INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

A SECURE MULTI-KEYWORD ENCRYPTED CLOUD DATA FOR RANKED SEARCH

Syed Firdose[1]

[1]*M.Tech Student Department of CSE, Madina Engineering College, Kadapa, Andhra Pradesh, India*

Email:sayedfirdoseit@gmail.com[1]

Abstract: - The importance of cloud computing, the complex data management systems are motivated to cloud storage from a local systems it has great flexibility and economic savings on public cloud. But the privacy of public cloud is very low because of the data will be hacked by considering these we implement a new approach MERS (Multi-keyword Encrypted Ranked Search).By this technique we maintain the privacy data and different number of keyword search along with ranked. The consideration of the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Similar works on searchable encryption it focuses on single keyword search or Boolean keyword search, and rarely sort the search results. We establish a set of strict security requirements for such a secure cloud data utilization system. Among various multi-keyword synonymous, we consider the efficient measure of "matching-coordination," i.e., as many matches as possible, to capture the relevance of data documents to the search query and the use of "inner product possibility" to evaluate s similarity measure. For improved search experience of the data search service, extend these two schemes to support more search semantics. By the investigation of analysis privacy and efficiency guarantees of the proposed schemes is given. The real world data experimental data sets schemas indeed found low computation and communication.

Keywords: Cloud computing, SSE, privacy-preserving, keyword ranked search, KNN technique

1. Introduction

Cloud computing is vital technology for industrial and economical and educational and commercial growth in every technology [7] .Generally cloud is a Meta data of internet cloud computing. Cloud computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources great flexibility and economic savings are motivating both individuals and enterprises to outsource their local complex data management system into the cloud. To protect data privacy and combat unsolicited accesses in the cloud and beyond, sensitive data, for example, e mails, personal health records, photo albums, tax documents, financial transactions, and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The trivial solution of downloading all the data and decrypting locally is clearly impractical, due to the huge amount of bandwidth cost in cloud scale systems. Moreover, aside from eliminating

the local storage management, storing data into the cloud serves no purpose unless they can be easily searched and utilized. Thus, exploring privacy-preserving and effective search service over encrypted cloud data is of paramount importance.

Considering the potentially large number of on-demand data users and huge amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability. On the one hand, to meet the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance 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. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-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 considering these we implement a new approach MERS (Multi-keyword Encrypted Ranked Search)[1].By this technique we maintain the privacy data and different number of keyword search along with ranked. The consideration of the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Similar works on searchable encryption it focuses on single keyword search or Boolean keyword search, and rarely sort the search results. We establish a set of strict security requirements for such a secure cloud data utilization system. Among various multi-keyword synonymous, we consider the efficient measure of "matching-coordination," i.e., as many matches as possible, to capture the relevance of data documents to the search query and the use of "inner product possibility" to evaluate s similarity measure. For improved search experience of the data search service, extend these two schemes to support more search semantics. By the investigation of analysis privacy and efficiency guarantees of the proposed schemes is given. The real world data experimental data sets schemas indeed found low computation and communication.

2. Proposed System:

In proposed system we design the MERS with searchable encryption, privacy-preserving, keyword ranked search, KNN technique

2.1 Cloud computing:

Cloud computing is an umbrella term used to refer to Internet based development and services. The cloud is a metaphor for the Internet. A number of characteristics define cloud data, applications services and infrastructure:

- Remotely hosted: Services or data are hosted on someone else's infrastructure.
- Ubiquitous: Services or data are available from anywhere.
- Commoditized: The result is a utility computing model similar to traditional that of traditional utilities, like gas and electricity. You pay for what you would like.
- Over time many big Internet based companies (Amazon, Google…) have come to realize that only a small amount of their data storage capacity is being used. This has led to the renting out of space and the storage of information on remote servers or "clouds". Information is then temporarily cached on desktop computers, mobile phones or other internet-linked devices. Amazon's Amazon Elastic Compute Cloud (EC2) and Simple Storage Solution (S3) are the current best known facilities.
- Cloud Services can also be used to hold structured data. There has been some discussion of this being a potentially useful notion possibly aligned with the Semantic Web, though concerns, such as this resulting in data becoming undifferentiated, have been raised.

