International Journal On Engineering Technology and Sciences – IJETS™
ISSN(P): 2349-3968, ISSN (O): 2349-3976
Volume IV, Issue II, February - 2017

MULTI-KEYWORD RANKED SEARCH IN CLOUD COMPUTING WITHOUT BURDEN OF NETWORK BANDWIDTH

Mrs. Janani
Department of Computer Science and Engineering
PPG Institute Of Technology
Coimbatore, Tamil Nadu, India
rjanani14@gmail.com

Mr.C.Nandhakumar
Senior Assistant Professor, Department of Computer Science and Engineering
PPG Institute Of Technology, Coimbatore, Tamil Nadu, India
ndk.apcse@gmail.com

ABSTRACT

Cloud computing a promising pattern for data outsourcing and high-quality data services. However, concerns of secure information on cloud potentially cause security problems. Data encryption protects data security to some extent, but at the cost of compromised efficiency. In this work, we develop the searchable encryption for multi-keyword ranked search over the cloud data. Today's mail servers such as IMAP servers, file servers and other data storage servers typically must be fully authorized they have access to the cloud data, and hence must be trusted not to reveal it without authentication which introduces undesirable security and privacy risks in applications. Previous work shows how to build encrypted file systems and secure servers, but typically one must sacrifice functionality to ensure security. By considering the huge amount of cloud data that are outsourced, we are using the relevance score and *k*-nearest neighbor techniques to develop an efficient multi-keyword search scheme that can return the ranked search results based on the accuracy. We leverage an efficient index to further to maximize the search efficiency, and adopt the blind storage system to conceal access pattern of the user. Security analysis shows that our method can achieve confidentiality of data's and index, privacy, unlinkability, and concealing access pattern of the search user. Finally we show that our work can reach much improved efficiency in terms of search functionality and search time compared with the existing methods.

Keywords

Cloud Computing, Ranking, Ranked Search, Privacy Preserving.

1. INTRODUCTION

Now a day's Cloud Computing becomes more and more sensitive information are being centralized into the cloud, such as emails, personal health records, company finance data, and government documents, etc. The fact that data owners and cloud server are no longer in the same secured domain may put the outsourced unencrypted data at risk the cloud server may leak data information to unauthorized users or even be hacked. It follows that sensitive data has to be encrypted prior to outsourcing for data privacy and combating unsolicited accesses. Data

encryption makes effective data utilization a very challenging task given that there could be a huge amount of outsourced data files. Cloud Computing, data owners may share their outsourced data with a large amount 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. Such keyword search technique allows data users to retrieve files of interest and has been widely applied in plaintext search scenarios.

Data encryption, which demands user's ability to perform keyword search and further restricts the protection of keyword privacy, makes the traditional normal text search methods fail for encrypted cloud data. Traditional methods allows searchable encryption schemes to a cloud client to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean search, without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud mechanism, they may suffer from the following two main drawbacks. Firstly for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file, which demands possibly large amount of post processing over- head; Another one is invariably sending back all files solely based on presence/absence of the keyword further incurs huge unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm. Lacking of effective mechanisms to ensure the file retrieval accuracy is a significant drawback of proposed searchable encryption schemes in the context of Cloud Computing. The state of the art in information retrieval (IR) community has already been utilizing various scoring methods to quantify and rank-order the relevance of files in response to any given search query. Although the importance of ranked search has received attention for a long history in the context of plaintext searching by IR community, surprisingly, it is still being overlooked and remains to be studied in the context of encrypted cloud data search.

