ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

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

Ajaykumar Narayankar[1], Gajanan Rathod[2], Sanket Londhe[3], Ashish Wankhade[4], .M.A. Ansari[5]

Student, Dept. of IT, Rajarshi Shahu College of Engineering, Tathawade, Savitribai Phule Pune University, Pune, India[1234]

Prof. Dept. of IT, Rajarshi Shahu College of Engineering, Tathawade, Savitribai Phule Pune University, Pune, India[5]

ABSTRACT: With the advent 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 their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we 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. We further use "inner product similarity" to quantitatively evaluate such similarity measure. We 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 requirements in two different threat models. To improve search experience of the data search service, we 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.

KEYWORDS: sensitive data; multi-keyword ranked search; latent semantic

I. INTRODUCTION

Computing resources are shared by many users. The benefits of cloud can be extended from individual users to organizations. The data storage in cloud is one among them. The virtualization of hardware and software resources in cloud nullifies the financial investment for owning the data warehouse and its maintenance. Many cloud platforms like Google Drive, cloud; SkyDrive, Amazon S3, Dropbox and Microsoft Azure provide storage services. Security and privacy concerns have been the major challenges in cloud computing. The hardware and software security mechanisms like firewalls etc. have been used by cloud provider. These solutions are not sufficient to protect data in cloud from unauthorized users because of low degree of transparency. Since the cloud user and the cloud provider are in the different trusted domain, the outsourced data may be exposed to the vulnerabilities. Thus, before storing the valuable data in cloud, the data needs to be encrypted [2]. Data encryption assures the data confidentiality and integrity. To preserve the data privacy we need to design a searchable algorithm that works on encrypted data. Many researchers have been contributing to searching on encrypted data. The search techniques may be single keyword search or multi keyword search. In huge database the search may result in many documents to be matched with keywords. This causes difficulty for a cloud user to go through all documents and have most relevant documents. Search based on ranking is another solution, wherein the documents are ranked based on their relevancy to the keywords [3]. Economical searchable encryption techniques help the cloud users especially in pay-as-you use model. The researchers combined the rank of documents with multiple keyword searches to come up with efficient economically viable searchable encryption techniques. In searchable encryption related literature, computation time and computation overhead are the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

two most frequently used parameters by the researchers in the domain for analysing the performance of their schemes. Computation time (also called "running time") is the length of time required to perform a computational process for example searching a keyword, generating trapdoor etc. Computation overhead is related to CPU utilization in terms of resource allocation measured in time.

II. LITERATURE SURVEY

With the advantage of storage as a service many enterprises are moving their valuable data to the cloud, since it costs less, easily scalable and can be accessed from anywhere any time. The trust between cloud user and provider is paramount. We use security as a parameter to establish trust. Cryptography is one way of establishing trust. Searchable encryption is a cryptographic method to provide security. In literature many researchers have been working on developing efficient searchable encryption schemes. In this paper we explore some of the effective cryptographic techniques based on data structures like CRSA and B-Tree to enhance the level of security, hence trust. We tried to implement the search on encrypted data using Azure cloud platform.[2]
Cloud computing is generating lot of interest to provide solution for data outsourcing and high quality data services. More and more institution, organizations and corporations are exploring the possibility of having their applications, data and their IT assets in cloud. As the data and there by the cloud's size increases searching of the relevant data is expected to be a challenge. To overcome this challenge, search index is created to aid in faster search. However, search Index creation and computation has been complex and time consuming, leading to cloud-down time there by hindering the swiftness in reacting to data request for mission critical requirements. Focus of this paper is to explain how in the proposed system, reusability of search index is helping to reduce the complexity of search index computation. Search index is proposed to be created using parameters like similarity relevance, user ranking and scheme robustness. User ranking helps to guarantee why a phrase or a sentence or a key word is used frequently in the uploaded data. The proposed system ensures that the reusability of search index concept, highly reduces cloud down time while maintaining the security using searchable symmetric encryption (SSE).The user requested file is retrieved from the cloud, using Two-round searchable encryption (TRSE) scheme that supports top-k multi-keyword retrieval. [1]

Nowadays, more and more people are motivated to outsource their local data to public cloud servers for great convenience and reduced costs in data management. But in consideration of privacy issues, sensitive data should be encrypted before outsourcing, which obsoletes traditional data utilization like keyword-based document retrieval. In this paper, we present a secure and efficient multi-keyword ranked search scheme over encrypted data, which additionally supports dynamic update operations like deletion and insertion of documents. Specifically, we construct an index tree based on vector space model to provide multi-keyword search, which meanwhile supports flexible update operations. Besides, cosine similarity measure is utilized to support accurate ranking for search result. To improve search efficiency, we further propose a search algorithm based on "Greedy Depth-first Traverse Strategy". Moreover, to protect the search privacy, we propose a secure scheme to meet various privacy requirements in the known ciphertext threat model. Experiments on the real-word dataset show the effectiveness and efficiency of proposed scheme. [3]

