IPHV8I3021X
International Journal Of Advanced Research and Innovation -Vol.8, Issue .III
ISSN Online: 2319 – 9253
Print: 2319 – 9245

EFFICIENT DATA SHARING SCHEME OVER ENCRYPTED CLOUD DATA

[#1]UMME SALMA, M.Tech Student,
[#2]Dr.PRABAHARAN, Associate Professor,
[#3]Dr.M.SUJATHA, Associate Professor,
Department Of CSE,
JYOTHISHMATHI INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR T.S.INDIA.

ABSTRACT: Cloud computing is very popular for its computing and storage capacity at lower cost. More and more data are being moved to the cloud to reduce storage cost. On the other hand, since the cloud is not fully trustable, in order to protect data privacy against third-parties and even the cloud server, they are usually encrypted before uploading. However, many operations, such as searching, are hard to perform on encrypted data. To solve this problem, searchable encryption has emerged. Searchable encryption in multi-user setting is much less efficient than in single-user setting. In order to address this problem, we propose a multi-owner to multi-user searchable encryption scheme based on attribute-based encryption. Our scheme keeps data secure in the cloud even against the cloud server. It allows users with appropriate authorizations to perform search operations on encrypted data. In addition, search tokens are generated by users instead of data owners. We prove that token privacy and index privacy are well protected in our scheme. The cloud server and illegimate users are not able to get any useful information about search tokens and ciphertexts. Ciphertexts of our scheme are constant-size, which reduce the time-complexity and bandwidth overhead of our scheme.

Keywords: *Cloud storage, cloud computing, attribute-based encryption, multi users searchable encryption.*

I. INTRODUCTION

With the development of the information technology, more and more information exchanges on the Internet. Therefore, not only the computers but also the mobile devices produce high-volume of data. As a result, cloud storage has become more and more popular with the low cost and vast capacity. While the information uploaded to the cloud may be sensitive and data owners want to keep them secret and prevent them from being exposed. What concerns data owners is that the cloud server is capable of manipulating their data in the cloud storage, so a method of encryption on the important data before uploading has been adopted. On the other hand, the goal of file storage is that we can find data when we need them. Although encryption helps in keeping data safe, it makes performing computation on the data difficult, such as search operations. One of naive methods is to download all encrypted files and decrypt them, then we can perform search operations on the plaintext files. This results in the huge bandwidth overhead and extra cost on the storage of downloaded files.

The searchable encryption(SE) can solve the problem of searching on the ciphertext and take full advantage of the cloud computing. Let's consider a scenario: In a hospital, all patients' information are uploaded to the hospital system in the form of ciphertexts, which include their names, telephone numbers, disease types, record time, the name of their doctors and so on. As a patient, he should be able to search all the information about himself and nothing about other people. As a doctor, to track the patients' recovery, he can search the names and telephone numbers of his patients. Even the doctor of a patient A changed for some reason, the new doctor still can search the name and telephone number of A without encrypting A's information again. The statistical department can search the number of the patients who affected some type of diseases during a certain period of time. Therefore, in this scenario, the different authorities should be given to different people so that they can search different information according to their roles. However, most existing schemes are not able to achieve this. In the general symmetric searchable encryption(SSE) setting, when a data user wants to search on the files, he should request a search token from the data owner. After sending the search token to the cloud, the data users will receive search results computed by the cloud server. In this case, we can see that the keyword that the data user searched is known by the data owner which is not desirable. What's more, multi-owners to multi-users scheme model is more practical than one-owners to one users. And, the computing capacity of mobile phones and wearable devices are so week that they only perform efficient algorithms. The problem of the computation complexity and space complexity of is worth noting. And the response speed of search operations should be improved.

Libraries are at the brink of accepting the idea of cloud computing because of its both economic and technological advantages. Sharing resources among various academic

libraries through Cloud reduces the overall cost and escalate the efficiency. While the list of the above uses of cloud computing is not exhaustive, it certainly gives reasons to use the cloud while considering the traditional alternatives to increase IT infrastructure flexibility, as well as influence on big data analytics and mobile computing.

