ISSN:0975-9646

A Secured and High Octane Rank Based Analysis in Cloud Computing Environment

Poojitha Koneru, Dr. S.Prabakaran

Dept of CSE, SRM University, Chennai

***Abstract*— In commercial public cloud the cloud computing alters the model of data service outsourcing. In this project, we define and solve the problem of ranked keyword search throughout cloud data. Ranked search greatly raises system usability by altering search result relevance ranking instead of sending undifferentiated results and further ensures the file retrieval accuracy. We explore the statistical measure approach i.e. relevance score, from information retrieval with focus on ranking and several components of an information retrieval system to build a searchable index.**

***Index Terms*— Ranked search, searchable encryption, order-preserving mapping, confidential data, cloud computing.**

1 INTRODUCTION

As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud, such as emails, personal health records, government documents, etc. By storing their data into the cloud, the information owners can be relieved from the burden of data storage and maintenance so as to enjoy the on-demand high quality data storage service. Cloud Computing is the dreamed vision of computing as a utility, where cloud customers can store their information into the cloud so that we achieve on-demand high quality applications and services from a shared pool of computing resources [1]. The benefits are not limited : relief of the burden for storage management, information access with independent geographical locations, and avoidance of investment on hardware, software, and personnel maintenances, etc [2]. As Cloud Computing becomes widespread, more sensitive information are being under control into the cloud, such as company finance information, government documents, , personal health records, and emails etc. The fact that cloud server and information owners are no longer in the same trusted domain may put the outsourced unencrypted data at risk [3]: the cloud server may even be hacked [5] or may dispose data information to unauthorized entities [4]. It follows that sensitive data has to be encrypted before outsourcing for data privacy and combating unsolicited accesses. data encryption makes efficient data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in Cloud Computing, data owners share their data with users, who might want to only retrieve specific data files they are interested in during a given particular session. One of the most familiar ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been enforced in plaintext search [6]. Unfortunately, data encryption, restricts user's ability to perform keyword search and demands the protection for keyword privacy, makes the traditional plaintext search methods fail for encrypted cloud data. Although traditional searchable encryption schemes (e.g. [7]–[11], to list a few) allow a user to securely search over encrypted data through keywords without decrypting it, these techniques support conventional Boolean keyword search, without capturing any relevance of the files in the search result. When enforced in large collaborative data outsourcing cloud platform, they may suffer from the following two main drawbacks. for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to match ones interest which demands large amount of post processing over-head; On the other hand, invariably sending back all files solely based on presence/absence of the keyword incurs 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 existing searchable encryption schemes in Cloud Computing. the state of the art in information retrieval (IR) community has already been utilizing various scoring mechanisms [13] to quantify and rank-order the relevance of files in response to any given search query. 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 is addressed in encrypted data search. Therefore, how to alter a searchable encryption system with support of secure ranked search, is the problem tackled in this paper. 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 the context of Cloud Computing. To achieve our design goals on both system security and usability, we propose to bring together the advance of both crypto and IR community to design the ranked searchable symmetric encryption scheme, in the spirit of "as-strong as-possible" security guarantee. we explore the statistical measure approach from IR and text-mining to embed weight information (i.e. relevance score) of each file during the establishment of searchable index before outsourcing the encrypted file collection. As directly outsourcing relevance scores will dispose lots of sensitive frequency information against the keyword privacy, we then integrate a recent crypto primitive [14] order preserving symmetric encryption (OPSE) and properly modify it to develop a one-to-many order-preserving

mapping technique for our purpose to protect those sensitive weight information, while providing efficient ranked search functionalities. Our contribution can be summarized as follows:

1) For the first time, we define the problem of secure ranked keyword search over encrypted cloud data, and provide such an effective protocol, which fulfills the secure ranked search functionality with little relevance score information leakage against keyword privacy.
2) Thorough security analysis shows that our ranked searchable symmetric encryption scheme indeed enjoys "as-strong-as-possible" security guarantee compared to previous SSE schemes.
3) We investigate the practical considerations and enhancements of our ranked search mechanism,including the efficient support of relevance score dynamics, the authentication of ranked search results, and the reversibility of our proposed one-to many order-preserving mapping technique.
4) Extensive experimental results demonstrate the effectiveness and efficiency of the proposed solution.

