Journal of Cloud Computing:
Advances, Systems and Applications

RESEARCH

Open Access

BGNBA-OCO based privacy preserving attribute based access control with data duplication for secure storage in cloud

M. Pavithra[1*], M. Prakash[2] and V. Vennila[3]

Abstract

Cloud computing technology offers flexible and expedient services that carry a variety of profits for both societies as well as individuals. De-duplication techniques were developed to minimize redundant data in the cloud storage. But, one of the main challenges of cloud storage is data deduplication with secure data storage.To overcome the issue, we propose Boneh Goh Nissim Bilinear Attribute-based Optimal Cache Oblivious (BGNBA-OCO) access control and secure de-duplication for data storage in cloud computing in this paper. The proposed method achieves fne-grained access control with low computation consumption. We design Boneh Goh Nissim Privacy Preserving Revocable Attribute-based Encryption that reinforces attribute revocation and averts the discharge of sensitive information. Furthermore, we utilize Optimal Cache Oblivious algorithm to prevent disclosure of access patterns to hide the access patterns in cloud storage via rand pattern matching. We support updating both encrypted data and access control policies to minimize communication and computation overhead of data duplication and encryption processes concurrently. We perform secure data sharing to achieve higher data confidentiality and integrity. Finally, we conducted the extensive experiments in cloud and the results illustrated that our proposed BGNBA-OCO method is more efficient than related works.

Keywords Data duplication, Access control, Attribute-based encryption, Cloud computing, Boneh Goh Nissim, Bilinear, Optimal cache oblivious

Introduction

Cloud computing deeply facilitates different types of services for data providers who want to store their sensitive data. A secure data deduplication approach is employed in cloud storage to reduce the storage space while eradicating the data copies. Several secure data sharing and deduplication have been developed. However, existing deduplication methods failed to offer sufficient security for sensitive data. Hence, a proposed secure deduplication technique is needed that combines encryption and access control to guarantee the confidentiality and privacy of data.With the rapid growth of cloud computing and sharing the of large-volume data, secure sensitive information sharing with data deduplication is a significant process in cloud-assisted systems. Several secure data sharing and deduplication have been developed.

*Correspondence:
M. Pavithra
pavithra.m79@gmail.com
[1] Department of Computer Science and Engineering, S.S.M College of Engineering, Namakkal (Dt), Tamil Nadu 638183, India
[2] Department of Data Science and Business Systems, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India
[3] Department of Computer Science and Engineering, K. S. R. College of Engineering, Tiruchengode, Tamil Nadu, India

A secure data-sharing scheme was developed in [1] with data deduplication to protect sensitive information confidentiality and improve deduplication efficiency. However, the data integrity rate was not improved. An efficient and secure attribute-based access control scheme was designed in [2] for multiple data distribution to considerably minimize the computation cost. But the communication overhead in the data storage remained a challenging task. A ciphertext-policy attribute-based encryption method was developed in [3] based on attribute revocation procedure by proxy re-encryption. However, it failed to implement the proposed scheme with better feasibility. An attribute-based storage system was developed in [4] to improve the duplicate detection. But it failed to improve the confidentially of the shared data with specifying an access policy.

A novel Modified Elliptic Curve Cryptography (MECC) method was introduced in [5] for secure deduplication on cloud storage. However, the MECC method was not efficient for increasing security with minimum computation time and storage in the cloud environment. A secure and scalable data deduplication approach was designed [6] for dynamic user managing securely and to avoid unauthorized cloud users from sensitive data. Though the approach reduces the unnecessary communication overhead, the computation overhead was not reduced. A multi-user updatable encryption approach was designed [7] for storing the remote ciphertext in the cloud. But the performance of data confidentiality remained unaddressed.

A verifiable deduplication approach named VeriDedup was developed in [8] to ensure the accuracy of duplication and provide flexible integrity over encrypted data. But the overhead analysis in the duplication was not analyzed. A novel encrypted deduplication storage method was introduced in [9], by analyzing the deduplication to metadata. But the integrity level was not enhanced. An effective secure deduplication approach was developed in [10] to provide authorized data access and maximally eliminate duplicates without affecting the security and privacy of cloud users. The main contribution of the BGNBA-OCO Method is summarized in the following.

To enhance the security of data sharing and data deduplication efficiency, a novel BGNBA-OCO method is introduced.

First, the Boneh Goh Nissim Privacy Preserving Revocable Attribute-based Encryption is performed to store the multiple patient files on cloud storage. To minimize the communication overhead in the deduplication, Optimal Cache Oblivious algorithm is designed with the help of rand matching coefficient. This helps to minimize the repeated data storage on cloud server.

Next, the attribute based access control policy is applied for authorization to download data from server. This inturn improve data confidentiality rate.

Additive Linear Secret Sharing matrix is applied to avoid the data modification and alteration for enhancing the data integrity level. Finally, the extensive experiments are conducted to evaluate the performance of our BGNBA-OCO method and other existing works.

Related works

In this section, previous Data deduplication technique was developed with cloud storage providers. Several secure data deduplication methods were introduced for different scenarios to ensure security of users' sensitive data. But, the high amount of data in storage was not considered. A lightweight rekeying-aware encrypted deduplication approach (REED) was designed in [11] for a re-encryption deduplication storage system. The one-way hash function was applied in the REED method to resist the stub-reserved attack and enhance the data privacy of data owners' sensitive data. Moreover, REED effectively reduced the computation overhead of the system but the communication overhead and time were not effectively reduced.

In order to minimum computation time overhead, a deduplication strategy based on the Merkle hash tree was developed in [12] for secure data sharing against brute-force attacks. The designed system efficiently supports the file- and block-level dedup to improve the rate of data deduplication for access control. But the designed strategy failed to apply large-scale cloud computing for reducing the resource wastage rate.

Enhanced Symmetric Key Encryption Algorithm (ESKEA) was developed in [13] for secure data storage with data deduplication to enhance data confidentiality. The designed Convergent Encryption (CE) algorithm to verify the service provider duplicates copies of data. In addition, ESKEA algorithm utilized a Spider Monkey Optimization Algorithm (SMOA) for optimally selecting the secret key. Lastly, the recovered key was employed to retrieve original data. However, the integrity of data storage was not improved.

A new privacy-preserving, revocable ciphertext policy attribute-based encryption (PR-CP-ABE) approach was developed in [14] to preserve sensitive outsourced data. The designed integrated secure deduplication method was used to enhance the storage efficiency of cloud service provider while preserving the data privacy. However, the approach consumed more time for attribute-based encryption.

A secure aggregation-based tag deduplication method (ATDS) was designed in [15] for side channel attack recognition through the public verification. The Lagrangian interpolation-based aggregation method was developed to attain tag deduplication. By applying this technique, content-associated public key was used to achieve auditing. However the higher performance data integrity level was not achieved.

An efficient and privacy-preserving big data deduplication method was developed in [16] to achieve both privacy-preserving and data availability also resists the brute-force attacks. The designed method utilized a three-tier cross-domain structural design for secure and privacy-preserving deduplication with big data in cloud storage. However, the proposed method failed to protect the duplicate information in healthcare industry.

The healthcare industry faces demand regarding the security and efficiency of data organization for patient health records. Hence, a secure and efficient medical data sharing method was developed in [17] with the purpose of deduplication over the cloud encrypted storage. The designed method achieved a viable less cloud storage space overhead, and enhances the system performance with high fault-tolerance capability. However, communication overhead in the deduplication over the cloud-encrypted storage was not minimized. A secure encrypted data deduplication model was developed in [18] to find out the different encrypted data created from similar plaintext. The ciphertext policy attribute-based encryption was performed to preserve the tags for data uploader through the cloud server in an offline manner. But it has a high computational overhead.

A modified ramp secret sharing (MRSS) approach was designed in [19] to minimize the storage overhead with privacy-preserving data processing applications. The designed MRSS scheme was utilized for any privacy preserving data processing that performs linear operations on the data. However, the confidentiality of privacy-preserving data processing was not improved. An Elliptical Curve Cryptographic and to generate key the Chinese Remainder Theorem (ECC-CRT) based deduplication model was developed in [20] to remove the frequent data on cloud storage. The cosine similarity checking was employed to perform the Deduplication it avoids malicious upload and downloads in storage space. However, the efficient algorithm was not implemented for focusing on the time consumption while generating the keys.

A blockchain-assisted data sharing method was proposed in [21] to perform the detection of deduplicated data and end the group of ciphertext after deduplication. However, the designed method was failed to improve the computational overhead ciphertext deduplication. In [22], Data de-duplication and recovery method was proposed into public key searchable encryption by involves the public key searchable encryption and proxy re-encryption. But, the proposed method was efficiently minimized the users storage fixed cost and enhance the work efficiency.

