ISSN 2347 - 3983

Volume 8. No. 10, October 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter258102020.pdf

https://doi.org/10.30534/ijeter/2020/258102020

# Efficient Data Retrieval and Fine Grained Access Control on Secure Hybrid Cloud

**Sridhar Reddy Vulapula[1], Srinivas Malladi[2]**

[1]Research Scholar –Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram,
AP, India., vsridharreddy19@gmail.com

[2]Professor –Department of Computer Science and Engineering,
Koneru Lakshmaiah Education Foundation, Vaddeswaram,
AP, India., srinu_cse@kluniversity.in

## ABSTRACT

Healthcare data has many private attributes to be secured from leakage due to inference either direct or indirect. Hybrid cloud is a model which is a combination of both private and public clouds. This model is proposed as a way of storing healthcare data. The data is carefully distributed between these two clouds to bring about security of private attributes. While substantial work has already been around for a while for distributing healthcare data, they do not seem to possess greater efficiency in terms of both data retrieval and consideration for fine coarse access control on the given information. The present work proposes an amicable solution for a more secure distribution of data using geometric data perturbation of healthcare data over hybrid clouds. It is based on inherent analysis of data. The distribution enforces effectively to gain a control over the data using an encryption that is attribute-based one. Besides, this particular solution also discusses a strategy to retrieve relevant information efficiently from hybrid clouds

**Key words:** Attribute-based Encryption, Fine grained access control, Geometric Data Perturbation and Hybrid Cloud

## 1. INTRODUCTION

Of late, several enterprises have been in the process of shifting to cloud storage of data keeping in view of factors such as cost efficiency, availability, redundancy etc. Using cloud storage in healthcare enterprises facilitates sharing environment of inter-organizational medical data. Data security and its privacy are the two critical factors to be considered with respect to the cloud data storage. Possible leakage of sensitive medical data may result in compromising with the individual privacy. Though encryption methods are used, leakage of sensitive data may still take place through inference possibly leading to anti-social activities like blackmailing, defaming etc. Hence, it is of paramount importance to ascertain the security as well as privacy of cloud storage data. All the methods currently being used for privacy protection can be categorized into following:

- Anonymization
- Randomization
- Cryptographic techniques
- Diversification
- Aggregation

But these methods have certain deficiencies related to scale of operations. Of late, hybrid clouds have been brought forward by many as a means to provide enhanced security and privacy. The data is obfuscated or transformed using specific parameters. The obfuscated data is stored in un-trusted public cloud while fully trusted private cloud is used to store the parameters used for obfuscation. By distributing obfuscated information and the parameters used for the purpose to different data stores, the security of the obfuscated data is ensured even if it is leaked. Certain open issues related to the hybrid cloud-based storage solutions are listed here:

- Efficiency in storage and retrieval process.
- Effective fine grained access control

Differential handling of security with the attributes must be done based on the sensitivity levels. Also, the fine grained access control must be done on those for different groups of users. The perturbed data must be indexed for an efficient retrieval. The retrieval process must also be protected against inference attacks. Besides data protection, the index must also be secured as privacy information can be accessed through inference. With these requirements, a secure geometric data perturbation method for healthcare data using hybrid clouds is proposed in this work. The perturbation is controlled through attribute-based encryption. The method also proposes a fine-grained access control on perturbed data with efficient secure indexing and retrieval of information.

