International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.3, No.12, December 2014 DOI:10.15693/ijaist/2014.v3i12.24-27

Survey on multi keyword ranked Search for cloud data

BalaGanesh.M
Associate Professor/Department of CSE
Anna University, Chennai, India

Karthik Elango.R.S
PG Scholar/Department of CSE
Anna University, Chennai

***Abstract*— With the arrival of cloud computing, knowledge homeowners are driven to source their complicated knowledge management systems from native sites to the industrial public cloud for excellent flexibility and economic savings. Except for protective knowledge privacy, sensitive knowledge needs to be encrypted before outsourcing that obsoletes ancient knowledge utilization supported plaintext keyword search. Thus, enabling AN encrypted cloud knowledge search service is of predominant importance. Considering the massive range of information users and documents within the cloud, it's necessary to permit multiple keywords within the search request and documents within the order of their relevancy to those keywords. Connected works on searchable encoding specialize in single keyword search or mathematician keyword search, and infrequently kind the search results. A tendency to outline and solve the difficult drawback of privacy preserving multi-keyword hierarchal search over encrypted cloud knowledge (MRSE).**

Index terms - Data privacy, cloud service provider(csp), cloud data.

I. INTRODUCTION

Cloud Computing is associated with a new paradigm for the provision of computing infrastructure. This paradigm shifts the location of this infrastructure to the network to reduce the costs associated with the management of hardware and software resources. The Cloud is drawing the attention from the Information and Communication Technology (ICT) community[1], thanks to the appearance of a set of services with common characteristics, provided by important industry players. However, some of the existing technologies the Cloud concept draws on (such as virtualization, utility computing or distributed computing) are not new The variety of technologies in the Cloud makes the over-all picture confusing. Moreover, the hype around Cloud Computing further muddies the message. Of course, the Cloud is not the first technology that falls into hype. Gartner's Hype Cycle characterizes how the hype about a technology evolves "from over enthusiasm through a period of disillusionment to an eventual understanding of the technology relevance and role in a market or domain". Arguably, Cloud Computing is now in the first stage of this hype cycle, labeled as 'Positive Hype'[2].This reinforces the overall confusion about the paradigm and its capacities, turning the Cloud into an excessively general term that includes almost any solution that allows the outsourcing of all kinds of hosting and computing resources. Yet, the notions of transparent access to resources on a pay-per-use basis, relying on an infinitely and instantly scalable infrastructure managed by a third-party, are a recurrent idea.

The example of what has happened with the Grid illustrates the need of a crisp definition for Clouds: although there are well-known Grid definitions (probably Foster's is the most widely accepted), none of them are widely accepted. A clear Grid definition may have helped to disseminate what the term 'Grid' actually means and what business benefits can be obtained from it. Thus, it is important to find a unified definition of what Cloud Computing is, delimiting the scope of research and emphasizing the potential business benefits. There are many definitions of Cloud Computing, but they all seem to focus on just certain aspects of the technology. This paper tries to give a more comprehensive analysis of all the features of Cloud Computing, to reach a definition that encompasses them. This paper proceeds as follows. First, we present an overview of the Cloud scenario. Analyzes present Cloud definitions, extracting relevant Cloud features and combining them to form both an integrative and a basic Cloud definition. we present the different approaches of grids and Clouds to clearly distinguish these two technology.

II. RELATED WORK

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 encouraging both individuals and enterprises to outsource their local complex data management system into the cloud. To protect data privacy and combat voluntary accesses in the 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 the 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 eliminate 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 the cloud, this problem is particularly challenging as it is extremely tricky to meet also the requirements of performance, system usability and scalability. On the one hand, to meet the effective data retrieval[11] need, the large amount of documents demand the cloud server to perform result consequence ranking, instead of returning undifferentiated results. 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 related data, which is highly desirable in the "pay-as-you use" cloud paradigm. 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. 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 recover 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 similarity measure among such multi-keyword semantics to refine the result relevance, and has been widely used in the plaintext information recuperation (IR)[11] community. However, how to apply it in the encrypted cloud data search system remains a very demanding task because of inbuilt security and privacy obstacles, including various strict necessities like the data privacy, the index privacy, the keyword privacy, and many others.