A data redundancy method designed in [23] to save cloud storage resources by minimizing the duplicate data in cloud server storage. However, more secure deduplication in a cloud storage system with better throughput and reduced deduplicate elimination ratio to save cloud storage. A cloud data auditing scheme was carried in [24] to supporting file and authenticator deduplication. The proposed scheme is initially sensible one to accurately achieve low-entropy security. But, also reduce the cloud's storage overhead considerably.

Point-of-interest (POI) category recommendation model was developed in [25] to preserve users' check-in records. Locality-sensitive hashing (LSH) was designed for categorizing similar users into identical groups. POI category was utilized to improve mine the user's interests. A new privacy-preserving POI recommendation model was employed in [26] to ensure user privacy protection. However, the communication overhead was not minimized.

Summary

In the related works discussed above, most of the schemes rely on data duplication. Some of the schemes achieved secure data duplication, but their using of public key encryption caused significant overhead to the process. In this paper, we propose BGNBA-OCO achieves secure deduplicaion of data for data storage in CC. Boneh Goh Nissim Bilinear Attribute-based cryptography and Optimal Cache Oblivious algorithm utilized for minimizing storage overhead, computation overhead.

Boneh Goh Nissim Bilinear Attribute-Based Optimal Cache Oblivious (BGNBA-OCO) Methods

Data sharing plays a vital role in cloud-supported electronic medical systems. Electronic medical records consist of disease-related information about patients. In order to conceal the patient's sensitive data, security is a key factor prior to uploading the electronic medical record to the server. During the sensitive data uploading, there are large amounts of duplicate data in electronic medical records, which incurs unnecessary storage in the cloud server.

In order to overcome the above problems, an efficient secure data sharing with an access control scheme with data deduplication and sensitive information hiding is required in cloud-assisted electronic medical systems. Based on this motivation, a novel method called Boneh Goh Nissim Bilinear Attribute-based Optimal Cache Oblivious (BGNBA-OCO) is developed for access control and secure de-duplication efficiency for data storage by protecting the sensitive information.

Figure 1 depicts the architecture diagram of the proposed BGNBA-OCO method for secure data sharing and de-duplication in cloud computing environment.

System model

The system model BGNBA-OCO method involves four entities such as patients, hospital, cloud server, and Medical practitioner.

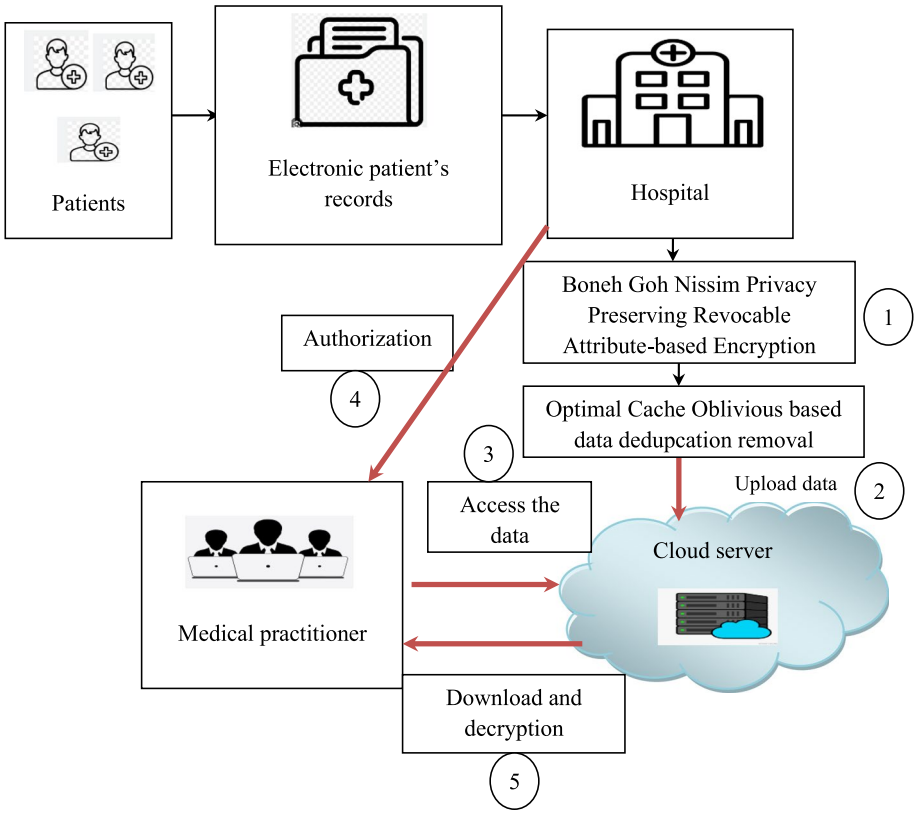

Fig. 1 Architecture diagram of proposed BGNBA-OCO method

- *Patients or cloud users:* The architecture includes a number of users or patients '$P = \{p_1, p_2, \ldots, p_n\}$ who dynamically generates the sensitive patient file $PI = \{PI_1, PI_2, \ldots, PI_m\}$.
- *Hospital:* The hospital is an entirely confidential place. The hospital generates vast amounts of electronic patient records day by day that need to be transferred to the cloud server for sharing with medical practitioners. To ensure security, the hospital needs to encrypt the patient file and upload the data to the cloud server for further processing. If medical practitioners want to obtain patient information, he/she first needs to get authorization from the hospital. Therefore, it uses the access control policy for authorization.
- *Cloud server:* The cloud server possesses enormous storage space and powerful computing capability. After receiving the ciphertext of the patient information, the cloud server performs a deduplication operation to improve the efficiency of storage space. Meanwhile, the cloud server recognizes the authorized user identity and performs deduplication. Besides, the hospital distributes patient information to medical practitioners via the cloud server under the condition of the access control policy.
- *Medical practitioner:* he/she wants to access the patient information from the server, they first need to obtain authorization from the hospital and then download information from the cloud server using the Additive Linear Secret Sharing matrix. Then the authorized practitioner decrypts and gets the original patient information under the conditions. This in turn improves the data confidentiality as well as integrity.

Security model

A Security Model exhibits the security of data transmission is designed between the server over the network to prevent the attackers and enhance the confidentiality or authenticity of the information that is being transmitted through the network. Let us consider the number of patient file $PI_1, PI_2, \ldots PI_m$, Hospital or data owner '*HS*, cloud server (*CS*), Medical practitioner. The secure data transmission is performed between server and medical practitioner.

Figure 2 illustrates the network security model of BGNBA-OCO method. Design an Boneh Goh Nissim Privacy Preserving Revocable Attribute-based algorithm for performing the security-related transformation between sender and Receiver. Before data storage in cloud server, Optimal Cache Oblivious algorithm is designed for data deduplication to minimize the storage space. Attribute secret key based user authorization is performed to validate the user authenticity. Finally, the authorized receiver performs decryption and gets the original patient file.

Boneh Goh Nissim Privacy Preserving Revocable Attribute-based Encryption

First process of the proposed BGNBA-OCO method is an attribute-based encryption before uploading the patient information into cloud server. The proposed technique uses the Boneh Goh Nissim cryptosystem to perform data encryption. The hospitals receive the patient's information $PI = \{PI_1, PI_2, \ldots, PI_m\}$.

Figure 3 illustrates the flow process of the Boneh Goh Nissim Revocable Attribute-based Encryption to guarantee the confidential access control over encrypted data in cloud environment. The hospital or data owner first collects the patient information $PI = \{PI_1, PI_2, \ldots, PI_m\}$. The Boneh Goh Nissim based Revocable Attribute-based Encryption is applied a public key cryptography for performing the attribute based encryption to ensure data confidentiality. In the attribute based encryption approach, the attribute secret key of a user depends on the certain attributes. In such a system, the decryption is obtained only if the set of attributes of the user attribute key matched.

Let us consider the set of system attributes $A = \{A_1, A_2, \ldots, A_k\}$. By applying a Boneh Goh Nissim cryptosystem, the private and public key is generated through the bilinear map concept.

The system selects an additive cycle group Q_a and a multiplicative cycle group Q_m of prime order p, p is a generator of Q_a. Therefore, the bilinear map is structure as,

$$M : Q_a * Q_a \rightarrow Q_m \tag{1}$$

A bilinear map 'M' is a function that combining elements of cycle groups' $Q_a * Q_a$' to yield an element of a third cycle groups'Q_m, and is linear.

