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

Multi-Keyword Ranked Search over Encrypted Cloud Data

Kolisetti Lakshmi Pravallika
PG Scholar,
Department of CSE,
Sri Chundi Ranganayakulu Engineering College,
Chilakaluripet, Guntur, AP, India.

K Swaroopa Rani
Assistant Professor,
Department of CSE,
Sri Chundi Ranganayakulu Engineering College,
Chilakaluripet, Guntur, AP, India.

ABSTRACT:

With the advent of cloud computing, data owners are motivated to outsource their complex data management systemsfrom local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitivedata 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 anddocuments in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of theirrelevance 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-preservingmulti-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements forsuch 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.

Wefurther use “inner product similarity” to quantitatively evaluate such similarity measure. We first propose a basic idea for the MRSEbased on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringentprivacy requirements in two different threat models. To improve search experience of the data search service, we further extendthese two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees ofproposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce lowoverhead on computation and communication.

INTRODUCTION:

CLOUD computing is the long dreamed vision ofcomputing as a utility, where cloud customers canremotely store their data into the cloud so as to enjoy theon-demand high-quality applications and services from ashared pool of configurable computing resources [2], [3]. Itsgreat flexibility and economic savings are motivating bothindividuals and enterprises to outsource their local complexdata management system into the cloud. To protectdata privacy and combat unsolicited accesses in the cloud\ and beyond, sensitive data, for example, e-mails, personalhealth records, photo albums, tax documents, financialtransactions, and so on, may have to be encrypted by dataowners before outsourcing to the commercial public cloud[4]; this, however, obsoletes the traditional data utilizationservice based on plaintext keyword search. The trivialsolution of downloading all the data and decrypting locallyis clearly impractical, due to the huge amount ofbandwidth cost in cloud scale systems.

Moreover, asidefrom eliminating the local storage management, storingdata into the cloud serves no purpose unless they can beeasily searched and utilized. Thus, exploring privacypreservingand effective search service over encryptedcloud data is of paramount importance. Considering thepotentially large number of on-demand data users andhuge amount of outsourced data documents in the cloud,this problem is particularly challenging as it is extremelydifficult to meet also the requirements of performance,system usability, and scalability.On the one hand, to meet the effective data retrievalneed, the large amount of documents demand the cloudserver to perform result relevance ranking, instead ofreturning undifferentiated results. Such ranked searchsystem enables data users to find the most relevantinformation quickly, rather than burdensomely sortingthrough every match in the content collection [5]. Rankedsearch can also elegantly eliminate unnecessary networktraffic by sending back only the most relevant data, whichis highly desirable in the “pay-as-you-use” cloud paradigm.

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

For privacy protection, such ranking operation,however, should not leak any keyword related information.On the other hand, to improve the search result accuracy aswell as to enhance the user searching experience, it is also necessary for such ranking system to support multiplekeywords search, as single keyword search often yields fartoo coarse results. As a common practice indicated bytoday's web search engines (e.g., Google search), data usersmay tend to provide a set of keywords instead of only oneas the indicator of their search interest to retrieve the mostrelevant data. And each keyword in the search request isable to help narrow down the search result further."Coordinate matching" [6], i.e., as many matches aspossible, is an efficient similarity measure among suchmulti-keyword semantics to refine the result relevance,and has been widely used in the plaintext informationretrieval (IR) community.

However, how to apply it in theencrypted cloud data search system remains a verynchallenging task because of inherent security and privacyobstacles, including various strict requirements like thedata privacy, the index privacy, the keyword privacy, andmany others (see Section 3.2).In the literature, searchable encryption [7], [8], [9], [10],[11], [12], [13], [14], [15] is a helpful technique that treatsencrypted data as documents and allows a user to securelysearch through a single keyword and retrieve documents ofinterest. However, direct application of these approaches tothe secure large scale cloud data utilization system wouldnot be necessarily suitable, as they are developed ascryptoprimitives and cannot accommodate such highservice-level requirements like system usability, usersearching experience, and easy information discovery.

