INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE

ISSN: 2057-5688

An Efficient Multi-User Searchable Encryption Scheme without Query Transformation over Outsourced Encrypted Data

YAMPATI HARIKA [1], Dr. SHAIK MOHAMMAD RAFI [2]

[1]Assistant professor, [2] Professor

CSE Department, Sri *Mittapalli College* of *Engineering*, *Guntur*, *Andhra Pradesh*-522233

Abstract— Searchable Encryption (SE) schemes provide security and privacy to the cloud data. The existing SE approaches enable multiple users to perform search operation by using various schemes like Broadcast Encryption (BE), Attribute-Based Encryption (ABE), etc. However, these schemes do not allow multiple users to perform the search operation over the encrypted data of multiple owners. Some SE schemes involve a Proxy Server (PS) that allows multiple users to perform the search operation. However, these approaches incur huge computational burden on PS due to the repeated encryption of the user queries for transformation purpose so as to ensure that users' query is searchable over the encrypted data of multiple owners. Hence, to eliminate this computational burden on PS, this system proposes a secure proxy server approach that performs the search operation without transforming the user queries. This approach also returns the top-k relevant documents to the user queries by using Euclidean distance similarity approach. Based on the experimental study, this approach is efficient with respect to search time and accuracy.

***Index Terms* — Cloud computing, privacy preserving, data encryption, keyword search, Searchable Encryption.**

I. INTRODUCTION

In recent years, many users have uploaded data to the cloud for easy storage and sharing with other users. At the same time, security and privacy concerns for the data are growing. Attribute-based encryption (ABE) enables both data security and access control by defining users with attributes so that only those users who have matching attributes can decrypt them. For real-world applications of ABE, revocation of users or their attributes is necessary so that revoked users can no longer decrypt the data. In actual implementations, ABE is used in hybrid with a symmetric encryption scheme such as the advanced encryption standard (AES) where data is encrypted with AES and the AES key is encrypted with ABE. The hybrid encryption scheme requires re-encryption of the data upon revocation to ensure that the revoked users can no longer decrypt that data. To re-encrypt the data, the data owner (DO) must download the data from the cloud, then decrypt, encrypt, and upload the data back to the cloud, resulting in both huge communication costs and computational burden on the DO depending on the size of the data to be re- encrypted. In this paper, we propose an attribute-based proxy re-encryption method in which data can be re-encrypted in the cloud without downloading any data by adopting both ABE and Syalim's encryption scheme. Our proposed scheme reduces the communication cost between the DO and cloud storage. Experimental results show that the proposed method reduces the communication cost by as much as one quarter compared to that of the trivial solution. Searchable Encryption (SE) schemes provide security and privacy to the cloud data by storing data in encrypted form while enabling search over encrypted data. SE schemes in [1], [2] support search operation efficiently only in a single owner and single user environment. Various schemes like BE [3], ABE [4] support search operations in a single owner and multi-user environment. However, these schemes do not allow to perform search operation over the data owned of multiple owners i.e. do not support search operation in a multi-owner and multi- user environment. Some SE schemes involve a third party entity like Proxy Server (PS) (it is also referred to as an Administrative or Intermediate Server) [5], whose job is to transform each individual data owner's index into a common index and also to transform each users query into a common query such that any user can search over any owner's data. The queries in this approach are required to be encrypted again and again for each data user whenever they issue the queries. This huge computational burden on PS due to the repeated encryption of queries makes it infeasible to adapt it in a real world environment. Hence, we propose a proxy server based SE approach to support search operation over the encrypted data of multiple owners without causing computational burden on third party entities like PS.

The first SE scheme [1] was proposed by using symmetric key encryption algorithm. SE by public key based approach

was proposed by using Identity-Based Encryption (IBE) [2]. These approaches support search operation in a single owner and a single user environment, which allows only a single user to perform the search operation over the data of single owner. BE scheme allows multiple users to perform the search operation over the encrypted data [6]. Another scheme supporting multiple users' search operation is proposed by using CP-ABE [7]. Keyword authorization based approach in

[8] supports search operation by multiple users. All these schemes support search operation in a single owner and multiuser environment, which allows the multiple users to perform the search operation over the encrypted data of a single owner. Multi-Keyword Ranked Search approach over the data of multiple owners is proposed [5]. This approach supports search operation in a multi-owner and multi-user environment, which allows multiple users to perform the search operation over the data of multiple owners. It incorporates Proxy Server (PS), which is responsible for transforming each owner's encrypted index into a common index and also each user's trapdoor into a common trapdoor such that any user can search over any owner's index. As the queries frequently undergo transformation each time the user issues them, this approach incurs huge computational burden on PS due to the repeated transformation of queries again and again.