Storage and Retrieval techniques: Cloud computing is a collection of computing resources used for storing or accessing data from any distant place. Organizations outsource their data on the cloud. Of the many benefits of cloud computing, of which mainly are relief in storage management, global data access and avoidance of capital expenditure on hardware, software and maintenances [6, 7], all these are attributed to the features of on-demand resource availability and pay-as-use concept. The common fact is that many of the organizations encounter obstacles in secure information storage and retrieval on cloud. For resolving these, data comprising of highly confidential data like email, health records, financial transaction and government documents etc has to be encrypted prior to being outsourced to cloud.The possibilities remain that the cloud provider and unauthorized person can breach the security of data stored on the untrusted cloud and obtain it. Data loss and privacy breaches cloud computing systems are reported.

As a result, organizations, health care centres [10] and government are sending the confidential files onto the cloud storage space since they are facing difficulties in maintaining the hardware infrastructure on premises. Many enterprises like Windows Azure, Amazon, IBM etc.. supply cloud services established on basis of IaaS(Infrastructure-as-a-Service). Hence for privacy apprehensions, data intended to be stored is encrypted form; thereafter the owner of the data uploads the data that is encrypted onto the cloud server and later it is retrieved whenever the need arises. Efficient utilization of data stands as a challenge for a enormous number of outsourced data files. An array of data sharing and retreival schemes as shown in Table 1 are available for user accessing data on cloud. Search based on keyword can be titled as one of widespread technique that is made usefor investigating files on encrypted cloud data.Most commonly in plain-text scenarios, keyword search procedures are extensively used and the user is allowed to retrieve chosen files from the storage space.

All the conventional Searchable Symmetric Encryption (SSE)(e.g., [11–15]) paradigms permit a user to examine over cipher text and extract the cipher text securely from the encrypted cloud data by using keywords and not decrypting the stored files.This provisions only techniques like Boolean keyword search devoid of the consideration of any relevance of the document. In a case where enormous number of documents are concerned, Boolean keyword search faces a key disadvantage. It occurs particularly when an user intends to extract matching document for each search request with no prior knowledge of the encrypted cloud data and wishes to examine the entire list of retrieved files, then in such a case it (i) requires huge amount of post-processing in times when examining unrelated files thus producing massive network traffic. (ii) suffers communication overhead.The shortcomings considered above can be resolved with top-k single keyword retrieval techniques [5, 16, 17] and multi-keyword retrieval techniques.

II. RELATED WORK

Many search schemes have been proposed after the symmetric searchable encryption was first introduced in [1] and asymmetric searchable encryption was introduced in [2]. Searchable Encryption SE technology solves the problem of searching on the encrypted files and improves the practicability of cloud storage and cloud computing. Not only the functionality but also efficiency of the searchable encryption have improved a lot. In [3], the scheme which supported searching multi-keyword at the same time was proposed. [5] introduced a searchable scheme with dynamic updating. And, the time of updating, which includes addition and deletion, was as much as searching. To improve the efficiency, utilizing the useful RAM as a solution like [6] was deployed. Since Abdalla et al. [4] constructed asymmetric searchable encryption scheme from identity based encryption(IBE) and proved it secure, some works transformed the efficient IBE into efficient PEKS [23]. The multi-participants searchable encryption allowed data share among many people [21, 22], where the authorised data-users can search on the files uploaded by according data-owners. Our work emphasises the research on multi-users model in searchable schemes.

Index The index of the files is very important in search. Different indexes have both advantages and disadvantages. Curtmola et al. [1] proposed the first encrypted searchable index basing on the inverted index. Some searches base on inverted index because of the efficiency while it is not convenient for the files updating. The index basing on a bloom filter was introduced by Goh et al. [7]. While Chang et al. proposed a vector index in [8]. Diverse kinds of index are proposed to assist in perfecting the scheme.

Attribute-Based Searchable Encryption To remove the Trusted Authority(TA) in identity-based encryption schemes, Sahai and Waters [15] presented FIBE as a solution which we considered as the prototype of Attribute-Based Encryption(ABE). In practical scenes, ciphertext-policy ABE(CP-ABE) as an effective method to solve the data share in safety between multiple data-owners and multiple data-users, was very popular and discussed by many papers [16–18]. ABE family is diversified a lot. ABE plays an important role in fine-grained access control. The searchable encryption scheme with data users' attributes as the search secret key is more practical. Attribute-based searchable encryption(ABSE) and its application were proposed in [14,24]. As scheme in [14], data owners

encrypted their index using different access policy, then data users were able to search keywords that they were interested in if their attributes satisfied the associated access policy. Besides, [14] proposed verifiable attribute-based searchable encryption on the malicious cloud. Most existing works has implemented additional function of ABE in ABSE, like revocability of data users' search right. However, the computation complexity and communication efficiency of the above schemes were not taken into account. In this paper, we use ABE as a smoothly tools to implement the search on ciphertext. Therefore, we require the ABE scheme with enough efficiency. We adopt constant-ciphertext extension ABE [19, 20] to prevent the time complexity and space complexity increase obviously with the data growing on the cloud. Besides, we reduce the communication cost between data-owners and cloud when upload the constant ciphertext.

