ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data

Gadekari Geetha Rani
P.G. Scholar (M. Tech),
Department of CSE,
Modugula Kalavathamma Institute of Technology for Women, Rajampet, Kadapa District.

P.Vijaya Raghavulu
Associate Professor,
Department of CSE,
Modugula Kalavathamma Institute of Technology for Women, Rajampet, Kadapa District.

Abstract:

With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system.

Among various multi-keyword 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. We further use "inner product similarity" to quantitatively evaluate such similarity measure. We first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

Index Terms:

Cloud computing, searchable encryption, privacy-preserving, keyword search, ranked search.

I.INTRODUCTION:

Now-a-days thousands of information is common everyday online. Daily new and additional information is outsourced due to growth in storage plus requirements of users, then essentially semi-trusted servers. Cloud computing is a Web-based model, where cloud clients can supply their information into the cloud[1]. By loading information into the cloud, the data owners stay unbound after the capacity of storage. Thus, to safeguard sensitive information integrity is an essential task. To safeguard information privacy in the cloud, the data owner has to be outsourced in the encoded system to the public cloud and the data operation is founded on plaintext keyword search. We select the efficient measure of "coordinate matching". Coordinate matching is used to measure the parallel amount. Coordinate matching captures the significance of data documents to the search query keywords. The search facility and privacy protective over encrypted cloud data are essential. If we study huge amount of data documents and data users in the cloud, it is hard for the necessities of performance, usability, plus scalability. Concerning to encounter the real data recovery, the huge amount of data documents in the cloud server achieve to outcome relevant rank instead of returning undistinguishable outcomes. Ranking scheme cares multiple keyword search to recover the search correctness.Today's Google network search devices, data users offer set of keywords instead of unique keyword search importance to retrieve the maximum significant data. Coordinate matching is a synchronize pairing of query keywords which are relevance to that document to the query. Due to inherence safety and privacy, it remains the interesting job on behalf of how to relate the encrypted cloud search.

The difficult of multi-keyword ranked search over encrypted cloud data is resolved by using stringent privacy necessities then numerous multi-keyword semantics. Among numerous multi-keyword ranked semantics, we choose coordinate matching. 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, an experiments on the real-world dataset further show the proposed schemes indeed introduce low overhead on computation and communication.

II.PROBLEM STATEMENT :

Actually large number of on-demand data users and huge amount of data documents in the cloud, this difficulty is challenging. It is essential for the search facility to permit multi keyword search query and make available result comparison ranking to see the effective data retrieval requirement. To develop the search result accuracy as well as to enrich the user searching experience, it is also essential for such ranking system to support multiple keywords search, as single keyword search regularly yields extreme coarse results. The searchable encryption method supports to give encrypted data as documents and agrees a user to firmly search over single keyword and retrieve documents of concern.

III. PROPOSED SOLUTION :

We propose an effective system where any authorized user can do a search on an encrypted data with multiple keywords, without revealing the keywords he searches for, nor the data of the documents that match by the query. Authorized users can make search processes by definite keywords on the cloud to retrieve the relevant documents. Our proposal system facilitates that a group of users can query the database provided that they possess so called trapdoors for the search terms that authorize the users to include them in their queries. Our proposed system is able to perform multiple keyword search in a single query and ranks the results so the user can retrieve only the most relevant matches in an ordered manner. And we establish a set of strict privacy requirements. Among numerous multi keyword semantics, we select the effective principle of "coordinate matching".

IV. SYSTEM OVERVIEW :

The system architecture is concerned by creating a simple structural framework for a system. It defines the overall frame of the project which briefly describes the functioning of the structure and the purpose of the project phase is to plan a solution of the problem identified by the necessity file. The below Figure 1 shows the outline of the structure. We consider three parts in our system architecture: Data Owner, Data user and Cloud Server.

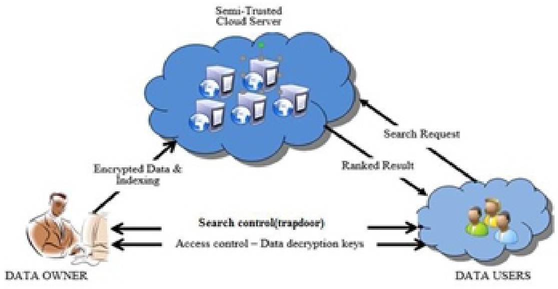

Fig. 1: Search over encrypted cloud data

Data Owner is responsible for the creation of the database.

- Data Users are the followers in a group who are able to use the files of the database.
- Cloud Server deals information facilities to certified users. It is necessary that server be insensible to content of the database it keeps.

