EXECUTION OF PRIVACY - PRESERVING MULTI-KEYWORD POSITIONED SEARCH OVER CLOUD INFORMATION

Sunitha. N[1] and Prof. B. Sakthivel[2]
sunithank.dvg@gmail.com and everrock17@gmail.com
[1]PG Student and [2]Professor & Head, Department of Computer Science and Engineering
P.S.V. College of Engineering & Technology, Krishnagiri

---***---

Abstract - *With the coming of distributed computing, information proprietors are spurred to re-appropriate their perplexing information the board frameworks from nearby locales to the business open cloud for extraordinary adaptability and financial funds. Be that as it may, for securing information protection, touchy information must be encoded before re-appropriating, which obsoletes customary information usage dependent on plaintext catchphrase seek. Subsequently, empowering an encoded cloud information look administration is of principal significance. Thinking about the expansive number of information clients and archives in the cloud, it is important to permit numerous watchwords in the hunt solicitation and return reports in the request of their significance to these catchphrases. Related takes a shot at accessible encryption center around single watchword look or Boolean catchphrase seek, and once in a while sort the indexed lists.*

In this paper, out of the blue, we characterize and take care of the testing issue of protection saving multi-watchword positioned look over encoded cloud information (MRSE).We build up a lot of severe protection necessities for such a safe cloud information usage framework. Among different multi-watchword semantics, we pick the productive closeness proportion of "organize coordinating", i.e., whatever number matches as could be expected under the circumstances, to catch the pertinence of information records to the hunt inquiry. We further use "internal item similitude" to quantitatively assess such likeness measure. We initially propose a fundamental thought for the MRSE dependent on secure inward item calculation, and after that give two altogether improved MRSE plans to accomplish different stringent protection necessities in two distinctive danger models. Careful examination exploring protection and effectiveness certifications of proposed plans is given. Trials on this present reality dataset further show proposed plots for sure present low overhead on calculation and correspondence.

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. Its great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. 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. For privacy protection, such ranking operation, however, should not leak any keyword related information.

On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results. How to design an efficient encrypted data search mechanism that supports multi-keyword semantics without privacy breaches still remains a challenging open problem. Our contributions are summarized as follows,

1) For the first time, we explore the problem of multi- keyword ranked search over encrypted cloud data, and establish a set of strict privacy requirements for such a secure cloud data utilization system.

2) We propose two MRSE schemes based on the similarity measure of "coordinate matching" while meeting different privacy requirements in two different threat models.

3) Thorough analysis investigating privacy and efficiency guarantees of the proposed schemes is given, and experiments on the real-world dataset further show the proposed schemes indeed introduce low overhead on computation and communication.

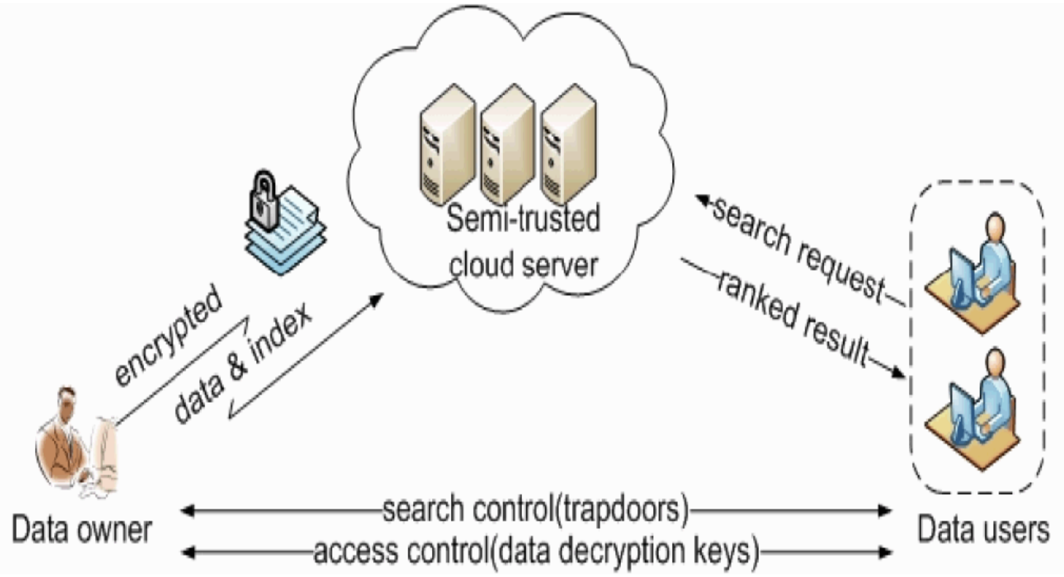
Semi-trusted cloud server
encrypted data & index
search request
ranked result
Data owner
Data users
search control(trapdoors)
access control(data decryption keys)

