AN ENHANCED TBAHIBE-LBKQS TECHNIQUES FOR PRIVACY PRESERVATION IN CLOUD

VALARMATHI G
Assistant Professor
SSR College of Arts, Commerce and Science
Silvassa, U.T of Dadra & Nagar Haveli

JINIL JACOB
Assistant Professor & HOD
SSR College of Arts, Commerce and Science
Silvassa, U.T of Dadra & Nagar Haveli

DOI: 10.17632/czdwpv35s.1

Abstract — In recent days, providing security to the data stored in cloud is an important and challenging task. For this purpose, several existing privacy preservation and encryption algorithms are proposed in the existing works. But, it has some drawbacks such as, high cost, required more amount of time for execution and low level security. In order to overcome all these drawbacks, this paper proposes a novel techniques such as, Tiered Blind and Anonymous Hierarchical Identity Based Encryption (TBAHIBE) and Location Based Keyword Query Search (LBKQS) for providing privacy preservation to the data stored in cloud environment. In this work, the privacy is provided to the medical data stored in the Electronic Health Record (EHR).

This paper proposed a novel techniques, namely, Tiered Blind and Anonymous Hierarchical Identity Based Encryption (TBAHIBE) and Location Based Keyword Query Search (LBKQS) for data privacy preservation in cloud. The main intention of this techniques are to provide privacy for the data stored in the cloud and, to enable the search based on the location and keyword. This work includes two modules, such as, secure data storage and location based search in cloud. In the first module, the medical data of the egg and sperm donor, receptor, doctor and lab technician are stored in the Electronic Health Record (EHR). Here, the authenticated persons can view the medical data, for instance, the doctor can view the donor and receptor medical details. In the second module, the location based search is enabled based on the keyword and query. Here, the doctor, patient and other users can fetch the medical details in a filtered format.

Cloud computing is an emerging and demanding technology that provides different cloud services to the user. It requires the improvement in security performance for efficient data transmission. The cloud provides several advantages [1], including:
- Resource elasticity
- On-demand self service
- Shared pool access

The advantages are graphically represented in the following Fig 1. The main aim of cloud is to share the data over the scalable nodes such as, user computers, cloud services and data centers. The secure data storage plays an essential in cloud, where the large number of cloud users store their data in a remotely located servers. It reduces the affliction of managing and data storing within the local storage. Moreover, it can be demanded on requirement from the distributed cloud computing. Thus, this paper selects the cloud environment for secure data privacy preservation. Before storing the data into cloud, it must be encrypted by the data owners to protect the data privacy. So, exploring privacy preserving and effective search is an

Fig 1. Advantages of cloud
this key, the filtered results are obtained by using the proposed LBKQS technique. In experiments, the performance of the proposed techniques are evaluated in terms of computation cost, communication cost.

The remaining sections of this paper is organized as follows: Section II reviews some of the existing works related to privacy preservation and secure data storage in cloud. Section III provides the detailed description for the proposed TBAHIBE mechanism. Section IV evaluates the performance of both existing and proposed techniques. Finally, this paper is concluded and the future work to be carried out is stated in Section V.

II. RELATED WORK

This section presents some of the existing works related to encryption, data storage and local based search in cloud. Yang, et al [2] suggested a hybrid solution to preserve the medical data during the process of data sharing in cloud. This work includes the following components:

- Vertical data partition using medical data publishing
- Integrity checking
- Data merging for medical data
- Hybrid search across plain text and cipher text

Lin, et al [3] introduced a new cloud based framework to implement a self-caring service, namely, home diagnosis for big medical data. In this paper, the Hadoop cluster was implemented for both offline data storage and index building. The speed of medical record retrieval was improved, which was the major advantage of this paper. Dou, et al [4] developed a new model, namely, History record-based Service optimization method (HireSome-II) to improve the credibility of a composition plan. In this work, the k-means clustering. The advantage of this method was, it significantly reduced the time complexity. Wang, et al [5] proposed a new cryptographie technique, namely, Hierarchical Identity Based Encryption (HIBE) for multi-linear maps. Here, the secret keys were delegated at lower levels to prove the security level of the proposed technique. Wang, et al [6] introduced an efficient framework, namely, Constant-size Cipher text Policy Comparative Attribute Based Encryption (CCP-CABE) to embed the attributes into the user’s key. The main objectives of this paper were listed as follows:

- It provided a secure and efficient access control in a cloud environment.
- It handled more expressive types of access control by integrating the wildcards and negative attributes.
- It minimized the computation overhead on the resource constrained data owners.