Although some recent designs have been proposed tosupport Boolean keyword search [16], [17], [18], [19], [20],[21], [22], [23], [24] as an attempt to enrich the searchflexibility, they are still not adequate to provide users withacceptable result ranking functionality (see Section 7). Ourearly works [25], [26] have been aware of this problem, andprovide solutions to the secure ranked search overencrypted data problem but only for queries consisting ofa single keyword. How to design an efficient encrypteddata search mechanism that supports multi-keywordsemantics without privacy breaches still remains a challengingopen problem. In this paper, for the first time, we define and solve theproblem of multi-keyword ranked search over encryptedcloud data (MRSE) while preserving strict systemwiseprivacy in the cloud computing paradigm.

Among variousmulti-keyword semantics, we choose the efficient similaritymeasure of "coordinate matching," i.e., as many matches aspossible, to capture the relevance of data documents to thesearch query. Specifically, we use "inner product similarity"[6], i.e., the number of query keywords appearing in adocument, to quantitatively evaluate such similarity measureof that document to the search query. During the indexconstruction, each document is associated with a binaryvector as a subindex where each bit represents whethercorresponding keyword is contained in the document. Thesearch query is also described as a binary vector where eachbit means whether corresponding keyword appears in thissearch request, so the similarity could be exactly measuredby the inner product of the query vector with the datavector.

However, directly outsourcing the data vector or thequery vector will violate the index privacy or the searchprivacy. To meet the challenge of supporting such multikeywordsemantic without privacy breaches, we proposea basic idea for the MRSE using secure inner productcomputation, which is adapted from a secure k-nearestneighbor (kNN) technique [27], and then give two significantlyimproved MRSE schemes in a step-by-stepmanner to achieve various stringent privacy requirementsin two threat models with increased attack capabilities. Ourcontributions are summarized as follows:

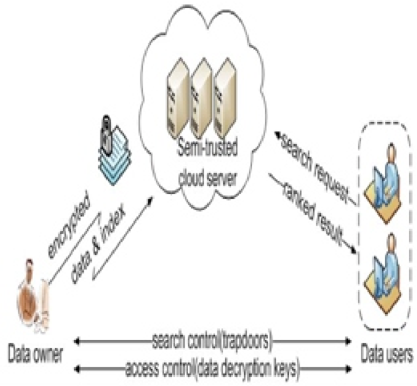


Fig. 1. Architecture of the search over encrypted cloud data.

1. For the first time, we explore the problem of multikeywordranked search over encrypted cloud data,and establish a set of strict privacy requirements forsuch a secure cloud data utilization system.2. We propose two MRSE schemes based on thesimilarity measure of "coordinate matching" whilemeeting different privacy requirements in twodifferent threat models.3.

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

We investigate some further enhancements of ourranked search mechanism to support more searchsemantics and dynamic data operations.4. Thorough analysis investigating privacy and efficiencyguarantees of the proposed schemes is given,and experiments on the real-world data set furthershow the proposed schemes indeed introduce lowoverhead on computation and communication.Compared with the preliminary version [1] of this paper,this journal version proposes two new mechanisms tosupport more search semantics. This version also studiesthe support of data/index dynamics in the mechanismdesign. Moreover, we improve the experimental works byadding the analysis and evaluation of two new schemes. Inaddition to these improvements, we add more analysis onsecure inner product and the privacy part.The remainder of this paper is organized as follows: InSection 2, we introduce the system model, the threat model,our design goals, and the preliminary. Section 3 describesthe MRSE framework and privacy requirements, followed bySection 4, which describes the proposed schemes. Section 5presents simulation results. We discuss related work onboth single and Boolean keyword searchable encryption inSection 6, and conclude the paper in Section 7.