Fig. 1 Cloud Information

II. PROBLEM FORMULATION

A. System Model

Considering a cloud data hosting service involving three different entities, as illustrated in the data owner, the data user, and the cloud server. The data owner has a collection of data documents F to be outsourced to the cloud server in the encrypted form C. To enable the searching capability over C for effective data utilization, the data owner, before outsourcing, will first build an encrypted searchable index I from F, and then outsource both the index I and the encrypted document collection C to the cloud server. Moreover, to reduce the communication cost, the data user may send an optional number k along with the trapdoor T so that the cloud server only sends back top-k documents that are most relevant to the search query. Finally, the access control mechanism is employed to manage decryption capabilities given to users.

B. Threat Model

The cloud server is considered as "honest-but-curious" in our model, which is consistent with related works on cloud security. Specifically, the cloud server acts in an "honest" fashion and correctly follows the designated protocol specification. However, it is "curious" to infer and analyze data (including index) in its storage and message flows received during the protocol so as to learn additional information.

Known Ciphertext Model In this model, the cloud server is supposed to only know encrypted dataset C and searchable index I, both of which are outsourced from the data owner.

Known Background Model In this stronger model, the cloud server is supposed to possess more knowledge than what can be accessed in the known ciphertext model.

C. Design Goals

To enable ranked search for effective utilization of out- sourced cloud data under the aforementioned model, our system design should simultaneously achieve security and performance guarantees as follows.

- **Multi-keyword Ranked Search**: To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning undifferentiated results.

- **Privacy-Preserving:** To prevent the cloud server from learning additional information from the dataset and the index, and to meet privacy requirements specified in section III-B.

- **Efficiency:** Above goals on functionality and privacy should be achieved with low communication and computation overhead.

D. Notations

• F – the plaintext document collection, denoted as a set of m data documents F = (F1, F2, . . . , Fm). • C – the encrypted document collection stored in the cloud server, denoted as C = (C1, C2, . . . , Cm). • W – the dictionary, i.e., the keyword set consisting of n keyword, denoted as W = (W1, W2, . . . , Wn).

E. Preliminary on Coordinate Matching

As a hybrid of conjunctive search and disjunctive search, "coordinate matching" [4] is an intermediate similarity measure which uses the number of query keywords appearing in the document to quantify the relevance of that document to the query.

III. FRAMEWORK AND PRIVACY REQUIREMENTS FOR MRSE

In this section, we define the framework of multi-keyword ranked search over encrypted cloud data (MRSE) and establish various strict system-wise privacy requirements for such a secure cloud data utilization system.

A. MRSE Framework

For easy presentation, operations on the data documents are not shown in the framework since the data owner could easily employ the traditional symmetric key cryptography to encrypt and then outsource data.

- **Setup:** Taking a security parameter l as input, the data owner outputs a symmetric key as SK.

- **BuildIndex(F, SK)** Based on the dataset F, the data owner builds a searchable index I which is encrypted by the symmetric key SK and then outsourced to the cloud server.

- **Trapdoor(Wf)** With t keywords of interest in Wf as input, this algorithm generates a corresponding trapdoor TWf.

- **Query(TWf, k, I)** When the cloud server receives a query request as (TWf, k), it performs the ranked search on the index I with the help of trapdoor TWf, and finally returns FWf, the ranked id list of top-k documents sorted by their similarity with Wf.

B. Privacy Requirements for MRSE

The representative privacy guarantee in the related literature, such as searchable encryption, is that the server should learn nothing but search results. With this general privacy description, we explore and establish a set of strict privacy requirements specifically for the MRSE framework. Therefore, the searchable index should be constructed to prevent the cloud server from performing such kind of association attack.

Keyword Privacy As users usually prefer to keep their search 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 corresponding trapdoor.

Trapdoor Unlinkability The trapdoor generation function should be a randomized one instead of being deterministic. In particular, the cloud server should not be able to deduce the relationship of any given trapdoors, e.g., to determine whether the two trapdoors are formed by the same search request.