## 2. RELATED WORK

Authors in [1] proposed a novel means of inter-organizational data sharing for healthcare data. The solution is designed to cater to the security and privacy needs on the patient data for semi-trusted clouds. Attribute based encryption is proposed in this work for selective access. Distribution of data across multiple clouds is done using cryptographic secret sharing. Retrieval of the required data becomes slightly inefficient in this method because of the involvement of more players in the area of providing cloud storage services. A scalable anonymization technique for data is proposed in [2]. A proximity-privacy model is proposed to address the issue of privacy breaches along with semantic proximity of sensitive values and also multiple sensitive attributes. An agglomerative clustering algorithm which is proximity-aware segregates related records into hierarchical groups and differential privacy is proposed to handle them. Attribute based encryption (ABE) is used for providing improved security of electronic health care records in [3]. A two-fold advantage of reduction in communication cost and fine-grained access control is achieved by employing ABE method. The authors analyzed the performance of four different ABE schemes – CP-ABE, KP-ABE, HABE, DABE. Authors in [4] proposed a solution to ABE key distribution issues when it is used for securing electronic healthcare records in cloud. Further, the key distribution process is also simplified using attributes and implicit authentication. The scheme is built on the assumption of centralized key issue authority which becomes an obstacle during failure. Attribute based encryption and searchable encryption are proposed for keyword based fine grained information retrieval in [5]. It is multi-authority scheme and user secret key distribution is proposed to solve the key leakage problem. This scheme is effective for resource constrained devices and most suitable for fog computing nodes. A hybrid cloud model is proposed in [6] for privacy preservation of the shared information. Hybrid solution has four important concepts. Partitioning the data vertically before publishing, access based on data merging, checking the integrity of the data, statistical and cryptography-based hybrid search are four concepts used in this work for effective data utilization without compromising on privacy protection. Authors in [7] made uses of an algorithm of reversible privacy contrast mapping (RPCM) for the purpose of data perturbation. This algorithm involves two stages – the first one is data perturbation while data recovery is the nest step. The first stage is performed by grouping two adjacent data values. This is in addition to embedding of a watermark. In the second stage i.e. data recovery, restoration of perturbed data is achieved. Embedded watermarking is incorporated to test and validate an integrity of altered data. A tree structure based fast perturbation algorithm is proposed in [8]. The perturbation time is reduced using unique tree traversal strategy that involves specifically defined tree and table structures. A hybrid architecture, proposed in [9], involves private cloud as an access interface between the two parties i.e. the data owner/user and the public cloud. Fuzzy keyword search, in addition to fine grained access control scheme, is proposed over encrypted data. Computational cost at the user end may also be considerably reduced by migrating ABE to the private cloud. A privacy preserving data publishing system called Cocktail for hybrid clouds is proposed in [10]. An extended quasi-identifier-partitioning approach is proposed that segments the data publishing phase. Differential privacy strategy is used at data querying stage, to protect from privacy breaches. This solution entails data privacy besides reducing loss of information. Application independent data partition strategy is proposed in [11]. Sensitive data is kept in private cloud while the public cloud keeps insensitive data in this scheme. Authors in [12] proposed a data perturbation scheme, performed in two stages, called RG+RP. The user perturbs the data using a nonlinear Repeated Gompertz (RG) and then projects the data to lower dimension in distance preserving manner using random project matrix (RP). Use of these two-stage schemes protects the data perturbation from estimation attacks and independent component analysis attacks. Due to distance preservation, fuzzy c mean clustering can be done on perturbed data with same result as that applied on raw data. An attack resilient geometric data perturbation is proposed in [13]. The scheme is able to provide a fine balance between quality of the data as well as privacy protection. The perturbation has 3 different stages: random rotation, translation and noise addition. Geometric properties of multi-dimensional dataset are preserved even after perturbation. A privacy evaluation framework of multi column for the purpose of analyzing the attack impact on geometric perturbation is proposed in [14]. Multidimensional geometric perturbation method called random projection perturbation is proposed in this work. Authors in [15] proposed a fast indexing for data retrieval in a secure cloud. Compression sensing is used for data sampling, compression and recovery. An encrypted high-performance index is constructed to retrieve the data. A new anonymization scheme, secure against the identity disclosure attack is proposed in [16]. The scheme splits the total database into regular slots of fixed size, coverts the same values of each slot into an average value. The data transformation is one way and it cannot be brought back to original state. An anonymization scheme based on classification capability of data is proposed in [17]. The classification capability of data is measured using mutual information. Two K-anonymity algorithm is proposed to transform the data without losing the classification capability. A privacy preservation scheme for the association rules data is proposed in [18]. Sensitive rules are hidden by removing certain items in the database and thereby removing the support and confidence values for those rules. Authors proposed a heuristics scheme to identify minimal transactions to hide the sensitive rules. Joint entropy with adaptive optimization process is employed for privacy preservation. The optimal value for entropy is found using the technique of particle swarm optimization. Attributes with higher entropy are treated as sensitive and are transformed [19]. By merging two techniques clustering and geometric data perturbation into a model to enrich privacy preservation characteristics of health care system in hybrid cloud by using Geometric Data Perturbation (GDP) algorithm and K-means. Data sets and attributes information is incorporated integral part of this research work [21]. The optimal value for entropy is found using the technique of particle swarm optimization. Attributes with higher entropy are treated as sensitive and are transformed. An attribute based encryption scheme proposes a public key cryptography technique where key selection will be based on the user attributes. The attributes used are biometrics of user who is going to upload the data [22]. Authors presented a method in which Cloud Service suppliers will deal with security for the information put away by the Remote clients in cloud. Further in this application security can be provided to the general population without knowing them personally. In this manner the secrecy of the client's is accomplished. Every data stored in cloud need some kind of security. This is commonly called as cloud security. In this paper, Hash Counter Hash (HCH) is the new security is given by the administration suppliers and information is approved by the information proprietors lastly scrambled information sought by the clients and unscramble the information for use[23]. A Scalable Attribute Based Encryption (SABE) is a decentralized coarse access control technique proposed to achieve flexible and scalable access control in cloud computing for secure distributed cloud storage. Writers of [25] projected about how to secure the data which is stored in the cloud. Data user stores in the cloud can either be a public data which requires minimal security or a highly confidential data which requires high security. Due to the distribution of the system, it becomes essential to search the misbehaving server which helps the user to identify the particular server and can retrieve the sensitive information safely. Along with this, it works to recover from server attack and data crashes effectively. Number of users who are using the cloud for the storage of data increases and also issues related with security. Hence security of data is considered as the salient factor upon making the client believe that their data is stored safely. This can be done by performing the client authentication. Along with this there exists a number of security and privacy issues [29] associated that come under two broad categories: Security and Privacy issues faced by cloud providers and by their customers. With the available algorithms that are utilized to alter the regular text to the cipher, Addition of the concept of the steganography to the cipher text and make the security more efficient and protect the data from the unauthorized access[30]. We need to have a secured mechanism where it should address security problem at the time of processing the data in cloud. Hence authors in [26] attempting to find out the best mechanism for accessing cloud data by comparing all the attribute based encryption algorithms namely KP-ABE, CP-ABE along with HASBE while considering many features of these ABE techniques. Considered features such as policy for accessing, Attribute fine coarse access control, computation overhead, Efficiency[31], End user Revocation, Scalability and Resistance for Collision discussed in detailed by including Advantage and limitations. In [27] biometric access scheme is discussed in which biometric data is encipher and stored in cloud.