The public key is generated as given below,

$$P_{pb} = [K, Q_m, z, b] \tag{2}$$

Where, P_{pb} denotes a public key,K indicates a multiplication of two large prime numbers, 'P', Q_m denotes a cycle groups' of order 'M', z indicates a generator of the cycle groups 'Q_m'.

$$K = x_1 x_2 \tag{3}$$

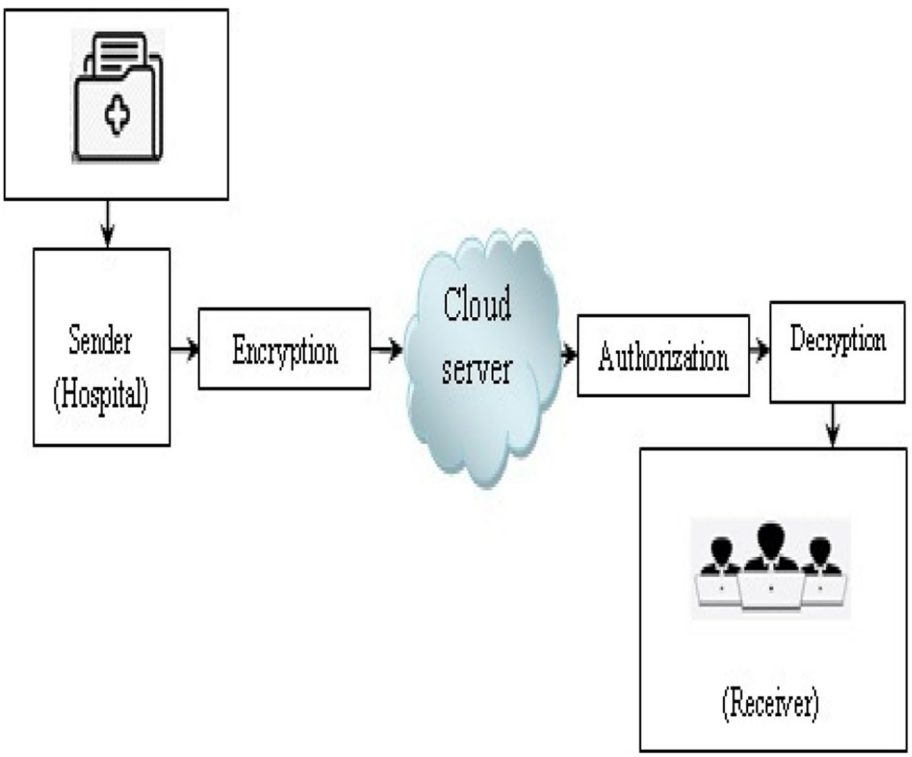

Fig. 2 Network security models

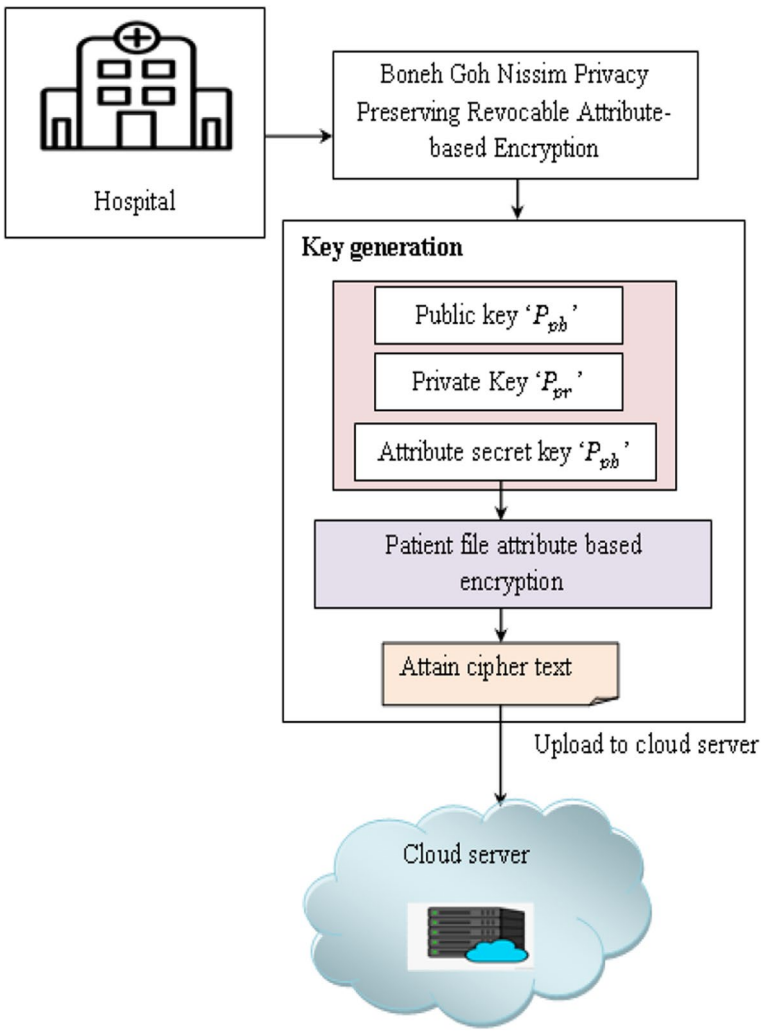

Fig. 3 Flow process of Boneh Goh Nissim revocable attribute-based encryption

$$b = a.x_2 \tag{4}$$

Where, x_1 and x_2 are the two large prime numbers, 'a' indicates a generator of the cyclic groups 'Q_m'.

Followed by, the private key is generated as given below,

$$P_{pr} = x_1 \tag{5}$$

Where, P_{pr} indicates a private key, x_1 denotes a large prime number. Then the attribute secret key is generated depends on the set of user attributes $A = \{At_1, At_2, \ldots, At_k\}$. Here the user attributes are name, mail ID, etc.

$$Sk = A \tag{6}$$

Where, Sk denotes an attribute secret key related with an attribute list 'A'. This key is used at the time of data decryption. When the decryption is performed, an access policy is defined with the attribute secret key. A ciphertext is decrypted by a user only if the user's attribute list matched. This helps to ensure the confidentiality of patient information.

The data owner performs data encryption with receiver public key. Encryption is the significant process of altering the original patient health information into an indecipherable form. The original data is called as plaintext and the encrypted data represented as cipher text.

Definition 1 (encryption)

Let us the patient's file $PI = \{PI_1, PI_2, \ldots, PI_m\}$ to be stored in a cloud server (CS). The encryption is the process of altering the original information into cipher text is given below,

$$CT = (PI * z) + (r * b) \tag{7}$$

Where, CT indicates a cipher text, PI indicates a patient's file, The symbol of z, b specified as a public key. The two keys are used for encryption with receiver public key. The public key is used to encrypt data securely before it is sent over the internet, r indicates a random number $r \in [0, 1, 2 \ldots n-1]$. The cipher text of the input data and is transferred into cloud server for avoiding unauthorized access. This in turn helps to prevent discharge of sensitive information.

Optimal Cache Oblivious algorithm based data deduplication

Once the data gets encrypted, the hospital or data owner needs to store their file on the cloud server. The server verifies the data deduplication ratio to minimize the computation and communication overhead. The proposed BGNBA-OCO method uses the Optimal Cache Oblivious algorithm for increasing the deduplication efficiency of sensitive patient file storage on the cloud server.

The cloud server permits to store the sensitive patient file if it has a high duplicate ratio. The optimal Cache Oblivious algorithm executes well on a multilevel memory hierarchy without identifying any parameters of the hierarchy, only identifying the existence of a data structure. The cache in the cloud server holds a large volume of data. Then the cloud server performs the caching to analyze the patient records in the memory hierarchy. During the analysis, it verifies whether the memory hierarchy with similar content is present in the existence of a data structure through the rand matching coefficient.

***Definition 2 data dedepulication*:**

Rand matching coefficient is a statistical technique used to checks whether patient records exist in the cache hierarchy.

$$RMC = \frac{Matching\ files}{Available\ files\ stored\ in\ cache} \tag{8}$$

Where RMC indicates a Rand matching coefficient that returns '1' for accurate matching and '0' for not matched. If the file is already presents, it indicates a 1' and the cloud server abort the file storage. Otherwise, it permits to upload the file in cache of cloud server for further processing.

This helps to avoid the data deduplication and also minimizes the excessive storage space. The algorithmic process of attribute-based Encryption and deduplication is described as given below.

Input: Dataset, Patient file $PI_1, PI_2, \ldots PI_m$, Hospital or data owner 'HS, cloud server (CS).