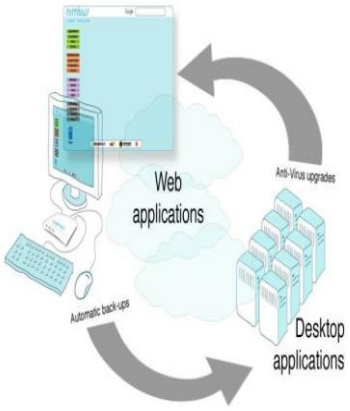

Figure 1: software deployment model

- SaaS is a model of software deployment where an application is hosted as a service provided to customers across the Internet as shown in figure 1. SaaS is generally used to refer to business software rather than consumer software, which falls under Web 2.0. By removing the need to install and run an application on a user's own computer it is seen as a way for businesses to get the same benefits as commercial software with smaller cost outlay. Saas also alleviates the burden of software maintenance and support but users relinquish control over software versions and requirements. The other terms that are used in this sphere include Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

2.2 SSE:

Searchable symmetric encryption (SSE) allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it[10]. This problem has been the focus of active research and several security definitions and constructions have been proposed as shown in figure 2. In this paper we review existing security definitions, pointing out their short-comings, and propose two new stronger definitions which we prove equivalent [3]. We then present two Constructions that we show secure under our new definitions. Interestingly, in addition to satisfying stronger security guarantees, our constructions are more efficient than all previous constructions. Further, prior work on SSE only considered the setting where only the owner of the data is capable of submitting **s**earch queries

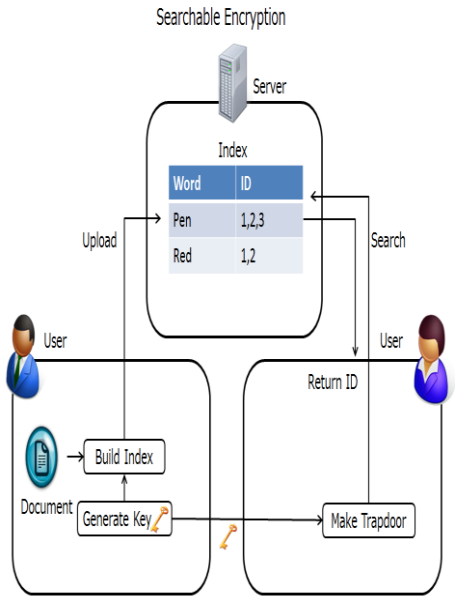

Figure 2: Active research and several securities

We consider the natural extension where an arbitrary group of parties other than the owner can submit search queries. We formally define SSE in this multi-user setting, and present an efficient construction.

3. Privacy-Preserving:

It is the long dreamed vision of computing as a utility, where users 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. By data outsourcing, users can be relieved from the burden of local data storage and maintenance. Thus, enabling public audit ability for cloud data storage security is of critical importance so that users can resort to an external audit party to check the integrity of outsourced data when needed. To securely introduce an effective third party auditor (TPA), the following two fundamental requirements have to be met: [1] TPA should be able to efficiently audit the cloud data storage without demanding the local copy of data, and introduce no additional on-line burden to the cloud user. Specifically, our contribution in this work can be summarized as the following three aspects, as shown in figure 3 below.

1) We motivate the public auditing system of data storage security in Cloud Computing and provide a privacy-preserving auditing protocol[2]l, i.e., our scheme supports an external auditor to audit user's outsourced data in the cloud without learning knowledge on the data content.

2) To the best of our knowledge, our scheme is the first to support scalable and efficient public auditing in the Cloud Computing. In particular, our scheme achieves batch auditing where multiple delegated auditing tasks from different users can be performed simultaneously by the TPA.

3) We prove the security and justify the performance of our proposed schemes through concrete experiments and comparisons with the state-of-the-art.

