International Journal on Applications in Information and Communication Engineering
Volume 1: Issue 10: October 2015, pp 19-22. www.aetsjournal.com ISSN (Online) : 2394-6237

Privacy-Preserving Multi-Keyword Ranked Search Over Encrypted Cloud Data

Mithilesh kumar Sharma, S.Karpagam

Abstract— Coming era of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data 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 and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of the irrelevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In the project for the first time, define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). The establish set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, thus 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. Thus further use "inner product similarity" to quantitatively evaluate such similarity measure. The first propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, as further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

Index Terms— *Cloud Computing, Data Security, Cloud Data Cryptography, Hybrid Cloud Architecture.*

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. 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, 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, the problem is particularly challenging project 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, the 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 retrieve the most relevant data. And each keyword in the search request is able to help narrow down the search result further. Co-ordinate matching 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 retrieval (IR) community. However, how to apply encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements the keyword privacy and many others.

Searchable encryption 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

Mithilesh kumar Sharma, PG Scholar, Department of MCA, St. Michael College of Engineering & Technology, Kalayarkoil, TamilNadu.

S.Karpagam, Professor, Department of IT and MCA, St. Michael College of Engineering & Technology, Kalayarkoil, TamilNadu.

the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot accommodate such high service-level requirements like system usability, user searching experience, and easy information discovery. Although some recent designs have been proposed to support Boolean keyword search as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality. Our early works have been aware of problem and provide solutions to the secure ranked search over encrypted data problem but only for queries consisting of a single keyword. How to design an efficient encrypted data search mechanism that supports multi-keyword semantics without privacy breaches still remains a challenging open problem.

For the first time, thus defines and solves the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system wise privacy in cloud computing paradigm. Among various multi-keyword semantics, thus choose the efficient similarity measure of coordinate matching as many matches as possible, to capture the relevance of data documents to the search query. Specifically, The inner product similarity the number of query keywords appearing in a document, to quantitatively evaluate such similarity 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 corresponding 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 search request, so the similarity could be exactly measured by the inner product 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, and propose a basic idea for the MRSE using secure inner product computation, which is adapted from a secure k-nearest neighbor technique, and then give two significantly improved MRSE schemes in a step-by-step manner to achieve various stringent privacy requirements in two threat models with increased attack capabilities.

II. Problem Statement

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. 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.

- ❖ Single-keyword search without ranking
- ❖ Boolean- keyword search without ranking
- ❖ Single-keyword search with ranking

III. Proposed Scheme

For the first time, here 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. To propose two MRSE schemes based on the similarity measure of "coordinate matching" while meeting different privacy requirements in two different threat models. Here investigate some further enhancements of our ranked search mechanism to support more search semantics and dynamic data operations.

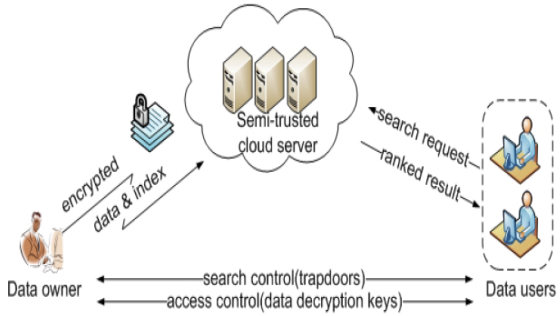

Fig 1. System Architecture

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

- ❖ Multi-keyword ranked search over encrypted cloud data (MRSE).
- ❖ Coordinate matching by inner product similarity.

IV. Implementation

The general-purpose, high-level Java programming language is also a powerful software platform. Using the generous API, you can write many types of programs. An application is a standalone program that runs directly on the Java platform. A special kind of application known as a *server* serves and supports clients on a network. Examples of servers are Web servers, proxy servers, mail servers, and print servers. Another specialized program is a *servlet*. A servlet can almost be thought of as an applet that runs on the server side. Java Servlets are a popular choice for building interactive web applications, replacing the use of CGI scripts. Servlets are similar to applets in that they are runtime extensions of applications.

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a

successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

1. User Login and Registration
2. File Upload and Key set with Encryption
3. User Request
4. Approved Key
5. Search Key
6. File Download and Decrypt

A. User Login and Registration

An authorize user in to the system. To adds security to the user data. The login credentials are secured by encryption and they are decrypted back by the server to avoid eavesdropping.

