identification details. The server processes the user data and provides the login details

The procedure is as follows:

1. User (A) provides the username, password, and personal details.
2. Server (S) verifies the availability of user details.
3. Server confirms the username, password, and secret code.
4. Server compute

$$J = h(ID \oplus h(PW \oplus x)) \tag{13}$$

where J signifies the user and h refers to the hash function in this step SHA-256 Cryptographic Hash Algorithmis used. A cryptographic hash (called 'digest') is a sort of 'signature' aimed at a text or else a data file. SHA-256 produces a nearly-unique 256-bit (32-byte) signature aimed at a text. ID Means identification, PW means password then x means secret key.

3.5.2 Login

This phase is summoned when the user desires to login into the clouds. The users are confirmed before getting access to the cloud.

The procedure is depicted below:

1. The user enters the username, password together with the secret code.
2. Waiting for Authentication.

3.5.3 Authentication

This phase is processed in the server where the server will make a choice whether A should be allowed to log in or not. The authentication phase process is as given below:

1. Server (S) compute,

$$J_1 = h(ID \oplus h(PW \oplus x)) \tag{14}$$

and check if

$$J_1 = J \tag{15}$$

where h refers to the hash function in this step we used SHA-256 Cryptographic Hash Algorithm. A cryptographic hash (called 'digest') is a sort of 'signature' aimed at a text or else a data file. SHA-256 produces a nearly-unique 256-bit (32-byte) signature aimed at a text. ID Means identification, PW means password then x means secret key then go on to the subsequent step, otherwise, terminate the data.

Begin

1. Input selection from the patient dataset (i.e.,) $S_1, S_2, \ldots\ldots, S_n$
2. Take particular data.
3. Data is to be encrypted and anonymized.
4. Data is to be encrypted using Triple DES algorithm.
   Keying option1:
   All three keys are independent triple-length key $3\times56=168$ independent key bits.
   Keying option2:
   K_1 & K_2 Are independent.
   $K_3 = K_1$ Is a double-length keys.
   It provides shorter key length $2\times56=112$ bits.
   Keying option3:
   All three keys are identical $K_1 = K_2 = K_3$
5. Used modified noise anonymization technique.
   For modification used PSO optimization algorithm for choosing string location.
6. Encrypted data's are stored in big data.
7. Data sharing
8. Conditional Sharing
   $J = h(ID \oplus h(PW \oplus x))$
   $J_1 = h(ID \oplus h(PW \oplus x))$
   If $J_1 = J$
   Proceed to next step.
   Else
   Abort selected data.

End

Fig. 5 Pseudo Code for A3DES algorithm

4 A3DES Algorithm Steps (Fig. 5)

4.1 Algorithm Steps

1. The Input is selected from the dataset followed by encryption and anonymization.
2. Data encryption is done utilizing Triple DES algorithm as explained in Fig. 4.
3. At the Modified anonymization technique stage PSO optimization algorithm is employed for the selection of string position.
4. Finally, conditional sharing operation is done. If the condition is satisfied the data is shared. Otherwise, the selected data is aborted.

5 Result and Discussion

The intended approach is employed to lessen the security issue in the Big data healthcare storage operation. The method which is proposed is employed in the JAVA with CloudSim utilizing the database which is regarded as the yardstick for basic scheduling

troubles. The intended approach implementation is contrasted to the other conventional algorithms.

5.1 Performance Analysis

5.1.1 Security in Encryption

The security level in encryption of the proposed and the existing techniques are calculated using (16).

$$Encryption\ (\%) = \frac{File\ size\ (in\ MB)}{Percentage\ (100\%)} \tag{16}$$

5.1.2 Performance

The performance of the proposed and the existing methods are calculated using (17)

$$Performance = \frac{No.\ of\ operation}{Percentage} \tag{17}$$

5.2 Comparative Analysis

5.2.1 Security Comparison in Encryption

The results exhibiting the supremacy of A3DES algorithm over other prevailing algorithms in considering the Security in encryption is displayed in Fig. 6. The outcomes are compared with algorithms like Single DES along with AES. Security in encryption is compared centered upon File Size (MB). A3DES algorithm offers better security as it encrypts the data 3 times. For further security, the anonymization technique is