Idris, et al [7] suggested a Big Data Service Engine (BISE) to provide the processing and storage services for human centric wellness data. The main objectives of this work were listed as follows:

- It offered the health care services by exploiting the big data technology.
- It maintained a physical, social and mental health services by using the service engine.

International Journal of Computer Application (2250-1797)
Volume 6- No.5, September- October 2016

- It performed analytics and present services by handling variety of data.

Zhang, et al [8] introduced an Extended Quasi Identifier (EIQ) partitioning technique to improve the privacy preservation of set valued data in hybrid cloud. The main intention of this paper was to protect the privacy of the data during the process of data publishing. In addition, the data publication solutions were provided in this work to ensure the confidentiality of the cloud data. The new query analysis tool was designed in this work to optimize the data query on hybrid cloud. Nepal, et al [9] identified a solution to ensure both the data integrity and confidentiality, when storing it on a hybrid cloud environment.

In this paper, the privacy breaches from counting query, linear query and batch linear query were employed in the data querying stage. Forkan, et al [10] suggested a new model, namely, Big Data for Context aware Monitoring (BDCaM) to discover the framework for assisted healthcare. Here, the Context Management System (CMS) was utilized to store the context histories of patients. Almrri and Lee [11] proposed a new sharing system to attain the privacy and utility of desirable features. The four main entities utilized in the system are patients, Cloud Service Provider (CSP), users and enterprise system. In this work, the security was provided against unauthorized access by removing the unique attributes in the outsourced data. In addition, two important issues such as, privacy protection and oblivious processing were analyzed within the untrusted domain. Hassanalieragh, et al [12] introduced an Internet-of-Things (IoT) based sensing method to monitor and manage the health of patients in cloud. It includes the following processes, such as,

- Data acquisition
- Data transmission
- Cloud processing

The major disadvantage of this technique was, it provided the training data for machine learning with little burden. Rajaei and Haghjoo [13] introduced an improved ambiguity based anonymization technique to preserve the data utility. In this work, different operations such as generalization, suppression, anonymization, permutation and perturbation were employed for anonymization. Moreover, the generalization based anonymization technique was applied to reduce the information while preserving the privacy requirements. Papadopoulos, et al [14] introduced a new privacy mechanism based on k Nearest Neighbor (kNN) queries to determine the notion of location privacy. The authors also suggested the following concepts to solve the problem of location privacy, which includes:

- Data transformation
- Location Obfuscation
- Private Information Retrieval (PIR)

Azaoui, et al [15] suggested a searching mechanism based on the publicly verifiable conjunctive keyword in the outsourced databases. In this work, the data owners ensured the properties of public delegatability and verifiability. Moreover, the polynomial based accumulators were utilized to represent the keywords in the database. The main contributions of this work were as follows:
\begin{itemize}
  \item It verified the correctness of the server’s response in a logarithmic time.
  \item It efficiently enabled the search of outsourced database.
\end{itemize}

Kuo [16] analyzed the issues and challenges in cloud to improve the health care services. The four different aspects such as, management, technology, security and legal were discussed to evaluate the services. Zhang, et al [17] suggested a Priority Based Health Data (PHDA) aggregation system with privacy preservation to aggregate the efficiency of health data in cloud. Based on the data priority, various forwarding strategies were utilized in this work. The advantages of this concept were, it provided lower communication overhead. Zhang, et al [18] introduced a scalable Top Down Specialization (TDS) approach to anonymize the large scale data in cloud using MapReduce framework. A group of innovative jobs were integrated to accomplish the computation in a scalable way. Prasad, et al [19] proposed a new data sharing and security mechanism to provide privacy preservation for cloud data. It utilized the Key Distribution Centre (KDC) mechanism to distribute and maintain the attributes and secret keys. Liu, et al [20] suggested a shared authority based authentication protocol to enable the privacy preservation in cloud. It attained shared access authority with the security and privacy considerations. Moreover, the proxy based re-encryption mechanism was applied to provide data sharing in cloud. Zhang, et al [21] analyzed the key research issues to improve the process of privacy protection and preservation. Based on the cloud service levels, the current cloud protection and preservation processes were determined. Moreover, the issues in both customer side and server side were classified by using the key research.