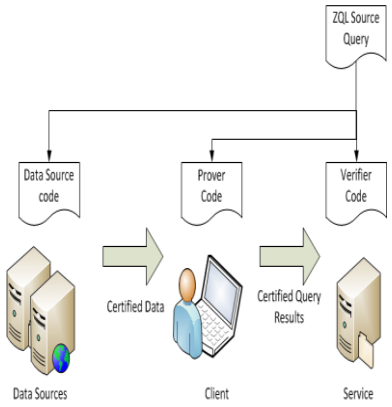

Figure 3: Data storage security in Cloud Computing

To enable privacy-preserving public auditing for cloud data storage under the aforementioned model, our protocol design should achieve the following security and performance guarantee [5].

3.1 Public audit ability: to allow TPA to verify the correctness of the cloud data on demand without retrieving a copy of the whole data or introducing additional on-line burden to the cloud users.

3.2 Storage correctness: to ensure that there exist no cheating cloud servers that can pass the audit from TPA without indeed storing users' data intact.

3.3 Privacy-preserving: to ensure that there exists no way for TPA to derive users' data content from the information collected during the auditing process [9].

3.4 ***Batch auditing:*** to enable TPA with secure and efficient auditing capability to cope with multiple auditing delegations from possibly large number of different users simultaneously.

3.5 Lightweight: to allow TPA to perform auditing with minimum communication and computation overhead.

4. Keyword Ranked Search:

Most of the data owners and the web service providers uses the Cloud technology for store and share their data. Because the cloud computing will provide an efficient way to Store and Share the data without compromising the privacy. Security is enforced by encrypting the data. Encryption will complicate some basic properties of data set such as search over the data. In earlier the searching is start done by exact query matching

where the searching is done for a given search word inside the unencrypted document [12]. Then searchable encryptions are developed for search over encrypted data [12]. The index based scheme will solve the problem of security and provides the search over encrypted data. But the ranked search based on frequency of search key word in document challenges the security even if we are using the index based scheme.

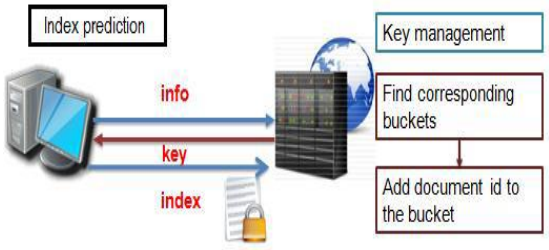

Figure 4: The multi-server secure scheme

The multi-server secure scheme in which the data is divided into leaked information and data set and are separately stored in different servers will give extra security, as shown in figure 4. But the servers must never integrate together. We cannot promise this over a long period of time. The continuous monitoring to the document ID, URL and the document indexes allows the hackers to break the system.

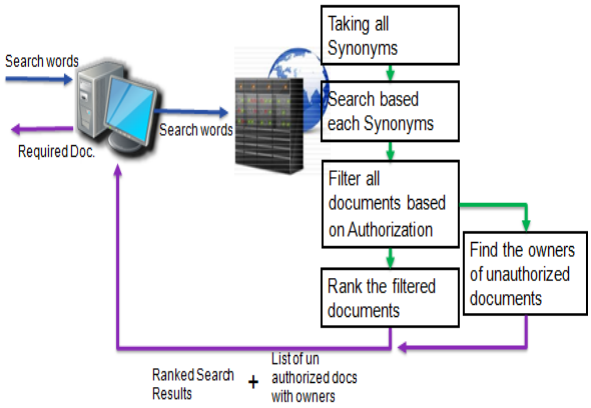

Figure 5: Fast data searching scheme

Here we propose an efficient, secure and fast data searching scheme which will also capable of handling efficient data management in cloud servers as explain in the figure 5. The index scheme is used to enable fast search while ensuring the security. Ranking of the document for displaying the document will improve the efficiency.

Figure 6: User-unaware-key-management

The efficiency of searching can also be improved by synonym based search. The security can be enhanced by encrypting each of the documents with different keys [11]. By the use of user-unaware-key-management we can improve confidentiality.

5. KNN Technique:

The *k*-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbors, with the object being assigned the class most common amongst its *k* nearest neighbors. *k* is a positive integer, typically small. If *k* = 1, then the object is simply assigned the class of its nearest neighbor. In binary (two class) classification problems, it is helpful to choose *k* to be an odd number as this avoids difficulties with tied votes, as shown in the figure 7.