Authors in [32] proposed a novel system for safe storage of private healthiness data in cloud computing. In order to enable fine grain entry, patients shall have complete control over the security of PHR data. In order to allow patients access by not just their private users, then similarly specific users in public networks with various qualified positions, credentials and associations, we encode the PHR records built on the system ABE (attribute dependent encryption).

Authors in [33] proposed an algorithm that is helpful to select the specific algorithm for certain input files to share between the two different nodes on schedule.

Authors in [28] proposed an CSHQS (Cloud Security Hybrid Querying System) algorithm processes data efficiently with data security in hybrid cloud and sub query mechanism handles different components.

## 3. SEARCHABLE FINE ACCESS CONTROL ON SECURE HYBRID CLOUDS(SFAC-SHC)

The architecture of the present searchable fine access control on secure hybrid clouds (SFAC-SHC) is given in figure 1. The proposed solution involves the following stages of:

1. CP-ABE based key generation
2. Fine grained access control
3. Geometric perturbation
4. Secure Retrieval

### 3.1 CP-ABE based key generation

Of the four fundamental algorithms of setup, key generation, encryption and decryption, only two steps of setup and key generation are used in this work. A key authority or key generation center generates both the public and private keys. When a data owner uploads the file, it requests the key authority with access policy for each user in terms of attributes to the key authority.

A single public key and private key for every single access policy is created by competent key authority. Public key/all private keys matching to policy are forward to data owner and private key is issued to corresponding data user when they request for information.

The access policy used in this work has two sets of information:

1. Attribute – value pairs of the user
2. Fields or Column names in the dataset allowed for view for the user.

The key generation is based on the attribute value pairs in the policy. Column name for access is controlled by fine grained access control stage.

### 3.2 Fine grained access control

The health care dataset uploaded by data owner is in the table format as shown in Table 1 with each row for a patient and each column representing a field. The dataset has 3 different classes of information: - Class 1, 2 and 3.

**Table 1:** Tabular format of healthcare dataset

| | |
|---|---|
| Class 1 | Very sensitive information that cannot be shared by data owner |
| Class 2 | Sensitive information that can be shared and fine grain controlled for the accessing users. |
| Class 3 | In sensitive information that can be shared without any security concerns |

The class 1 information is enciphered by the data owner by using public key received from the key authority using any symmetric key algorithm and generates a processed dataset. Each row has an identifier which can be a unique string or a number.

An association map (figure 1) is created between the private key for the policy and the field or column names allowed for view for the user satisfying this policy. The processed dataset and the association map are sent to private cloud for geometric perturbation.

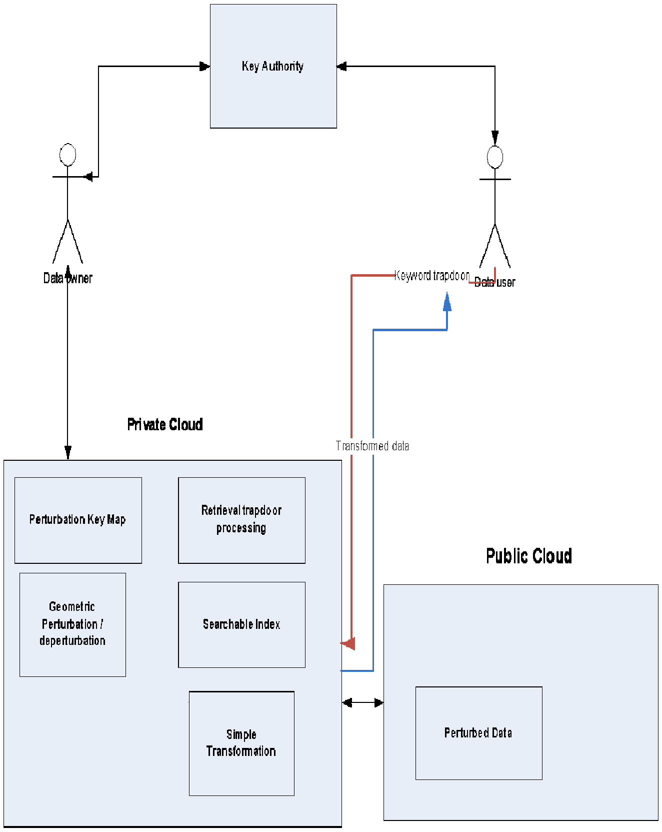


**Figure 1:** Architecture of SFAC-SHC

### 3.3 Geometric perturbation

For each mapping in the association map, the field or column names to be controlled are collected. A random geometric perturbation key is generated, which is a sequence of multiplicative transformation (TP), translation (Vs) and distance perturbation (D).

$$GDP(P) = TP + V_s + D$$

Where D is given as

$$D = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

Where Random projection matrix is represented by T, the matrix to be transformed by P, translation matrix by Vs and random Gaussian noise by D. The advantage in this perturbation is that even after applying the perturbation, the geometric properties like distance are maintained in the transformed dataset.