III. PRELIMINARIES

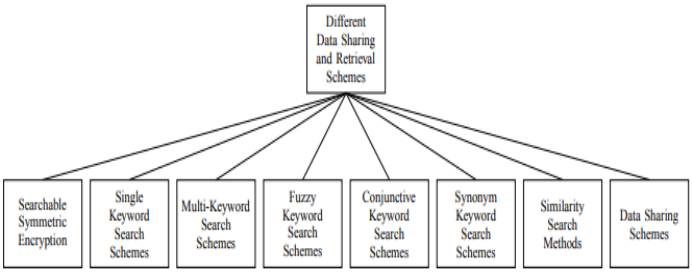
A. Vector Space Model

As mentioned earlier and in [5] for a single keyword search, the ranking is done using the TF-IDF scheme while for multi keyword search, it is employed using vector space model to score a file. The vector space model [23] is an algebraic model for representing a file as a vector. Every separate term represents a dimension of the vector; for example, in a file when a term occurs, its weight in the vector is non-zero otherwise is zero. Vector Space Model Scheme supports features like allowing an extent of striking similarity between files and queries and then ranks the files based on their relevance and also supporting multi term and non-binary presentation. Based on the weight or score of a file, files are ranked and presented in the top-k ranked order for a given search.

B. Inverted Index

Inverted index is a kind of indexing structure that contains a list of mappings from the set of keywords to the corresponding set of files containing the keyword in the file collection uploaded in the cloud server. For Ranked keyword search scheme, the task of determining the files that are most relevant is typically done by assigning a numerical score, which can be pre-compiled, to each file based on some ranking function.

TABLE I
DIFFERENT DATA SHARING AND RETRIEVAL SCHEMES



C. Ranking Function

The Ranked Searchable Symmetric Encryption Scheme(RSSE) is based on keyword search. The search generates a ranked order of the files containing a keyword based on the keyword search request sent to the server by a data user; a ranking function or relevance criteria is required to sort the retrieved encrypted files. The most widely used technique for evaluating relevance score is TF x IDF rule, where TF is term frequency which is the number of times a term or keyword is present in a file and IDF is Inverse Document Frequency that is calculated by dividing the number of the files in the entire collection to the number of files containing the particular keyword being searched by the user. Many variations of the TF x IDF Scheme is available, but none of the variant schemes of TF × IDF rule overshadow the other in terms of outcome [24].

IV. SEARCHABLE SYMMETRIC ENCRYPTION

With the concept of data storage in Cloud, Private-key storage is often opted for confidentiality and purpose of data security. In Private-key storage [25–27] when a data user has limited resources, implementation of outsourcing reduces the cost to a minimum for them to store and distribute large amounts of data that is symmetrically encrypted. The reason being that, in regular private-key encryption, data user is not able to retrieve only a select segment of their data as search over encrypted data is prevented. In order to solve this issue, proposals have been made for techniques on symmetric encryption with search abilities enabled [11, 28–30]. This paradigm is known as searchable symmetric encryption. The area of searchable encryption has been recognized by DARPA as one of the technical advances that is of significance for security of data for national and private information systems [31].

Another methodology for facilitating symmetric encryption with search abilities is by usage of a secure index [28]. Index can be defined as a data structure where document collections can be stored while supporting efficient keyword search i.e., when provided with a keyword, a pointer to the documents that contain it is given by the index. An index is deemed to be secure if only the data user possessing the trapdoor can perform the search operation for a keyword and also if only the trapdoor can be produced with help of a secret key. If not for the information about trapdoors, no breach of contents by index takes place.

