A Secure Query Processing on Cloud Data with Synonym Search

B. Mahesh[1] P. Munisekhar[2]

[1] Department of CSE, J.N.T.U Anantapur, Tirupathi, Andhra Pradesh, India, Email-hesh100.ma@gmail.com

[2] Departments of CSE, J.N.T.U Anantapur, Hyderabad, Andhra Pradesh, India, Email-munisekhar.prudhvi@gmail.com

Abstract

During this paper, for the first time, we tend to outline and solve the difficult downside of privacy preserving multi-keyword hierarchic search over encrypted cloud data (MRSE).We establish a collection of strict privacy needs for such a secure cloud knowledge utilization system. Among numerous multi-keyword semantics, we decide the economical similarity live of "coordinate matching", i.e., as several matches as potential, to capture the connection of knowledge documents to the search question. We further use "inner product similarity" to quantitatively value such similarity live. we tend to initial propose a basic plan for the MRSE supported secure scalar product computation, then provide two considerably improved MRSE schemes to realize numerous stringent privacy needs in 2 completely different threat models. Thorough analysis work privacy and potency guarantees of planned schemes are given. Experiments on the real-world dataset additional show planned schemes so introduce low overhead on computation and communication.

Keywords: *Trapdoor, Query Processing, MRSE, Keyword, Ranking, Cloud.*

1. Introduction

Cloud computing is that the long unreal vision of computing as a utility, wherever cloud customers will remotely store their

data into the cloud thus on relish the on-demand top quality applications and services from a shared pool of configurable computing resources [1].It's nice flexibility and economic savings area unit motivating each people and enterprises to outsource their native complicated knowledge management system into the cloud. to guard knowledge privacy and combat uninvited accesses within the cloud and on the far side, sensitive knowledge, e.g., emails, personal health records, icon albums, tax documents, financial transactions, etc., might need to be encrypted by knowledge owners before outsourcing to the industrial public cloud [2]; this, however, obsoletes the standard knowledge utilization service based on plaintext keyword search. The trivial resolution of downloading all the information and decrypting domestically is clearly impractical, thanks to the massive quantity of information measure value in cloud scale systems. Moreover, except for eliminating the local storage management, storing knowledge into the cloud serves no purpose unless they'll be simply searched and used. Thus, exploring privacy-preserving and effective search service over encrypted cloud knowledge is of predominant importance. Considering the potentially sizable amount of on-demand knowledge users and huge quantity of outsourced knowledge documents within the cloud, this problem is especially difficult because it is very tough to meet conjointly the necessities of performance, system usability and quantifiability.

1.1 Existing System

The large range of knowledge users and documents in cloud, it's crucial for the search service to permit multi-keyword question and supply result similarity ranking to satisfy the effective information retrieval want. The searchable cryptography focuses on single keyword search or Boolean keyword search, and barely differentiates the search results. To meet the effective knowledge retrieval want, the massive quantity of documents demand the cloud server to perform result connectedness ranking, rather than returning uniform results. Such stratified search system permits knowledge users to search out the foremost relevant data quickly, instead of burdensomely sorting through each match within the content assortment [5]. Stratified search can even elegantly eliminate excess network traffic by causation back solely the foremost relevant knowledge that is extremely fascinating within the "pay-as-you-use" cloud paradigm. For privacy protection, such ranking operation, however, shouldn't leak any keyword connected data. On the opposite hand, to boost the search result accuracy additionally on enhance the user looking expertise, it's conjointly necessary for such ranking system to support multiple keywords search, as single keyword search typically yields way too coarse results. As a typical follow indicated by today's internet search engines (e.g., Google search), knowledge users could tend to produce a collection of keywords rather than only one because the indicator of their search interest to retrieve the foremost relevant knowledge. And every keyword within the search request is in a position to assist slim down the search result more. “Coordinate matching” [6], i.e., as several matches as attainable, is associate economical similarity live among such multi-keyword linguistics to refine the result connectedness, and has been wide employed in the plaintext data retrieval (IR) community. However, the way to apply it within the encrypted cloud knowledge search system remains a really difficult task thanks to inherent security and privacy obstacles, together with varied strict needs just like the knowledge privacy, the index privacy, the keyword privacy, and lots of others.

1.1 Disadvantages of Existing System

- Single-keyword search while not ranking.
- Boolean- keyword search while not ranking.
- Single-keyword search with ranking.
- It still not up to offer users with acceptable result ranking practicality.
- It cannot accommodate such high service-level necessities like system usability, user looking expertise, and straightforward data discovery.
- Shared information won't be secure.