PROBLEM FORMULATION
2.1 SYSTEM MODEL:

Considering a cloud data hosting service involving thredifferent entities, as illustrated in Fig. 1: the data owner, thedata user, and the cloud server. The data owner has acollection of data documents F to be outsourced to thecloud server in the encrypted form C. To enable thesearching capability over C for effective data utilization,the data owner, before outsourcing, will first build anencrypted searchable index I from F, and then outsourceboth the index I and the encrypted document collection Cto the cloud server. To search the document collection for tgiven keywords, an authorized user acquires a correspondingtrapdoor T through search control mechanisms, forexample, broadcast encryption [10]. Upon receiving T froma data user, the cloud server is responsible to search theindex I and return the corresponding set of encrypteddocuments. To improve the document retrieval accuracy,the search result should be ranked by the cloud serveraccording to some ranking criteria (e.g., coordinate matching,as will be introduced shortly). Moreover, to reduce thecommunication cost, the data user may send an optionalnumber k along with the trapdoor T so that the cloud serveronly sends back top-k documents that are most relevant tothe search query.

Finally, the access control mechanism [28] is employed to manage decryption capabilities given tousers and the data collection can be updated in terms ofinserting new documents, updating existing documents,and deleting existing documents.The cloud server is considered as “honest-but-curious” inour model, which is consistent with related works on cloudsecurity [28], [29]. Specifically, the cloud server acts in an“honest” fashion and correctly follows the designatedprotocol specification. However, it is “curious” to inferand analyze data (including index) in its storage andmessage flows received during the protocol so as to learnadditional information. Based on what information thecloud server knows, we consider two threat models withdifferent attack capabilities as follows. Known ciphertext model. In this model, the cloud server issupposed to only know encrypted data set C and searchableindex I, both of which are outsourced from the data owner.Known background model. In this stronger model, the cloudserver is supposed to possess more knowledge than what canbe accessed in the known ciphertext model. Such informationmay include the correlation relationship of given searchrequests (trapdoors), as well as the data set related statisticalinformation. As an instance of possible attacks in this case,the cloud server could use the known trapdoor informationcombined with document/keyword frequency [30] todeduce/identify certain keywords in the query.

DESIGN GOALS:

To enable ranked search for effective utilization of outsourcedcloud data under the aforementioned model, oursystem design should simultaneously achieve security andperformance guarantees as follows.. Multi-keyword ranked search. To design searchschemes which allow multi-keyword query andprovide result similarity ranking for effective dataretrieval, instead of returning undifferentiated results.. Privacy-preserving. To prevent the cloud server fromlearning additional information from the data setand the index, and to meet privacy requirementsspecified in Section 3.2.. Efficiency. Above goals on functionality and privacyshould be achieved with low communication andcomputation overhead.

FRAMEWORK AND PRIVACY REQUIREMENTS FORMRSE:

In this section, we define the framework of multi-keywordranked search over encrypted cloud data (MRSE)

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

andestablish various strict systemwise privacy requirements forsuch a secure cloud data utilization system. 3.1

MRSE FRAMEWORK:

For easy presentation, operations on the data documentsare not shown in the framework since the data owner couldeasily employ the traditional symmetric key cryptographyto encrypt and then outsource data. With focus on theindex and query, the MRSE system consists of four algorithms as follows:. Setup ð1‘Þ. Taking a security parameter ‘ as input, thedata owner outputs a symmetric key as SK.. BuildIndexðF; SKÞ. Based on the data set F, the dataowner builds a searchable index I which is encrypted bythe symmetric key SK and then outsourced to the cloudserver. After the index construction, the documentcollection can be independently encrypted and outsourced.. Trapdoor ðfWÞ. With t keywords of interest infW as input,this algorithm generates a corresponding trapdoor TeW.. Query ðTeW; k; IÞ. When the cloud server receives a queryrequest as (TeW, k), it performs the ranked search onthe index I with the help of trapdoor TeW, and finallyreturns FeW, the ranked id list of top-k documents sortedby their similarity with fW.Neither the search control nor the access control is withinthe scope of this paper. While the former is to regulate howauthorized users acquire trapdoors, the later is to manageusers' access to outsourced documents.

PRIVACY REQUIREMENTS FOR MRSE:

The representative privacy guarantee in the related literature,such as searchable encryption, is that the server shouldlearn nothing but search results. With this general privacydescription, we explore and establish a set of strict privacyrequirements specifically for the MRSE framework.As for the data privacy, the data owner can resort to thetraditional symmetric key cryptography to encrypt the databefore outsourcing, and successfully prevent the cloudserver from prying into the outsourced data. With respect tothe index privacy, if the cloud server deduces any associationbetween keywords and encrypted documents from index, itmay learn the major subject of a document, even the contentof a short document [30]. Therefore, the searchable indexshould be constructed to prevent the cloud server fromperforming such kind of association attack. While data andindex privacy guarantees are demanded by default in therelated literature, various search privacy requirements involvedin the query procedure are more complex anddifficult to tackle as follows.