Output: avoid deduplication and minimize the storage space

Begin

1. For each patient P_i

2. Hospital generate private key 'P_{pr}', public key 'P_{pb}' and attribute secret key 'Sk'

3 End for

4. Hospital perform data encryption to obtain cipher text ' CT ' using (7)

5. **For each incoming** CT

6. CS uses the optimal cache oblivious concept to search the patient file using (8)

7. ***If*** $(RMC = 1)$ **then**

8. file already exist

9. CS abort to upload the file

10. else

11. file not exist

12. CS permit to upload the file

13. End if

14. End for

15. End for

End

Algorithm 1: Boneh Goh Nissim Revocable Attribute-based Encryption and deduplication

The above algorithm [1] illustrates the procedure of secure medical data deduplication and data storage

in the cloud environment. For each patient, the cloud server generates the private, public, and attributes secret keys. With the generated keys, the hospital performs data encryption to convert the original data into cipher text. Before data uploading, the server verifies the data duplication. The server verifies the incoming cipher text into its cache hierarchy. If the content exists, the cloud server aborts the data storage to avoid data deduplication. Otherwise, it allows storing the data in the cache for further processing to minimize the storage overhead.

Authorization and decryption

The final process of the proposed BGNBA-OCO is to perform the authorization and decryption. In the cloud, whenever the medical practitioner needs to access the patient file from the cloud server, they first sent the authorization request to the hospital or data owner. After getting the request from the medical practitioner, the hospital verifies the medical practitioner's authenticity with the help of the access control policy based on a set of attributes. Based on the attribute verification, the hospital permits to download the patient file from the cloud server.

The *RMC* methods gives either or output. Rand matdefiing coefficient that returns '1' for accurate matching and '0' for not matched. If the file is already presents, it indicates a 1' and the cloud server abort the file storage. Otherwise, it permits to upload the file in cache of cloud server for further processing.

Figure 4 depicts the flow process of the authorization and decryption to obtain the original patient file from the cloud server. First, the medical practitioner send authorization request to hospital.

$$MP \overset{Authorization_{Req}}{\rightarrow} HS \tag{9}$$

Where, MP denotes medical practitioner, $Authorization_{Req}$ denotes an authorization request to hospital 'HS'. After getting the request, hospital performs authenticity verification through an attribute secret key created at the time of the key generation. In other words, an access policy is defined with the attribute secret key.

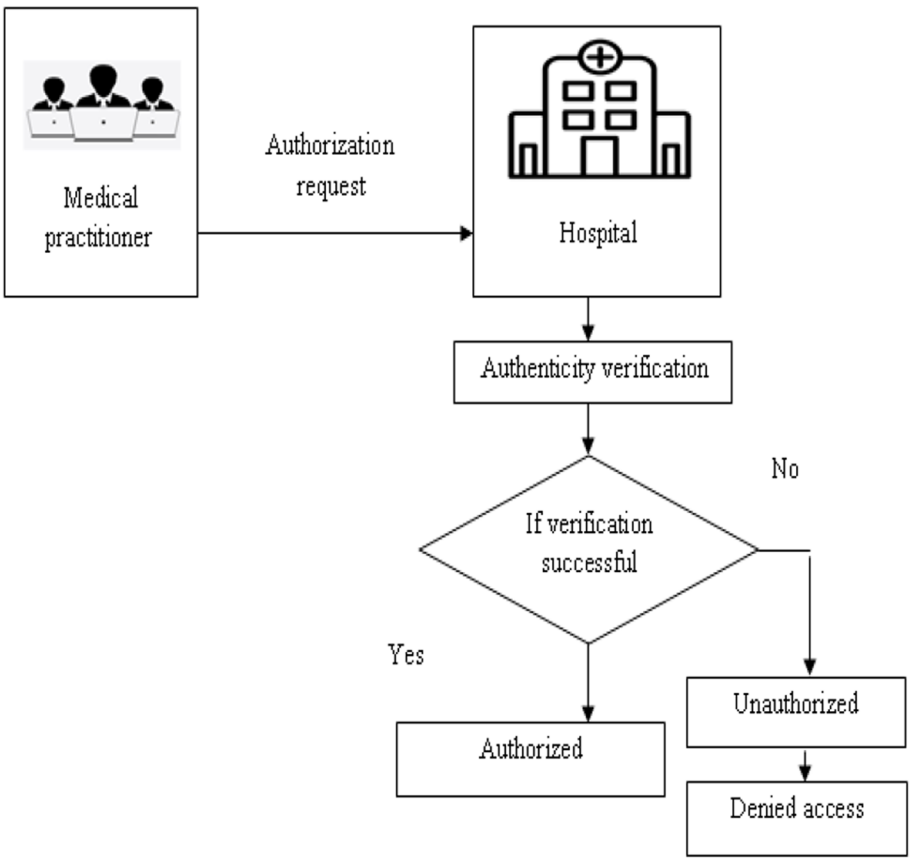

Fig. 4 Flow process of authentication and decryption

***Definition 3 authorization*:**
The medical practitioner enters the attribute secret key for proving their authenticity.

$$MP \xrightarrow{Sk\prime} HS \tag{10}$$

After receiving the attribute secret key '*Sk*ı', '*HS*' verifies whether the key is matched $Sk = Sk\imath$ or not $Sk \neq Sk$. If both the attribute secret keys are matched (i.e.$Sk = Sk\imath$), the hospital permits to access the patient file.

Definition 4 decryption
The medical practitioner accesses the data after the decryption with their private key. The decryption is the process of converting the cipher text into its original patient file. The authorized medical practitioner decrypts the data with their private secret key as give below,

$$PI = log_{x_1 z}[x_1 CT] \tag{11}$$

Where, '*PI*' designates a patient medical file, '*CT*' indicates a cipher text, 'x_1' represents a private key, '*z*' indicates a generator of the cyclic groups 'Q_m'. Finally, the medical practitioner performs decryption to obtain the original patient medical file.

Proof of corrections
The equation [11] is used for verifying the security proof,

$$PI = log_{x_1 z}[x_1 CT]$$

By applying the base of the logarithm rule,

$$log_{x_1 z} x_1 CT = \frac{\log x_1 CT}{\log x_1 z} = \frac{x_1 CT}{x_1 z}$$

$$PI = \frac{x_1 CT}{x_1 a}$$

$$PIx_1 z = x_1 CT$$

In order to prove that the above equation [7], the right hand side (RHS) functions is considered

$$\begin{gathered} x_1.CT = x_1.[(PI * z) + (r * b)] \\ \text{Where, } [CT = (PI * z) + (r * b)] \\ x_1.CT = (x_1 \times PI \times (x_1 \times r \times b) \\ = x_1 PIzx_1 \mathrm{r}(a.x_2) \\ (\text{whereb} = a.x_2) \\ x_1 PIz + ra(x_1 x_2) \\ x_1 PIz + raK(sinceK = x_1 x_2) \end{gathered}$$

Additive linear secret sharing matrix based access structure
Additive linear secret sharing scheme is a particular kind of secret sharing system where the secret value satisfies a linear relationship. It helps to perform a fine-grained access by the authorized user and also improve the efficiency of decryption with minimum time.

An Additive linear secret sharing matrix involves the process of dividing a secret value i.e. ciphertext into multiple blocks in the specified organization. It helps to avert a single block from having entire information of the original secret value. To obtain the secret information, all blocks must group their secret value collectively to expose the original secret value.

Figure 5 depicts the flow process of additive linear secret sharing based access structure to avoid illegal user access or modify the secret data.

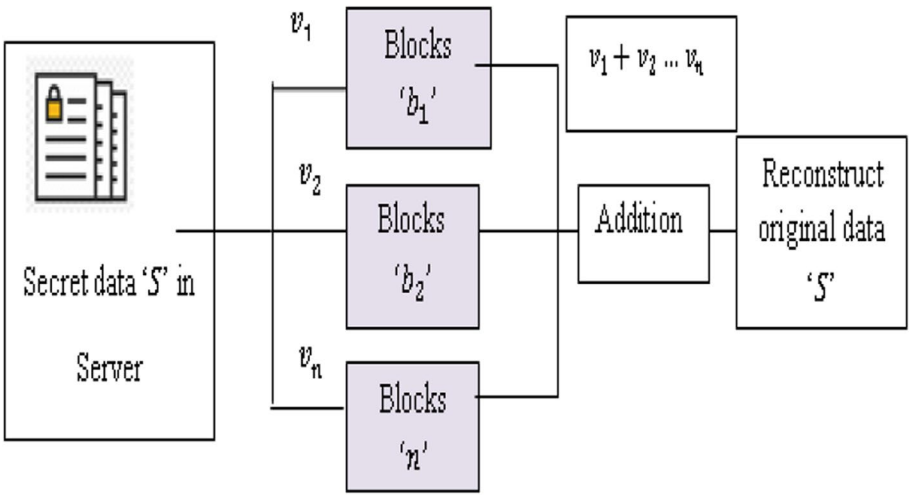

Fig. 5 Additive linear secret sharing based access structure

Definition 5 (Access structure):
Let the input secret data is divided into number of sub values $v_1, v_2, \ldots v_n$ and it distributed into medical practitioner with multiple blocks $b_1, b_2 \ldots, b_l$. The matrix 'H' with 'l' rows and n column is constructed with the blocks and the secret value.

$$b = [b_1, b_2 \cdots, b_l] \cdots where\,'l'\,\text{rows} \cdots \tag{12}$$

$$s = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \text{where} n\text{column} \tag{13}$$

Where, S denotes a secret value. In order to reconstruct the secret value, all the sub values are summed.

$$S = v_1 + v_{2 \cdots v_n \ldots} \tag{14}$$

In this way, access structure of the authorized user is defined to enhance the data integrity in cloud.