2. Proposed System

We outline and solve the difficult downside of privacy-preserving multi-keyword graded search over encrypted cloud knowledge (MRSE), and establish a collection of strict privacy necessities for such a secure cloud knowledge utilization system to become a reality. Among varied multi-keyword linguistics, we decide the economical principle of “coordinate matching”.

In this paper, for the primary time, we tend to outline and solve the matter of multi-keyword stratified search over encrypted cloud information (MRSE) whereas conserving strict system wise privacy within the cloud computing paradigm. Among numerous multi-keyword linguistics, we elect the economical similarity live of “coordinate matching,” i.e., as several matches as potential, to capture the connectedness of information documents to the search question. Specifically, we tend to use “inner product similarity”, i.e., the amount of question keywords showing in a very document, to quantitatively measure such similarity live of that document to the search question. throughout the index construction, every document is related to a binary vector as a sub index wherever every bit represents whether or not corresponding keyword is contained within the document. The search question is additionally delineate as a binary vector wherever every bit means that whether or not corresponding keyword seems in this search request, that the similarity may be precisely measured by the real number of the question vector with the info vector. However, directly outsourcing the info vector or the question vector can violate the index privacy or the search privacy. to satisfy the challenge of supporting such multikeyword linguistics while not privacy breaches, we tend to propose a basic plan for the MRSE victimization secure real number computation, that is customized from a secure k-nearest neighbor (kNN) technique, so provide two considerably improved MRSE schemes in a very bit-by-bit manner to realize numerous demanding privacy necessities in two threat models with enhanced attack capabilities.

2.1 Advantages of Proposed System

- Multi-keyword ranked search over encrypted cloud data (MRSE)
- “Coordinate Matching” by inner product similarity.
- It projected schemes so introduce low overhead on computation and communication.
- It uses stratified search mechanism to support a lot of search linguistics and dynamic information operations.
- It is safer and economical.

2.2 System Architecture

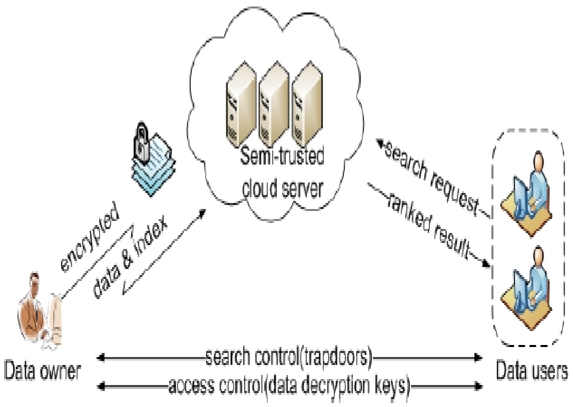

Semi-trusted cloud server
encrypted data & index
search request
ranked result
Data owner
search control(trapdoors)
access control(data decryption keys)
Data users

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

2.3 Contributions

Our contributions area unit summarized as follows,

1. For the primary time, we tend to explore the matter of multi-keyword ranked search over encrypted cloud knowledge, and establish a collection of strict privacy necessities for such a secure cloud knowledge utilization system.
2. We tend to propose two MRSE schemes supported the similarity measure of "coordinate matching" whereas meeting completely different privacy necessities in two completely different threat models.
3. Thorough analysis work privacy and potency guarantees of the projected schemes is given, and experiments on the real-world dataset any show the proposed schemes so introduce low overhead on computation and communication.

2.4 Design Goals

To change stratified seek for effective utilization of outsourced cloud information below the same model, our system style ought to at the same time win security and performance guarantees as follows.

Multi-keyword stratified Search: to style search schemes which permit multi-keyword question and supply result similarity ranking for effective information retrieval, instead of returning uniform results.

Privacy-Preserving: To forestall the cloud server from learning extra info from the dataset and therefore the index and to satisfy privacy needs per section III-B.

Efficiency: higher than goals on practicality and privacy should be achieved with low communication and computation overhead.

3. Related Work

3.1 Single Keyword Searchable Encryption

Traditional single keyword searchable encoding schemes [5] sometimes build associate degree encrypted searchable index such its content is hidden to the server unless it's given applicable trapdoors generated via secret key(s) [2].It is initial studied by Song et al. [5] within the interchangeable key setting, and enhancements and advanced security definitions are given in Goh [6], Yangtze River et al. [7] and Curtmola et al. [8].Our early work solves secure hierarchical keyword search which utilizes keyword frequency to rank results rather than returning dedifferentiated results. However, it solely supports single keyword search. within the public key setting, Boneh et al. [9] gift the primary searchable encoding construction, where anyone with public key will write to the info hold on server however solely approved users with personal key will search. Public key solutions square measure sometimes terribly computationally expensive but. Moreover, the keyword privacy may not be protected within the public key setting since server may encrypt any keyword with public key then use the received trapdoor to judge this ciphertext.