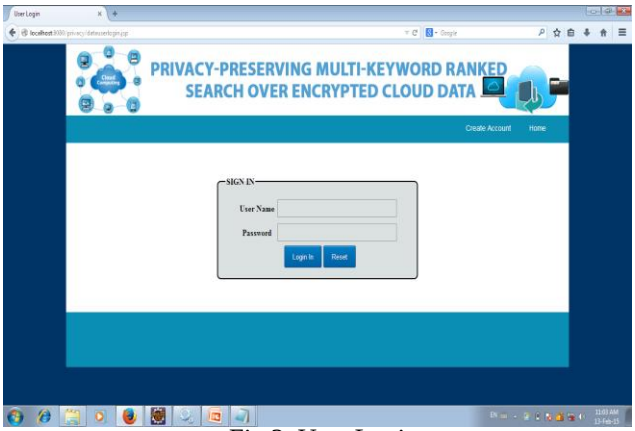

Fig 2. User Login

Users are having authentication and security to access the detail which is presented in the ontology system. Before accessing or searching the details user should have the account in that otherwise they should register first.

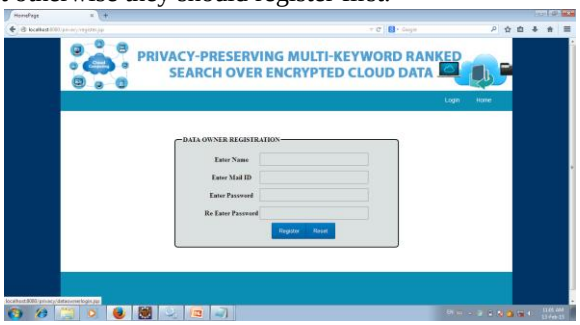

Fig 3. User Registration

B. File Upload and Key set with Encryption

Data owner after logging into system to added the data from crawling web and the data are stored in the structure so it can be accessed easily. The data will be large so that it should be stored in the proper structure.

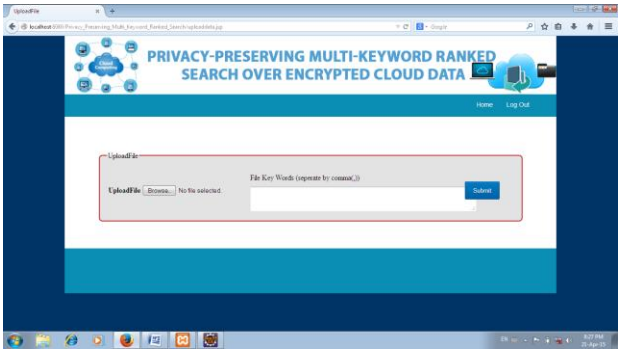

Fig 4. FileUpload

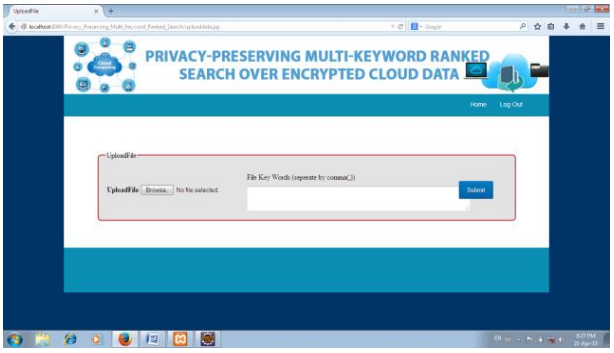

Fig 5. Key set with Encryption

C. User Request

The access to the data from the data owner is will be achieved by the access token.

These access token are the random number generated by the owner and send through the mail.

Fig 6. User Request

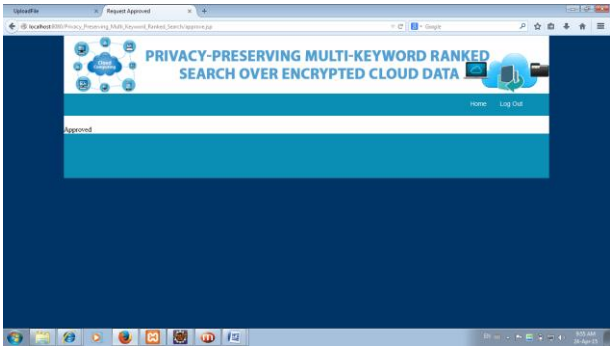

Fig 7. Approved Key

D.Approved and Search Key

The data in the server is grouped in such a way the user can easily query the data and thus approach also provides instant search assist so user can easily get the required information.