Data owner has amount of data records that he wishes to outsource on cloud server in encrypted form. Before outsourcing, data owner will first construct a secure searchable index from a set of diverse keywords removed from the file collection and store both the index and the encrypted file on the cloud server. We undertake the approval between the data owner and users is done. To search the file collection for a given keyword, certified user creates and submits a search request in a secret form-a trapdoor of the keyword to the cloud server. Upon getting the search request, the server is in charge to search the index and return the matching set of files to the user. We study the secure ranked keyword search problematic as follows: the search result must be returned giving to definite ranked relevance principles, to develop file retrieval correctness for users. Though, cloud server must study unknown or little about the important principles themselves as they reveal major sensitive information against keyword privacy. To decrease bandwidth, the user may send possible value k along with the trapdoor and cloud server only sends back the top-k most appropriate files to the user's concerned keyword.

Design Goals:

To allow ranked search for operative use of outsourced cloud data under the aforesaid model, our system design should instantaneously achieve security and performance assurances 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. • Efficiency: Above goals on functionality and privacy should be achieved with low communication and computation over head.

Coordinate Matching:

"Coordinate matching" [2] is an intermediate similarity measure which uses the number of query keywords appearing in the document to quantify the relevance of that document to the query. When users identify the exact subset of the dataset to be regained, Boolean queries achieve well with the exact search necessity stated by the user. It is more elastic for users to identify a list of keywords indicating their concern and regain the most relevant documents with a rank order..

5. Privacy Requirements for MRSE In the related literature, such as searchable encryption is that the server should study nothing but search results. With this general privacy picture, we discover and create a set of strict privacy necessities specially for the MRSE framework.

Data privacy, the data owner can resort to the traditional symmetric key cryptography to encrypt the data before outsourcing, and effectively prevent the cloud server into the outsourced data.

Index privacy, if the cloud server infers any association between keywords and encrypted documents from index. Therefore, the searchable index should be built to prevent the cloud server from acting such kind of association attack.

Keyword Privacy, as users generally wish to have their search from existence showing to others like the cloud server, the most vital concern is to hide what they are searching, i.e., the keywords specified by the corresponding trapdoor. The trapdoor can be generated in a cryptographic way to protect the query keywords.

Trapdoor, the trapdoor generation function should be a randomized one instead of being deterministic. The cloud server should not be able to deduce the connection of any given trapdoors, i.e, to determine whether the two trapdoors are formed by the same search request. Otherwise, the deterministic trapdoor generation would give the cloud server benefit to collect frequencies of dissimilar search requests concerning different keyword(s), which may further disturb the aforesaid keyword privacy requirement. . Access Pattern, within the ranked search, the access pattern is the sequence of search results where every search result is a set of documents with rank order.

V.Multi-Keyword Ranked Search over Encrypted (MRSE):

CLOUD computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, for example, emails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud; this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of bandwidth cost in cloud scale systems. Moreover, aside from eliminating the local storage management, storing data into the cloud serves no purpose unless they can be easily searched and utilized. Thus, exploring privacy preserving and effective search service over encrypted cloud data is of paramount importance. Considering the potentially large number of on-demand data users and huge amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability. On the one hand, to meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance ranking, instead of returningundifferentiated results. Such ranked search system enables data users to find the most relevant information quickly, rather than burdensomely sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information.

On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results. As a common practice indicated by today's web search engines (e.g., Google search), data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data. And each keyword in the search request is able to help narrow down the search result further. "Coordinate matching", as many matches as possible, is an efficient similarity measure among such multi-keyword semantics to refine the result relevance, and has been widely used in the plaintext information retrieval (IR) community.However, how to apply it in the encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements like the data privacy, the index privacy, the keyword privacy, and many others. Encryption 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 as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality. Our early works have been aware of this problem, and provide solutions 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 the project, for the first time, 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-keyword semantics, 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, inner product similarity 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, 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. We investigate some further enhancements of our ranked search mechanism to support more search semantics and dynamic data operations.

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

VI.RESULT AND CONCLUSION:

Figure 2

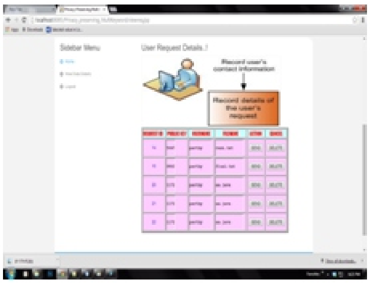
Sidebar Menu
User Request Details..!
Record user's contact information
Record details of the user's request

Figure 3

CONCLUSION:

In this paper, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we propose a basic idea of MRSE using secure inner product computation. Then, we give two improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. We also investigate some further enhancements of our ranked search mechanism, including supporting more search semantics, i.e., TF _ IDF, and dynamic data operations. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world data set show our proposed schemes introduce low overhead on both computation and communication.

In our future work, we will explore checking the integrity of the rank order in the search result assuming the cloud server is untrusted.

REFERENCES:

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, pp. 829-837, Apr, 2011.

[2] 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.

[3] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.

[4] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc. 14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.

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

[6] I.H. Witten, A. Moffat, and T.C. Bell, Managing Gigabytes: Compressing and Indexing Documents and Images. Morgan Kaufmann Publishing, May 1999.

[7] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

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

[9] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.

[10] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.

[11] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.

[12] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and Efficiently Searchable Encryption," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.