3.2 Boolean Keyword Searchable Encryption

To enrich search functionalities, conjunctive keyword search over encrypted knowledge are planned. These schemes incur giant overhead caused by their basic primitives, like computation value by additive map, e.g. [6], or communication value by secret sharing, e.g. [5]. As a lot of general search approach, predicate encoding schemes are recently planned to support both conjunctive and dividing search. Conjunctive keyword search returns "all-or-nothing", which suggests it solely returns those documents within which all the keywords fixed by the search question appear; dividing keyword search returns uniform results, which suggests it returns each document that contains a set of the precise keywords, even just one keyword of interest. In short, none of existing Boolean keyword searchable encoding schemes support multiple keywords ranked search over encrypted cloud knowledge whereas conserving privacy as we tend to propose to explore during this paper. Note that, inner product queries in predicate encoding solely predicate whether 2 vectors ar orthogonal or not, i.e., the inner product price is hid except once it equals zero. Without providing the potential to check hid inner merchandise, predicate encoding isn't qualified for activity hierarchal search. Moreover, most of those schemes are engineered upon the expensive analysis of pairing operations on elliptic curves. Such unskillfulness disadvantage conjointly limits their sensible performance when deployed within the cloud. On a special front, the analysis on top-k retrieval [7] in information community is also loosely connected to our downside.

4. Literature Survey

4.1 Study about A break in the clouds: towards a cloud definition

Cloud Computing is related to a brand new paradigm for the provision of computing infrastructure. This paradigm shifts the placement of this infrastructure to the network to reduce the prices related to the management of hardware and package resources [6]. The Cloud is drawing the attention from the knowledge and Communication Technology (ICT) community, because of the looks of a group of services with common characteristics, provided by necessary industry players. However, a number of the present technologies the Cloud construct attracts on (such as virtualization, utility computing or distributed computing) aren't new. The variety of technologies within the Cloud makes the picture confusing [8]. Moreover, the ballyhoo around Cloud computing any muddies the message. Of course, the Cloud isn't the primary technology that falls into ballyhoo. Gartner's ballyhoo Cycle [2] characterizes however the ballyhoo regarding a technology evolves "from over enthusiasm through a amount of edification to Associate in Nursing ultimate understanding of the technology relevance and role in a very market or domain".

4.2 Study about Cryptographic cloud storage

Advances in networking technology and a rise within the want for computing resources have prompted several organizations to source their storage and computing wants. This new economic and computing model is commonly observed as cloud computing and includes numerous sorts of services such as: infrastructure as a service(IaaS), wherever a client makes use of a service provider's computing, storage or networking infrastructure; platform as a service (PaaS), wherever a client leverages the provider's resources to run custom applications; and finally code as a service (SaaS), wherever customers use code that's run on the provider's infrastructure. Cloud infrastructures are roughly categorized as either non-public or public. During a non-public cloud, the infrastructure is managed and in hand by the client and placed on-premise (i.e., within the customer's region of control). In particular, this suggests that access to client knowledge is below its management and is simply granted to parties it trusts. In a public cloud the infrastructure is in hand and managed by a cloud service supplier and is found off-premise (i.e., in the cloud service provider's region of control). This suggests that client knowledge is outside its management and will potentially be granted to untrusted parties. Storage services supported public clouds like Microsoft's Azure storage service and Amazon's S3 give customers with climbable and dynamic storage. By moving their knowledge to the cloud customers will avoid the prices of building and maintaining a non-public storage infrastructure, opting instead to pay a service supplier as a operate of its needs. For many customers, this provides many edges together with availableness (i.e., having the ability to access knowledge from anywhere) and reliableness (i.e., not having to fret concerning backups) at a comparatively low price.

5. Simulated Result

We demonstrate a simulated experimental evaluation of the projected technique on a real-world dataset:

5.1 Trapdoor Generation

Fig. 2(a) shows that the time to generate a trapdoor is greatly plagued by the quantity of keywords within the wordbook. Like index construction, every trapdoor generation incurs 2 multiplications of a matrix and a split question vector, wherever the spatiality of matrix or question vector is completely different in 2 planned schemes and becomes larger with the increasing size of wordbook. Fig. 2(b) demonstrates the trapdoor generation value within the MRSE II scheme is regarding twenty percentages larger than that within the MRSE I scheme. just like the sub index generation, the distinction of prices to generate trapdoors is major ally caused by the various dimensionality of vector and matrices within the 2 MRSE schemes. a lot of significantly, it shows that the quantity of question keywords has very little influence on the overhead of trapdoor generation, that may be a important advantage over connected works on multi-keyword searchable encoding.