Access Pattern Within the ranked search, the access pattern is the sequence of search results where every search result is a set of documents with rank order. Specifically, the search result for the query keyword set Wf is denoted as FWf, consisting of the id list of all documents ranked by their relevance to Wf

IV. PRIVACY-PRESERVING AND EFFICIENT MRSE

To efficiently achieve multi-keyword ranked search, we propose to employ “inner product similarity” to quantitatively evaluate the efficient similarity measure “coordinate matching”. Specifically, Di is a binary data vector for document Fi where each bit Di[j] ∈ {0, 1} 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] ∈ {0, 1} represents the existence of the corresponding keyword Wj in the query Wf..

A. MRSE I: Privacy-Preserving Scheme in Known Ciphertext Model

1) Secure kNN Computation: In the secure k-nearest neighbor (kNN) scheme [23], Euclidean distance between a database record pi and a query vector q is used to select k nearest database records. The major reason is that the only randomness involved is the scale factor r in the trapdoor generation, which does not provide sufficient nondeterminacy in the overall scheme as required by the trapdoor unlinkability requirement as well as the keyword privacy requirement. To provide a more advanced design for the MRSE, we now provide our MRSE I scheme as follows.

2) MRSE I Scheme: In our more advanced design, instead simply removing the extended dimension in the query vector as we plan to do at the first glance, we preserve this dimension extending operation but assign a new random number t to the extended dimension in each query vector.

- **Setup** The data owner randomly generates a (n + 2)-bit vector as S and two (n+ 2)×(n+ 2) invertible matrices {M1, M2}. The secret key SK is in the form of a 3-tuple as {S, M1, M2}.

- **BuildIndex(F, SK)** The data owner generates a binary data vector Di for every document Fi, where each binary bit Di [j] represents whether the corresponding keyword Wj appears in the document Fi

- **Trapdoor(Wf)** With t keywords of interest in Wf as input, one binary vector Q is generated where each bit Q[j] indicates whether Wj ∈ Wf is true or false.

- **Query(TWf, k, I)** With the trapdoor TWf, the cloud server computes the similarity scores of each document Fi as in equation 1. WLOG, we assume r > 0. After sorting all scores, the cloud server returns the top-k ranked id list FWf.

3) Analysis: We analyze this MRSE I scheme from three aspects of design goals described in section II. Functionality and Efficiency Assume the number of query keywords appearing in a document Fi is xi = Di. Q. From equation 1, the final similarity score as yi = Ii TWf =r(xi+εi)+t is a linear function of xi, where the coefficient r is set as a positive random number.

B. MRSE II: Privacy-Preserving Scheme in Known Background Model When the cloud server has knowledge of some background information on the outsourced dataset, e.g., the correlation relationship of two given trapdoors, certain keyword privacy may not be guaranteed anymore by the MRSE I scheme.

2) MRSE II Scheme: The privacy leakage shown above is caused by the fixed value of random variable εiin data vector Di. To eliminate such fixed property in any specific document, more dummy keywords instead of only one should be inserted into every data vector Di . All the vectors are extended to (n + U + 1)-dimension instead of (n + 2), where U is the number of dummy keywords inserted.

V. PERFORMANCE ANALYSIS

In this section, we demonstrate a thorough experimental evaluation of the proposed technique on a real-world dataset: the Enron Email Dataset . We randomly select different number of emails to build dataset. The whole experiment system is implemented by C language on a Linux Server with Intel Xeon Processor 2.93GHz. The public utility routines by Numerical Recipes are employed to compute the inverse of matrix.

A. Precision and Privacy

As presented in Section IV, dummy keywords are inserted into each data vector and some of them are selected in every query. Therefore, similarity scores of documents will be not exactly accurate.

B. Efficiency

1) Index Construction: To build a searchable subindex Ii for each document Fi in the dataset F, the first step is to map the keyword set extracted from the document Fi to a data vector Di, followed by encrypting every data vector.

VI. RELATED WORK

Single Keyword Searchable Encryption

Traditional single keyword searchable encryption schemes 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)B. Boolean Keyword Searchable Encryption To enrich search functionalities, conjunctive keyword search over encrypted data have been proposed.

VII. CONCLUSION

In this paper, for the first tisme 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 similarity measure of "coordinate matching", i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we propose a basic idea of MRSE using secure inner product computation. Then we give two improved MRSE schemes to achieve various stringent privacy requirements in two different threat models.

REFERENCES

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, 2009.

[2] S. Kamara and 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, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.

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

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

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

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