NPV6I1011

International Journal Of Technical & Scientific Research -Vol.6, Issue .1

*ISSN Online: 2319-9245*

# AN EFFICIENT AND SECURE RANKED KEYWORD SEARCH OVER ENCRYPTED CLOUD DATA

[#1]ACHYUTH THANUKU , M.Tech Student,

[#2]Dr. RAVINDRA BABU KALLAM, HOD, Dept of CSE,

KAMALA INSTITUTE OF TECHNOLOGY AND SCIENCES, T.S., INDIA.

**ABSTRACT—** As cloud computing become more flexible & effective in terms of economy, data owners are motivated to outsource their complex data systems from local sites to commercial public cloud. But for security of data, sensitive data has to be encrypted before outsourcing, which overcomes method of traditional data utilization based on plaintext keyword search. Considering the large number of data users and documents in cloud, it is necessary for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Retrieving of all the files having queried keyword will not be affordable in pay as peruse cloud paradigm. In this paper, we propose the problem of Secured Multikeyword search (SMS) over encrypted cloud data (ECD), and construct a group of privacy policies for such a secure cloud data utilization system. From number of multi-keyword semantics, we select the highly efficient rule of coordinate matching, i.e., as many matches as possible, to identify the similarity between search query and data , and for further matching we use inner data correspondence to quantitatively formalize such principle for similarity measurement. We first propose a basic Secured multi keyword ranked search scheme using secure inner product computation, and then improve it to meet different privacy requirements. The Ranked result provides top k retrieval results. Also we propose an alert system which will generate alerts when un-authorized user tries to access the data from cloud, the alert will generate in the form of mail and message.

***Keywords—*** *Encryption, Inner product similarity, Multikeyword search, ranking.*

## I.INTRODUCTION

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. . Cloud Computing means a remote server that access through the internet which helps in business applications and functionality along with the usage of computer software. Cloud computing saves money that users spend on annual or monthly subscription. Due to advantage of cloud services, more and more sensitive information are being centralized into the cloud servers, such as emails, personal health records, private videos and photos, company finance data, government documents, etc. To protect data privacy, confidential data has to be encrypted before outsourcing, so as to provide end-to-end data confidentiality assurance in the cloud. Data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. This keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Unfortunately, data encryption, which restricts user's ability to perform keyword search and further demands the protection of keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data. Ranked search greatly improves system usability by normal matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency).

In the literature, searchable encryption [5]–[13] is a helpful technique that treats encrypted data as documents and allows a user to securely search through a single keyword and retrieve documents of interest. However, direct application of these approaches to the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot accommodate such high service-level requirements like system usability, user searching experience, and easy information discovery. Although some recent designs have been proposed to support Boolean keyword search [14]–[21] as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality (see section VI). Our early work [22] has been aware of this problem, and provided a solution to the secure ranked search over encrypted data problem but only for queries consisting of a single keyword. How to design an efficient encrypted data search mechanism that supports multi-keyword semantics without privacy breaches still remains a challenging open problem. In this paper, for the first time, we define and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. Among various multi key word semantics, we choose the efficient similarity measure of “coordinate matching”, i.e., as many matches as possible, to capture the relevance of data documents to the search query. Specifically, we use “inner product similarity” [4], i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query. During the index construction, each document is associated with a binary vector as a sub index where each bit represents whether corresponding keyword is contained in the document. The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured by the inner product of the query vector with the data vector. However, directly outsourcing the data vector or the query vector will violate the index privacy or the search privacy. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, we propose a basic idea for the MRSE using secure inner product computation, which is adapted from a secure k-nearest neighbor (kNN) technique [23], and then give two significantly improved MRSE schemes in a step-by step manner to achieve various stringent privacy requirements in two threat models with increased attack capabilities. Our contributions are summarized as follows,

1) For the first time, we explore the problem of multi keyword ranked search over encrypted cloud data, and establish a set of strict privacy requirements for such a secure cloud data utilization system.