Keyword privacy. As users usually prefer to keep theirsearch from being exposed to others like the cloud server,the most important concern is to hide what they are searching, i.e., the keywords indicated by the correspondingtrapdoor. Although the trapdoor can be generated in acryptographic way to protect the query keywords, the cloudserver could do some statistical analysis over the searchresult to make an estimate. As a kind of statisticalinformation, document frequency (i.e., the number of documentscontaining the keyword) is sufficient to identify thekeyword with high probability [31]. When the cloud serverknows some background information of the data set, thiskeyword specific information may be utilized to reverseengineerthe keyword.Trapdoor unlinkability. The trapdoor generation functionshould be a randomized one instead of being deterministic.In particular, the cloud server should not be able to deducethe relationship of any given trapdoors, for example, todetermine whether the two trapdoors are formed by thesame search request. Otherwise, the deterministic trapdoorgeneration would give the cloud server advantage toaccumulate frequencies of different search requests regardingdifferent keyword(s), which may further violate theaforementioned keyword privacy requirement. So thefundamental protection for trapdoor unlinkability is tointroduce sufficient nondeterminacy into the trapdoorgeneration procedure.Access pattern.

PRIVACY-PRESERVING AND EFFICIENT MRSE:

To efficiently achieve multi-keyword ranked search, wepropose to employ "inner product similarity" [6] toquantitatively evaluate the efficient similarity measure"coordinate matching." Specifically, Di is a binary datavector for document Fi where each bit Di½j_ 2 f0; 1grepresents the existence of the corresponding keywordWj in that document, and Q is a binary query vector indicating the keywords of interest where each bit Q½j_ 2f0; 1g represents the existence of the corresponding keywordWj in the query fW. The similarity score of documentFi to query fW is therefore expressed as the inner productof their binary column vectors, i.e., Di _ Q. For the purposeof ranking, the cloud server must be given the capability tocompare the similarity of different documents to thequery. But, to preserve strict systemwise privacy, datavector Di, query vector Q and their inner product DiQshould not be exposed to the cloud server. In this section,we first propose a basic idea for the MRSE using

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

secureinner product computation, which is adapted from asecure kNN technique, and then show how to significantlyimprove it to be privacy-preserving against different threatmodels in the MRSE framework in a step-by-step manner. We further discuss supporting more search semantics anddynamic operation.4.1 Secure Inner Product ComputationIn the secure kNN scheme [27], euclidean distance betweena data record pi and a query vector q is used to select knearest database records. The secret key is composed ofone ðd þ 1Þ-bit vector as S and two d þ 1Þ _ d þ 1Þinvertible matrices as fM1;M2g, where d is the number offields for each record pi. First, every data vector pi andquery vector q are extended to ðd þ 1Þ-dimension vectors as~pi and ~q, where the ðd þ 1Þth dimension is set to _0:5kp2ikand 1, respectively. Besides, the query vector ~q is scaled bya random number r > 0 as ðrq; rÞ. Then, ~pi is split into tworandom vectors as f~pi0; ~pi00g, and ~q is also split into tworandom vectors as f~q 0;~q 00g. Note here that vector Sfunctions as a splitting indicator. Namely, if the jth bit ofS is 0, ~pi0½j_ and ~pi00½j_ are set as the same as ~pi½j_, while ~q 0½j_and ~q 00½j_ are set to two random numbers so that their sumis equal to ~q½j_; if the jth bit of S is 1, the splitting process issimilar except that ~pi and ~q are switched. The split datavector pair f~pi0; ~pi00g is encrypted as fMT1 ~pi0;MT2 ~pi00g, andthe split query vector pair f~q 0;~q 00g is encrypted asfM_11 ~q 0;M_12 ~q 00g.