The algorithmic process of authorization and data decryption are given below,

Input: Cipher text ' CT'
Output: Improve the data confidentiality

Begin
1. **For each incoming** CT
2. **Apply** Additive linear secret sharing based access structure
3. ***If*** medical practitioner access data **then**
4. Medical practitioner send request to'HS'
5. HS verifies the authenticity with attribute secret key 'Sk'
6. ***If*** (Sk'=Sk'$_n$) **then**
7. Medical practitioner is said to be authorized
8. Access the original data
9. **else**
10. Medical practitioner is said to be unauthorized
11. Access denied
12. **End if**
13. **End if**
14. **If user access data**
15. **Divide** input secret data into number of subvalues $v_1, v_2, \ldots v_n$
16. Distribute into medical practitioner with several blocks
17. Sum all blocks value $v_1, v_2, \ldots v_n$
18. Obtain the original data
19. **End if**
20. Authorized user receives the patient file 'PI'
End

Algorithm 2: Authorization and data decryption

The algorithm of the authorization and data decryption is performed to improve data confidentiality and integrity. Whenever the medical practitioner needs to access the data from the server, they first verify the authorization by means of attribute secret key matching. When the secret key gets matched correctly, the cloud server permits to access the stored secret patient file. Otherwise, the cloud server denied access. Then the authorized entity performs decryption with their private key to get the original patient file. Then the proposed BGNBA-OCO technique also uses the additive linear secret sharing scheme for access structure policy in order to guarantee the integrity of data access.

Implementation setup

Experiment evaluation of BGNBA-OCO, existing secure data sharing scheme [1], efficient and secure attribute based access control scheme [2] are implemented using java and cloudsim simulator for secured data sharing and deduplication in cloud environment. In order to conduct the implementation, Heart failure clinical records dataset collected from UCI repository https://archive.ics.uci.edu/ml/datasets/Heart+failure+clinical+records. The dataset includes a 299 Heart patients records with 13 features. The features information is listed in Table 1.

Performance analysis and comparisons

In this section, comparative analysis of BGNBA-OCO, existing secure data sharing scheme [1], efficient and secure attribute based access control scheme [2] are presented for cloud environments with different valuation metrics such as,

- Communication overhead
- Computation overhead
- Data confidentiality rate
- Data Integrity rate

Communication overhead

It is measured a server side during the data storage. The overhead is an amount of memory space require for storing the patient files into the cache of the cloud server.

$$comm_{OH} = \sum_{i=1}^{n} PI_i * MS\,(PI) \tag{15}$$

Where, $comm_{OH}$ is a communication overhead, PI_i patient file, $MS(PI)$ is a memory space for storing the one patient file. It is measured as Mega bytes (MB). Below Table 2 shows the comparison data results of communication overhead.

Table 1 features description

S.No	Feature	Explanation	Measurement	Range
1	Age	Age of the patient	Years	(40, …, 95)
2	Anaemia	Decrease of red blood cells or hemoglobin (haematocrit levels were lower than 36%)	Boolean	0, 1
3	creatinine_ phos phokinase	Level of the CPK enzyme in the blood	mcg/L	(23, …, 7861)
4	Diabetes	If the patient has diabetes	Boolean	0, 1
5	ejection_ fraction	Percentage of blood leaving the heart at each contraction	Percentage	(14, …, 80)
6	high_ blood_ pressure	If a patient has hypertension	Boolean	0, 1
7	Platelets	Platelets in the blood	Kiloplatelets /mL	(25.01, …, 850.00)
8	serum_ creatinine	Level of creatinine in the blood	mg/dL	(0.50, …, 9.40)
9	serum_ sodium	Level of sodium in the blood	mEq/L	114, …, 148
10	Sex	Woman or man	Binary	0, 1
11	Smoking	If the patient smokes	Boolean	0, 1
12	Time	Follow-up period	Days	(4, …, 285)
13	DEATH_ EVENT	If the patient died during the follow-up period	Boolean	0, 1

Computation overhead

It is measured as an amount of time taken for secret sharing the patient files into the authorized client.

$$comp_{OH} = \sum_{i=1}^{n} PI_i * t(PI) \tag{16}$$

Where, $comp_{OH}$ indicates a computation overhead, PI_i patient file, $t(PI)$ denotes a time for storing one patient file. It is measured in terms of milliseconds (ms). Table 3 illustrates the data results value of computation overhead of this scheme.

Data confidentiality rate

It is the measure of number of data that are accessed by authorized access and denied access for unauthorized user in loud environment. The confidentiality rate is calculated using the following equation [17],

$$DCR = \sum_{i=1}^{n} \left[\frac{PICA}{PI_i} \right] * 100 \tag{17}$$

Where, DCR represent the data confidentiality rate, PI_i denotes the number of patient file, $PICA$ patient file correctly accessed. Data confidentiality rate is measured in percentage (%). Table 4 illustrates the data results value of data confidentiality rate values comparison of this scheme.

Data integrity rate

It measured as the number of data that are not altered or changed by any unauthorized users. The performance of integrity rate is mathematically calculated as given below,

$$DIR = \sum_{i=1}^{n} \left[\frac{PINA}{PI_i} \right] * 100 \tag{18}$$

Where, DIR indicates a data integrity rate, PI_i denotes the number of patient file, $PINA$ denotes a number of patient file not altered. Data integrity rate is measured in percentage (%). Table 5 illustrates the data results value of data integrity rate values of this scheme.

Figure 6, shows the comparison results of communication overhead. The more patient files the cloud server stores, the minimum communication overhead is. Figure 6 shows clearly that the communication overhead of cloud servers is influenced by deduplication efficiency as well as the number of files. In Fig. 6, the communication overhead of the BGNBA-OCO method is best with four methods. Let us consider the number of patient files improves, the communication overhead among these four methods becomes better. By the four methods, the deduplication efficiency of the BGNBA-OCO method is maximized, so the communication overhead of our method is reduced. This BGNBA-OCO method enhances the security of the server side. The cloud server uses an

Table 2 Comparison of communication overhead

No of patient files	Communication overhead (MB)			
	BGNBA-OCO	Secure data sharing scheme [1]	Efficient and secure attribute based access control scheme [2]	PR-CP-ABE [3]
25	11.25	13	15	17
50	13	14.1	16.5	18.5
75	16.5	18	19.5	21.75
100	20	22	24	26
125	22.5	25	27.5	30
150	24	27	29.25	31.8
175	26.25	28.87	31.5	33.95
200	27	29.8	32	35.2
225	28.8	31.05	33.75	36.67
250	31.25	33	35	37.5

Table 3 Comparison of computation overhead

Number of patient files	Computation overhead (ms)			
	BGNBA-OCO	Secure data sharing scheme [1]	Efficient and secure attribute based access control scheme [2]	PR-CP-ABE [3]
25	100	120	132.5	147.5
50	110	125	137.5	157.5
75	120	131.25	144.75	166.5
100	130	142	152	170
125	138.75	156.25	168.75	187.5
150	153	166.5	180	202.5
175	164.5	175	192.5	210
200	168	180	200	222
225	180	200.25	212.62	236.25
250	195	212.5	230	250

Table 4 Comparison of data confidentiality rate

Number of patient files	Data confidentiality rate (%)			
	BGNBA-OCO	Secure data sharing scheme [1]	Efficient and secure attribute based access control scheme [2]	PR-CP-ABE [3]
25	96	88	84	76
50	94	90	86	78
75	96	90.66	88	81.33
100	95	89	86	81
125	97.6	88.8	84.8	80
150	96.66	91.33	86.66	81.33
175	97.71	89.14	85.71	81.14
200	95.5	90.5	87.5	82
225	97.33	91.11	85.77	82.22
250	96	90	84	80.8

Table 5 Comparison of data integrity rate

Number of patient files	Data integrity rate (%)			
	BGNBA-OCO	Secure data sharing scheme [1]	Efficient and secure attribute based access control scheme [2]	PR-CP-ABE [3]
25	92	84	80	72
50	90	88	84	76
75	94.66	89.33	86.66	78.66
100	94	88	85	80
125	96.8	87.2	83.2	79.2
150	95.33	90	85.33	80.66
175	96.57	88.57	84.57	80.57
200	94.5	89.5	86	85.5
225	96.44	89.77	84.44	81.77
250	95.2	88.4	83.2	80