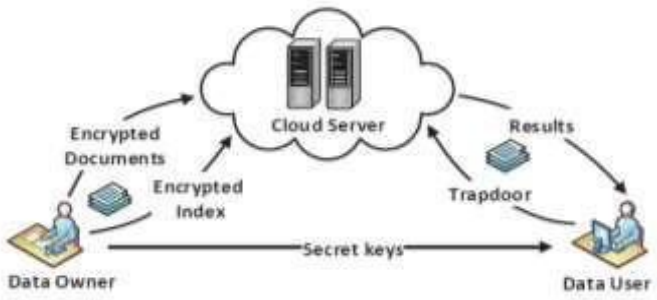


Fig.1 Proposed System Design

A. Existing System Summary

Revocation of users or their attributes is an indispensable feature of ABE for real-world applications. In real-world situations, users and their attributes change over time within the system. For example, users may be found to be malicious, may simply leave the system, or their attributes may change. Therefore, revoking users or their attributes accordingly so that they can no longer decrypt data is important. Existing revocation methods of ABE are proposed based on the notion of using ABE to encrypt the data entirely, whereas in actual implementations, hybrid encryption of ABE and symmetric encryption, specifically the advanced encryption standard (AES), are used for efficiency. In hybrid encryption, data is encrypted with AES and the AES key is encrypted with ABE. Because existing revocation methods affect only ABE ciphertext, this fact introduces a problem in which users can keep the AES key prior to revocation and use it to decrypt data even after the users are revoked. Therefore, although existing revocation methods can be applied to revoke users from ABE, re-encrypting data with a new AES key is necessary so that the old AES key can no longer be used.

II. RELATED WORKS

In the year of 2000, the authors "D. X. Song, D. Wagner, and A. Perrig", proposed a paper titled "Practical techniques for searches on encrypted data", in that they described such as: the proposed system is desirable to store data on data storage servers such as mail servers and file servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query, without loss of data confidentiality. We describe our cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems.

Our techniques have a number of crucial advantages. They are provably secure: they provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext when only given the ciphertext; they provide query isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot search for an arbitrary word without the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The algorithms presented are simple, fast (for a document of length n, the encryption and search algorithms only need O(n) stream cipher and block cipher operations), and introduce almost no space and communication overhead, and hence are practical to use today.

In the year of 2004, the authors "D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano", proposed a paper titled "Public key encryption with keyword search", in that they described such as: We study the problem of searching on data that is encrypted using a public key system. Consider user Bob who sends email to user Alice encrypted under Alice's public key. An email gateway wants to test whether the email contains the keyword "urgent" so that it could route the email accordingly. Alice, on the other hand does not wish to give the gateway the ability to decrypt all her messages. We define and construct a mechanism that enables Alice to provide a key to the gateway that enables the gateway to test whether the word "urgent" is a keyword in the email without learning anything else about the email. We refer to this mechanism as Public Key Encryption with keyword Search. As another example, consider a mail server that stores various messages publicly encrypted for Alice by others. Using our mechanism Alice can send the mail server a key that will enable the server to identify all messages containing some specific keyword, but learn nothing else. We define the concept of public key encryption with keyword search and give several constructions.

In the year of 2006, the authors "V. Goyal, O. Pandey, A. Sahai, and B. Waters", proposed a paper titled "Attribute- based encryption for fine-grained access control of encrypted data", in that they described such as: more sensitive data is shared and stored by third-party sites on the Internet, there will be a need to encrypt data stored at these sites. One drawback of encrypting data, is that it can be selectively shared only at a coarse-grained level (i.e., giving another party your private key). We develop a new cryptosystem for fine-grained sharing of encrypted data that we call Key-Policy Attribute-Based Encryption (KP-ABE). In our cryptosystem, ciphertexts are labeled with sets of attributes and private keys are associated with access structures that control which ciphertexts a user is able to decrypt. We demonstrate the applicability of our construction to sharing of audit-log information and broadcast encryption. Our construction supports delegation of private keys which subsumes Hierarchical Identity-Based Encryption (HIBE).

In the year of 2016, the authors "W. Zhang, Y. Lin, S. Xiao,
J. Wu, and S. Zhou", proposed a paper titled "Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing", in that they described such as: With the advent of cloud computing, it has become increasingly popular for data owners to outsource their data to public cloud servers while allowing data users to retrieve this data. For privacy concerns, secure searches over encrypted cloud data have motivated several research works under the single owner model. However, most cloud servers in

practice do not just serve one owner; instead, they support multiple owners to share the benefits brought by cloud computing. In this paper, we propose schemes to deal with privacy preserving ranked multi-keyword search in a multi-owner model (PRMSM).

