Performance of SKSE and MRSE in Cloud Cache

[1]M.Sandhya, [2]CH. Raja Jacob
[1,2]Dept. of CSE, NOVA College of Engineering & Tech., Jangareddygudem, AP, India

Abstract
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 principle of "coordinate matching", i.e., as many matches as possible, to effectively capture similarity between query keywords and outsourced documents, and use "inner product similarity" to quantitatively formalize such a principle for similarity measurement. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we first propose a basic MRSE scheme using secure inner product computation, and significantly improve it to achieve privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset show our proposed schemes introduce low overhead on both computation and communication. As our future work, we will explore supporting other multi-keyword semantics (e.g., weighted query) over encrypted data, integrity check of rank order in search result and privacy guarantees in more stronger threat model.

Keywords
SKSE, MRSE, Multi-Keyword Ranked Search, Cloud Cache

I. Introduction

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 [1]. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud, especially when the data produced by them that need to be stored and utilized is rapidly increasing. To protect data privacy and combat unsolicited accesses in cloud and beyond, sensitive data, e.g., emails, personal health records, photo albums, tax documents, financial transactions, etc., may have to be encrypted by data owners before outsourcing to commercial public cloud [2]; 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 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, large amount of documents demand cloud server to perform result relevance ranking, instead of returning undifferentiated result. 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 [3]. 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-youuse" cloud paradigm. For privacy protection, such ranking operation, however, should not leak any keyword related information. On the other hand, to improve search result accuracy as well as enhance user searching experience, it is also crucial for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse result. 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" [4], i.e., as many matches as possible, is an efficient principle 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 data privacy, index privacy, keyword privacy, and many others . In the literature, searchable encryption [5–13] is a helpful technique that treats encrypted data as documents and allows a user to securely search over it through single keyword and retrieve documents of interest. However, direct application of these approaches to deploy 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 in mind. 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 . Our early work [22], has been aware of this problem, and solves the secure ranked search over encrypted data with support of only single keyword query. But how to design an efficient encrypted data search mechanism that supports multikeyword semantics without privacy breaches still remains an 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 cloud computing paradigm. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents. Specifically, we use "inner product similarity" [4], i.e., the number of query keywords appearing in a document, to quantitatively evaluate the similarity of that document to the search query in "coordinate matching" principle. During 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 inner product of query vector with data vector. However, directly outsourcing data vector or query vector will violate index privacy or search privacy. To meet the challenge of supporting such multikeyword semantic without privacy breaches, we propose a basic MRSE scheme using secure inner product computation, which is adapted from a secure

k-Nearest Neighbor (kNN) technique [4], and then improve it step by step to achieve various privacy requirements in two levels of threat models. 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 to become a reality.
2. We propose two MRSE schemes following the principle of "coordinate matching" while meeting different privacy requirements in two levels of threat models.
3. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

II. Existing System

Single Keyword Searchable Encryption Traditional single keyword searchable encryption schemes [5–13, 22], 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. [5] in the symmetric key setting, and improvements and advanced security definitions are given in Goh [6], Chang et al. [7] and Curtmola et al. [8]. Our early work [22] solves secure ranked keyword search which utilizes keyword frequency to rank results instead of returning undifferentiated results. However, it only supports single keyword search. In the public key setting, Boneh et al. [9] present the first searchable encryption construction, where anyone with public key can write to the data stored on server but only authorized users with private key can search. Public key solutions are usually very computationally expensive however. Furthermore, the keyword privacy could not be protected in the public key setting since server could encrypt any keyword with public key and then use the received trapdoor to evaluate this cipher text. Boolean Keyword Searchable Encryption To enrich search functionalities, conjunctive keyword search [14–18] 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. [16], or communication cost by secret sharing, e.g. [15]. As a more general search approach, predicate encryption schemes [19–21] are recently proposed to support both conjunctive and disjunctive search. Conjunctive keyword search returns "all-or-nothing", which means it only returns those documents in which all the keywords specified by the search query appear; disjunctive keyword search returns undifferentiated results, which means it returns every document that contains a subset of the specific keywords, even only one keyword of interest. In short, none of existing 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. Note that, inner product queries in predicate encryption only predicates whether two vectors are orthogonal or not, i.e., the inner product value is concealed except when it equals zero. Without providing the capability to compare concealed inner products, predicate encryption is not qualified for performing ranked search. Furthermore, most of these schemes are built upon the expensive evaluation of pairing operations on elliptic curves. Such inefficiency disadvantage also limits their practical performance when deployed in cloud. On a different front, the research on top-k retrieval [24] in database community is also loosely connected to our problem.