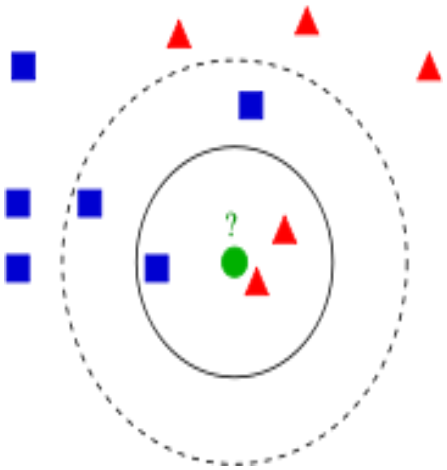

Figure 7: The *k*-nearest neighbor Technique

The same method can be used for regression, by simply assigning the property value for the object to be the average of the values of its *k* nearest neighbors. It can be useful to weight the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones.

The neighbors are taken from a set of objects for which the correct classification (or, in the case of regression, the value of the property) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required. In order to identify neighbors, the objects are represented by position vectors in a multidimensional feature space. It is usual to use the Euclidean distance, though other distance measures, such as the Manhattan distance could in principle be used instead. The *k*-nearest neighbor algorithm is sensitive to the local structure of the data.

6. Result Analysis:

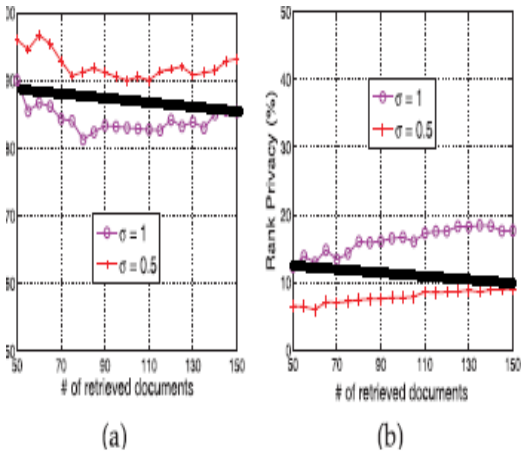

Figure 8: With different choice of standard deviation for the random variable ", there exists tradeoff between (a) Precision, and (b) Rank Privacy

Figure 8 shows that the precision in MRSE scheme is evidently affected by the standard deviation of the random variable". From the consideration of effectiveness, standard deviation is expected to be smaller so as to obtain high precision indicating the good purity of retrieved documents.

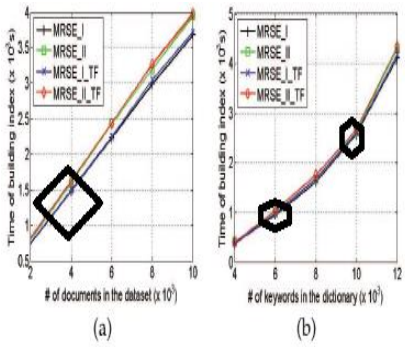

Figure 9: Time cost of building index. (a) For the different size of data set With the same dictionary, n¼4; 000. (b) For the same data set with different size of dictionary, m¼1; 000

Figure 9 shows that the number of keywords indexed in the dictionary determines the time cost of building a sub-index.

7. Conclusion:

In this paper, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient similarity measure of " matching-coordination," 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 MERS using secure inner product computation.

8. Future Enhancement:

In our future work, for un-trusted cloud servers we improvise the integrity in rank order search result.

9. References:

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi-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, "A Break 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-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.
[4] A. Singhal, "Modern Information Retrieval: A Brief Overview," IEEE Data Eng. Bull., vol. 24, no. 4, pp. 35-43, Mar. 2001.
[5] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.
[6] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.
[7] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.
[8] D. Boneh, G.D. Cr.escenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.
[9] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and Efficiently Searchable Encryption," Proc. 27th Ann. Int'l Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.

10. Biography

Syed Firdose is a student in Master of Technology in the Department of Computer Science and Engineering, Madina Engineering College, Kadapa, Andhra Pradesh, India.