To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, we systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files, we propose a novel additive order and privacy preserving function family. To prevent the attackers from eavesdropping secret keys and pretending to be legal data users submitting searches, we propose a novel dynamic secret key generation protocol and a new data user authentication protocol. Furthermore, PRMSM supports efficient data user revocation. Extensive experiments on real-world datasets confirm the efficacy and efficiency of PRMSM.

In the year of 2006, the authors "R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky", proposed a paper titled "Searchable symmetric encryption: improved definitions and efficient constructions", in that they described such as: Searchable symmetric encryption (SSE) allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it. This problem has been the focus of active research and several security definitions and constructions have been proposed. In this paper we begin by reviewing existing notions of security and propose new and stronger security definitions. We then present two constructions that we show secure under our new definitions. Interestingly, in addition to satisfying stronger security guarantees, our constructions are more efficient than all previous constructions. Further, prior work on SSE only considered the setting where only the owner of the data is capable of submitting search queries. We consider the natural extension where an arbitrary group of parties other than the owner can submit search queries. We formally define SSE in this multi-user setting, and present an efficient construction.

III. PROPOSED SYSTEM SUMMARY

The following steps are required to ensure that multiple users could search over the encrypted data of multiple owners without query transformation:

(a) Building Index: The data owners create an index for each of their documents as follows: Initially the stop words and non-alphabetic characters in documents are identified and ignored. Unique keywords in each document are listed and corresponding TF-IDF values are noted. The TF-IDF is a keyword scoring mechanism, which conveys the importance of the keyword in the entire data set. Hence, they are referred to as the relevance score information. The TF-IDF includes two attributes called Term Frequency and Inverse Document Frequency.

TF refers to the number of occurrences of term't' in a document 'd'. The IDF is obtained by dividing the total number of documents by the Document Frequency (DF). The DF is the number of documents that contain the term 't'. To avoid any direct possible inferences from the TFIDF values by the cloud server, each TF-IDF value is raised to the power of a random number. For example, suppose TF-IDF values of two keywords are 3, 2 and the random number is chosen as 3. The new TF-IDF values will become 27 and 9. The information as to which users are permitted to search an index is also included in the index by maintaining a list of all the user ids permitted to access the index.

(b) Index Encryption: At the Data owner side:

- For each keyword of the index, determine its length (n).
- If the length is even, randomly select n/2 number of positions within the keyword and encrypt the characters located at those positions using RSA algorithm with the private key of the data owner.
- For odd length keyword, encrypt n/2+1 characters randomly using the private key of owner.

- Once the above is done for every keyword in an index, then the partially encrypted index is sent to the proxy server.
- ***At the Proxy server side:***
  - The proxy server receives the partially encrypted index. It has its own secret key called as the common key is used to encrypt the remaining unencrypted characters of each keyword in the index. Thus it completes the index encryption fully.

(c) Search Operation: To retrieve the relevant documents, the data user issues his/her user id and a query, which is required to be encrypted. The data user randomly selects the positions of the characters within each keyword for encryption. The encryption of each keyword of the query follows the same procedure as explained in index encryption. The queries after encryption are termed as trapdoors. The trapdoors and the corresponding user ids along with the parameter 'k' are sent to the cloud server. The cloud server then checks if the user is authorized to search an index. If the user is authorized, trapdoors are matched with the keywords of the index and a match score is calculated, which is explained below.

(d) Matching Score Calculation: Every keyword in query is matched with each keyword in each index. This matching is done by making use of the Euclidean distance similarity approach. The least match score implies the highest match. The matching score is calculated as follows:

- Initially the lengths of keywords (index keyword and query keyword) that are to be matched are found.
- If the lengths are found to be different, current index keyword is ignored. When the lengths are same, sum of squares of differences in ASCII values of letters in the first keyword and corresponding letters in the second keyword is found. The square root of this value is called the match score.

Once the matching scores are calculated and assigned, most matching keyword in each document index for query keyword is noted by selecting the keyword having least score in that index. The TF-IDF value corresponding to this keyword is noted. Then, the cloud server determines the total score of each document by adding the TF-IDF values of each matching query keyword in that document. The scores of those documents are then sorted in descending order. Finally, the first 'k' documents among them are returned to the users.