III. Proposed Work

To efficiently achieve multi-keyword ranked search, we propose to employ "inner product similarity" [4] to quantitatively formalize the efficient ranking principle "coordinate matching". Specifically, Di is a binary data vector for document Fi where each bit Di[j] 2 f0; 1g represents the existence of the corresponding keyword Wj in that document, and Q is a binary query vector indicating the keywords of interest where each bit Q[j] 2 f0; 1g represents the existence of the corresponding keyword Wj in the query fW. The similarity score of document Fi to query fW is therefore expressed as the inner product of their binary column vectors, i.e., Di ¢ Q. For the purpose of ranking, cloud server must be given the capability to compare the similarity of different documents to the query. But, to preserve strict system-wise privacy, data vector Di, query vector Q and their inner product Di ¢Q should not be exposed to cloud server. In this section, we first propose a basic MRSE scheme using secure inner product computation, which is adapted from a secure k-nearest neighbor (kNN) technique, and then show how to significantly improve it to be privacy-preserving against different levels of threat models in the MRSE framework in a step-by-step manner.

A. MRSE I: Basic Scheme

1. Secure kNN Computation: In the secure k-nearest neighbor (kNN) scheme [26], Euclidean distance between a database record pi and a query vector q is used to select k nearest database records. The secret key is composed of one (d+1)- bit vector as S and two (d + 1) £ (d + 1) invertible matrices as fM1;M2g, where d is the number of fields for each record pi. First, every data vector pi and query vector q are extended to (d + 1)-dimension vectors as ~pi and ~q, where the (d + 1)- th dimension is set to ¡0:5jjp2i jj and 1, respectively. Besides, the query vector ~q is scaled by a random number r > 0 as (rq; r). Then, ~pi is split into two random vectors as f~pi 0; ~pi 00g,and ~q is also split into two random vectors as f~q 0; ~q 00g. Note here that vector S functions as a splitting indicator. Namely, if the j-th bit of S is 0, ~pi 0[j] and ~pi 00[j] are setas the same as ~pi[j], while ~q 0[j] and ~q 00[j] are set to two random numbers so that their sum is equal to ~q[j]; if the j- th bit of S is 1, the splitting process is similar except that ~pi and ~q are switched. The split data vector pair f~pi 0; ~pi 00g is encrypted as fMT 1 ~pi 0;MT 2 ~pi 00g, and the split query vector pair f~q 0; ~q 00g is encrypted as fM¡1 1 ~q 0;M¡1 2 ~q 00g. In the query step, the product of data vector pair and query vector pair, i.e., ¡0:5r(jjpijj2¡2pi ¢q), is serving as the indicator of Euclidean distance (jjpijj2¡2pi ¢ q+jjqjj2) to select k nearest neighbors. Without prior knowledge of secret key, neither data vector nor query vector, after such a series of processes, can be recovered by analyzing their corresponding ciphertext. As MRSE is using inner product similarity instead of Euclidean distance, we need to do some modifications on the data structure to fit the MRSE framework. By eliminating imension extension, the final result changes to be the inner product as rpi ¢ q, and it seems that an efficient inner product computation scheme can be directly achieved. However, this approach hides the product only by a scale factor r which will leak the relationship among different queries. For example, if two queries are taking for the same keywords, denoted as rq and r0q, similarity scores in two queries will satisfy the scale relationship, i.e., (pi ¢rq)=(pi ¢r0q) = (pj ¢rq)=(pj ¢r0q) = r=r0. As a consequence, the search pattern of data user is leaked via examining whether similarity scores for all documents in two queries hold such relationship.