III. PROPOSED METHOD

This section presents the detailed description for the proposed Tiered Blind and Anonymous Hierarchical Identity Based Encryption (TBAHIBE) based privacy preservation mechanism. The main intention of this work is to securely store the data in cloud with privacy preservation. Here, the application of egg and sperm donor is considered. This work includes the following modules:

- Secure storage in cloud
- Location privacy search in cloud

At first, the users such as, doctor, patient and lab technician registered their accounts. Then, the personal profile will be created for them and it will be stored in the Cloud Electronic Health Record (HER) database. Here, a novel encryption technique, namely, Tiered Blind and Anonymous Hierarchical Identity Based Encryption (TBAHIBE) mechanism is proposed for data encryption, which generates the individual key for the encrypted data. After encryption, the secrecy will be maintained for each data stored in the database. Moreover, the Quasi and sensitive identifiers are utilized to provide the privacy for the preserved data. In the second module, the doctor, patient and other users (visitors) login into their account using the unique ID and password. If it is an authenticated account, the normalized data will be retrieved by using the proposed Location Based Keyword Query Search (LBKQS) technique. Here, the data are searched based on the location. Finally, the filtered data will be obtained based on the similarity search. The detailed description is provided in the following stages:

A. Secure Storage in Cloud

In this paper, the medical application of egg and sperm donor, receptor is taken for analysis. The main aim of this process is to hide the privacy data of both the donor and receptor. The flow of the secure storage in cloud is shown in the following Fig 2. Initially, the registration process is performed by creating the login account for doctor, patient and lab technician. Then, the personal profile is generated for each cloud user, who registered their accounts. Here, the data is stored in the encrypted format by using the proposed TBAHIBE. It generates the individual key for user and it stores the encrypted data in the cloud EHR database. Here, the lab technician will update the medical details of the patient with anonymity. Then, the Quasi and sensitive identifiers are utilized to retrieve the data from the EHR. Normally, the quasi identifier is used to uniquely identify the tuples in the table. Moreover, it links the anonymized dataset other datasets. The sensitive identifiers reveal the sensitive information of individuals by linking with the external data. The main intention of blinding is to offer perfect confidentiality of a message and its corresponding signature.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{Fig2.png}
\caption{Secure storage in cloud}
\end{figure}

Algorithm 1 –Tiered Blind and Anonymous Hierarchical Identity Based Encryption (TBAHIBE)

\begin{verbatim}
// Register
Users (U_i = 1 to N)
For each U_i \leftarrow Register into the cloud database (C_d):
Hierarchical Identity = \{Hosp_D, D_I, U_LOC, U_ID\};
Key \leftarrow Identity;
Data (P_D, M_D, M_key);
C_d \leftarrow Load (Dec)
// Encryption
For any (U_i || D || L_i) // Where, user or doctor or lab technician
\end{verbatim}
The Algorithm I shows the clear procedure of TBAHIBE encryption technique. It includes two processes such as, registration and encryption. The medical data of each user is stored in the cloud database. These information are arranged in a hierarchical manner like hospital ID, doctor ID, user location and user ID. After loading the data, it will be stored in the encrypted format. Algorithm II shows the process of data accessing from cloud database. After storing the data, it will be accessed by the authenticated persons, who have the unique ID and password. Here, the authenticated key is generated with the location and time stamp for login verification. Based on the user ID and password, the data can be retrieved from the cloud database.

