Privacy-Preserving Multi-keyword Ranked Search Over Encrypted Cloud Data

Jyothi Koodi[1], G. Srinivasachar[2]

[1] M.tech , Dept. of CSE, Atria Institute of Technology , Bengaluru, Karnataka, India
[2] Assistant Professor Dept. of CSE, Atria Institute of Technology , Bengaluru, Karnataka, India

---***---

Abstract: The innovation in cloud computing has encouraged the data owners to outsource their data managing system from local sites to profitable public cloud for excessive flexibility and profitable savings. But people can like full benefit of cloud computing, if we are able to report very real secrecy and security concerns that come with loading sensitive personal information. Allowing an encrypted cloud data search facility is of great significance. In view of the huge number of data users, documents in the cloud, it is important for the search facility to agree multi keywords query and arrange for result comparison ranking to meet the actual need of data recovery search and not regularly distinguish the search results. Related mechanisms on searchable encryption emphasis on single keyword search or Boolean keyword search, and often sort the search outcomes.

In this system, we explain and solve the interesting problem of privacy preserving multi keywords ranked search over encrypted cloud data, and create a set of strict privacy necessities for such a safe cloud data application system to be effected in real. We first offer a basic idea for the multi keyword ranked search over encrypted cloud data (MRSE) based on effective comparison measure of coordinate matching, i.e. as many matches as possible, in order to capture the significance of data documents to the search query. Then we give two considerably developed multi keywords ranked search encryption schemes to reach many tough privacy requirements in two differ threat models.

Key Words: Multi-Keyword search, Coordinate Matching, Keywords, Index Generation, Trapdoor

1. 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.

2. 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.

3. 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".

4. 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.

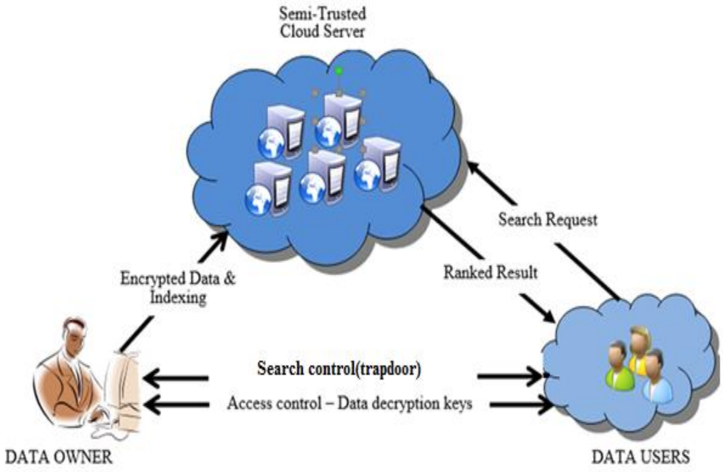


Fig. 1: Search over encrypted cloud data

- 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.

6. Modules

Our proposed system consists of the following modules:

- Data User Module
- Data Owner Module
- File Upload Module with Encryption
- File Download Module with Decryption
- Rank Search Module

Data User Module

Data users are users on this system, who will be able to download files from the cloud that are uploaded by the data owners. Since the files stored on the cloud server could be in huge numbers, there is a search facility provided to the user. The user should be able to do a multi-keyword search on the cloud server. Once, the result appears for the specific search, these users should be able to send a request to the respective data owners of the file through the system (also called trap-door request) for downloading these files. The data users will also be provided a request approval screen, where it will notify if the data owner has accepted or rejected the request. If the request has been approved, the users should be able to download the decrypted file.

Data Owner Module

In this module, the data owners should be able to upload the files. The files are encrypted before the files are uploaded to the cloud. The data owners are provided an option to enter the keywords for the file that are uploaded to the server. These keywords are used for the indexing purpose which helps the search return values very quickly. These files when once available on the cloud, the data users should be able search using the keywords. The data owners will also be provided with a request approval screen so they are able to approve or reject the request that are received by the data users.

File Upload & Encryption Module

In this module, the data owners should be able to upload the files. The files are encrypted before the files are uploaded to the cloud. The data owners are provided an option to enter the keywords for the file that are uploaded to the server. These keywords are used for the indexing purpose which helps the search return values very quickly. These files when once available on the cloud, the data users should be able to search using keywords. The data owners will also be provided with a request approval screen so that they are able to approve or reject the request that are received by the data users.