B. MRSE I Scheme

To provide a guarantee against isolation on search pattern clearly presented above, we have to eliminate the scale relationship among similarity scores in different queries. To do so, instead of simply removing the extended dimension as we plan to do at the first glance, we preserve this dimension extending operation but assign a random number to the extended dimension in each query vector. The whole scheme to achieve ranked search with multiple keywords over encrypted data is as follows. ² Setup After extracting the distinct keywords set W from the document collection F, data owner randomly generates a (n+1)-bit vector as S and two (n+1)£(n+1) invertible matrices fM1;M2g. The secret key SK is in the form of a 3-tuple as fS;M1;M2g. ² Build Index(F; SK) Data owner generates a binary data vector Di for every document Fi, where each binary bit

IV. Results

In order to implement these three algorithms on the basis on time complexity and space complexity.

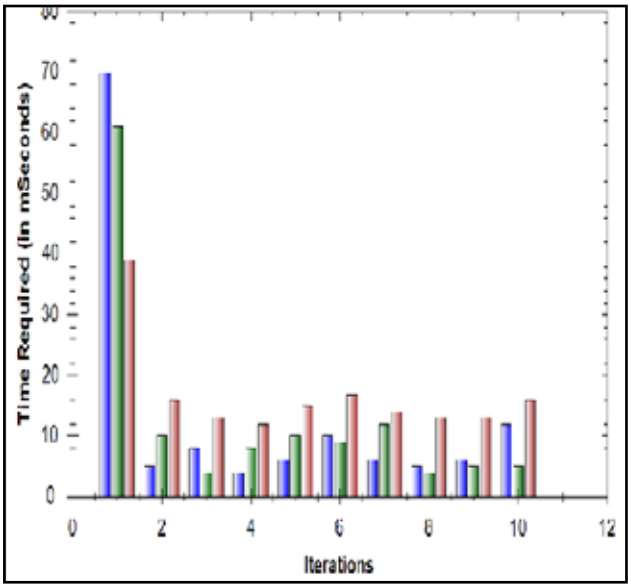

Fig. 1: Bar Chart on Basis on Time Complexity

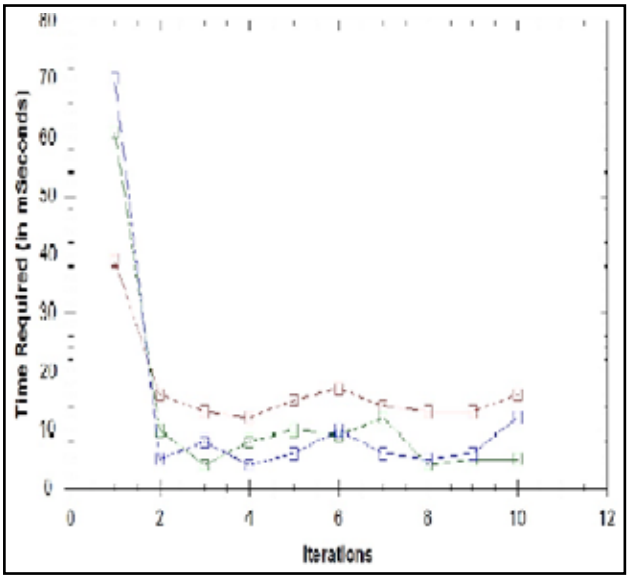

FIg. 2: Line Chart on Basis on Time Complexity

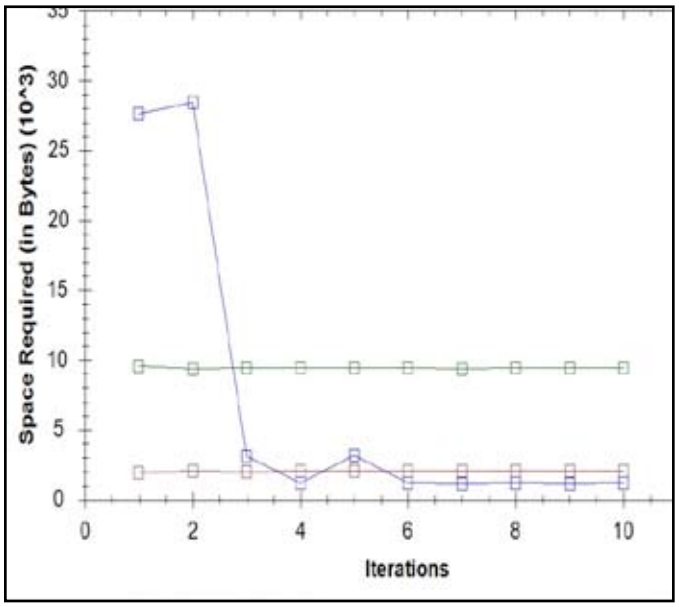

Fig. 3: Line Chart on Basis on Space Complexity

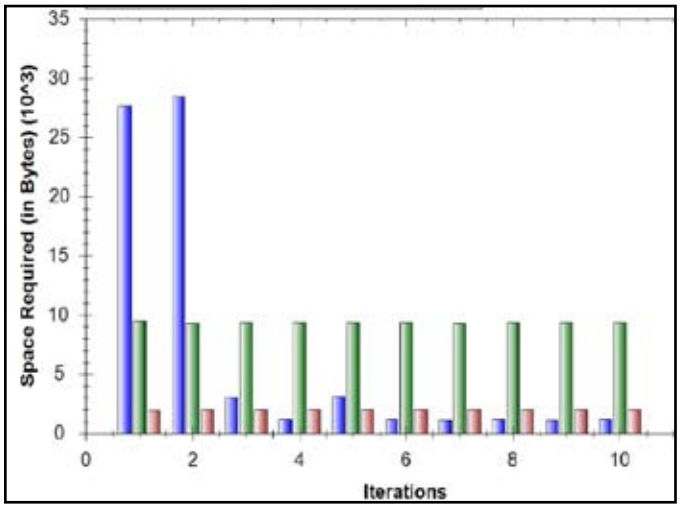

Fig. 4: Bar Chart on Basis on Space Complexity

V. Conclusions

With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has 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 cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate 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 cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize

such principle for similarity measurement. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

References

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, 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, 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, T. C. Bell,"Managing gigabytes: Compressing and indexing documents and images", Morgan Kaufmann Publishing, San Francisco, May 1999.

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

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

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

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

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