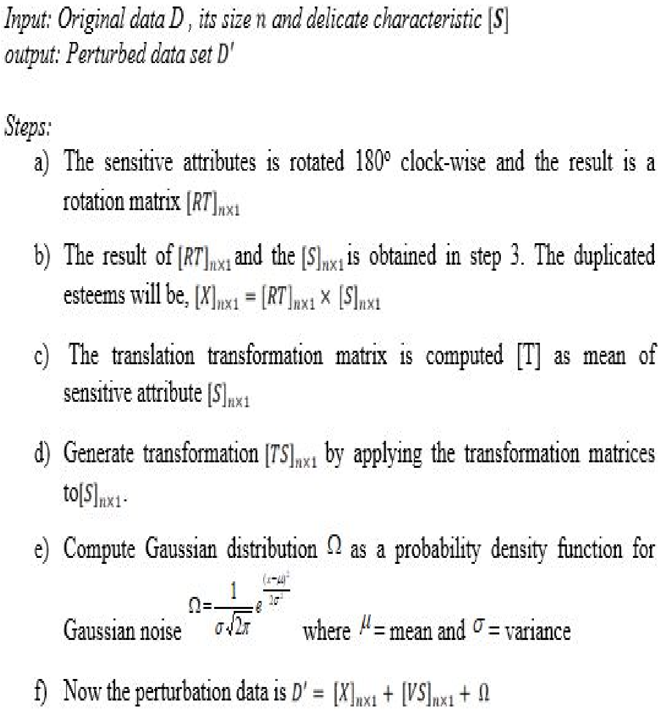
The fields in the data to be controlled are copied to a separate table along with corresponding identifier. Leaving the identifier, the rest of the columns in the separate table are perturbed using the generated geometric perturbation key. A random file name is generated for this perturbed data and this file is moved to public cloud for storage. An entry is added to the perturbation key map mapping between the hash of the private key and following information.

| | |
|---|---|
| Hash (Private key) | Field name perturbed<br>Perturbation key<br>Perturbed file name (saved in public cloud)<br>Private key |

Data owner transmit the hashing function to be used to the private cloud. Data owner also sends this hashing function to key authority to distribute to the data users.

Data Perturbation algorithm is used for geometric perturbation.

### 3.3.1 Data Perturbation Algorithm



*Input: Original data D, its size n and delicate characteristic* $[S]$

*output: Perturbed data set* $D'$

*Steps:*

a) The sensitive attributes is rotated 180° clock-wise and the result is a rotation matrix $[RT]_{n\times1}$

b) The result of $[RT]_{n\times1}$ and the $[S]_{n\times1}$ is obtained in step 3. The duplicated esteems will be, $[X]_{n\times1} = [RT]_{n\times1} \times [S]_{n\times1}$

c) The translation transformation matrix is computed [T] as mean of sensitive attribute $[S]_{n\times1}$

d) Generate transformation $[TS]_{n\times1}$ by applying the transformation matrices to$[S]_{n\times1}$.

e) Compute Gaussian distribution $\Omega$ as a probability density function for Gaussian noise $\Omega = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$ where $\mu$ = mean and $\sigma$ = variance

f) Now the perturbation data is $D' = [X]_{n\times1} + [VS]_{n\times1} + \Omega$

It is detailed in our earlier work [20].

In addition to perturbation, searchable index is constructed between the field values to the row index of the dataset.

### 3.4 Secure Retrieval

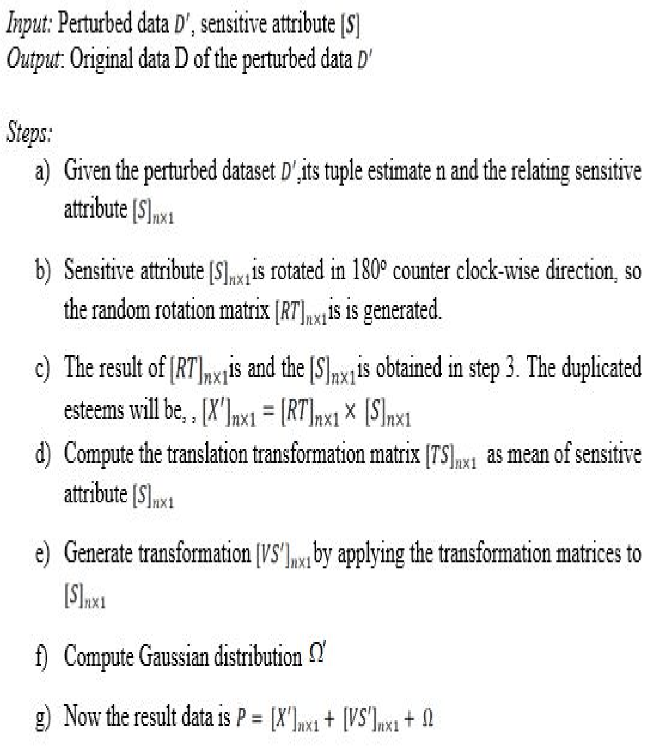
The data user can retrieve the data in two modes - all data or matching a particular field value pair.

For retrieval in all data mode, data user first fetches the private key for its matching attributes from the key authority. The private key and the hashing function are returned by the key authority. The private key is hashed and then sent to the private cloud. At the private cloud, lookup is done on the perturbation key mapping to find the match for the hashed private key. If a match could not be found, retrieval fails. If a match is found, following information is retrieved from the mapping:

1. Field name perturbed
2. Perturbation key
3. Perturbed file name (saved in public cloud)
4. Private key

The perturbed file is retrieved from the public cloud and using the perturbation key, deperturbation is done on the data. We use the Data Deperturbation algorithm given below for geometric deperturbation.