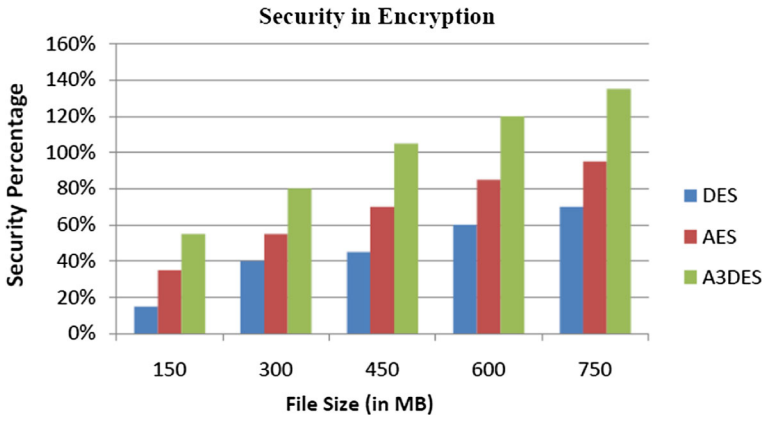

Fig. 6 Security comparison in encryption

Table 3 Security comparison in decryption

File size (in MB)	DES (%)	AES (%)	A3DES (%)
200	20	40	60
400	30	50	90
600	45	80	110
800	60	95	120
1000	80	115	145

employed. Therefore, it is the well-secured technique on contrasting with all other algorithms. Those comparisons are clarified in the graph displayed below:

In Fig. 6 the file size is taken on the horizontal axis as well as the security percentage is taken on the vertical axis. The graph is obtained utilizing (16). The security percentage for data encryption for file sizes 150 MB, 300 MB, 450 MB, 600 MB, and 750 MB were appeared to be 58%, 80%, 105%, 120% and 135% respectively. This is advanced than the existing techniques.

5.2.2 Security Comparison in Decryption

In the proposed work, the utmost famous modern encryption decryption algorithms are compared. In Table 3, AES, DES along with 3DES algorithms is compared with Security in percentage and file size (MB). The A3DES algorithm decrypts the data 3 times thus it's more secure compared with other decryption algorithms. In this technique symmetric key generation is utilized thus it is better. So, it is the well secured technique contrasted with other algorithms. The comparison of the decryption shown in Table 3.

The percentage of security after decryption for file sizes 200 MB, 400 MB, 600 MB, 800 MB, 1000 MB were appeared to be 60%, 90%, 110%, 129%, 145% respectively which are superior than the exiting techniques.

5.2.3 Anonymization Comparison

The noise anonymization technique is modified. That is used for more security power when added to the big data storage. The result is contrasted with that of the existing works and is provided in Fig. 7. Proposed technique gives the better result.

In Fig. 7, tasks are taken along the horizontal axis and accuracy is taken on the vertical axis. The anonymizations for tasks 10, 20, 30, 40, 50 were appeared to be 7, 115, 6, 20, and 45 respectively. Thus shows the supremacy over other techniques

5.2.4 Comparing Performance Result

At the finish of the complete process, after encryption as well as decryption, the output files are contrasted with that of the other prevailing algorithms in percentage wise. This performance comparison in Fig. 8 clearly shows the security of the proposed system.