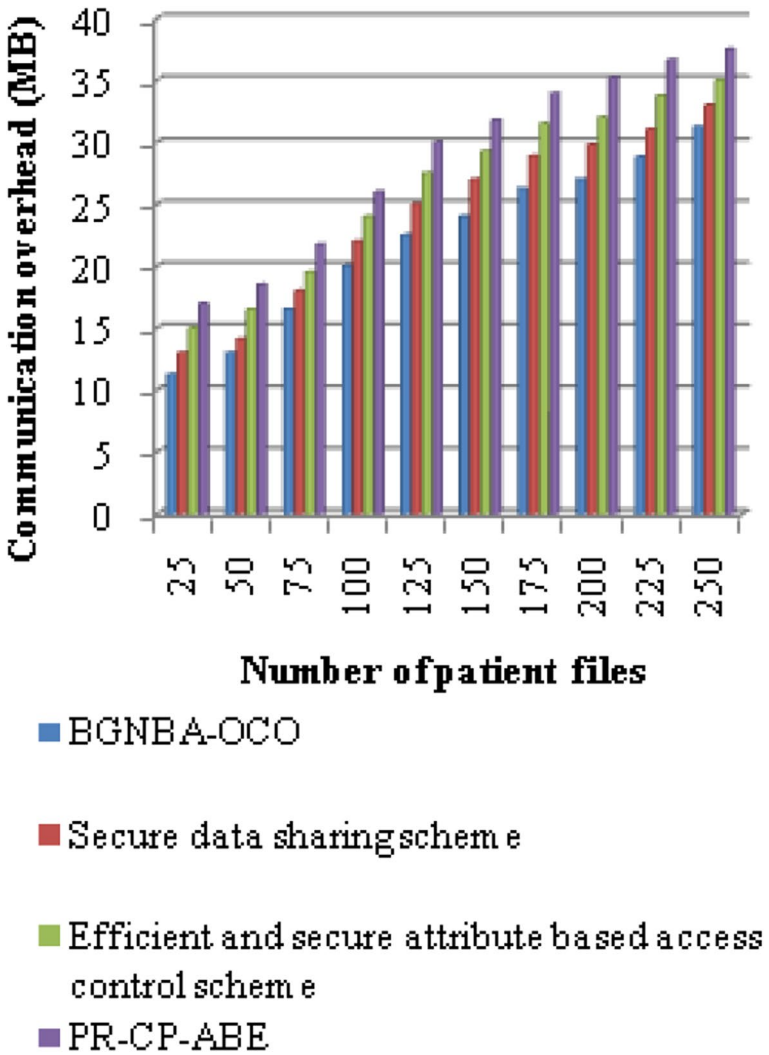

Fig. 6 Performance of communication overhead versus number of patient files

optimal cache-oblivious algorithm to find whether the patient file exists in the cache of cloud storage by means of rand pattern matching. This process avoids data deduplication and minimizes storage space.

The experiment is conducted with 25 patient files in the first iteration. The performance of communication overhead using the BGNBA-OCO method was found to be 11.25*ms*, whereas the communication overhead was found to be 13MB, 15MB, and 17MB using [1] [2] [3] correspondingly. Similarly, different performance results of communication overhead were observed. The experimental results of the BGNBA-OCO method are compared to conventional methods. The overall comparison results inferred that the communication overhead involved in data storage using the BGNBA-OCO method is minimized by 9%, 17%, and 24% when compared to existing [1] [2] [3] respectively.

In Fig. 7, describe the performance analysis of computation overhead using four schemes, BGNBA-OCO, existing [1, 2 and 3]. From the Fig. 7 it is conditional to computation overhead improved with number of patient files for all four methods. Also, the computation overhead by applying BGNBA-OCO is found to be comparatively smaller than the other two existing methods. This is due to the reason BGNBA-OCO technique effectively performs a key generation process for data encryption, decryption, and authentication using *Boneh Goh Nissim Bilinear* Attribute-based cryptography. This process minimizes the time consumption of secure data sharing between cloud servers and medical practitioners. The experiment is conducted with 25 number of patient files in the first iteration. The performance of computation overhead using computation overhead was found to be 100*ms*, whereas the computation overhead was found to be 120 *MB*, 132.5*MB* and 147.5*MB* using [1] [2] [3] respectively. From this result, it is inferred that the computation overhead of computation overhead was found to be comparatively smaller. Therefore the average results indicate that the performance analysis of computation overhead using BGNBA-OCO technique is decreased by 10%, 17%, and 26% when compared to [1] [2] and [3] respectively.

From above Fig. 8, the performance results of data confidentiality rate Vs. number of patient files. The data