### 3.4.1 Data Deperturbation Algorithm



```
Input: Perturbed data $D'$, sensitive attribute $[S]$
Output: Original data D of the perturbed data $D'$

Steps:
  a) Given the perturbed dataset $D'$,its tuple estimate n and the relating sensitive attribute $[S]_{nx1}$
  b) Sensitive attribute $[S]_{nx1}$is rotated in 180° counter clock-wise direction, so the random rotation matrix $[RT]_{nx1}$is is generated.
  c) The result of $[RT]_{nx1}$is and the $[S]_{nx1}$is obtained in step 3. The duplicated esteems will be, , $[X']_{nx1} = [RT]_{nx1} \times [S]_{nx1}$
  d) Compute the translation transformation matrix $[TS]_{nx1}$ as mean of sensitive attribute $[S]_{nx1}$
  e) Generate transformation $[VS']_{nx1}$by applying the transformation matrices to $[S]_{nx1}$
  f) Compute Gaussian distribution $\Omega'$
  g) Now the result data is $P = [X']_{nx1} + [VS']_{nx1} + \Omega$
```

The data after deperturbation must not be sent to the user directly. The private key along with the current hour is hashed to a numeric code and the simple transformation operation is done on the field's values with the numeric code (like progressive addition). This transformation helps to prevent from network capture attacks.

At the data user end, inverse of simple transformation (like progressive subtraction) is done using the private key and the current hour to get the original data.

Due to communication of only transformed data between private cloud and the user, the retrieval process is secure against network capture attacks. The retrieval data passed from cloud to user end is masked with simple transformation. Without the information of private key and the parameter used for hashing (here it is current hour), removal of the mask cannot happen. Thereby, even if network capture attack is launched, the retrieved data is still masked and secure.

The proposed scheme also supports retrieval by field's value. The field and the corresponding value to be searched is encrypted with private key using a symmetric cryptographic algorithm and sent to private cloud. This encrypted value is called as trapdoor for search. Since the field name and corresponding value is encrypted, it is difficult for network capturing attacks to correlate between the search information and the result.
At the private cloud, the filed name and the corresponding value is decrypted. Lookup is done on the perturbation key mapping to find the match. If a match could not be found, retrieval fails. If a match is found, following information is retrieved from the mapping table:

1. Field name perturbed
2. Perturbation key
3. Perturbed file name (saved in public cloud)
4. Private key

If the field name provided for search in the list of field name perturbed, the search continues, else error is returned. Thereby fine-grained access control is enforced even in search. The field value given for search is looked up in the searchable index of field for match. If no matching row index is found, then error is returned. If indexes of a matching row are found, those particular row indexes are retrieved from the public cloud. Deperturbation is done on the received row indexes. The data after deperturbation must not be sent directly to the user. The private key along with current hour is hashed to a numeric code and the simple transformation operation is done on the field's values with the numeric code (like progressive addition). At the data user end, inverse of simple transformation (like progressive subtraction) is done using the private key and the current hour to get the original data.

## 4. PROPOSED SOLUTION

The proposed solution has the following novel aspects:

1. Data owner has more control on very sensitive information even though data is uploaded to cloud. This is enabled by moving very sensitive information to class 1 and encrypting with public key of the data owner. Without owner sharing this key, it is not possible to get this information.

2. The data communicated from cloud to user is simply transformed with private key and current hour. So, it is difficult for network attackers to capture and decipher information from it.

3. Two modes of retrieval are supported for the users. Retrieval can be either the whole file or certain records matching a criterion.
4. Fine grained access control at field level is enforced for the users in both the retrieval modes.
5. Data owner has more control on which users he wants to share data with based on the attributes of the users.

## 5. RESULTS

The performance of the proposed searchable fine access control on secure hybrid clouds (SFAC-SHC) is compared in different aspects of

1. Perturbation efficiency
2. Data storage and retrieval efficiency
3. Security against attacks

Arrhythmia dataset from UCI machine learning repository is used for evaluation [21].

### 5.1 Perturbation Efficiency

The perturbation efficiency of the proposed solution is compared against RG+RP algorithm proposed in [12]. K-Means clustering is done on original data as well as on the perturbed data got using the proposed and RG+RP. The clustering accuracy is calculated between

1. Clusters got using proposed and cluster of original data set
2. Clusters got using ElGamal's and the cluster of original set of data

Clustering accuracy is deliberated as

$$ACC = \frac{1}{N}\sum_{i=1}^{k}(|Cluster_i(P)| - |Cluster_i(P')$$

Where P is the raw data, $P'$ is changed data, number of clusters is denoted with k and number of items in set of data is denoted with N. The result of clustering accuracy is measured for different k values and the result is below (Table 2)

**Table 2:** Results of clustering accuracy for different k values

| K | Clustering accuracy in RG+RP [12] | Clustering accuracy in Proposed |
|---|---|---|
| 2 | 66.23 | 65.89 |
| 3 | 66.56 | 71.25 |
| 4 | 73.33 | 74.59 |
| 5 | 67.22 | 79.58 |

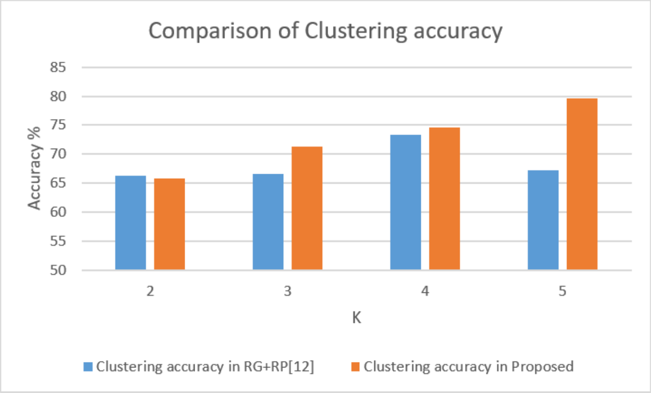


**Figure 2:** Comparision of clustering accuracy

The clustering accuracy as shown in figure 2 lies more in the proposed solution as the transformation method adopted retains the geometrical properties even after transformation.