Background on Searchable Symmetric Encryption

Searchable encryption allows data owner to outsource their data in an encrypted manner while maintaining the selectively-search capability over the encrypted cloud data. Searchable encryption can be achieved in its full functionality using an oblivious RAMs. Although encrypting everything during the search from a malicious server, utilizing oblivious RAM usually brings the cost of logarithmic amount of interactions between the user and the cloud server for each search request. In order to achieve more efficient solutions, almost all the existing works on

searchable encryption literature resort to the weakened security guarantee, i.e., revealing the access pattern and search pattern but nothing else. Here access pattern refers to the outcome of the search result, i.e., which files have been retrieved. The search pattern includes the equality pattern among the two search requests, and any information derived thereafter from this statement. Having a correct intuition on the security guarantee of existing SSE literature is very important for us to define our ranked searchable symmetric encryption problem.

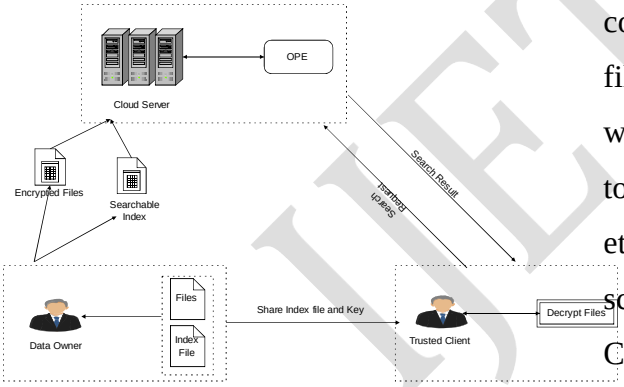

Figure 1 Architecutre diagram

In later we will demonstrate that following the exactly same security guarantee of existing SSE scheme, it would be very inefficient to achieve ranked keyword search, which motivates us to further weaken the security guarantee of previous SSE appropriately and realize an "as-strong-as-possible" ranked searchable symmetric encryption. This notion has been employed by cryptographers in much recent work where efficiency is preferred over security.

2 RELATED WORKS

Searchable Encryption: Existing searchable encryption has been widely discussed as a cryptographic primitive, with a focus on security definition formalizations and efficiency improvements. Song et al. first introduced the notion of searchable encryption. They proposed a scheme in the symmetric key setting, where each word in the file is encrypted independently under a special two-layered encryption construction. A searching overhead is linear to the whole file collection length. Goh developed a Bloom filter based per-file index, minimizing the work load for each search request proportional to the number of files in the collection. Chang et al. also developed a similar per-file index scheme. To further enhance search efficiency, Curtmola et al. proposed a per-keyword based approach, where a single encrypted hash table index is built for the entire file collection, with every entry consisting of the trapdoor of a keyword and an encrypted set of related file identifiers. Searchable encryption has also been considered in the public-key setting. Boneh et al. studied the first public-key based searchable encryption scheme, with an analogous scenario to that of. In their work, anyone with the public key can write to the data stored on the cloud server but only authorized users with the private key can

search. As an attempt to enrich query predicates, conjunctive keyword search over encrypted data have also been proposed.

Our work on secure ranked search over encrypted cloud data, very recently, Cao et al. propose a privacy-preserving multi-keyword ranked search scheme, which extends our existing work in with support of multi-keyword query. They choose the principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between a multi keyword search query and data documents, and later quantitatively formalize the principle by a secure inner product computation mechanism. One drawback of the scheme is that cloud server has to linearly traverse the whole index of all the documents for each search request, while ours is as efficient as existing SSE schemes with only constant search cost on cloud server.

Secure top-k **retrieval from Database Community:** from database community is the most related work to our proposed RSSE. The idea of uniformly distributing posting elements using an order-preserving cryptographic function was first discussed. The order-preserving mapping function proposed doesn't support score dynamics, i.e., any insertion and updates of the scores in the index will result in the posting list completely rebuilt. Uses a different order-preserving mapping based on pre-sampling and training of the relevance scores to be outsourced, which is not as efficient as our proposed schemes. When scores following different distributions need to be inserted, their score transformation function still needs to be rebuilt. In our scheme the score dynamics can be gracefully handled, which is an important benefit inherited from the original OPSE. This can be observed from where the same score will always be mapped to the same random-sized non-overlapping bucket, given the same encryption key. We note that supporting score dynamics, which can save quite a lot of computation overhead when file collection changes, is a significant advantage in our method. Both methods above do not exhibit thorough security analysis which we do in the work.