2) We propose two MRSE schemes based on the similarity measure of “coordinate matching” while meeting different privacy requirements in two different threat models.

3) Thorough analysis investigating privacy and efficiency guarantees of the proposed schemes is given, and experiments on the real-world dataset further show the proposed schemes indeed introduce low overhead on computation and communication.

The remainder of this paper is organized as follows. In Section II, we introduce the system model, the threat model, our design goals, and the preliminary. Section III describes Fig. 1: Architecture of the search over encrypted cloud data the MRSE framework and privacy requirements, followed by section IV, which describes the proposed schemes. Section V presents simulation results. We discuss related work on both single and Boolean keyword searchable encryption in Section VI, and conclude the paper in Section VII.

## II. PROBLEM FORMULATION

### A. *System Model*

Considering a cloud data hosting service involving three different entities, as illustrated in Fig. 1: the data owner, the data user, and the cloud server. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable the searching capability over C for effective data utilization, the data owner, before outsourcing, will first build an encrypted searchable index I from F, and then outsource both the index I and the encrypted document collection C to the cloud server. To search the document collection for t given keywords, an authorized user acquires a corresponding trapdoor T through search control mechanisms, e.g., broadcast encryption [8]. Upon receiving T from a data user, the cloud server is responsible to search the index I and return the corresponding set of encrypted documents. To improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria (e.g., coordinate matching, as will be introduced shortly). Moreover, to reduce the communication cost, the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query. Finally, the access control mechanism [24] is employed to manage decryption capabilities given to users.

### B. *Threat Model*

The cloud server is considered as “honest-but-curious” in our model, which is consistent with related works on cloud security [24], [25]. Specifically, the cloud server acts in an “honest” fashion and correctly follows the designated protocol specification. However, it is “curious” to infer and analyze data (including index) in its storage and message flows received during the protocol so as to learn additional information. Based on what information the cloud server knows, we consider two threat models with different attack capabilities as follows.

***Known Cipher text***

Model In this model, the cloud server is supposed to only know encrypted dataset C and searchable index I, both of which are outsourced from the data owner.

***Known Background Model***

In this stronger model, the cloud server is supposed to possess more knowledge than what can be accessed in the known cipher text model. Such information may include the correlation relationship of given search requests (trapdoors), as well as the dataset related statistical information. As an instance of possible attacks in this case, the cloud server could use the known trapdoor information combined with document/keyword frequency [26] to deduce/identify certain keywords in the query.

### C. *Design Goals*

To enable ranked search for effective utilization of outsourced cloud data under the aforementioned model, our system design should simultaneously achieve security and performance guarantees as follows.

- Multi-keyword Ranked Search: To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning undifferentiated results.
- Privacy-Preserving: To prevent the cloud server from learning additional information from the dataset and the index, and to meet privacy requirements specified in section III-B.
- Efficiency: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