A. Public-key searchable encryption

In the case of searching for data on cloud containing public-key-encrypted data, owner of the decryption key for the encrypted data can be a different person from the other data users who encrypt the data (and direct it to the server).Typically, for a general application a data-user publishes a public key while multiple users send e-mails to the mail server [12, 32]. Any data user with access to the

public key can add words to the index, but trapdoors can be generated only by the user who have access to the private key. Trapdoors are used to test for the occurrence of a keyword. The original work on public-key encryption with keyword searches (PEKS) have shown how to build a public-key encryption scheme that hides even the access pattern [12].This construction, however, has an overhead in search time that is proportional to the square root of the database size, which is far less efficient then the best private-key solutions[33].

B. Private-key searchable encryption

In the case of exploring a private-key-encrypted data, the data user encodes the data; so that it can be organized in a random manner way (before encryption) and additional data structures are included so as to provide an efficient access of required data. For allowing an user with the private key to access data, the data and data structures are encrypted and stored on the server. In this method, the rudimentary job of the user of pre-processing data is at least as large as the data, but the work following that of accessing the data is quite trivial in comparison to the extent of the data for both to the user and the server. Moreover, the entire information about the user's access pattern can be concealed.

Song et al., [11] have described different practical techniques for search on encrypted data. The Crypto Systems are secure for encrypted data and untrusted server cannot learn anything about the plain text based on the search results. Two techniques viz, Hidden queries and query isolation are introduced in this work. The hidden queries searches word without revealing the information to the server and query isolation server learns nothing except the search results. The algorithms are simple, fast, without space and communication overhead. Sequential scan is not efficient and is slow for a large number of documents.

Wang et al., [36] have proposed keyword search encryption technique to resolve the problem of encrypted data through query limitation. The suggested methodology combines the fine-grained access control and keyword search encryption to make available access controls of several users in the cloud setting characterized by encrypted data security. The downside of this scheme is that as the number of access categories of the search files increases, the number of query tokens escalates. The paradigm provides data fortification with high secure strength.

Liu et al., [37] have investigated an efficient privacy preserving keyword search scheme in cloud computing. The cloud server provider does not know any information about specified keywords and encrypted emails. It is able to protect user data and user queried keyword during search process. The construction is based on bilinear maps on elliptic curve to build an efficient Identity-Based Encryption (IBE)[38] and security is based on Bilinear Diffe-Hellman(BDH) assumption. The scheme is semantically secure but the experiment is not performed on the encrypted data

Jiang et al., [39] have presented a new approach to construct efficient Disjunctively Oblivious Keyword Search (DOKS) protocol which permits fast search and short cipher-text. It provides strong privacy on users side and cloud storage providers. The computation and storage space is less compared to previous Oblivious Keyword Search (OKS) protocols. The privacy and efficiency are better in DOKS protocol. The user submits two search keywords that are not distinguishable and need not know the relation between the cipher-text of the document and search keywords. The matching documents retrieve without revealing statistical information on the search query but the scheme does not support multi-keyword search.

V. PERFORMANCE AND EFFICIENCY ANALYSIS

In the practical scenes, the users prefer the scheme with quicker response and less bandwidth cost. Therefore, we considered the efficiency of the scheme when we designed the scheme. In our scheme, we achieved the constant-size ciphertext. Basing on the computation on the group of prime order p, we mainly evaluate the exponentiation operation, multiplication operation and pairing operation in time complexity and the size of group G and G_T in space complexity. It is worth mentioning that multiplication operation is much more efficient than the exponentiation. Compared to the CP-ABKS works in [14], we analyse the efficiency of the scheme from the two aspects of time complexity and space complexity. The results are showed in below tables Table1 and Table2. E denotes the exponentiation operation on the element in group G, E_T denotes the exponentiation operation on the element in group G_T . Similarly, M denotes the multiplication operation on the element in group G, M_T denotes the multiplication operation on the element in group G_T . And Pair is the symbol of the pairing operation. We use |G| and $|G_T|$ as the remarks of the size of G and G_T respectively. At last, we use N to represent the number of attributes which satisfy the access policy and S to represent the number of attributes which owned by the data-user. (In our scheme, S=N.)