Other Related Methods: Allowing range queries over encrypted data in the public key settings has been studied where advanced privacy preserving schemes were proposed to allow more sophisticated multi-attribute search over encrypted files while preserving the attributes' privacy. Though these two methods provide provably strong security, they are generally not efficient in our settings, as for a single search request, a full scan and expensive computation over the whole encrypted scores corresponding to the keyword posting list are required. The two methods do not support the ordered result listing on the server side. They cannot be effectively utilized in our method since the

user still doesn't know which retrieved files would be the most relevant.

Difference from Conference Version: We have revised the article a lot and improved many technical details as compared. The mechanism design incurs negligible overhead on data users, and further enhances the quality of data search service. We extend our existing result on order-preserving one to - many mapping and demonstrated that this mapping process is indeed reversible, which can be very useful in many practical applications. We provide the corresponding algorithm for the reverse mapping and also include its performance results. The related work has been substantially improved, which now faithfully reflects much recent advancement on privacy-preserving search over encrypted data.

3 CHALLENGES AND CONTRIBUTIONS

To design a secure and well functioning search scheme over encrypted data, one has to make three important design choices that are closely inter-related and largely determine the performance of the resulting search scheme, 1). Data structure used to build secure indexes and trapdoors; 2). Effective search algorithm that can quantify the level of match between keywords in the query and keywords in the index with high efficiency, and 3). Security and privacy mechanisms that can be integrated in the above two design choices thus the index privacy and search privacy can be protected.

We reveal a relevance score in searchable encryption to achieve multi-keyword ranked search over the encrypted mobile cloud data. We construct an efficient index to improve the efficiency of search.

_ By modifying the blind storage system in the EMRS, we solve the trapdoor unlinkability problem and conceal access pattern of the search user from the cloud server.

_ We give thorough security analysis to show that the EMRS can get a high security level including confidentiality of documents and index, privacy, unlinkability, and concealing access pattern of the cloud search user. We implement extensive experiments, which show that the EMRS can achieve enhanced efficiency in the terms of functionality and search efficiency compared with existing methods.

4. METHODOLOGY OVERVIEW

Authenticating Ranked Search Result

Cloud servers sometimes behave beyond the semi-honest model. This will happen either because cloud server intentionally wants to do so for saving cost when handling large number of search requests, or there may be software bugs, or

internal or external attacks. Enabling a search result authentication mechanism that can find such unwanted behaviors of cloud server is also of practical interest and worth further investigation. To authenticate a ranked search result the relevance sequence among the results are not disrupted. To reach these goal we propose to utilize the one way hash chain technique, which can be added directly on top of the previous RSSE design.

Threat Model

The secure search schemes adopts to consider the cloud server to be "honest-but-curious", that is the cloud server "honestly" follows the designated protocol specification, but it is "curious" to infer and analyze data in its storage and message flows received during the protocol in order to learn additional information.

Search Privacy

In the literature, many privacy requirements are defined for PKC-based and SKCbased search schemes. We briefly introduce these *search privacy* requirements as follows.

1. **Keyword Privacy:** One of the major privacy concerns is how to protect the keywords of interest in a user's trapdoor against the cloud server. Cloud server is not able to infer what the data user is searching. This fundamental privacy requirement should be satisfied for any valid encrypted data search scheme. Although trapdoor generation can be performed in a cryptographic way to protect the query keywords, the cloud server could identify the searched keywords by other side channel attacks, such as frequency analysis attack. Given the keyword-specific document frequency information or the keyword frequency distribution information in a particular dataset, it is sufficient for an attacker to reverse-engineer the keyword in a trapdoor. Notice that this privacy requirement is referred to as *predicate privacy* in the PKC-based search scenario and it cannot be protected inherently for any asymmetric secure search scheme.