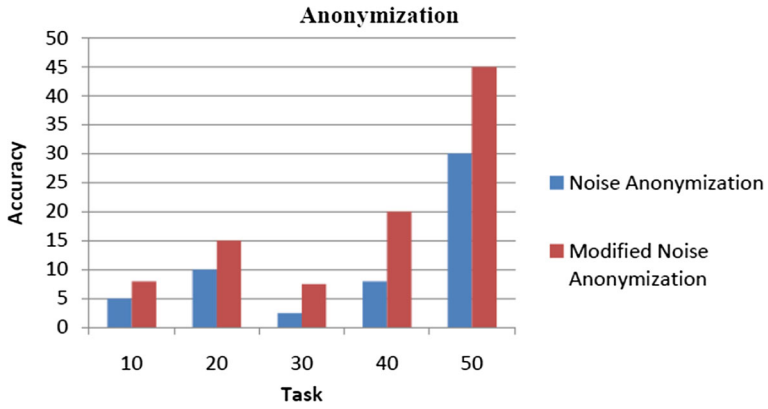

Fig. 7 Anonymization comparison

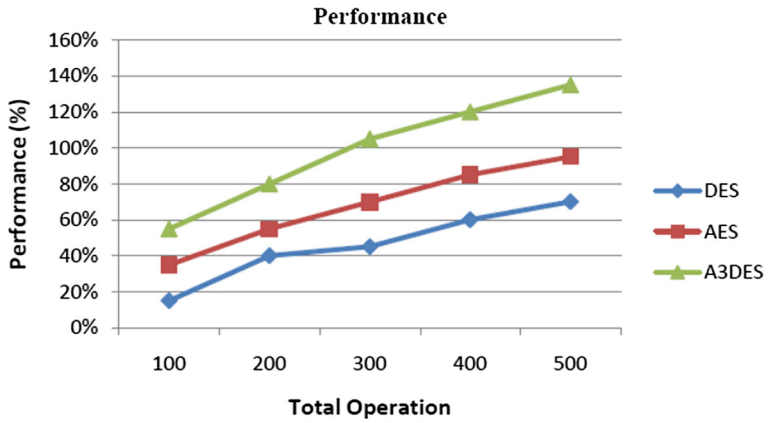

Fig. 8 Performance result comparison

So, A3DES algorithm is a better-secured algorithm for the big data healthcare storage operation.

In Fig. 8 total operations is taken along the horizontal axis and performance in percentage is taken along the vertical axis. The graph is plotted using (17). The percentage of performance for 100, 200, 300, 400, 500 were observed to be 59%, 80%, 105%, 129%, 135% respectively

6 Conclusion

In given paper, A3DES algorithm is introduced; it is the combination of Anonymization and triple DES algorithm. It is developed for security purpose in big data healthcare. The security is improved using A3DES algorithm. Implementation of the proposed secure big data storage is performed in the working platform of JAVA with CloudSim. This proved that advantages of the triple DES technique that includes better security

during encryption as well as decryption, better accuracy. The security percentage aimed at data encryption for file sizes 150 MB, 300 MB, 450 MB, 600 MB, and 750 MB were observed to be 58%, 80%, 105%, 120% and 135% respectively. The percentage of security after decryption for file sizes 200 MB, 400 MB, 600 MB, 800 MB, 1000 MB were appeared to be 60%, 90%, 110%, 129%, 145% respectively. The anonymizations for tasks 10, 20, 30, 40, 50 were appeared to be 7, 115, 6, 20, and 45 respectively. The percentage of performance for 100, 200, 300, 400, 500 were observed to be 59%, 80%, 105%, 129%, 135% respectively. These results were greater than the existing techniques. Thus the experimental product demonstrated that the proposed system is better and well secured than all other existing ones. This works could be focused towards improving the accuracy in future.