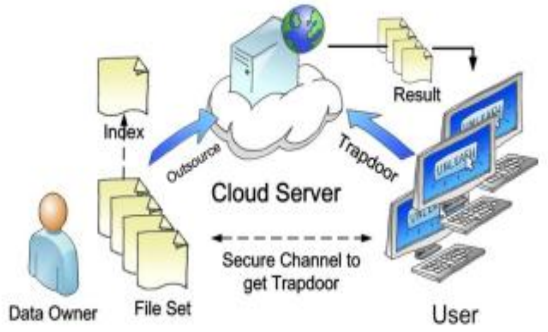

Fig. 1. System architecture of search over encrypted data in cloud computing

## III. LITERATURE SURVEY

***A. Secured Multi-keyword Ranked Search over Encrypted Cloud Data:***

In cloud computing data possessor are goaded to farm out their complex data management systems from local sites to the commercial public cloud for greater flexibility and economic savings. To ensure safety of stored data, it is must to encrypt the data before storing. It is necessary to invoke search with the encrypted data also. The specialty of cloud data storage should allow copious keywords in a solitary query and result the data documents in the relevance order. In [1], main aim is to find the solution of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. A variety of multi- keyword semantics are available, an efficient similarity measure of “coordinate matching” (as many matches as possible), to capture the data documents' relevancy to the search query is used. Specifically “inner product similarity”, i.e., the number of query keywords appearing in a document, to quantitatively evaluate such similarity measure of that document to the search query is used in MRSE algorithm. The main limitation of this paper was, the user’s identity (ID) is not kept hidden. Due to this, whoever puts the data on Cloud Service Provider was known. This may be risky in some situations where confidentiality of data need to be maintained. Hence, this drawback is overcome in the proposed system.

***B. Privacy Preserving Keyword Searches on Remote Encrypted Data:***

Consider the problem: a user U wants to store his files in an encrypted form on a remote file server S. Later the user U wants to efficiently retrieve some of the encrypted files containing specific keywords, keeping the keywords themselves secret and not to endanger the security of the remotely stored files. For example, a user may want to store old e-mail messages encrypted on a server managed by Yahoo or another large vendor, and later retrieve certain messages while travelling with a mobile device. In [2], solutions for this problem under well-defined security requirements are offered. The schemes are efficient as no public-key cryptosystem is involved. Indeed, the approach is independent of the encryption method chosen for the remote files. They are incremental too. In that, user U can submit new files which are secure against previous queries but still searchable against future queries. From this, the main theme taken is of storing data remotely on other server and retrieving that data from anywhere via mobile, laptop etc.

***C. Cryptographic Cloud Storage:***

When the benefits of using a public cloud infrastructure are clear, it introduces significant security and privacy risks. In fact, it seems that the biggest obstacle to the adoption of cloud storage (and cloud computing in general) is concern over the confidentiality and integrity of data. In [3], an overview of the benefits of a cryptographic storage service, for example, reducing the legal exposure of both customers and cloud providers, and achieving regulatory compliance is provided. Besides this, cloud services that could be built on top of a cryptographic storage service such as secure backups, archival, health record systems, secure data exchange and e-discovery is stated briefly.

***D. Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data:***

On one hand, users who do not necessarily have prior knowledge of the encrypted cloud data, have to post process every retrieved file in order to find ones most matching their interest; On the other hand, invariably retrieving all files containing the queried keyword further incurs unnecessary network traffic, which is absolutely undesirable in today’s pay-as-you-use cloud paradigm. This paper has defined and solved the problem of effective yet secure ranked keyword search over encrypted cloud data [4]. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency) thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. For the first time, the paper has defined and solved the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. The proposed ranking method proves to be efficient to return highly relevant documents corresponding to submitted search terms. The idea of proposed ranking method is used in our proposed system in order to enhance the security of data on Cloud Service Provider.

### *D. Providing Privacy Preserving in Cloud Computing:*

Privacy is an important issue for cloud computing, both in terms of legal compliance and user trust and needs to be considered at every phase of design. The [5] paper tells the importance of protecting individual's privacy in cloud computing and provides some privacy preserving technologies used in cloud computing services. Paper tells that it is very important to take privacy into account while designing cloud services, if these involve the collection, processing or sharing of personal data. From this paper, main theme taken is of preserving privacy of data. This paper only describes privacy of data but doesn't allow indexed search as well as doesn't hide user's identity. Thus, these two drawbacks are overcome in our proposed system.

### *E. Privacy Preserving Data Sharing With Anonymous ID Assignment:*

In this paper, an algorithm for anonymous sharing of private data among N parties is developed. This technique is used iteratively to assign these nodes ID numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the other members of the group. In [6], existing and new algorithms for assigning anonymous IDs are examined with respect to trade-offs between communication and computational requirements. These new algorithms are built on top of a secure sum data mining operation using Newton's identities and Sturm's theorem. The main idea taken from this paper is of assigning anonymous ID to the user on the cloud.

