Keyword Search For Encrypted Data With Both Side Verification Using Blockchain

[1]Raju.S, [2]Vasudeva.R, [3]Shilpa.V

[1] M.Tech, [2]Assistant Professor, [3]Assistant Professor

[1, 2] Department of Computer Science and Engineering, C.Byregowda Institute of Technology, Kolar, Karnataka, India

[3]School of Computing and Information Technology, REVA University, Bangalore, Karnataka, India

Abstract : Now a days the cloud computing is popular because they reduce the cost and complexity of owning operating computers, network resources. The benefits of cloud are low up-front costs, rapid return of investment, rapid deployment, customization, flexible use and solutions that can make use of new innovations. It seems that cloud hacks are current fashion in this age of access and therefore mainstream cloud services comes with a certain amount of risk. In order to protect both data user and data owner needs to look for each provider's explanation of how can trust that user data is secure as send it to the cloud or download it. In this paper, we introduce keyword search for encrypted with both side verification using blockchain Technology. Data owner first encrypts the data before uploading to the cloud using index based on digital signature and has the rights to give the permission to the user, search over the outsourced data and check whether cloud fulfills the pre-specified condition or not.

***IndexTerms* - Blockchain, cloud computing, fair payment, searchable encryption, verifiability.**

I. Introduction

Cloud computing services cover a vast range of options now, from the basics of storage, networking and processing power through to natural language processing and artificial intelligence as well as standard office applications. Pretty much any service that doesn't require you to be physically close to the computer hardware that you are using can now be delivered into the cloud.

Techniques are available in cryptography such as symmetric key and asymmetric key encryption. In symmetric method, the same keys used for encryption and decryption but different keys used in asymmetric key. Advance the asymmetric key encryption more adaptable technique. Furthermore, this is important to address cloud security issues; specifically, explore different cloud security characteristics including vulnerabilities, threats, risks, and attack models distinguish the security requirements including accessibility, secrecy, trustworthiness, transparency, and so forth recognize the included gatherings (clients, service provides, outsiders, insiders) and the part of each party to the attack-resistance cycle and understand the effect of security on different cloud deployment models (public, private, hybrid &community). The fundamental commitment to this paper is a general way to deal with secure the data privacy is to encrypt the data before outsourcing. Next, this will achieve an immense cost as far as information usability.

Blockchain technology has the core characteristics of decentralization, accountability and security. This technique can improve operational efficiency and save costs significantly [11]. The structure of blockchain technology is represented by a list of blocks of transactions in a particular order. These lists can be stored as a flat file (txt. format) or in the form of a simple database. Two vital data structures used in blockchain include.

- **Pointers** - variables that keep information about the location of another variable. Specifically, this is pointing to the position of another variable.

- **Linked lists** - a sequence of blocks where each block has specific data and links to the following block of the help of a pointer.

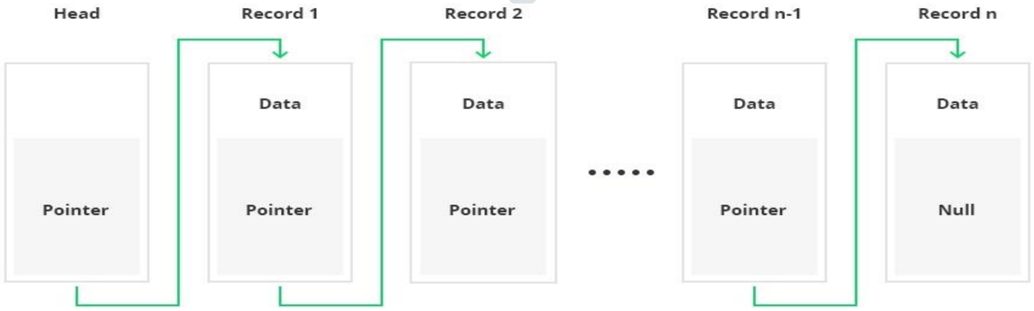

Figure 1: Blockchain hashing[11]

Logically, the first block does not contain the pointer since this one is the first in a chain. At the same time, there is potentially going to be a final block of the blockchain database that has a pointer towards no value as shown in the Figure 1.