2. **Trapdoor Unlinkability:** It is required that the trapdoor should be generated in a random manner. Otherwise, given any two trapdoors, the attacker can easily determine the relationship of them, such as whether they contain the same set of keywords. So that sufficient non determinacy should be introduced into the trapdoor generation algorithm. It is worth noting that violation of this privacy requirement can further compromise the keyword privacy in that it allows the cloud server to accumulate frequencies of

different search requests with respect to different keyword(s).

3. **Access Pattern:** It is defined to be the sequence of returned documents. Note that protecting access pattern by using private information retrieval technique is extremely expensive since the algorithm has to "touch" the whole dataset outsourced on the cloud server which is inefficient in the large scale cloud system. Thus for efficiency concerns, most of the search over encrypted data schemes do not aim to protect it.

4.2 Our Contributions

- Introduce a relevance score in searchable encryption to achieve multi-keyword ranked search over the encrypted mobile cloud data. In addition to that, construct an efficient index to improve the search efficiency.
- The blind storage system in the EMRS, to solve the trapdoor unlinkability problem and conceal access pattern of the search user from the cloud server.
- To give thorough security analysis to demonstrate that the EMRS can reach a high security level including confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability, and concealing access pattern of the search user. Which show that the EMRS can achieve enhanced efficiency in the terms of functionality and search efficiency compared with existing methods.

- We consider the cloud server to be curious but honest which means it performs the task assigned by the data owner and the search user correctly.

CONCLUSION AND FEATURE WORKS

We have proposed a multi-keyword ranked search method to enable accurate, efficient and secure search over encrypted mobile cloud data. Security analysis have showed that proposed method can effectively achieve confidentiality of documents and index, privacy, unlinkability, and concealing access pattern of the search user. Extensive performance evaluations have reveals that the proposed method can achieve better efficiency in terms of the functionality and computation overhead compared with existing ones. We will investigate on the authentication and access control issues in searchable encryption technique in future.

And also "whole" checking, rank checking greatly reduces the workload of audit services, while still achieves an effective detection of misbehaviors'. A probabilistic audit on rank checking is preferable to realize the anomaly detection in a timely manner and finally to check the result correct or not through TPA analysis. This auditing process will improve

efficient when a auditing the rank and perfectly to measure untrusted cloud returned result over high performance way.

REFERENCES

[1] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, ``An SMDP-based service model for interdomain resource allocation in mobile cloud networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 5, pp. 2222_2232, Jun. 2012.

[2] M. M. E. A. Mahmoud and X. Shen, ``A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 10, pp. 1805_1818, Oct. 2012.

[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Y. Luo, ``Exploiting geodistributed clouds for a e-health monitoring system with minimum service delay and privacy preservation," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 2, pp. 430_439, Mar. 2014.

[4] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, ``A survey of mobile cloud computing: Architecture, applications, and approaches," *Wireless Commun. Mobile Comput.*, vol. 13, no. 18, pp. 1587_1611, Dec. 2013.

[5] H. Li, Y. Dai, L. Tian, and H. Yang, ``Identity-based authentication for cloud computing," in *Cloud Computing*. Berlin, Germany: Springer-Verlag, 2009, pp. 157_166.

[6] W. Sun, *et al.*, ``Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in *Proc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur.*, 2013, pp. 71_82.

[7] B.Wang, S.Yu,W. Lou, andY. T. Hou, ``Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 2112_2120.

[8] E. Stefanov, C. Papamanthou, and E. Shi, ``Practical dynamic searchable encryption with small leakage," in *Proc. NDSS*, Feb. 2014.

[9] Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, ``Secure dynamic searchable symmetric encryption with constant document update cost," in *Proc. GLOBECOM*, Anaheim, CA, USA, 2014.

[10] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Ro³u, and M. Steiner, ``Highly-scalable searchable symmetric encryption with support for Boolean queries," in *Proc. CRYPTO*, 2013, pp. 353_373.