Fig.4 Accept the User

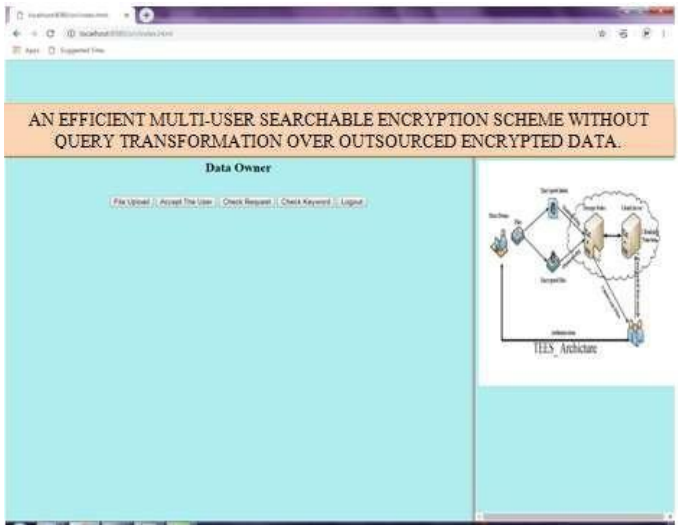
IV. RESULTS AND DISCUSSION

The following figure, Fig.2 illustrates the Data Owner Login page view of the proposed system.

Fig.2 Login Page

The following figure, Fig.3 illustrates the Data Owner Home page view of the proposed system.

Fig.3 Data Owner Home Page



The following figure, Fig.4 illustrates the User Acceptance Providing page view of the proposed system, in which the data owner can accept the user.

The following figure, Fig.5 illustrates the Check Request State view of the proposed system.

Fig.8 User Login

The following figure, Fig.9 illustrates the Data User Home Page view of the proposed system.

Fig.9 Data User Home Page

The following figure, Fig.10 illustrates the File Searching Option of the proposed system.

Fig.10 File Searching Option

The following figure, Fig.11 illustrates the Download View of the proposed system.

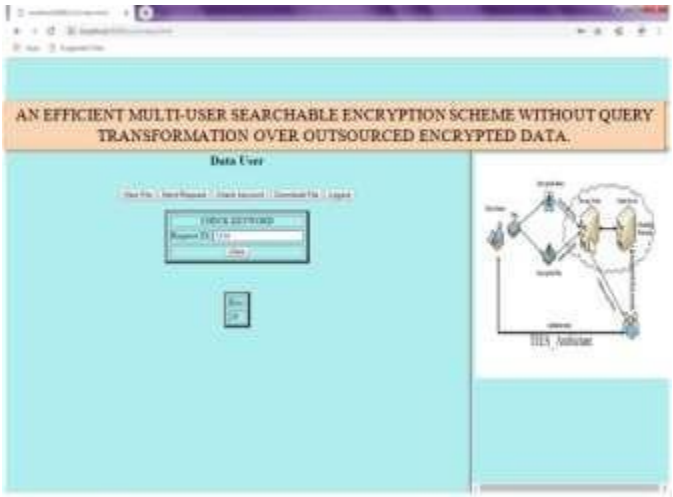


Fig.11 Download Page

V. Conclusion and Future Scope

A Proxy server based approach for supporting search operation over the data of multiple owners is proposed. Different from the existing approaches, the data user's query in this approach can be used to search over the multiple owners' data without transforming the query. In order to bypass the query transformation, the idea of partial encryption is used, i.e., half of each of the both index keyword and query keyword are encrypted by using the secret key of the data owner and the data user respectively and the other half of the index keyword and query keyword is encrypted by using common secret key of the proxy server. The experimental results confirm that the proposed approach is efficient. Future work could be to include a module for addition and revocation of data users and also to enhance the security functionalities of the proposed approach.

References

[1] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. S&P 2000. Proceedings. IEEE, 2000, pp. 44–55.

[2] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in International Conference on the Theory and Applications of Cryptographic Techniques. Springer, 2004, pp. 506–522.

[3] J. Lotspiech, "12 - broadcast encryption," in Multimedia Security Technologies for Digital Rights Management, W. Zeng, H. Yu, and C.-Y. Lin, Eds. Burlington: Academic Press, 2006, pp. 303 – 322.

[4] V. Goyal, O. Pandey, A. Sahai, and B. Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in Proceedings of the 13th ACM Conference on Computer and Communications Security, ser. CCS '06. New York, NY, USA: ACM, 2006, pp. 89–98.

[5] W. Zhang, Y. Lin, S. Xiao, J. Wu, and S. Zhou, "Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing," IEEE Transactions on Computers, vol. 65, no. 5, pp. 1566–1577, 2016.

[6] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," CCS-2006:ACM conference on Computers and Communications Security, pp. 79–88, 2006.

[7] Q. Wang, Y. Zhu, and X. Luo, "Multi-user searchable encryption with fine-grained access control without key sharing," in 2014 3rd International Conference on Advanced Computer Science Applications and Technologies, Dec 2014, pp. 145–150.

[8] Z. Deng, K. Li, K. Li, and J. Zhou, "A multi-user searchable encryption scheme with keyword authorization in a cloud storage," Future Generation Computer Systems, vol. 72, pp. 208–218, 2017.