II. Related Work

In the era of cloud computing it is easy to use it because of its elasticity and cost effective i.e., which is easy about managing the hardware and software so that the companies need not to buy it. By seeing these features the companies and user are planning to move for cloud. So that even power generation is changing their direction to cloud but the facility given by the 3rd party cloud has an extra security warning. The movement of data from one part of territory to the territory in the organization, due to number of users rapidly increasing the security should be concerned. This tells the solutions recently introduced to the work published to converse the security issues. In addition a highlight of short sight of security attacks in the mobile cloud computing. Finally at the end the issue that have been raised but yet not solved and oncoming probe direction is also submitted [1].

Cloud computing is a potential and emerging technology for next generation of computer applications. The secured data sharing methods provides security between data owner and user with the user's attribution. The obstacle and hurdles toward the rapid growth of cloud system affected by some major issues such as data security, privacy and data sharing. There are so many researchers have introduced different techniques for data defense also to attain the highest level of data security in the cloud computing system [2]. However, those have some issues, thus we required to solve in more effective way. In addition need to guarantee confidentiality of the outsourced information and the user should not burden by the local search. An advancement efficient file modified key policy ABE encryption scheme. The access to the layered structure has been united into single structures while the access to stored file encrypted by united structure. Then the secret data has encrypted by data owner using modified KP-ABE Encryption scheme. The cipher text part connected to attributes that is shared by the file. Since both cipher text storage time and encryption cost is stored.

To lessen threat and risk the data is kept in encoded form which is stored in mail as well as file server i.e., one data on another one by the server which is used to store. Normally implies to giving up for the purpose of security [3]. Let us consider an example that if customer wants to fetch particular words of the document where the data stored in the server carryout search and reply to query without loss of data confidentiality was not known before. The difficulty of search for encoded form of data and to furnish evidence of security for the crypto system result is done by the cryptographic schemes which have number of advantages that may be positive or negative. The encryption for provable secrecy is granted that is secure means the cautious server cannot learn no matter what is given in the plaintext. When the query is furnished with searching separately if they have only cipher text then the cautious server doesn't know about the plaintext but only the search result. The random word cannot be searched by the cautious server without the permission of the user because of the controlled search. The secret word search can be done by the user to the server where cautious server doesn't know the word by the support of hidden queries.

We study the problem of searching for data that is encrypted using a public key system [4]. For searchable symmetric encryption schemes [5] (or symmetric-key encryption with keyword search) the security against passive adversaries (i.e. privacy) has been mainly considered so far. In the past few years, there has been a significant growth of the interest to outsource data as well as operational services to clouds. Traditional cloud storage has come to rely almost exclusively on large storage providers acting as trusted third parties to transfer and store data. This system poses a number of shortcomings including the performance, availability, security, and high operational cost.

A decentralized cloud storage network has been introduced into many advantages over the datacenter-based storage. Similar to traditional solution, decentralized cloud storage network leverages client-side encryption to maintain data security. However, the management of encrypted data poses several challenges, the most important of which is data usability. More concretely, the data owner should have the capability to grant permission for others to search for the remotely encrypted dataset and obtain partial but useful content [6].

Fog computing can be viewed as an extension of cloud computing that enables transactions and resources at the edge of the network. In the paradigms of fog computing, the fog user (outsourcer) with resource-constraint devices can outsource the distributed computation tasks to the un-trusted fog nodes (workers) and pays for them [7].

The problem of simultaneously achieving fine-grainedness, high efficiency on the data owner's side and standard data confidentiality of cloud data sharing actually still remains unresolved, addresses this challenging issue by proposing a new attribute-based data sharing scheme suitable for resource-limited mobile users in cloud computing [8].

Linearly homomorphic signature schemes allow performing linear computations on authenticated data and the correctness of the computation can be publicly verified. The ID-based linearly homomorphic signature schemes can be applied for e-business and cloud computing [9].

Note that a symmetric balanced incomplete block design is utilized for key generation, which substantially reduces the burden on members to derive a common conference key. Both theoretical and experimental analyses demonstrate that the proposed scheme is secure and efficient for group data sharing among cloud computing [10].

III. PROPOSED WORK

3.1 System Architecture

The data owner has a collection of n files to outsource onto the cloud server in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other authorized users. To achieve this, the data owner needs to build a searchable index from a collection of keywords then outsources both the encrypted index and encrypted files onto the cloud server as shown in the Figure 2.