In the query step, the product of datavector pair and query vector pair, i.e., _0:5rðkpik2 _ 2pi _ qÞ,is serving as the indicator of euclidean distance ðkpik2 _2pi _ q þ kqk2Þ to select k nearest neighbors.As the MRSE is using the inner product similarityinstead of the euclidean distance, we need to do somemodifications on the data structure to fit the MRSEframework. One way to do that is by eliminating thedimension extension, the final result changes to be theinner product as rpi _ q. While the encryption of either datarecord or query vector involves two multiplications of ad _ d matrix and a d-dimension vector with complexityOðd2Þ, the final inner product computation involves twomultiplications of two d-dimension vectors with complexityOðdÞ. In the known ciphertext model, the splittingvector S is unknown, so ~pi0 and ~pi00 are considered as tworandom d-dimensional vectors. To solve the linear equationscreated by the encryption of data vectors, we have2dm unknowns in m data vectors and 2d2 unknowns infM1;M2g. Since we have only 2dm equations, which areless than the number of unknowns, there is no sufficientinformation to solve either data vectors or fM1;M2g.Similarly, ~q 0 and ~q 00 are also considered as two randomd-dimensional vectors. To solve the linear equationscreated by the encryption of query vectors, we have 2dunknowns in two query vectors and 2d2 unknowns infM1;M2g. Since we have only 2d equations here, which areless than the number of unknowns, there is no sufficientinformation to solve either query vectors or fM1;M2g.Hence, we believe that without prior knowledge of secretkey, neither data vector nor query vector, after such a seriesof processes like splitting and multiplication, can berecovered by analyzing their corresponding cipher texts.

PERFORMANCE ANALYSIS:

In this section, we demonstrate a thorough experimentalevaluation of the proposed technique on a real-world dataset: the Enron Email Data Set [35]. We randomly selectdifferent number of e-mails to build data set. The wholeexperiment system is implemented by C language on aLinux Server with Intel Xeon Processor 2.93 GHz. Thepublic utility routines by Numerical Recipes are employedto compute the inverse of matrix. The performance of ourtechnique is evaluated regarding the efficiency of fourproposed MRSE schemes, as well as the tradeoff betweensearch precision and privacy.

5.1 PRECISION AND PRIVACY:

As presented in Section 4, dummy keywords are insertedinto each data vector and some of them are selected inevery query. Therefore, similarity scores of documents willbe not exactly accurate. In other words, when the cloudserver returns top-k documents based on similarity scoresof data vectors to query vector, some of real top-k relevantdocuments for the query may be excluded. This is becauseeither their original similarity scores are decreased or thesimilarity scores of some documents out of the real top-kare increased, both of which are due to the impact ofdummy keywords inserted into data vectors. To evaluatethe purity of the k documents retrieved by user, we definea measure as precision Pk ¼ k0=k where k0 is number ofreal top-k documents that are returned by the cloud server..

ANALYSIS:

We analyze this MRSE_I scheme from three aspects ofdesign goals described in Section 2.Functionality and efficiency. Assume the number of querykeywords

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

appearing in a document Fi is xi ¼ Di _ Q. From(1), the final similarity score as yi ¼ Ii _ TeW¼ rðxi þ "iÞ þ tis a linear function of xi, where the coefficient r is set as apositive random number. However, because the randomfactor "i is introduced as a part of the similarity score, thefinal search result on the basis of sorting similarity scoresmay not be as accurate as that in original scheme. For theconsideration of search accuracy, we can let "i follow anormal distribution Nð_; _2Þ, where the standard deviation_ functions as a flexible tradeoff parameter amongsearch accuracy and security. From the consideration ofeffectiveness, _ is expected to be smaller so as to obtainhigh precision indicating the good purity of retrieved.