Algorithm II - Data accessing from cloud database

Input: Uniqid ($U_{ID}$) and pass word $P_w$;
Output: Key for authentication;
Generate the authentication key ($Key_{Au}$) with the location and the time stamp ($Loc, T$) for login verification;
Verify ($U_{ID}, P_w$);
$Key_{Au} \leftarrow Gen (U_{Loc}, Cur_T)$;
If ($U_{ID}$ in need of Donor)
$Gen (Dmr_{ID}, U_{ID})$;
$RL \leftarrow add (Dmr_{ID})$;
If ($U_{ID}$ in wish to Donate)
$Gen (Recp_{ID}, U_{ID})$;
$DL \leftarrow add Recp_{ID}$;
End if;
End if;
End for;

Algorithm III describes the procedure of best match identification of donor for a receptor. Here, the donor list and the receptor list are given as the input, based on this, the best match of donor to receptor is identified. At first, the donor with least count are extracted and arranged in an alphabetical order. Then, it will be prioritized with least count. For each receptor in RL list, the donors with least count is loaded. Based on the location, gender and race, the match will be identified.

Algorithm III - Best Match Identification of Donor for a Receptor

Input: Donor List (DL), Receptor List (RL);
Output: Best match of donor to receptor;
Step 1: Extract the donor with least count ($D_{count}$);
Step 2: $D_{min} = min count (DI)$;
$Dlp \leftarrow Ascending (DI)$ based on $D_{min}$;
Algorithm IV shows the procedure of location based keyword query search. In this stage, the visitors who have the valid ID and password can request the query to the database. If the keywords and the corresponding key stored in the database are matched, the results are retrieved from the database. Moreover, the authentication, correctness and freshness processes are performed to provide the data in a filtered format.

**Algorithm IV – Location based keyword query search**

```
For each user who visit Cd
    Register (UrVID, UrLoc);
    Load Query (Q);
    Identify (Kw) from Q with Dynamic Cuckoo Hashing (DCH); // Kw represents the keywords;
    Match (lKey, Kw);
    If Index matched
        Genkey ()
        Skey(Vid)_l ↦ (UrVID, UrLoc )
        Results (R) ↦ Process (Q, UrVID);
        Ddec (DecResults, Skey(Vid)_l)
        If (R ↦ Retrieval (Cd))
            Check (Correctness (R));
            Check (Completeness (R));
            Check Freshness (R);
            View Anyrst ↦ (UrVID, UrLoc):
        End if;
        //Correctness (Anyrst)
        If (Anyrst ↦ Cd) && UrLoc ↦ RegLoc)
            Ensure proof of correctness;
        Else
            Perform authentication;
        //Completeness (Anyrst)
        If (Anyrst ↦ Cd) && UrLoc ↦ RegLoc)
            Match (Q, Anyrst);
            If (Partial (Anyrst))
                Repeat search with more specific query;
            Else
                Ensure completeness;
        End if;
        //Freshness (Anyrst)
        Extract Anyrst ↦ Cd;
        Set Lmax with respect to CurT;
        If (Anyrst within Lmax)
            Ensure proof of freshness;
        Else
            Repeat query search;
```

### IV. PERFORMANCE ANALYSIS

This section presents the results of both existing and proposed techniques. Here, the medical application is taken for the analysis with own dataset. The data are stored into the dataset in the encrypted format. The existing encryption and search based techniques considered in this work are, target snapshot top k-queries, moving snapshot top k-queries, Secure Data Outsourcing (SDO), Baseline Method (BM), Reverse top-k Boolean Spatial Keyword (RKBKS), Enhanced RKBKS (ERKBSK), Traditional Rivest Shamir Adleman (RSA) and Modified RSA. The performance of these techniques are evaluated in terms of,

- Computation cost
- Communication cost
- Query evaluation
- Query time
- Computation time
- Encryption time
- Decryption time
- Key generation time

#### A. Computation Cost

The computation cost is the amount of cost required for the process of query search. Here, it is calculated by varying the number of keywords from 0 to 100, which is shown in below:

\[
\text{Computation Cost} = \frac{N_H}{T}
\]

Where, \(N_H\) represents the number of hash computation to verify query results and \(T\) indicates the required amount of time. Fig 4 graphically represents the amount of required computation cost for both existing and proposed location based searching techniques. When compared to the existing target snapshot top k-queries and moving snapshot top k-queries techniques [22], the proposed LBKQS requires the minimum computation cost.