The data user is the authorized ones that have the rights to access the documents of data owner and by the search control mechanism to fetch keyword encrypted document from cloud server. Next the user can download the document using the key by the decryption processes to the client system

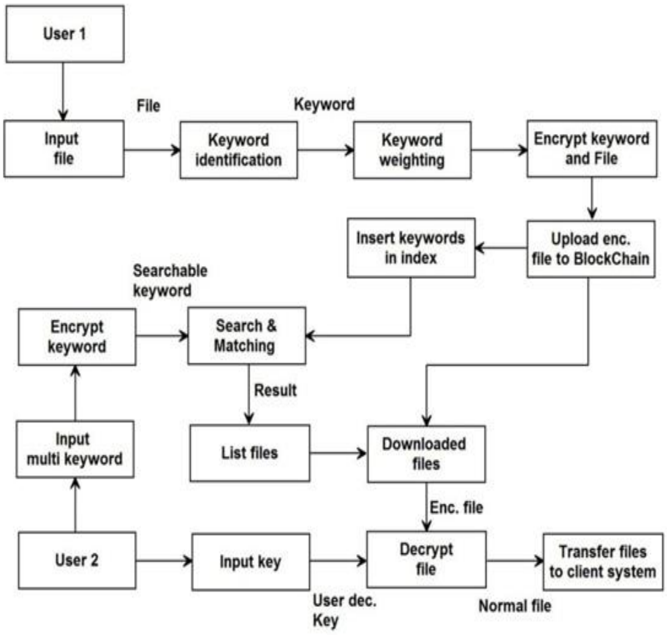

Figure 2: System Architecture

3.2 RSA Algoirthm

The *RSA algorithm* is the basis of a cryptosystem a suite of cryptographic *algorithms* that is used for specific security services or purposes which enables public key *encryption* and is widely used to secure sensitive data, particularly it is being sent over an insecure network such as the internet. It is an asymmetric algorithm. The RSA algorithm involves three steps key generation, encryption and decryption. Key generation RSA involves a public key and a private key. The public key can be known to everyone and is used for encrypting messages. Messages encrypted with the public key can only be decrypted using the private key. The keys of the RSA algorithm are generated the following way.

1. Choose two distinct prime numbers a and b.
2. Compute n = ab. n is used as the modulus for both the public and private keys
3. Compute $\varphi(n) = (a - 1)(b - 1)$, where φ is Euler's totient function.
4. Choose an integer e such that $1 < e < \varphi(n)$ and greatest common divisor of $(e, \varphi(n)) = 1$; i.e., e and φ(n) are co prime. e is released as the public key exponent having a short bit-length.

3.3 Keyword Search

The search method checks queried keywords exist on a file or not. If the user searches for a single or more keywords, there will possibly be many correct matches where some of them may not be useful for the user at all. Therefore, it is difficult to decide as to which documents are the most relevant add ranking capability to the system by adding extra index information on frequently occurring keywords in a file with ranking the user can retrieve only the top matches chosen by the user. In order to rank the file, a ranking function is required, which assigns relevancy scores to each document matching to a given search query.

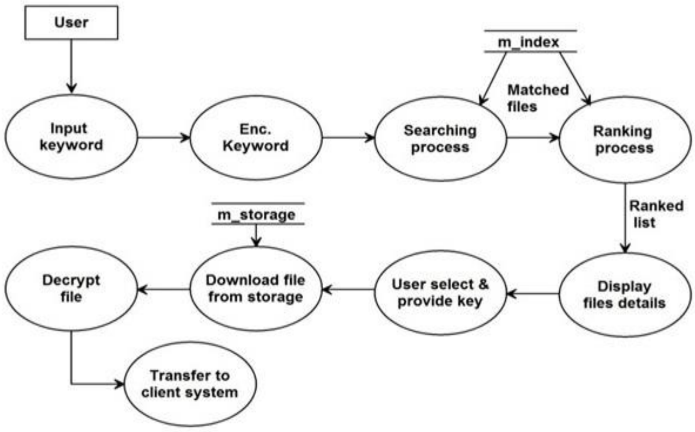

Figure 3: Data Flow Diagram

In the level1Data Flow Diagram, select a file and transfer that file to server. Server receives all the details and generates a Message Digest ,once MD file is generated it retrieves all the public key belongs to the user group i.e.(MD + Public key) generates a secure MD, Encrypt secure MD with user Private keys and generates Ring-Signature and send a mail to all users as shown in the Figure 3.