### *G. Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing:*

In this paper, main idea is to formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy [7]. This basic idea is taken but it is for multi-keyword raked search (MRSE scheme) in our proposed system. In [8], design of secure cloud storage service which addresses the reliability issue with near optimal overall performance is proposed.

### *H. Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing:*

Achieving finegrainedness, scalability, and data confidentiality of access control simultaneously is a problem which actually still remains unresolved. The paper [9] addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine-grained data access control to untrusted cloud servers without disclosing the underlying data contents. In [10], authors have proposed a privacy-preserving public auditing system for data storage security in Cloud Computing scheme is proposed. It utilizes the homomorphism linear authenticator and random masking to guarantee that the TPA would not learn any knowledge about the data content stored on the cloud server during the efficient auditing process, which eliminates the burden of cloud user from the tedious and possibly expensive auditing task, it also alleviates the user's fear of his/her outsourced data leakage.

## IV. PROPOSED SYSTEM

Considering a cloud data hosting service involving three different entities, the data owner, the data user along with his ID, and the cloud server. The data owner first registers on cloud using anonymity algorithm for cloud computing services. Before saving user registration information to database present on cloud anonymous algorithm process the data and then anonymous data is saved to registration database. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable searching capability over C for effective data utilization, the data owner, will first build an encrypted searchable index I from F before outsourcing , and then outsource both the index I and the encrypted document collection C to the cloud server. The work deals with efficient algorithms for assigning identifiers (IDs) to the users on the cloud in such a way that the IDs are anonymous using a distributed computation with no central authority. Given are N nodes, this assignment is essentially a permutation of the integers {1….N} with each ID being known only to the node to which it is assigned. Our main algorithm is based on a method for anonymously sharing simple data and results in methods for efficient sharing of complex data. To search the document collection for given keywords, an authorized user having an ID acquires a corresponding trapdoor T through search control mechanisms, for example, broadcast encryption. On receiving T from a data user, cloud server is responsible to search the index I and then returns the corresponding set of encrypted documents. In order to improve the document retrieval accuracy, the search result should be ranked by the cloud server according to some ranking criteria (e.g., coordinate matching) and assigning anonymous ID [6] to the user on cloud in order to make the data on cloud more secure. Moreover, to reduce the cost of communication the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query. At last, the access control mechanism is employed in order to manage decryption capabilities given to users and the data collection can be updated in terms of inserting new documents, updating existing ones, and deleting the existing documents.

## V. CONCLUSIONS

The previous work [1] mainly focused on providing privacy to the data on cloud in which using multi-keyword ranked search was provided over encrypted cloud data using efficient similarity measure of co-ordinate matching. The previous work [4] also proposed a basic idea of MRSE using secure inner product computation. There was a need to provide more real privacy which this paper presents. In this system, stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the user's data on cloud from the CSP and the third party user. Thus, by hiding the user's identity, the confidentiality of user's data is maintained.

## REFERENCES:

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, 2009.

[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS. Springer, Heidelberg.

[3] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001.

[4] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.

[5] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of S&P, 2000.

[6] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, 2003, http:// eprint.iacr.org/2003/216.

[7] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.

[8] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.

[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYPT, 2004.

[10] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficiently searchable encryption," in Proc. of CRYPTO, 2007.

[11] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, "Searchable encryption revisited: Consistency properties, relation to anonymous ibe, and extensions," J. Cryptol., vol. 21, no. 3, pp. 350–391, 2008.

[12] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.

[13] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. S. III, "Public key encryption that allows pir queries," in Proc. of CRYPTO, 2007.

[14] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Proc. of ACNS, 2004, pp. 31–45.

[15] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Proc. of ICICS, 2005.

[16] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Proc. of TCC, 2007, pp. 535–554.

[17] R. Brinkman, "Searching in encrypted data," in University of Twente, PhD thesis, 2007.

[18] Y. Hwang and P. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Pairing, 2007.

[19] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in Proc. of EUROCRYPT, 2008.

[20] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption," in Proc. of EUROCRYPT, 2010.