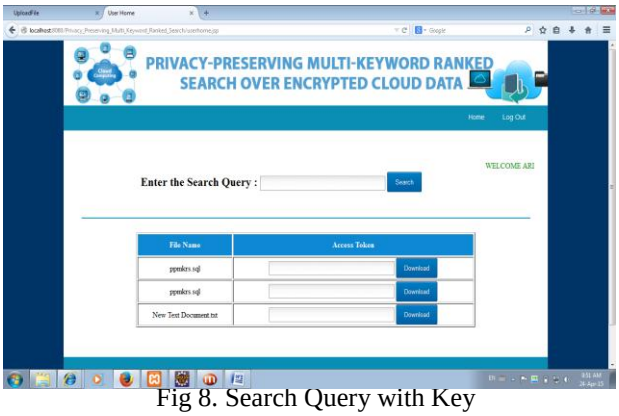

Fig 8. Search Query with Key

E. File Download and Decrypt

The results from the server are obtained by querying by the client. If client wants the full file he needs the access token to get the full file. To get that file client needs to request to the owner if owner accepts the request client will get access to the file.

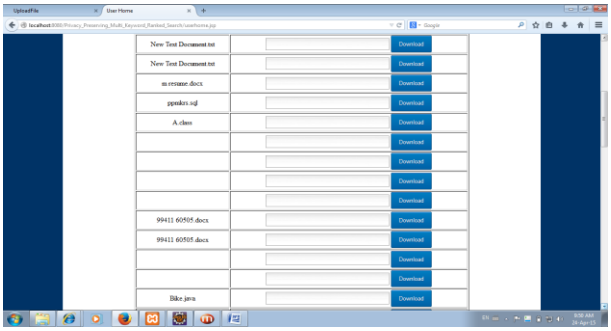

Fig 9. File Download

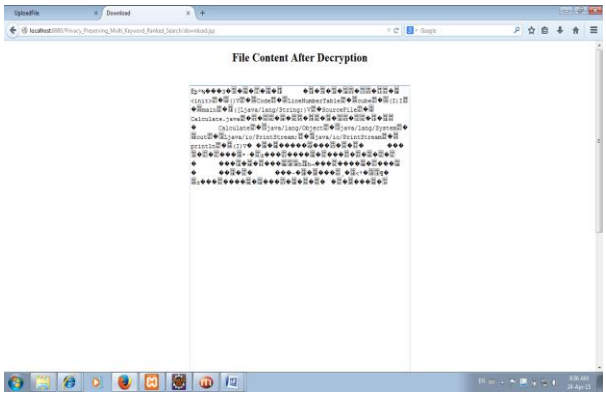

Fig 10. File Download

V. CONCLUSION

In this paper, 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, choose the efficient similarity measure of coordinate matching 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, then propose a basic idea of MRSE using secure inner product computation. Then give two improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. Also investigate some further enhancements of our ranked search mechanism, including supporting more search semantics TFIDF and dynamic data operations. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset show our proposed schemes introduce low overhead on both computation and communication.

VI. FUTURE WORK

Proposed the problem of multiple-keyword ranked search over encrypted cloud data, and construct a variety of security requirements. From various multi-keyword concepts, choose the efficient principle of coordinate matching. First propose secure inner data computation. Also achieve effective ranking result using k-nearest neighbour technique. System is currently work on single cloud, In future is will extended up to sky computing & Provide better security in multi-user systems.

REFERENCES

[1] L. M. Vaquero, L. Rodero-Merino, J. [2]. S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS. Springer, Heidelberg.
[2] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001.
[3] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.
[4] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of S&P, 2000.
[5] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, 2003, http://eprint.iacr.org/2003/216.
[6] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.
[7] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.
[8] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYPT, 2004.
[9] Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficientl searchable encryption," in Proc. of CRYPTO, 2007.
[10] J. Katz, A. Sahai, and B. Waters, "Predicate Encryption Supporting Disjunctions, Polynomial Equations, and Inner Products," Proc. 27th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2008.