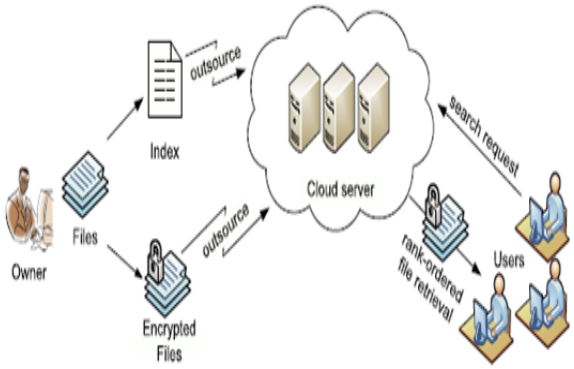

Fig 1: Architecture for search over encrypted cloud data

2 PROBLEM STATEMENT

Traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search without capturing any relevance of the files in the search result. When directly enforced in large collaborative data outsourcing cloud environment, they may suffer from the following two main drawbacks. On the one hand, for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing overhead; On the other hand, invariably sending back all files solely based on presence/absence of the keyword further incurs large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm.

Our work is among the first few ones to explore ranked search over encrypted data in Cloud Computing. 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 toward practical deployment of privacy-preserving data hosting services in the context of Cloud Computing. To achieve our design goals on both system security and usability, we propose to bring together the advance of both crypto and IR community to design the ranked searchable symmetric encryption (RSSE) scheme, in the spirit of "as-strong-as-possible" security guarantee.

A. The System and Threat Model

We consider an encrypted cloud data hosting service involving three different entities, as illustrated in Fig. 1: data owner, data user, and cloud server. Data owner has a collection of n data files C = (F1, F2,…,Fn) that he wants to outsource on the cloud server in encrypted form while still keeping the capability to search through them for effective data utilization reasons. To do so, before outsourcing, data owner will first build a secure searchable index I from a set of m distinct keywords W = (w1,w2,….,wm) extracted from the file collection C, and store both the index I and the encrypted file collection C on the cloud server.

We assume the authorization between the data owner and users is appropriately done. To search the file collection for a given keyword w, an authorized user generates and submits a search request in a secret form— a trapdoor Tw of the keyword w—to the cloud server. Upon receiving the search request T_w, the cloud server is responsible to search the index I and return the corresponding set of files to the user. We consider the secure ranked keyword search problem as follows: the search result should be returned according to certain ranked relevance criteria (e.g., keyword frequency based scores, as will be introduced shortly), to improve file retrieval accuracy for users without prior knowledge on the file collection C. However, cloud server should learn nothing or little about the relevance criteria as they exhibit significant sensitive information against keyword privacy. To reduce bandwidth, the user may send an optional value k along with the trapdoor T_w and cloud server only sends back the top-k most relevant files to the user's interested keyword w.

B. Design Goals

To alter ranked searchable symmetric encryption for effective utilization of outsourced and encrypted cloud data under the aforementioned model, our system design should achieve the following security and performance guarantee. Specifically, we have the following goals:

- Ranked keyword search: to explore different mechanisms for designing effective ranked search schemes based on the existing searchable encryption framework;
- Security guarantee: to prevent cloud server from learning the plaintext of either the data files or the searched keywords, and achieve the "as strong- as-possible" security strength compared to existing searchable encryption schemes;
- Efficiency: above goals should be achieved with minimum communication and computation overhead.