3.3.1 Term Frequency[12]

TF-IDF stands for *Term Frequency Inverse Document Frequency* and the TF-IDF weight is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection or corpus. The importance increases proportionally to the number of times a word appears in the document but is offset by the frequency of the word of the corpus. Variations in the TF-IDF weighting scheme are often used by search engines as a central tool for scoring and ranking a document's relevance given a user query. One of the simplest ranking functions is computed by summing the TF-IDF for each query term; many more sophisticated ranking functions are variants of this simple model.

How to compute

The TF-IDF weight is composed by two terms: the first compute the normalized Term Frequency (TF), the number of times a word appears in a document, divided by the total number of words in that document; the second term are the Inverse Document Frequency (IDF), computed as the logarithm of the number of the documents in the corpus divided by the number of documents where the specific term appears.

- **TF: Term Frequency**, which measures how frequently a term occurs in a document. Since every document is different from length, it is possible that a term would appear much more times for long documents than shorter ones. Thus, the term frequency is often divided by the document length (the total number of terms of the document) as a way of normalization:

  TF(t) = (Number of times term t appears in a document)
  / (Total number of terms in the document)

- **IDF: Inverse Document Frequency**, which measures how important a term is. While computing TF, all terms is considered equally important. However it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by computing the following:

  IDF(t) = log_e (Total number of documents /
  Number of documents with term t in it).

3.4 Sequence Diagram

The sequence of interaction is displayed in Figure 4. The sequence diagram tries to portray the entities, initially data owner uploads the files to the cloud server then the file containing unnecessary words is removed after the weightage process of the keyword and a hash code is generated for that keyword and stored in the index where acknowledgement sent to the gateway server then the file is encrypted before it stored in block and an acknowledgment is sent to the gateway server as well as confirmation to the data owner.

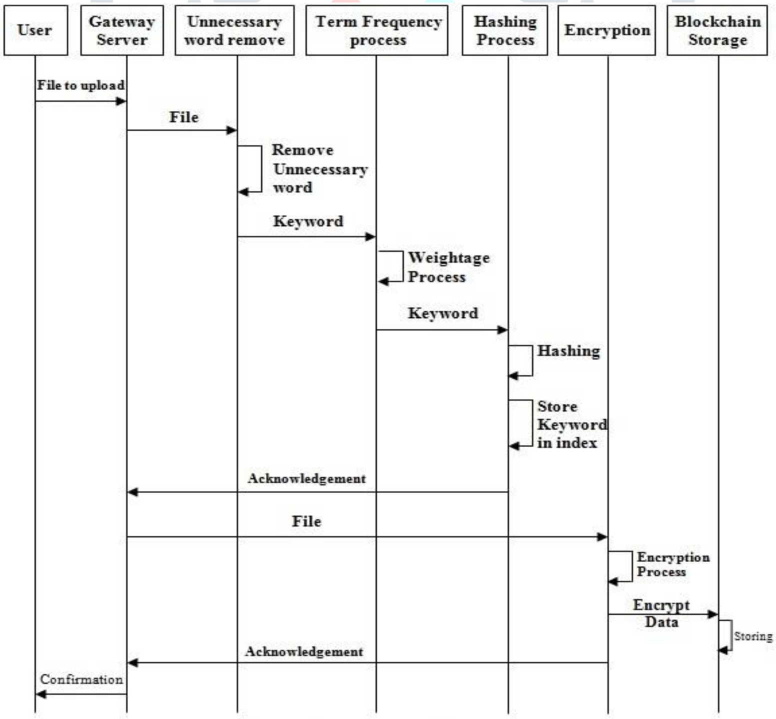

Figure 4: Sequence Diagram

IV. Experimental Setup

➢ **Data Admin**

The data owner registers the user details and also includes user details here the owner uploads the file with encryption this ensures the files to be protected from unauthorized user.

The Data admin page contains following links

- User creation: User is created by the admin consist of various details like name, password and email so on.
- User Deletion: Admin has the rights to delete the user.
- Upload File: The file is uploaded by the admin for the specific grade of the user
- User Request: admin can either accept or reject the user request. If it is accepted means public key is sent to the user otherwise no.

- **Data User**

This data user is used to search the file using the keywords and get the accurate result list based on the user query and user to download the file using the key to decrypt the downloaded data.

The data user has some features are as follows

- Send Request: The user here able to send the request for the key for decrypting the file.
- Search Keyword: If the user wants to search a specific keyword if the keyword is found it will show the related file that you can download or it will display keyword not found.