### 5.2 **Data storage and retrieval efficiency**

The performance of proposed technique is compared with quite similar approach used in the fine-grained searchable retrieval system proposed in [5]. Performance is compared in terms of the following parameters by varying the number of attributes:

1. Time utilized for Key generation
2. Time utilized for Index generation
3. Time utilized for Trapdoor generation
4. Time to Search

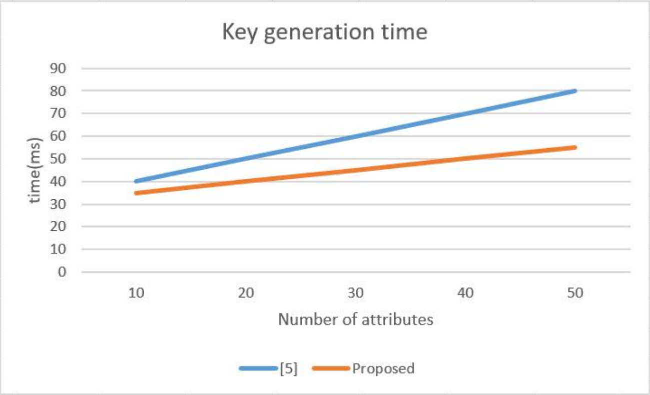


**Figure 3:** Key Generation Time

Figure 3 shows Time utilized for Key generation which is relatively lower in the proposed solution as key size is shorter (16 bytes) in the proposed solution compared to [5].

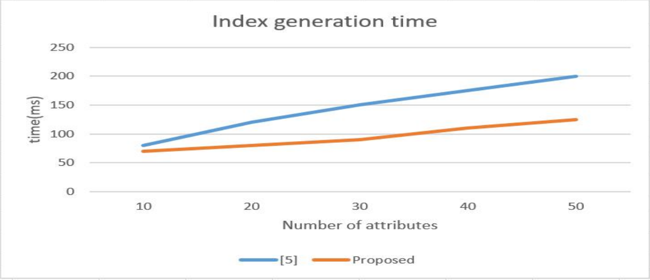


**Figure 4:** Index Generation Time

The index generation time in figure 4 is shorter in the proposed solution compared to [5] as the index is computed only on certain fields as demanded by the users. But in [5] index is constructed for all fields and this increases the index generation time.

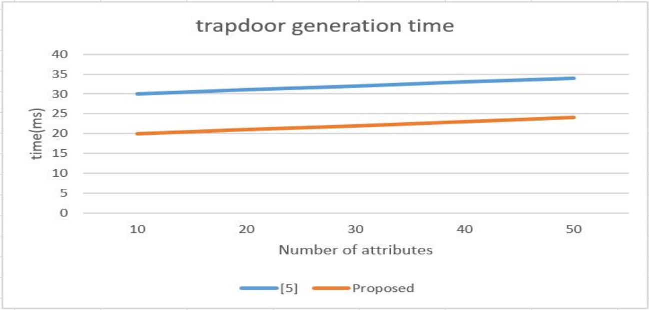


**Figure 5:** Trapdoor Generation Time

The trapdoor or the encrypted search keyword generation time from figure 5 is lower in the proposed solution compared to [5]. This reduction is due to reduced key size and AES with less rounds for generation of trapdoor in the proposed solution.

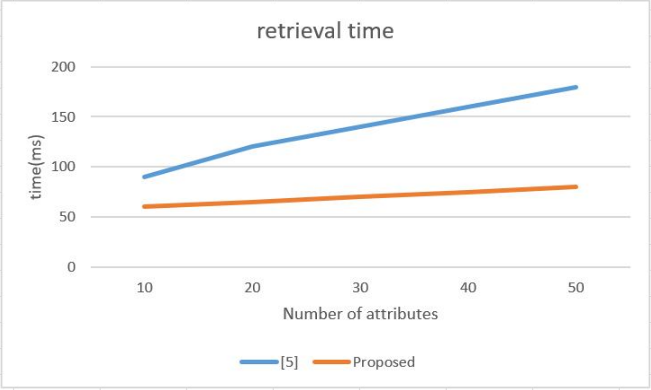


**Figure 6:** Retrieval Time

The retrieval time as shown in figure 6 is also shorter in the proposed solution compared in [5]. The reasons for the shorter retrieval time are due to lower levels of trapdoor decryption, lookup from index and lower time for deperturbation and simple transformation.

5.3**Security against Attacks**

Proposed solution security is assessed with regard to the level of difficulty of evaluating the actual data from the modified one by an attacker who steals the perturbed data from the cloud. There are two classes of fields in the dataset to be privacy preserved:

1. Class 1
2. Class 2

Class 1 data are the very sensitive ones. Class 2 are sensitive information that can be shared and fine grain controlled for the accessing users. In the proposed scheme, fields class 1 are encrypted with AES and then geometric perturbation is applied if it needs to be shared. Class 2 fields only undergo geometric perturbation.