3 EFFICIENT RANKED SEARCHABLE SYMMETRIC ENCRYPTION SCHEME

The above straightforward approach demonstrates the core problem that causes the inefficiency of ranked searchable encryption. That is how to let server quickly perform the ranking without actually knowing the relevance scores. To effectively support ranked search over encrypted file collection, we now resort to the newly developed cryptographic primitive – order preserving symmetric encryption (OPSE) [14] to achieve more practical performance. Note that by resorting to OPSE, our security guarantee of RSSE is inherently weakened compared to SSE, as we now let server know the relevance order. However, this is the information we want to tradeoff for efficient RSSE. We will first briefly discuss the primitive of OPSE and its pros and cons. Then we show how we can adapt it to suit our purpose for ranked searchable encryption with an "as-strong-as-possible" security guarantee. Finally, we demonstrate how to choose different scheme parameters via concrete examples.

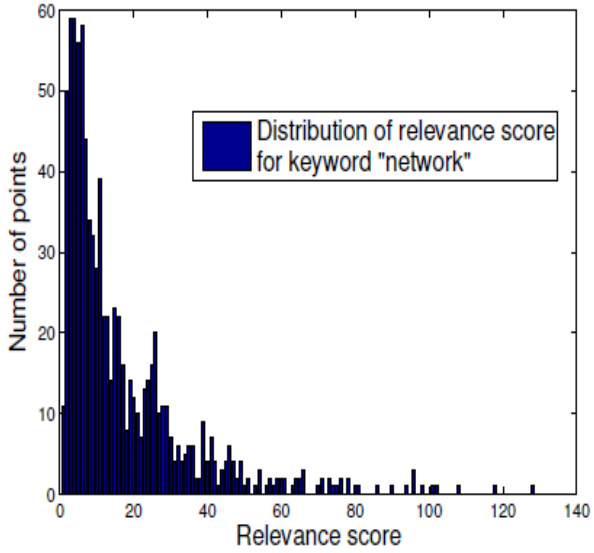

Fig 2: An example of relevance score distribution.

4 SECURITY ANALYSIS

We evaluate the security of the proposed scheme by analyzing its fulfillment of the security guarantee described in previous. Namely, the cloud server should not learn the plaintext of either the data files or the searched keywords. We start from the security analysis of our one-to-many order-preserving mapping. Then we analyze the security strength of the combination of one to many order-preserving mapping and SSE.

A. Security Analysis for One-to-many Mapping

Our one-to-many order-preserving mapping is adapted from the original OPSE, by introducing the file ID as the additional seed in the final ciphertext chosen process. Since such adaptation only functions at the final ciphertext selection process, it has nothing to do with the randomized plaintext-to-bucket mapping process in the original OPSE.

B. Security Analysis for Ranked Keyword Search

Compared to the original SSE, the new scheme embeds the encrypted relevance scores in the searchable index in addition to file ID. Thus the encrypted scores are the only additional information that the adversary can utilize against the security guarantee, i.e., keyword privacy and file confidentiality. Due to the security strength of the file encryption scheme, the file content is clearly well protected. Thus, we only need to focus on keyword privacy. From previous discussion, we know that as long as data owner properly chooses the range size R sufficiently large, the encrypted scores in the searchable index will only be a sequence of order-preserved numeric values with very low duplicates.

5 PERFORMANCE ANALYSIS

We conducted a thorough experimental evaluation of the proposed techniques on real data set: Request for comments database (RFC) [23]. At the time of writing, the RFC database contains 5563 plain text entries and totals about 277 MB. This file set contains a large number of technical keywords, many of which are unique to the files in which they are discussed. Our experiment is conducted using C programming language on a Linux machine with dual Intel Xeon CPU running at 3.0GHz. Algorithms use both openssl and MATLAB libraries. The performance of our scheme is evaluated regarding the effectiveness and efficiency of our proposed one-to many order-preserving mapping, as well as the overall performance of our RSSE scheme, including the cost of index construction as well as the time necessary for searches. Note that though we use a single server in the experiment, in practice we can separately store the searchable index and the file collections on different virtualized service nodes in the commercial public cloud, such as the Amazon EC2 and Amazon S3, respectively. In that way, even if data owners choose to store their file collection in different geographic locations for increased availability, the underlying search mechanism, which always takes place based on the searchable index, will not be affected at all.