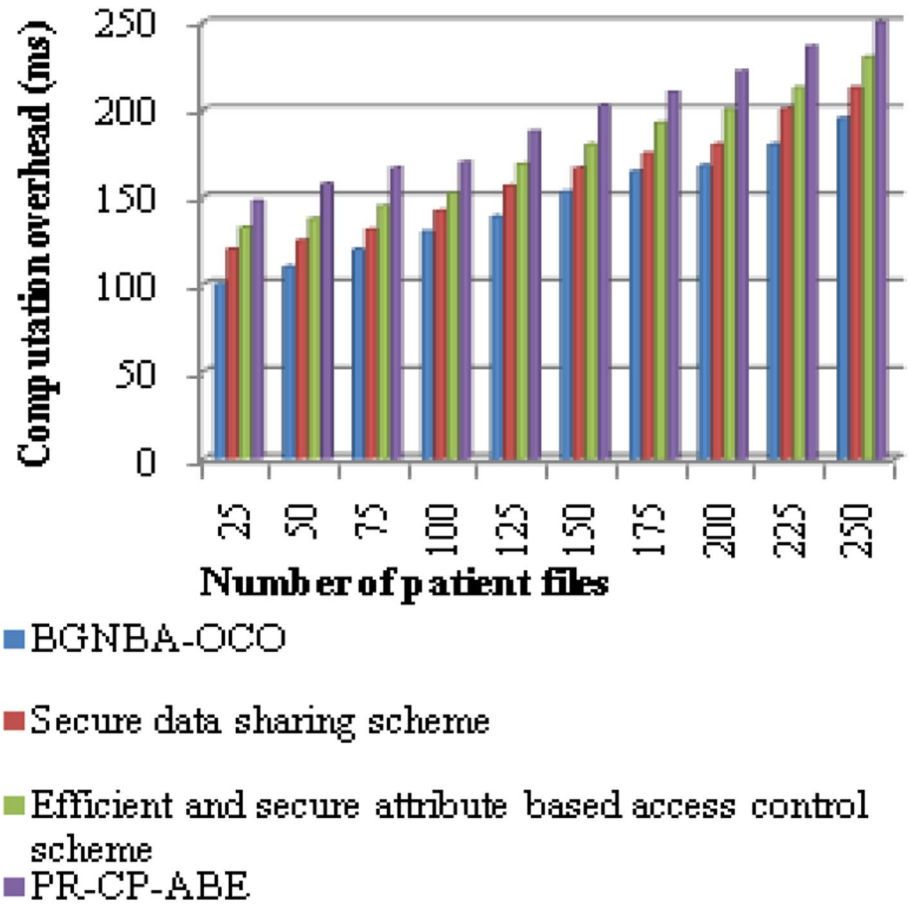

Fig. 7 Performance of computation overhead versus number of patient files

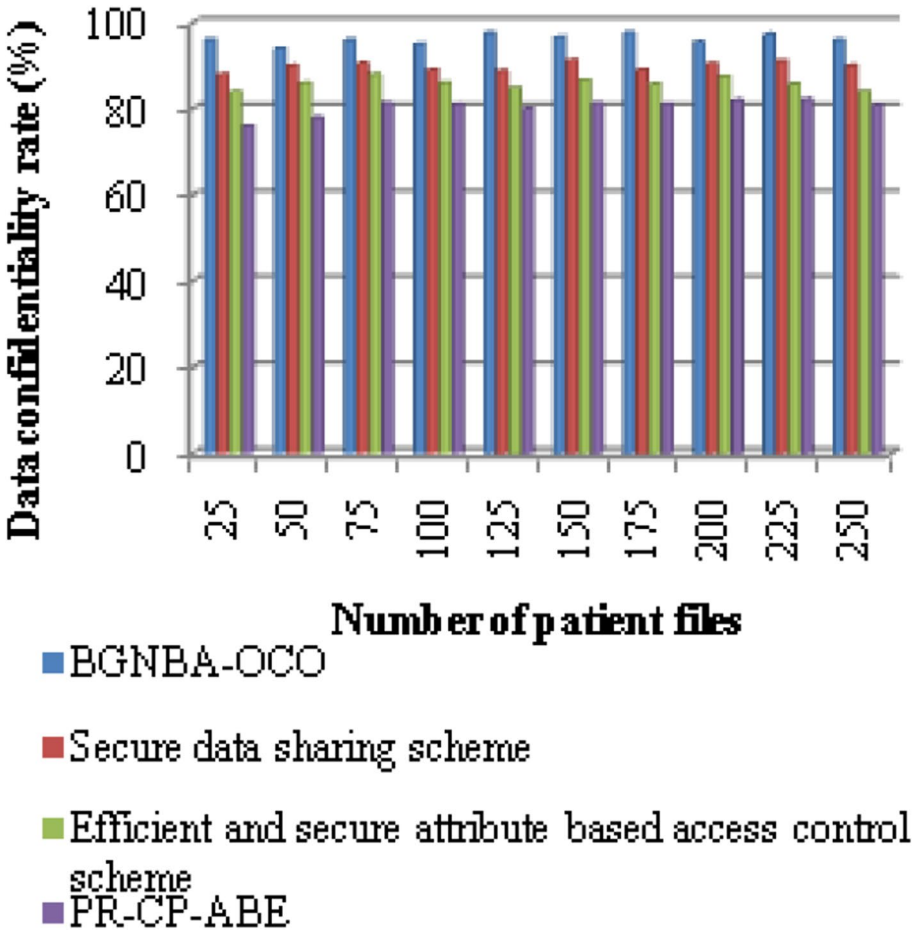

Fig. 8 Performance of data confidentiality rate versus number of patient files

confidentiality rate using the proposed BGNBA-OCO method is compared better than other existing methods. As shown in Table 4, let's assume the 25 patient files taken from the dataset to compute the data confidentiality rate. In the total number of patient files, the amount of patient files accessed by authorized users is 24. Therefore, the data confidentiality rate was found to be 96% using the BGNBA-OCO technique. Likewise, the data confidentiality rates of existing [1] [2] and [3] are 88%, 84% and 76% respectively. Similarly, dissimilar performance results are attained for each method. The overall performance of the BGNBA-OCO technique is compared to existing methods. The experimental results specified that the data confidentiality rate of the BGNBA-OCO technique is improved by 7%, 12%, and 20% compared to [1] [2], and [3] respectively. This is because of applying a *Boneh Goh Nissim Bilinear* Attribute-based encryption method to decrypt the given cipher text patient file. If the authorized user access data and the unauthorized user does not acquire any file from the cloud server. This enhances the confidentiality of patient file sharing from the hospitals to medical practitioners through the cloud server.

Figure 9, describes the performance analysis of the data integrity rate with respect to the number of patient files taken from 25 to 250 from the dataset. The performance of the data integrity rate is estimated by applying four various methods BGNBA-OCO, existing [1] [2] and [3]. With these four schemes, the BGNBA-OCO technique achieved improved data integrity rate results when compared to conventional methods. For example, 25 patient files are considered in the first iteration to determine the data integrity rate in secret sharing with hospitals and medical practitioners. The BGNBA-OCO is applied and observes the data integrity rate to be 92% and the observed integrity rate of existing [1] [2] [3] are 84%, 80%, and 72% respectively. Lastly, the overall experimental results specified that BGNBA-OCO better the performance results of data integrity rate by 7% when compared to [1], 12% when compared to [2], and 19% when compared to [3] respectively.

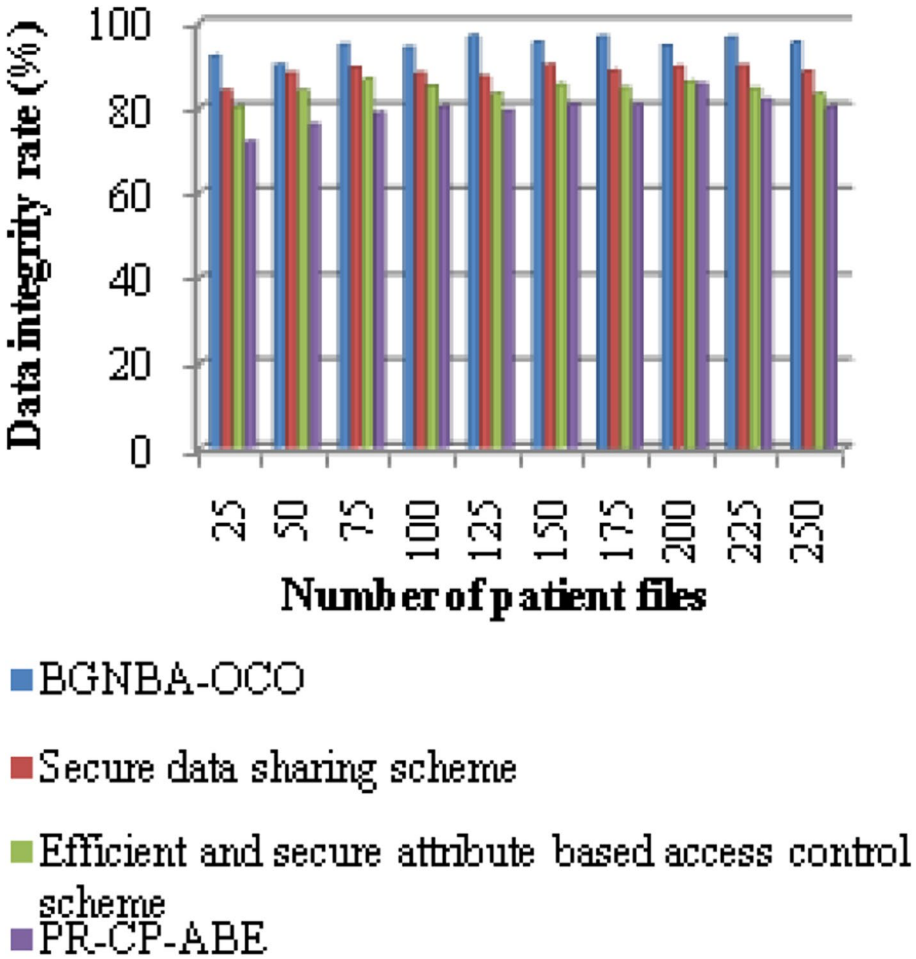

Fig. 9 Performance of data integrity rate versus number of patient files

Conclusion

In this paper, a secure data sharing called BGNBA-OCO is developed for deduplication and sensitive information hiding in the cloud environment. In the BGNBA-OCO method, first patient files are encrypted with the corresponding public key to hide the sensitive information. Before being uploaded into the server, first the data deduplication is verified using the Optimal Cache Oblivious algorithm to minimize the communication overhead and increase the deduplication efficiency. Then the file is uploaded to the server for further processing. Then the authorization and decryption process is performed. The authorized user accesses the file from the cloud server resulting in enhanced confidentiality and integrity. A comprehensive experimental assessment is carried out by means of the heart disease patient file sharing with different parameters such as communication overhead, computation overhead, data confidentiality rate, and data integrity rate. The quantitative performance result indicates that the proposed BGNBA-OCO method achieves a higher data confidentiality rate and data integrity rate and minimizes the communication overhead, as well as computation overhead. In future work, more encryption approaches will be developed to optimize storage space and security of data by using a hashing algorithm while performing efficient data sharing with a higher confidentiality rate and data integrity rate.

Acknowledgements

We (Mrs.M.Pavithra, Dr.M.Prakash, Dr.V.Vennila) hereby declare that this Research Paper titled as "BGNBA-OCO Based Privacy Preserving Attribute Based Access Control With Data Duplication For Secure Storage In Cloud" submitted by us is based on original work and actual work carried out by us. Any reference to this work done by any other person or institution or any material obtained from other sources have been duly cited and referenced.

We further certify that we have followed the norms of plagiarism of the journal. The article has not been published or submitted for publication anywhere else nor will be sending for publication in the future

Authors' contributions
Yes, All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. Yes, The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Funding
No funding applied for this article.

Availability of data and materials
The datasets investigated through present study are not publicly available.

Declarations

Ethics approval and consent to participate
This article does not contain any studies with human participants or animals carried out by some of the authors.
Yes, All authors have participated in (a) conception and design, or analysis and interpretation of data; (b) drafting the article or revising it seriously for significant intellectual content; and (c) approval of the final version.
Yes, The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the paper.
The article does not contain any studies with human participants carried out by any of the authors.

Consent for publication
We are the Authors, submitting this article for your consideration in the titled " BGNBA-OCO Based Privacy Preserving Attribute Based Access Control With Data Duplication For Secure Storage In Cloud". Kindly review the article soon and provide us the comments as early as possible. Since we are ready to clarify any issues pointed out by the editor and reviewer.