	Our Scheme	[14]
Enc	4E+(N+2)M	(2N+4)E+M
KeyGen	(S+1)E+2SM+SM_T	(2S+2)E+SM
TokenGen	6E+M	(2S+4)E+M
Search	4Pair+E_T+3M_T	(2N+3)Pair+NE_T+(N+2)M_T

Table 1: Time Complexity Analysis

	Our Scheme	[14]
Enc	3|G|	(2N+3)|G|
KeyGen	2|G|+|G_T|	(2S+1)|G|
TokenGen	4|G|+|G_T|	(2S+3)|G|

Table 2: Space Complexity Analysis

Considering the balance of time complexity and space complexity, our scheme aggregated σ1, σ2, · · · , σn into σuser and y1, y2, · · · , yn into yuser in KeyGen phase. This behavior reduced the sizes of secret key and token, thus saving the bandwidth cost. There is a transformation of our scheme. We can transmit σ1, σ2, · · · , σn and y1, y2, · · · , yn as parts of the secret key. Then the cloud server should firstly aggregate the tokens σ s i , ys i in the Search phase. The computation works is transferred to the cloud server from the client. However, the communication cost between cloud server and the data user rise.

VI. CONCLUSION

We introduced an efficient searchable scheme basing on the ciphertext-policy attribute-based encryption. The EABSE scheme allows secure data share in multi-owners and multi-users system model. The data owners upload the data wrapped with a certain access policy, and others cannot get any information about the data whose attributes cannot satisfy the access policy. While the authenticated data users are able to search a keyword w on the ciphertexts and get the encrypted files including w. Besides, the search token is generated by data users. This action prevents the keyword that users queried from being known by owners and improves the users' privacy. It is mentioned that, our scheme raises the efficiency of computation and reduces the cost of communication because of the constant-size ciphertext. Our performance and efficiency analysis illustrates this point effectively. The new requirements also rise with the times, so the efficient search on the dynamic dataset is still for future work.

REFERENCES

[1] Curtmola, J. Garay, S. Kamara and R. Ostrovsky, Searchable symmetric encryption: improved denitions and efficient constructions. Proc, ACM CCS, 2006, pp. 79-88.

[2] D. Boneh, G.D. Crescenzo, R. Ostrovsky and G. Persiano, Public key encryption with keyword search. Proc, EUROCRYPT, May 2004,pp. 506-522.

[3] P. Golle, J. Staddon, B. Waters, Secure conjunctive keyword search over encrypted data. ACNS 2004. Lecture Notes in Com-puter Science, vol 3089. Springer, Berlin, Heidelberg.

[4] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, Searchable encryption revisited: consistecy properties, relation to anony-mous IBE and extentions. Springer, 2004.

[5] S. Gajek, Dynamic Symmetric Searchable Encryption from Constrained Functional Encryption. In: Sako K. (eds) Topics in Cryptology-CT-RSA 2016. Lecture Notes in Computer Science, vol 9610. Springer,Cham, pp. 75-89.

[6] S. Garg, P. Mohassel, C. Papamanthou, Efficient Oblivious RAM in Two Rounds with Applications to Searchable Encryption. In:Robshaw M., Katz J. (eds) Advances in Cryptology C CRYPTO 2016.CRYPTO 2016. Lecture Notes in Computer Science, vol 9816. Springer,Berlin, Heidelberg

[7] E. Goh, Secure Indexes. In: IACR Cryptology ePrint Archive, vol 2003

[8] YC. Chang, M. Mitzenmacher, Privacy Preserving Keyword Searches on Remote Encrypted Data. In: Ioannidis J., Keromytis A., Yung M. (eds) Applied Cryptography and Network Security. ACNS 2005. Lecture Notes in Computer Science, vol 3531. Springer, Berlin, Heidelberg.pp, 442-455.

[9] C. Rongmao, M. Yi, Y. Guomin, G. Fuchun, H. Xinyi, W. Xiaofen, W. Yongjun, "Server-Aided Public Key Encryption With Keyword Search", Information Forensics and Security IEEE Transactions on, ISSN 1556-6013.vol. 11, 2016, pp. 2833-2842.

[10] D.E. Knuth, The art of computer programming,volume 1:Fundamental algorithms,2nd edition.Addison-Wesley (1973)

[11] D. X. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," IEEE Symposium on Security and Privacy, pp. 44–55, 2000.

[12] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," in Proceedings of the Advances in CryptologyEurocrypt 2004, pp. 506–522, 2004.

[13] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," in Proceedings of the 13th ACM Conference on Computer and Communications Security, pp. 79–88, 2006.

[14] A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35–43, 2001.

[15] C.-I. Fan and S.-Y. Huang, "Controllable Privacy Pre serving Search Based on Symmetric Predicate Encryption in Cloud Storage," Future Generation Computer Systems, vol. 29, no. 7, pp. 1716–1724, 2013.