In the literature, searchable encryption [5][12] is a helpful technique that treats encrypted data as documents and allows a user to securely search through a single keyword and retrieve documents of interest. However, direct application of these approaches to the secluded large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot accommodate. 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 the cloud computing paradigm. Among various multi keyword semantics, we choose the qualified similarity measure of "coordinate matching", i.e., as many matches as possible, to capture the application of data documents to the search query. Specifically, we use "inner product similarity" [4], i.e., the number of query keywords appear in a document, to quantitatively evaluate such resemblance measure of that document to the search query. During the index construction, each document is associated with a binary vector as a sub index where each bit represents whether consequent 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 the inner formulate of the query vector with the data vector. However, directly outsourcing the data vector or the query vector will violate the index privacy or the search privacy. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, we propose a basic idea for the MRSE using secure inner product computation, which is made to order from a secure k-nearest neighbor (kNN) technique [6], and then give two significantly improved MRSE schemes in a step-by-step manner to achieve various inflexible privacy requirements in two threat models with increased attack capabilities. We introduce the system model, the threat model, our design goals, and the preliminary. bearing in mind a cloud data hosting service involving three different entities, 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 effectual 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. To search the document collection for t given keywords, an authorized user acquires a corresponding trapdoor T through search control mechanisms, e.g., broadcast encryption [8][12]. Upon receiving T from a data user, the cloud server is responsible to search the index I and return the corresponding set of encrypted documents. To improve the document recovery accuracy, the search result should be ranked by the cloud server according to some ranking criteria (e.g., coordinate matching, as will be introduced shortly). 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 appropriate to the search query. Finally, the access control mechanism [7] is employed to manage decryption capabilities given to users. Within the ranked search, the access pattern is the sequence of search results where every search result is a set of credentials with rank order. Specifically, the search result for the query keyword set fW is denoted as FfW, consisting of the id list of all documents ranked by their relevance to fW. Then the access pattern is denoted as (FfW1 ,FfW2 , . . .) which are the results of sequential searches. Although a few searchable encryption works[12], e.g., has been proposed to utilize private information retrieval (PIR) technique , to hide the access pattern, our proposed schemes are not designed to protect the access pattern for the efficiency concerns. This is because any PIR based technique must

"touch" the whole dataset outsourced on the server which is inefficient in the large scale cloud system.

A. Efficiency improvements based on anonymity

We now turn the attention to the question of efficiency, attempting to identify natural cryptographic tasks for which anonymity can give rise to substantial efficiency gains. In contrast to the feasibility results discussed above, here we do not restrict ourselves to unconditional results. In particular, we would like to improve over the best known solutions under any cryptographic assumption [10].

A key observation, that underlies our protocols in this setting, is that local randomization of inputs, via secret-sharing, when combined with the global mixing of the shares, provided by anonymity, allows us to keep the inputs private and, at the same time, allows us to carry out some useful computations on the inputs. we elaborate below.

B. Non-interactive Private Statistics

We show how n ¸ 2 clients can privately compute statistics (such as mean, standard deviation, correlations) on their combined inputs by each sending few anonymous messages to a central server. Our protocols only require one-way anonymous communication and are private with respect to an adversary corrupting the server along with an arbitrary number of clients.5 Note that it is impossible to obtain such non-interactive protocols in the standard model, even if one settles for computational privacy. (It is possible to solve this problem in the non-interactive model of however, such a solution requires setup assumptions and provides a weaker security guarantee)

Achieving secure, scalable, and fine-grained data access control in cloud computing S. Yu, C. Wang, K. Ren, and W. Lou[1] Cloud computing is an emerging computing paradigm in which resources of the computing infrastructure are provided as services over the Internet. As promising as it is, this paradigm also brings forth many new challenges for data security and access control when users outsource sensitive data for sharing on cloud servers, which are not within the same trusted domain as data owners. To keep sensitive user data confidential against untrusted servers, existing solutions usually apply cryptographic methods by disclosing data decryption keys only to authorized users. However, in doing so, these solutions inevitably introduce a heavy computation overhead on the data owner for key distribution and data management when fine- grained data access control is desired, and thus do not scale well. The problem of simultaneously achieving fine-graininess, scalability, and data confidentiality of access control actually still remains unresolved. This paper addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine- grained data access control to untrusted cloud servers without disclosing the underlying data contents.