[10] M. Bellare, A. Boldyreva, 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, 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, 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, W. E. S. III, "Public key encryption that allows pir queries", in Proc. of CRYPTO, 2007.

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

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

[16] D. Boneh, 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, 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, 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, B. Waters,"Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption", in Proc. of EUROCRYPT, 2010.

[21] E. Shen, E. Shi, B. Waters,"Predicate privacy in encryption systems", in Proc. of TCC, 2009.

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

[23] S. Zerr, E. Demidova, D. Olmedilla, W. Nejdl, M. Winslett, S. Mitra,"Zerber: r-confidential indexing for distributed documents", in Proc. of EDBT, 2008, pp. 287–298.

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

[25] Y. Ishai, E. Kushilevitz, R. Ostrovsky, A. Sahai, "Cryptography from anonymity", in Proc. of FOCS, 2006, pp. 239–248.

[26] W. K. Wong, D. W. Cheung, B. Kao, N. Mamoulis, "Secure knn computation on encrypted databases", in Proc. of SIGMOD, 2009.

[27] W. W. Cohen,"Enron email dataset", [Online Available: http://www.cs.cmu.edu/»enron/.

M. Sandhya was born in Ponnur, Guntur Dt, Andhra Pradesh, India. She received B.Tech (CSE) in the year 2007 from JNT University, Hyderabad, Andhra Pradesh, India. Presently, she is pursuing M.Tech in S.E from Nova College of Engineering & Technology, JangareddyGudem, West Godavari Dt, Andhra Pradesh, India. Her Research interest includes Data-warehousing and Data-Mining and Ranking Spatial Databases by Quality Preferences.

Mr.Ch.Raja Jacob, well known Author and excellent teacher Received M.C.A and M.Tech (CSE) from Acharya Nagarjuna university is working as Associate Professor and HOD, Department of MCA, M.Tech Computer science engineering , Nova college of Engineering and Technology, He is an active member of ISTE. He has 7 years of teaching experience in various engineering colleges. To his credit couple of publications both national and international conferences /journals . His area of Interest includes Data Warehouse and Data Mining, information security, flavors of Unix Operating systems and other advances in computer Applications.