V. Results and Discussion

- Encryption and Decryption Result: The data is encrypted by applying the RSA algorithm and that data is stored in cloud by this user can access the file using the key for decryption process.
- Ranking details: The term frequency algorithm is used to compute ranking i.e., whenever the user searches for the data, the user gets the expected result of the query.

The proposed system uses the blockchain for data storage access, permission and searches. This mechanism for a user to identity the specific data rather than downloading entire files the user interacts with the blockchain content to get the specific data based on the keywords. Finally the user retrieves the encrypted file from the cloud. The technologies used for searchable symmetric encryption is blockchain.

Table 1: Feature Comparison

Schemes	User Side Verification	Server Side Verification	No TTP
[13]	×	×	-
[5]	✓	×	-
Our method	✓	✓	✓

Table 1 represents feature comparison of schemes the scheme [13] the "×" symbol represent the property doesn't performed and the "-"symbol mention the property not to be considered. The scheme [5] where the user side is accomplished but not the server side but in our method both side verification is supported and doesn't require any trusted third party.

VI. Conclusions

Intending to acknowledge secure keyword searched for encoded information on noxious clients and malevolent cloud providers a dependable keyword search conspires over encrypted distributed storage without requiring any trusted third party. So for this RSA algorithm is used for securing the file before it is uploaded to the cloud and keyword search is designed for searching the encrypted data. Our security investigation and execution assessment demonstrated that it is secure and productive. In future this can be adopted for Mobile cloud computing, Wireless body area network, Smart home system, Bitcoin based fair payment and Electronic health record system. Ensuring privacy protection and fast decryption for outsourced data security in cloud computing.

References

[1] M.Ali, S.U. Khan, and A. V.Vasilakos, ''Security in cloud computing: Opportunities and challenges,'' Inf. Sci., vol. 305, pp. 357–383, Jun. 2015.

[2] Z.Wan, J. Liu, and R. H. Deng, ''HASBE: A hierarchical attribute - based solution for flexible and scalable access control in cloud computing,'' IEEE Trans. Inf. Forensics Security, vol. 7, no. 2, pp. 743–754, Apr. 2012

[3] D. X. Song, D. Wagner, and A. Perrig, '' Practical techniques for searches on encrypted data, '' in Proc. IEEE Symp. Secur. Privacy, May 2000, pp. 44–55.

[4] D. Boneh, G.Di Crescenzo, R. Ostrovsky, and G. Persiano, ''Public key encryption with keyword search,'' in Eurocrypt, vol. 3027. Berlin, Germany: Springer, 2004, pp. 506–522.

[5] K. Kurosawa and Y. Ohtaki, ''UC - secure searchable symmetric encryption,'' in Financial Cryptography, vol. 7397. Berlin, Germany: Springer, 2012, pp. 285–298.

[6] H. G. Do and W. K. Ng, ''Blockchain - based system for secure data storage with private keyword search,'' in Proc. IEEE World Congr. Services (SERVICES), Jun 2017, pp. 90–93.

[7] H. Huang, X.Chen, Q. Wu, X. Huang, and J. Shen, ''Bitcoin - based fair payments for outsourcing computations of fog devices,'' Future Gener.Comput. Syst., vol. 78, pp. 850–858, Jan. 2018.

[8] J. Li, Y. Zhang, X.Chen, and Y. Xiang, ''Secure attribute-based data sharing for resource-limited users in cloud computing,'' Comput. Secur., vol. 72, pp. 1–12, Jan. 2018.

[9] Q. Lin, H. Yan, Z. Huang, W. Chen, J. Shen, and Y. Tang, ''An ID -based linearly homomorphic signature scheme and its application in blockchain,'' IEEE Access, vol. 6, pp. 20632–20640, Feb. 2018.

[10] J. Shen, T. Zhou, X. Chen, J. Li, and W. Susilo, ''Anonymous and traceable group data sharing in cloud computing,'' IEEE Trans. Inf. Forensics Security, vol. 13, no. 4, pp. 912–925, Apr. 2018.

[11] https://mlsdev.com/blog/156-how-to-build-your-own-blockchain-architecture.

[12] http://tfidf.com

[13] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, ''Searchable symmetric encryption: Improved definitions and efficient constructions,'' in *Proc.13th ACM Conf.* Comput. Commun. Secur., 2006, pp. 79–88.