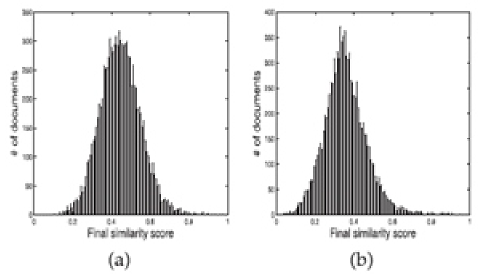


Fig. 2. Distribution of final similarity score with different standarddeviations, 10k documents, 10 query keywords. (a) _ ¼ 1. (b) _ ¼ 0:5.documents. To quantitatively evaluate the search accuracy,we set a measure as precision Pk to capture the fraction ofreturned top-k documents that are included in the real topklist. Detailed accuracy evaluation on the real-world dataset will be given in Section 5.As for the efficiency, our inner product-based MRSEscheme is an outstanding approach from the performanceperspective. In the steps like BuildIndex or Trapdoor, thegeneration procedure of each subindex or trapdoor involvestwo multiplications of a ðn þ 2Þ _ ðn þ 2Þ matrix and aðn þ 2Þ-dimension vector. In the Query, the final similarityscore is computed through two multiplications of twoðn þ 2Þ-dimension vectors.Privacy. As for the data privacy, traditional symmetric keyencryption techniques could be properly utilized here andis not within the scope of this paper. The index privacy iswell protected if the secret key SK is kept confidential sincesuch vector encryption method has been proved to besecure in the known ciphertext model [27]. Althoughwe add two more dimensions to the vectors compared to the adapted secure inner product computation, the numberof equations as 2ðn þ 2Þm is still less than the number ofunknowns as the sum of 2ðn þ 2Þm unknowns in m datavectors and 2d2 unknowns in fM1;M2g. With the randomnessintroduced by the splitting process and the randomnumbers r, and t, our basic scheme can generate two totallydifferent trapdoors for the same query fW. This nondeterministictrapdoor generation can guarantee the trapdoorunlinkability which is an unsolved privacy leakage problemin related symmetric key-based searchable encryptionschemes because of the deterministic property of trapdoorgeneration [10]. Moreover, with properly selected parameter_ for the random factor "i, even the final score resultscan be obfuscated very well, preventing the cloud serverfrom learning the relationships of given trapdoors and thecorresponding keywords.

Note that although _ is expectedto be small from the effectiveness point of view, the smallone will introduce small obfuscation into the final similarityscores, which may weaken the protection of keywordprivacy and trapdoor unlinkability. As shown in Fig. 2, thedistribution of the final similarity scores with smaller _ willenable the cloud server to learn more statistical informationabout the original similarity scores, and therefore _ shouldbe set large enough from the consideration of privacy..

TABLE 1

K_3 Appears in Every Document

Doc	Query for $\{K_1, K_2, K_3\}$	Query for $\{K_1, K_2\}$
1	$x_1 = 3, y_1 = r(3 + \varepsilon_1) + t$	$x'_1 = 2, y'_1 = r'(2 + \varepsilon_1) + t'$
2	$x_2 = 2, y_2 = r(2 + \varepsilon_2) + t$	$x'_2 = 1, y'_2 = r'(1 + \varepsilon_2) + t'$
3	$x_3 = 1, y_3 = r(1 + \varepsilon_3) + t$	$x'_3 = 0, y'_3 = r'(0 + \varepsilon_3) + t'$

MRSE_II:

Privacy-Preserving Scheme in KnownBackground ModelWhen the cloud server has knowledge of some backgroundinformation on the outsourced data set, forexample, the correlation relationship of two given trapdoors,certain keyword privacy may not be guaranteed anymore by the MRSE_I scheme. This is possible in theknown background model because the cloud server canuse scale analysis as follows to deduce the keywordspecific information, for example, document frequency,which can be further combined with background informationto identify the keyword in a query at high probability.

After presenting how the cloud server uses scale analysisattack to break the keyword privacy, we propose a moreadvanced MRSE scheme to be privacy-preserving in theknown background model.