Difficulty level is estimated with the help of a different-based technique which is essentially a variant of the same approach. The variation between the orignal and the estimated data is assumed as the random vaiable Di without knowing anything about the original data, the mean or the variance of the variation present in the quality of the estimation. Because the mean of the difference may easily be done away with when an intruder makes a wild guess of the original distribution of column, only the variance of the difference (VoD) may be employed as the primary metric to ascertain the level of difficulty in approximating the original data.

When $X_i$is assumed as a random variable standing for the column i, $X_i'$is assumed as the estimated result of $X_i$ and the difference $D_i = X' - X$. Let mean of D be $E(D_i)$ and variance be $Var(D_i)$. VOD for column i is $Var(D_i)$. VOD is measured for each column and average VOD is given as privacy measure(pm)

$$pm = \frac{\sum_{i=1}^{N} VOD_i}{N}$$

A guess is launched for 5 hours on the perturbed data and the privacy measure (pm) is measured for every 1-hour interval and plotted below

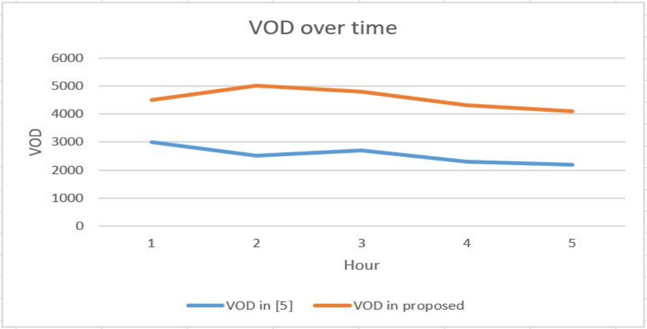


**Figure 7:** VOD Over Time

By considering all these results, it is clearly understood from figure 7 that VOD in the solution proposed is way higher than VOD in [5]. Higher VOD suggests that itis nearly impossible to make even an approximate estimation of the original data from the perturbed one. The VOD has increased in the proposed solution due to geometric perturbation combined with encryption for class 1 data fields.

## 6. CONCLUSION

In this work, searchable fine access control on secure hybrid clouds (SFAC-SHC) is proposed. The scheme uses multiple concepts of CP-ABE, fine grained access control, geometric perturbation and searchable indexing on perturbed data. Secure Perturbed data is stored in untrusted public cloud without any risk of leakage. The information needed for deperturbation of the data on the public cloud is kept at the private cloud. The proposed scheme is secure against network capturing attacks and un-authorized access attacks. Fine grained access control is enforced field wise, so that information is strictly controlled. The work is designed on the assumption of complete trust in private cloud. As future work, the work needs to be adapted for semi-trusted private cloud by offloading some of the functionalities to respective data owner or the data users.