6 CONCLUSIONS

In this paper, as an initial attempt, we motivate and solve the problem of supporting efficient ranked keyword search for achieving effective utilization of remotely stored encrypted data in Cloud Computing. We first give a basic scheme and show that by following the same existing searchable encryption framework, it is very inefficient to achieve ranked search. We then appropriately weaken the security guarantee, resort to the newly developed crypto primitive OPSE, and derive an efficient one-to-many order-preserving mapping function, which allows the effective RSSE to be designed. We also investigate some further enhancements of our ranked search mechanism, including the efficient support of relevancescore dynamics, the authentication of ranked search results, and the reversibility of our proposed one-to-many order-preserving mapping technique. Through thorough security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of our solution.

REFERENCES

[1] P. Mell and T. Grance, "Draft nist working definition of cloud computing," Referenced on Jan. 23rd, 2010 Online at http://csrc. nist.gov/groups/SNS/cloud-computing/index.html, 2010.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A berkeley view of cloud computing," University of California, Berkeley, Tech. Rep. UCBEECS- 2009-28, Feb 2009.

[3] Cloud Security Alliance, "Security guidance for critical areas of focus in cloud computing," 2009, http://www. cloudsecurityalliance.org.

[4] Z. Slocum, "Your google docs: Soon in search results?" http:// news.cnet.com/8301-17939 109-10357137-2.html, 2009.

[5] B. Krebs, "Payment Processor Breach May Be Largest Ever," Online at http://voices.washingtonpost.com/securityfix/2009/01/ payment processor breach may b.html, Jan. 2009.

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

[7] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy'00, 2000.

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

[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYP'04, volume 3027 of LNCS. Springer, 2004.

[10] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS'05, 2005.

[11] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS'06, 2006.

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

[14] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, "Orderpreserving symmetric encryption," in Proc. of Eurocrypt'09, volume 5479 of LNCS. Springer, 2009.

[15] J. Zobel and A. Moffat, "Exploring the similarity space," SIGIR Forum, vol. 32, no. 1, pp. 18–34, 1998.

[16] O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," Journal of the ACM, vol. 43, no. 3, pp. 431–473, 1996.

[17] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in Proc. of Crypto'07, volume 4622 of LNCS. Springer, 2007.

[18] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k retrieval from a confidential index," in Proc. of EDBT'09, 2009.

[19] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing," IEEE Transactions on Parallel and Distributed Systems (TPDS), vol. 22, no. 5, pp. 847–859, May 2011.

[20] C. Wang, Q. Wang, K. Ren, and W. Lou, "Towards Secure and Dependable Storage Services in Cloud Computing," IEEE Transactions on Service Computing (TSC), to appear.

[21] C. Wang, S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage," IEEE Transactions on Computers (TC), to appear.

[22] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A. L. Varna, S. He, M.Wu, and D.W. Oard, "Confidentiality-preserving rank-ordered search," in Proc. of the Workshop on Storage Security and Survivability, 2007.

[23] RFC, "Request For Comments Database," http://www.ietf.org/rfc.html.

[24] B. Waters, D. Balfanz, G. Durfee, and D. Smetters, "Building an encrypted and searchable audit log," in Proc. of NDSS'04, 2004.

[25] F. Bao, R. Deng, X. Ding, and Y. Yang, "Private query on encrypted data in multi-user settings," in Proc. of ISPEC'08, 2008.

[26] P. Golle, J. Staddon, and B. R. Waters, "Secure Conjunctive Keyword Search over Encrypted Data," in Proc. of ACNS'04, 2004, pp. 31–45.

[27] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Proc. Of ICICS'05, 2005.