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

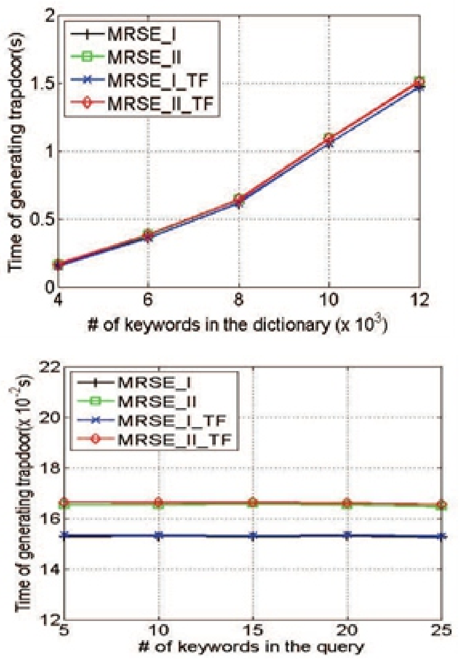
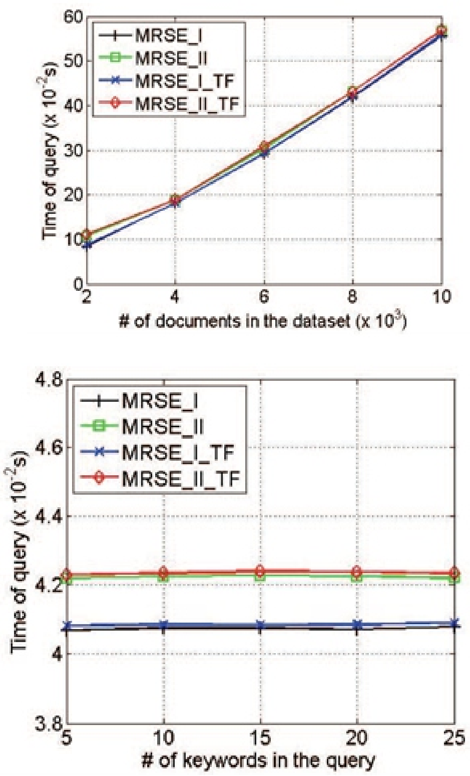


Fig. 3. With different choice of standard deviation _ for the randomvariable “, there exists tradeoff between (a) Precision, and (b) Rank



Privacy. Fig. 4. Time cost of building index. (a) For the different size of data setwith the same dictionary, n ¼ 4;000. (b) For the same data set withdifferent size of dictionary, m ¼ 1;000.

CONCLUSION:

In this paper, for the first time we define and solve theproblem of multi-keyword ranked search over encryptedcloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose theefficient similarity measure of “coordinate matching,” i.e., asmany matches as possible, to effectively capture therelevance of outsourced documents to the query keywords,and use “inner product similarity” to quantitativelyevaluate such similarity measure. For meeting the challengeof supporting multi-keyword semantic without privacybreaches, we propose a basic idea of MRSE using secureinner product computation. Then, we give two improvedMRSE schemes to achieve various stringent privacy requirementsin two different threat models. We also investigatesome further enhancements of our ranked search mechanism,including supporting more search semantics, i.e., TFIDF, and dynamic data operations. Thorough analysisinvestigating privacy and efficiency guarantees of proposedschemes is given, and experiments on the real-world dataset show our proposed schemes introduce low overhead onboth computation and communication. In our future work, we will explore checking theintegrity of the rank order in the search result assumingthe cloud server is untrusted.

REFERENCES:

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-PreservingMulti-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, “ABreak 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-BasedSecure 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 KaufmannPublishing, May 1999.

[7] D. Song, D. Wagner, and A. Perrig, "Practical Techniques forSearches on Encrypted Data," Proc. IEEE Symp. Security andPrivacy, 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 KeywordSearches 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, "SearchableSymmetric 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, "PublicKey Encryption with Keyword Search," Proc. Int'l Conf. Theory andApplications of Cryptographic Techniques (EUROCRYPT), 2004.

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

[13] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J.Malone-Lee, G. Neven, P. Paillier, and H. Shi, "SearchableEncryption Revisited: Consistency Properties, Relation to AnonymousIbe, and Extensions," J. Cryptology, vol. 21, no. 3, pp. 350-391, 2008.

[14] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "FuzzyKeyword Search Over Encrypted Data in Cloud Computing,"Proc. IEEE INFOCOM, Mar. 2010.

[15] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public KeyEncryption That Allows PIR Queries," Proc. 27th Ann. Int'lCryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.