## REFERENCES

1. Benjamin Fabian,"**Collaborative and secure sharing of healthcare data in multi-clouds**",Information systems,2014Stefanos A. Nikolidakis, Dimitrios D. Vergados, Christos Douligeris Algorithms, Energy Efficient Routing in Wireless Sensor Networks Through Balanced Clustering, 2013.
2. X. Zhang et al., "**Proximity-Aware Local-Recoding Anonymization with MapReduce for Scalable Big Data Privacy Preservation in Cloud**," in IEEE Transactions on Computers, vol. 64, no. 8, pp. 2293-2307, 1 Aug. 2015, doi: 10.1109/TC.2014.2360516.
3. Zahoor A. Khan, Shyamala Sivakumara, William Phillips, Bill Robertson, "**A QOS-aware Routing Protocols for Reliability Sensitive Data in Hospital Body Area Networks**", Trans. on ELSEVIER in proc. ANT, pp. 171-179, 2013.
4. Achampong, Emmanuel & Dzidonu, Clement. (2016). **Optimising Attribute-based Encryption to Secure Electronic Health Records System within a Cloud Computing Environment**. 27-34. 10.21742/ijcs.2016.3.2.04.
5. Jin Sun, Xiaojing Wang, "**A searchable personal health records framework with fine-grained access control in cloud-fog computing**",PLOS ONE,2018
6. J. Yang, J. Li, and Y. Niu, "**A hybrid solution for privacy preserving medical data sharing in the cloud environment**", Future Generation Computer Systems, Vol. 43-44, No. 2, pp. 7486, 2015.
7. Kao, Yuan-Hung & Lee, Wei-Bin & Hsu, Tien-Yu & Lin, Chen-Yi &Tsai, Hui-Fang & Chen, Tung-Shou. (2015). **Data Perturbation Method Based on Contrast Mapping for Reversible Privacy-preserving Data Mining**. Journal of Medical and Biological Engineering. 35. 10.1007/s40846-015-0088-6.
8. Yun, Unil & Kim, Jiwon. (2015). **A fast perturbation algorithm using tree structure for privacy preserving utility mining**. Expert Systems with Applications. 42. 1149–1165. 10.1016/j.eswa.2014.08.037.
9. J.Li, J.Li, X.Chen, Z.Liu, and C.Jia, **Privacy preserving data utilization in hybrid clouds**, Future Generation Computer Systems.2014;vol.30, pp.98- 106.
10. H. Zhang, Z. Zhou, L. Ye and X. Du, "**Towards Privacy Preserving Publishing of Set-Valued Data on Hybrid Cloud**," in *IEEE Transactions on Cloud Computing*, vol. 6, no. 2, pp. 316-329, 1 April-June 2018.
11. Z. Zhou, H. Zhang, X. Du, P. Li and X. Yu. Prometheus: **Privacy-Aware Data Retrieval on Hybrid Clouds**. In Proc. of INFOCOM, 2013.
12. Lyu, Lingjuan & Bezdek, James & Law, Yee Wei & He, Xuanli & Palaniswami, Marimuthu. (2018). **Privacy-preserving collaborative fuzzy clustering**.Data & Knowledge Engineering. 10.1016/j.datak.2018.05.002.
13. Chen, Keke & Sun, Gordon & Liu, Ling. (2007). **Towards Attack-Resilient GeometricData Perturbation**. 10.1137/1.9781611972771.8.
14. Chen, K., Liu, L. **Geometric data perturbation for privacy preserving outsourced data mining**. Knowl Inf Syst 29, 657–695 (2011).
15. X. Yuan, X. Wang, C. Wang, J. Weng and K. Ren, "**Enabling Secure and Fast Indexing for Privacy-Assured Healthcare Monitoring via Compressive Sensing**," in *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 2002- 2014, Oct. 2016
16. A. Majeed, "Attribute-centric anonymization scheme for improving user privacy and **utility of publishing e-health data**," Journal of King Saud University - Computer and Information Sciences, 2018.
17. Li, Jiuyong & Liu, Jixue & Baig, Muzammil & Wong, Raymond. (2011). "**Information based data anonymization for classification utility**. Data Knowl. Eng.. 70. 1030-1045. 10.1016/j.datak.2011.07.001.
18. P. Cheng, J. Roddick, S. Chu, and C. Lin, "**Privacy preservation through a greedy, distortion-based rule-hiding method,**" Applied Intelligence,vol. 44, no. 2, 2015, pp. 295-306.
19. Sabin Begum, R., Sugumar, R. **Novel entropy-based approach for cost- effective privacy preservation of intermediate datasets in the cloud**. Cluster Comput 22, 9581–9588 (2019).
20. Reddy, Vulapula Sridhar, and Barige Thirumala Rao. **"A combined clustering and geometric data perturbation approach for enriching privacy preservation of healthcare data in hybrid clouds."** International Journal of Intelligent Engineering and Systems 11.1 (2018): 201-210.
21. https://archive.ics.uci.edu/ml/datasets/Arrhythmia.
22. Ruth Ramya K., Saikrishna D.N.V., Sravya Nandini T., Tanmai Gayatri R," **A survey on using**

**biometrics for cloud security**”.International Journal of Engineering and Technology(UAE),2018

23. Vurukonda N., Thirumala Rao B.,”**Hash counter hash method for privacy and security in cloud computing with attribute-based encryption**”,Journal of Advanced Research in Dynamical and Control Systems,2017
24. Ranjeeth Kumar M., Srinivasu N., Reddy L.C.,”**Fine grained multi access control via group sharing in distributed cloud data**”,Journal of Theoretical and Applied Information Technology,2017
25. Wadhya R., Divya Harika B., Sandeep Reddy C., Krishna Reddy V.,”**Security for data storage in cloud**”,Journal of Advanced Research in Dynamical and Control Systems,2017
26. Vurukonda N., Thirumala Rao B.,”**Secure sharing of outsourced data in cloud computing with comparison of different attribute based encryption**”,Journal of Advanced Research in Dynamical and Control Systems,2017
27. Dr.V.Naresh, T.Gopi Venkata Ajay, T.Naga Sai Reddy, M.Srinivas,”**An Efficient And Privacy Preserving Biometric Authentication Scheme In Cloud Computing**”, International Journal Of Scientific & Technology Research Volume 9, Issue 01, January 2020 Issn 2277-8616.
28. Vulapula Sridhar Reddy, Malladi Srinivas," **Secure Data Accessing Over Cloud Computing Environment Using Hybrid Query**", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 14-Special Issue, 2018.
29. Vulapula, Sridhar Reddy, and Malladi Srinivas. "**Review on privacy preserving of medical data in cloud computing system**." Indian Journal of Public Health Research & Development 9.12 (2018): 2261-2269.
30. Dasari, Mr RakeshNag, Y. Prasanth, and O. NagaRaju. "**An Analysis of Most Effective Virtual Machine Image Encryption Technique for Cloud Security**." International Journal of Applied Engineering Research 12.24 (2017): 15501-15508.
31. Vurukonda N., Sai Teja K.V., Naveen C., Hemamadhuri K. **An efficient data loss prevention in cloud computing with data classification**",International Journal of Innovative Technology and Exploring Engineering, vol-8, Issue-7. May, 2019
32. Parikshith Nayaka S K1, Dayanand Lal.N2, Vasudev Shahapur3 , Saritha A K4 , Nida Kousar. **A Modern themed System for Patients Security of data exposure in semi-convinced Servers in the Cloud,** International Journal of Emerging Trends in Engineering Research, Volume 8. No. 8, August 2020.
33. M.Robinson Joel , V. Ebenezer , M. Navaneethakrishnan , N. Karthik **Encrypting and decrypting different files over different algorithm on Cloud Platform,** International Journal of Emerging Trends in Engineering Research, Volume 8. No. 4, April 2020.