The file before upload will have to be encrypted with a key so that the data users cannot just download it without this key. This key will be requested by the data users through the trap-door. The encryption of these files uses RSA algorithm so that unauthorized users will not be able to download these files.

File Download & Decryption Module

Data users are users on this system, who will be able to download files from the cloud that are uploaded by the data owners. Since the files stored on the cloud server could be in huge numbers, there is a search facility provided to the user. The user should be able to do a multi-keyword search on the cloud server. Once, the result appears for the specific search, the users should be able to send a request to the respective data owners of the file through the system (also called trap-door request) for downloading these files. The data users will also be provided a request approval screen, where it will notify if the data owner has accepted or rejected the request. If the request has been approved, the users should be able to download the decrypted file.

The file before download will have to be decrypted with a key. This key will be requested by the data users through the trap-door request. Once the key is provided during the download, the data users will be able to download the file and use them.

Rank-Search Module

This module allows the data users to search the files with multi-keyword rank searching. This model uses the frequently used rank searching algorithm for present the output for multi-keywords. "Coordinate Matching" principle will be adopted for the multi-keyword searching. This module also takes care of creating an index for faster search.

7. RESULTS

Multiple users are created at a centralized location for the data owners and data users. We can see that either of the users can access the system once they login. The exchange of communication between data owners and data users is strictly through E-mail system which enables the system to be secured. Since the contents are encrypted and kept in the cloud, public viewing of these files is impossible. The files or contents can be viewed only after the consent of the data owners, after getting the secret key.

8. RELATED WORK

Single keyword searchable encryption schemes [3]–[11], [18] usually build an encrypted searchable index such that its content is hidden to the server unless it is given appropriate trapdoors generated via secret key(s) [2]. It is first studied by Song et al. [3] in the symmetric key setting, and improvements and advanced security definitions are given in Goh [4], Chang et al. [5] and Curtmola et al. [6].Our early work [18] solves secure ranked keyword search which utilizes keyword frequency to rank results instead of returning undifferentiated results. However, it only supports single keyword search.To enrich search functionalities, conjunctive keyword search [12]–[15] over encrypted data have been proposed.These schemes incur large overhead caused by their fundamental primitives, such as computation cost by bilinear map, e.g. [13], or communication cost by secret sharing,e.g. [12]. As a more general search approach, predicate encryption schemes are recently proposed to support both conjunctive and disjunctive search. Boolean keyword searchable encryption schemes support multiple keywords ranked search over encrypted cloud data while preserving privacy as we propose to explore in this paper.

9. CONCLUSION

In this work, firstly we describe and resolve the difficult of multi-keyword ranked search over encrypted cloud data, and create a variety of privacy requirements. Between numerous multi-keyword semantics, we select the effective similarity measure of "coordinate matching", i.e., as various matches as likely, to effectively capture the relevance of outsourced documents to the query communication .In our future work, we will search supporting other multi keyword semantics over encrypted data and checking the integrity of the rank order in the search result keywords. For convention the challenge of supportive multi-keyword semantic without privacy breaks, we propose a basic idea of MRSE. Then we give two better MRSE outlines to realise many stringent privacy requirements in two dissimilar threat models. Detailed analysis studying privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world data set show our future systems introduce low overhead on both computation and communication.

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

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

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

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

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

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

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

[9] 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.

[10] 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.

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

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

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

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

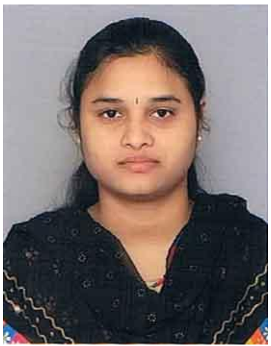
[15] R. Brinkman, "Searching in encrypted data," in University of Twente, PhD thesis, 2007.

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

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

[18] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS'10, 2010.

BIOGRAPHIES



Miss. Jyothi Koodi has completed B.E in Computer Science at KCT Gulbarga, India in 2011. She is currently pursuing M.Tech in Computer Science from Visvesvaraya Technological University, India. She is interested with areas of research related to Cloud computing.

G. Srinivasachar, M. S. (Comp. Science) from IIT Madras. He is currently working with Atria Institute of Technology, Bangalore as Asst Professor. He is interested in research in cloud computing.