Zerber: r-Confidential Indexing for Distributed Documents Sergej Zerr1 Elena Demidova, Daniel Olmedilla1 Wolfgang Nejdl1 Marianne Winslett2 and Soumyadeb Mitra2 [2]To carry out work assignments, little teams distributed at intervals a bigger enterprise usually have to be compelled to share documents among themselves whereas shielding those documents from others\' eyes. during this state of affairs, users would like associate degree classification facility that may quickly find relevant documents that they\'re allowed to access, while not (1) leaky info concerning the remaining documents, (2) imposing an outsized management burden as users, groups, and documents evolve, or (3) requiring users to agree on a central fully sure authority. to deal with this downside, we tend to propose the conception of r-confidentiality, that captures the degree of knowledge escape from associate degree index concerning the terms contained in inaccessible documents. Then we tend to propose the r-confidential Zerber classification facility for sensitive documents, that uses secret ripping and term merging to produce tunable limits on info escape, even underneath applied math attacks; needs solely restricted trust in an exceedingly central classification authority; and is extraordinarily straightforward to use and administer. Experiments with real-world knowledge show that Zerber offers glorious performance for index insertions and lookups whereas requiring solely a modest quantity of cupboard space and network information measure.

C. Wang, Q. Wang, K. Ren, and W. Lou. "Privacy-preserving public auditing for data storage security in cloud computing" [3] Using Cloud Storage, users will remotely store their information Associate in Nursing d get pleasure from the on-demand prime quality applications and services from a shared pool of configurable computing resources, while not the burden of native information storage and maintenance. However, the very fact that users not have physical possess particle of the outsourced information makes the information integrity protection in Cloud Computing a formidable task, particularly for users with forced computing resources. Moreover, users ought to be ready to simply use the cloud storage as if it's native, while not worry in anogram regarding the requirement to verify its integrity. Thus, facultative tavern lic audit ability for cloud storage is of vital importance in order that users will resort to a 3rd party auditor (TPA) to examine the integrity o f outsourced information and be worry-free. To firmly introduce a good T PA, the auditing method ought to herald no new vulnerabilities towards user information privacy, and introduce no extra on-line burden to user. during this paper, we tend to propose a secure cloud storage system supporting privacy-preserving public auditing. we tend to furthe r extend our result to change the TPA to perform audits for multiple users at the same time and with

efficiency. intensive security and performance analysis show the planned schemes are demonstrably se cure and extremely economical.

III. Existing System

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[11] need. The searchable encryption focuses on single keyword search or Boolean keyword search, and rarely differentiates the search results.

A problem of Existing system is Single-keyword search without ranking, Boolean- keyword search without ranking and Single-keyword search with ranking.

VI. Conclusion

A survey on MRSE, for the first time 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, to 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, to propose a basic idea of MRSE using secure inner product computation.

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: Compress-ing 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] W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in *Proceedings of the 35th SIGMOD international conference on Management of data*, 2009, pp. 139–152.

[7] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving secure, scalable, and fine-grained data access control in cloud computing," in *Proc. of INFOCOM*, 2010.

[8] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in *Proc. of INFOCOM*, 2010.

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

[10] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Cryptography from anonymity," in *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, 2006, pp. 239–248.

[11] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k retrieval from a confidential index," in *Proc. of EDBT*, 2009, pp. 439–449.

[12] Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, "Fully secure functional encryption: Attribute-based encryption and (hierarchical)inner product encryption," in Proc. of EUROCRYPT, 2010.

Authors Profile

R.S.KarthikElangoreceived the **B.E.** degree in computer science and engineering from the A.R.J College of Engineering, Mannargudi, Anna University, Chennai, India, in 2012.Currently doing **M.E.** in computer science and engineering in Anna University of technology, Chennai, India. His journal interest includes survey on multi keyword ranked search for cloud data. MRSE algorithm provides Data privacy, Data security and Cloud Service Provider (CSP) using encrypted data secure authentications.