Competing interests
Yes, All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. Yes, The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Received: 28 June 2023 Accepted: 7 November 2023
Published: 5 January 2024

References

1. Wang Z, Gao W, Yang M, Hao (2022) Enabling Secure Data sharing with data deduplication and sensitive information hiding in cloud-assisted Electronic Medical Systems. Cluster Computing, Springer, 1–16. https://doi.org/10.1007/s10586-022-03785-y.
2. Xue K, Gai N, Hong J, Wei DSL, Hong P, Yu N (2022) Efficient and Secure Attribute-Based Access Control With Identical Sub-Policies Frequently Used in Cloud Storage. IEEE Trans Dependable Secure Comput 19(1):635–646. https://doi.org/10.1109/TDSC.2020.2987903
3. Naruse T, Mohri M, Shiraishi Y (2015) Provably secure attribute-based encryption with attribute revocation and grant function using proxy re-encryption and attribute key for updating. Human-centric Computing and Information Sciences, Springer 5:1–13. https://doi.org/10.1186/s13673-015-0027-0
4. Cui H, Deng RH, Li Y, Wu G (2019) Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud. IEEE Transactions on Big Data 5(3):330–342. https://doi.org/10.1109/TBDATA.2017.2656120
5. Shynu PG, Nadesh RK, Menon VG, Venu P, Abbasi M, Khosravi MR (2020) A secure data deduplication system for integrated cloud-edge networks. Journal of Cloud Computing: Advances, Systems and Applications 9:1–12. https://doi.org/10.1186/s13677-020-00214-6
6. Yuana H, Chena X, Jianga T, Zhanga X, Yana Z, Xiang Y (2018) DedupDUM: Secure and scalable data deduplication with dynamic user management. Information Sciences, Elsevier 456:159–173. https://doi.org/10.1016/j.ins.2018.05.024
7. Wang Y, Miao M, Wang J, Zhang X (2021) Secure deduplication with efficient user revocation in cloud storage. Computer Standards & Interfaces, Elsevier 78:1–8. https://doi.org/10.1016/j.csi.2021.103523
8. Yu X, Bai H, Yan Z, Zhang R (2023) VeriDedup: A Verifiable Cloud Data Deduplication Scheme with Integrity and Duplication Proof. IEEE Trans Dependable Secure Comput 20(1):680–694. https://doi.org/10.1109/TDSC.2022.3141521
9. Li J, Huang S, Ren Y, Yang Z, Lee PPC, Zhang X, Hao Y (2022) Enabling Secure and Space-Efficient Metadata Management in Encrypted Deduplication. IEEE Trans Comput 71(4):959–970. https://doi.org/10.1109/TC.2021.3067326
10. Yang X, Lu R, Shao J, Tang X, Ghorbani AA (2022) Achieving Efficient Secure Deduplication With User-Defined Access Control in Cloud. IEEE Trans Dependable Secure Comput 19(1):591–606. https://doi.org/10.1109/TDSC.2020.2987793
11. Yuan H, Chen X, Li J, Jiang T, Wang J, Deng RH (2022) Secure Cloud Data Deduplication with Efficient Re-encryption. IEEE Trans Serv Comput 15(1):442–456. https://doi.org/10.1109/TSC.2019.2948007
12. Gang F, Wei D (2022) Dynamic Deduplication Algorithm for Cross-User Duplicate Data in Hybrid Cloud Storage. Security and Communication Networks, Hindawi 2022:1–9. https://doi.org/10.1155/2022/8354903
13. Ebinazer SE, Savarimuthu N, Bhanu SMS (2021) ESKEA: Enhanced Symmetric Key Encryption Algorithm Based Secure Data Storage in Cloud Networks with Data Deduplication. Wireless Personal Communications, Springer 117:3309–3325. https://doi.org/10.1007/s11277-020-07989-6
14. Xu R, Joshi J, Krishnamurthy P (2021) An Integrated Privacy Preserving Attribute-Based Access Control Framework Supporting Secure Deduplication. IEEE Trans Dependable Secure Comput 18(2):706–721. https://doi.org/10.1109/TDSC.2019.2946073
15. Tang X, Zhou L, Hu B, Wu H (2021) Aggregation-Based Tag Deduplication for Cloud Storage with Resistance against Side Channel Attack". Security and Communication Networks, Hindawi 2021:1–15. https://doi.org/10.1155/2021/6686281
16. Yang X, Lu R, Choo KKR, Yin F, Tang X (2022) Achieving Efficient and Privacy-Preserving Cross-Domain Big Data Deduplication in Cloud. IEEE Transactions on Big Data 8(1):73–84. https://doi.org/10.1109/TBDATA.2017.2721444
17. Fu Y, Xiao N, Chen T, Wang J (2022) Fog-to-MultiCloud Cooperative Ehealth Data Management with Application-Aware Secure Deduplication. IEEE Trans Dependable Secure Comput 19(5):3136–3148. https://doi.org/10.1109/TDSC.2021.3086089
18. Zhang S, Xian H, Li Z, Wang L (2020) SecDedup: Secure Encrypted Data Deduplication with Dynamic Ownership Updating. IEEE Access 8:186323–186334. https://doi.org/10.1109/ACCESS.2020.3023387
19. Lakshmi VS, Deepthi S, Deepthi PP (2021) Collusion resistant secret sharing scheme for secure data storage and processing over cloud. Journal of Information Security and Applications, Elsevier 60:1–16. https://doi.org/10.1016/j.jisa.2021.102869
20. Begum BR, Chitra P (2021) ECC-CRT: An Elliptical Curve Cryptographic Encryption and Chinese Remainder Theorem based Deduplication in Cloud". Wireless Personal Communications, Springer 116:1683–1702. https://doi.org/10.1007/s11277-020-07756-7
21. Zhang T, Wang C, Chandrasena U, Blockchain-assisted data sharing supports deduplication for cloud storage. Connection Science, 35:1, 2174081 DOI: https://doi.org/10.1080/09540091.2023.2174081
22. Li L, Zheng D, Zhang H, Qin B (2023) Data Secure De-Duplication and Recovery Based on Public Key Encryption With Keyword Search". IEEE Access 11:28688–28698. https://doi.org/10.1109/ACCESS.2023.3251370
23. Gund, Avinash and Mahadik, Prerna and Thorat, Ashvini R and Yevle, Ganesh K, Data De-Duplication Using Blockchain with Advanced Security in Cloud Computing, August 5, 2022. https://doi.org/10.2139/ssrn.4289505
24. Gao X, Yu J, Shen W-T, Chang Y, Zhang S-B, Yang M, Wu B (2021) Achieving lowentropy secure cloud data auditing with file and authenticator deduplication. Inf Sci 156:177–191. https://doi.org/10.1016/j.ins.2020.08.021
25. Qi L, Liu Y, Zhang Y, Xu X, Bilal M, Song H (2022) Privacy-Aware Point-of-Interest Category Recommendation in Internet of Things. IEEE Internet Things J 9(21):21398–21408. https://doi.org/10.1109/JIOT.2022.3181136
26. Liu Y, Zhou X, Kou H, Zhao Y, Xu X, Zhang X, Qi L (20213) Privacy-Preserving Point-of-Interest Recommendation based on Simplified Graph Convolutional Network for Geological Traveling. ACM Transactions on Intelligent Systems and Technology, 1–17. https://doi.org/10.1145/3620677

Publisher's Note

M. Pavithra received her Bachelor degree in BE Computer Science Engineering and Master degree in ME Computer Science Engineering from J.K.K. Nattraja College of Engineering and Technology, Komarapalayam, Tamilnadu. Currently part-time research scholar in Anna University, Chennai. She has 7 years of teaching experience and currently working as an Assistant Professor in SSM College of Engineering, Komarapalayam. Her research areas are mainly in Networking, Cloud Computing, and Cloud Security. She has published papers in reputed conferences and journals

M. Prakash received the Doctor of Philosophy from Anna University, Chennai, India. He received the Master of Technology degree in Information Technology from Anna University, Chennai, India. He is currently working as Associate Professor in the Department of Data Science and Business Systems, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India. He has 15 years of experience in teaching. His research interest includes Big data analytics, Machine Learning, Databases and Security, Information Management. He is a professional member of IEEE, ISTE, IE(I), IAENG, CSTA, IACSIT and UACEE

V. Vennila received her Ph.D degree from Anna University, Chennai, India. She is having 14 years of experience in research and teaching. She is the life member of Indian Society for Technical Education. So far she has published 40 research papers in various national, International conferences and journals. Her area of interest is Data Mining, Machine Learning, Artificial Intelligence, Big data analytics and Cloud Computing. She always engaged to fill the gap between the students and industry with latest technologies. Currently she is working as an Associate Professor in Department of Computer Science and Engineering, K.S.R. College of Engineering, Tiruchengode, Tamilnadu, India