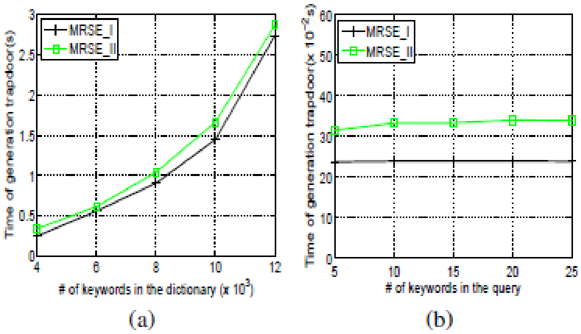


Fig. 2: Time price of generating trapdoor. (a) For a similar query keywords among totally different sizes of lexicon, t = 10. (b) For various numbers of question keywords among a similar dictionary, n = 4000.

5.2 Query

Query execution within the cloud server consists of computing and ranking similarity scores for all documents in the dataset. Fig. 3 shows the question time is dominated by the number of documents within the dataset whereas the quantity of keywords within the question has terribly slight impact thereon like the cost of trapdoor generation higher than. With relevance the communication price in question, the dimensions of the trapdoor is the same as that of the sub index listed within the Tab. 1, which keeps constant given a similar wordbook, regardless of what percentage keywords square measure contained in a very question. Whereas the computation and communication price within the question.

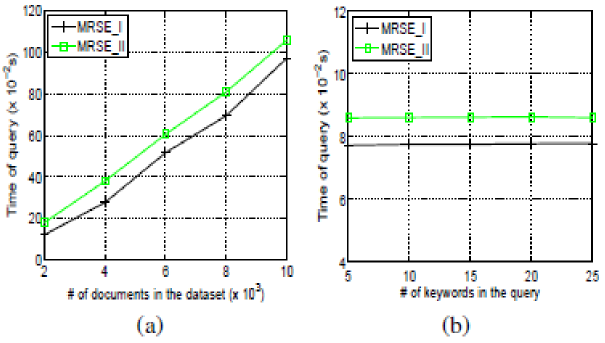
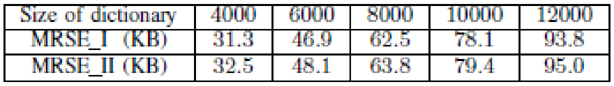


Fig. 3: Time price of question. (a) For an equivalent question keywords in different sizes of dataset, t = 10. (b) For various numbers of question keywords within the same dataset, m = 1000.

Table 1: Size of sub index/trapdoor



Size of dictionary	4000	6000	8000	10000	12000
MRSE_I (KB)	31.3	46.9	62.5	78.1	93.8
MRSE_II (KB)	32.5	48.1	63.8	79.4	95.0

6. Conclusion

In this paper, for the primary time we tend to outline and solve the problem of multi-keyword stratified search over encrypted cloud data, and establish a range of privacy necessities. Among various multi-keyword linguistics, we elect the economical similarity measure of “coordinate matching”, i.e., as several matches as doable, to effectively capture the connection of outsourced documents to the question keywords, and use “inner product similarity” to quantitatively judge such similarity live. For meeting the challenge of supporting multi-keyword linguistics without privacy breaches, we tend to propose a basic plan of MRSE using secure scalar product computation. Then we tend to provide 2 improved MRSE schemes to attain varied tight privacy requirements in 2 completely different threat models. Thorough analysis investigating privacy and potency guarantees of planned schemes is given, and experiments on the real-world dataset show our planned schemes introduce low overhead on each computation and communication.

7. Future Work

In future work, we'll explore supporting different multikeyword semantics (e.g., weighted query) over encrypted knowledge and checking the integrity of the ordering within the search result.

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.

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

[10] M. Bellare, A. Boldyreva, and A. ONeill, “Deterministic and efficiently searchable encryption,” in *Proc. of CRYPTO*, 2007.

B.Mahesh received the B.Tech Degree in Computer Science and Engineering from YITS college of Engineering, University of JNTUA in 2012.He is currently working towards the Master’s Degree in Computer Science and Engineering, in Shree Institute of Technology & Sciences University of JNTUA. She interest lies in the areas of Web Development Platforms, SQL, and Cloud Computing Technology.

P.Munisekhar Received M.Tech in Computer Science and Engineering from JNTUA University. Currently he is an Assistant Professor in the Department of Computer Science and Engineering at Shree Institute of Technology & Sciences-Tirupati.