![Fig 4. Computation cost of both existing and proposed techniques](image)

#### B. Communication Cost

Communication cost is defined as the cost required to retrieve the query results from the database. Fig 5 shows the communication cost required for both existing target snapshot top k-queries, moving snapshot top k-queries and proposed LBKQS techniques. The communication cost is calculated by
varying the number of keyword queries, which is shown in below:

\[ \text{Communication Cost} = \frac{Q_R}{N_K} \quad (2) \]

Where, \( Q_R \) represents the query region size to retrieve the query results and \( N_K \) indicates the number of keyword queries. From the results, it is observed that the proposed LBKQS technique requires low cost, when compared to the existing techniques.

C. Query Evaluation

Fig 6 shows the cost of query evaluation with respect to different key size (KB). Moreover, the cost is calculated for both existing SDO [23] and proposed LBKQS techniques. From this analysis, it is evaluated that the proposed LBKQS technique requires low cost, when compared to the other technique. It is calculated as follows:

\[ \text{cost} = \frac{Q_D}{T} \quad (3) \]

Where, \( Q_D \) represents the decryption of query results for the requested query region and \( T \) is the required amount of time.

D. Query Time

Fig 7 depicts the query time of existing BM, RKBSK, ERKBSK [24] and proposed LBKQS techniques. The time is estimated by varying the number of query keywords. The query time is calculated as follows:

\[ \text{Query time} = \frac{k_Q}{M_T} \quad (4) \]

Where, \( K_Q \) indicates the number of keyword queries and \( M_T \) indicates the maximum time to execute the query for all keywords. When compared to the existing techniques, the proposed LBKQS technique requires minimized query time.

E. Computation Time

Fig 8 shows the computation time required for both existing MDE [2] and proposed TBAHIBE encryption techniques. The amount of time required to process to number of EHRs from 50000 to 200000. Here, the colored lines indicates the number of attributes. From this evaluation, it is obtained that the proposed TBAHIBE requires less time, when compared to the existing MDE technique.
F. Encryption Time

Encryption is defined as the process of converting the original data into an unknown format. Encryption time is the amount of time required to encrypt the given data. In this paper, the medical details of the egg and sperm donor and receptor are stored in EHR in the encrypted format. Here, the encryption time is calculated for both traditional RSA, modified RSA, ESRKGS [25] and proposed TBAHIBE techniques.

\[
\text{Encryption time} = \frac{N_A}{T}
\]  

(4)

Where, \(N_A\) represents the number of attributes in \(N\) number of medical records and \(T\) indicates the amount of time taken to encrypt.

G. Decryption Time

Decryption is defined as the process of converting the encrypted text into the original text. Decryption time is the amount of required to convert the encrypted data into the original data. Fig 10 shows the decryption time of both existing and proposed techniques with respect to different key size (KB). When compared to the existing techniques, the proposed TBAHIBE requires minimized decryption time (ms).

V. CONCLUSION AND FUTURE WORK

This paper proposed a novel scheme, TBAHIBE-LBKQS techniques to provide privacy preservation for medical data. The main intention of this work is to provide high privacy to the data stored in EHR. Here, the medical data of the egg and sperm donor and receptor are stored in the cloud environment. This work includes two modules, such as, secure storage in
cloud and location based search. At first, the registration process is performed by updating the patient’s personal information in the encrypted format by using the TBAHIBE technique. In the second module, the user can query a keyword based on the location to retrieve the data by using the LBKQS technique. Finally, the authenticated users retrieved the filtered medical information. The main advantages of this work are, high security, privacy, low cost and minimum processing time. Moreover, the proposed techniques are compared with some of the existing techniques used for encryption and location based search. The results are analyzed and evaluated in terms of computation cost, computation cost, query evaluation, encryption time, decryption time, query response time and computation time. From this analysis, it is proved that the proposed TBAHIBE-LBKQS techniques provides better results, when compared to the existing techniques.

In future, this work will be extended to device specific with location coordination based searching.

REFERENCES

[1] Advantages of cloud computing. Available: https://www.google.co.in/search?q=advantages+of+cloud+computing&biw=1366&bih=667&tbm=isch&tbo=u&source=univ&sa=X&ved=0ahUKEwiF4P2Q_4DNAhVMY8KHDbNCIEQsAQIPg#imgrc=JNHCGZ8dQb29PiM%3A