References

1. Gulia, N., Singh, S., Sapra, L.: A study on different classification models for knowledge discovery. Int. J. Comput. Sci. Mob. Comput. **4**(6), 241–248 (2015)
2. Ngai, E.W.T., Xiu, L., Chau, D.C.K.: Application of data mining techniques in customer relationship management: a literature review and classification. Expert Syst. Appl. **36**(2), 2592–2602 (2009)
3. Nivetha, P.R., Thamaraiselvi, K.: A survey on privacy preserving data mining techniques. Int. J. Comput. Sci. Mob. Comput. **2**(10), 166–170 (2013)
4. Reddy, P.S., Ravi, C.: A novel technique for privacy preserving data publishing. Int. J. Comput. Sci. Mob. Comput. **3**(11), 156–163 (2014)
5. Samuel, S., Chen, S., Burr, D.L., Zhang, L.: A new data collection technique for preserving privacy. J. Priv. Confid. **7**(3), 99–129 (2017)
6. Punitha, N., Amsaveni, R.: Methods and techniques to protect the privacy information in privacy preservation data mining. Int. J. Comput. Technol. Appl. IJCTA **2**(6), 2091–2097 (2011)
7. Mohana Chelvan, P., Perumal, K.: On privacy preserving data mining and feature selection stability measures: a comparative analysis. Int. J. Comput. Eng. Technol. IJCET **9**(2), 1–15 (2018)
8. El Ouazzani, Z., El Bakkali, H.: A new technique ensuring privacy in big data: K-anonymity without prior value of the threshold k. Procedia Comput. Sci. **127**, 52–59 (2018)
9. Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H., Saadi, M.: Big data security and privacy in healthcare: a review. Procedia Comput. Sci. **113**, 73–80 (2017)
10. Qi, X., Zong, M.: An overview of privacy preserving data mining. Procedia Environ. Sci. **12**, 1341–1347 (2012)
11. Ganesh, D., Mahendran, S.K.: Privacy preservation for data mining security issues. In: International Journal of Computer Applications (0975–8887). International Conference on Current Trends in Advanced Computing (ICCTAC-2015), pp 33–39 (2015)
12. Vennila, S., Priyadarshini, J.: Scalable privacy preservation in big data a survey. Procedia Comput. Sci. **50**, 369–373 (2015)
13. Singh, M., Halgamuge, M.N., Ekici, G., Jayasekara, C.S.: A review on security and privacy challenges of big data. In: Sangaiah, A., Thangavelu, A., Meenakshi Sundaram, V. (eds.) Cognitive Computing for Big Data Systems Over IoT. Lecture Notes on Data Engineering and Communications Technologies, vol. 14. Springer, Cham (2018)
14. Li, S., Gao, J.: Security and privacy for big data. In: Yu, S., Guo, S. (eds.) Big Data Concepts, Theories, and Applications, pp. 281–313. Springer, Cham (2016)
15. Divecha, H., Mehta, S.: Privacy preserving based on geometric transformation using data perturbation technique. Int. J. Softw. Hardw. Res. Eng. **2**(5), 6–13 (2014)
16. Aldeen, Y.A.A.S., Salleh, M., Razzaque, M.A.: A comprehensive review on privacy preserving data mining. SpringerPlus **4**(1), 694 (2015)
17. Shorfuzzaman, M.: Leveraging cloud based big data analytics in knowledge management for enhanced decision making in organizations. Int. J. Distrib. Parallel Syst. **8**(1), 1–13 (2017)
18. Abouelmehdi, K., Beni-Hessane, A., Khaloufi, H.: Big healthcare data: preserving security and privacy. J. Big Data **5**(1), 1 (2018)

19. Xu, J., Wei, L., Wu, W., Wang, A., Zhang, Y., Zhou, F.: Privacy-preserving data integrity verification by using lightweight streaming authenticated data structures for healthcare cyber-physical system. Future Gener. Comput. Syst. (2018). https://doi.org/10.1016/j.future.2018.04.018
20. Kaur, H., Kumar, N., Batra, S.: An efficient multi-party scheme for privacy preserving collaborative filtering for healthcare recommender system. Future Gener. Comput. Syst. **86**, 297–307 (2018)
21. Yang, Y., Zheng, X., Guo, W., Liu, X., Chang, V.: Privacy-preserving smart IoT-based healthcare big data storage and self-adaptive access control system. Inf. Sci. (2018). https://doi.org/10.1016/j.ins.2018.02.005
22. Lu, R., Heung, K., Lashkari, A.H., Ghorbani, A.A.: A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT. IEEE Access **5**, 3302–3312 (2017)
23. Ara, A., Al-Rodhaan, M., Tian, Y., Al-Dhelaan, A.: A secure privacy-preserving data aggregation scheme based on bilinear ElGamal cryptosystem for remote health monitoring systems. IEEE Access **5**, 12601–12617 (2017)
24. Rahman, F., Bhuiyan, M.Z.A., Ahamed, S.I.: A privacy preserving framework for RFID based healthcare systems. Future Gener. Comput. Syst. **72**, 339–352 (2017)