III. PROBLEM STATEMENT

We define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we 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. We further use "inner product similarity" to quantitatively evaluate such similarity measure. As a special case of modification, the operation of deleting existing documents introduce less computation and communication cost since it only requires to update the document frequency of all the keywords contained by these documents

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

IV. OBJECTIVE

I. We 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.
II. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given.
III. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.
IV. The proposed schemes indeed introduce low overhead on computation and communication.
V. The proposed schemes introduce nearly constant overhead while increasing the number of query keywords.

Therefore, our schemes cannot be compromised by timing-based side channel attacks that try to differentiate certain queries based on their query time.

VI. EXISTING SYSTEM AAPROACH

The existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical. All these multi keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval.

VII.PROPOSED SYSTEM

A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data We construct a special tree-based index structure and propose a “Greedy Depth-first Search” algorithm to provide efficient multi-keyword ranked search. The proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

I. Abundant works have been proposed under different threat models to achieve various search functionality,
II. Recently, some dynamic schemes have been proposed to support inserting and deleting operations on document collection.
III. This paper proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi keyword ranked search and dynamic operation on the document collection.

VIII. PROPOSED SYSTEM ALGORITHMS

1. Algorithm to provide efficient multi-keyword ranked search .
2. The secure kNN algorithm is utilized to encrypt the index and query vectors.
3. Propose a “Greedy Depth-first Search” algorithm based on this index tree.
4. Algorithm achieves better-than-linear search efficiency but results in precision loss.
5. The LSH algorithm is suitable for similar search but cannot provide exact ranking.
6. {I′s ; ci} ← GenUpdateInfo (SK; Ts; i; up type)) This algorithm generates the update information {I′s ; ci} which will be sent to the cloud server.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

IX. SYSTEM ARCHITECTURE

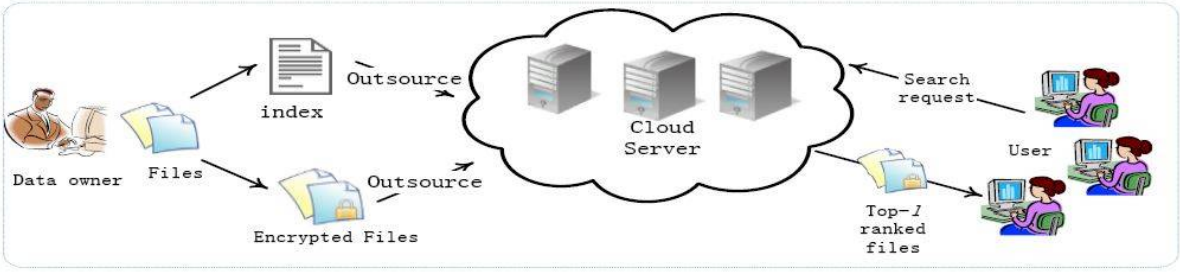

Fig No 01. System Architecture

The cloud server both follows the designated protocol specification but at the same time analyzes data in its storage and message flows received during the protocol so as to learn additional information.
The designed goals of our system are following:
Latent Semantic Search: We use statistical techniques to estimate the latent semantic structure, and get rid of the obscuring "noise" [5].
Multi-keyword Ranked Search: It supports both multi-keyword query and support result ranking.
Privacy-Preserving: Our scheme is designed to meet the privacy requirement and prevent the cloud server from learning additional information from index and trapdoor.
1) Index Confidentiality. The TF values of keywords are stored in the index. Thus, the index stored in the cloud server needs to be encrypted;
2) Trapdoor Unlinkability. The cloud server should not be able to deduce relationship between trapdoors.
3) Keyword Privacy. The cloud server could not discern the keyword in query, index by analyzing the statistical information like term frequency.

X. EXPERIMENTAL SETUP

In this section, we show a thorough experimental evaluation of the proposed technique on a real dataset: the MED dataset.
F-measure that combines precision and recall is the harmonic mean of precision and recall[8]. Here, we adopt F-measure to weigh the result of our experiments.

$$F = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

For a clear comparison, our proposed scheme attains score higher than the original MRSE in F-measure. Since the original scheme employs exact match, it must miss some similar words which is similar with the keywords. However, our scheme can make up for this disadvantage, and retrieve the most relevant files. Fig .2 shows that our method achieves remarkable result.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

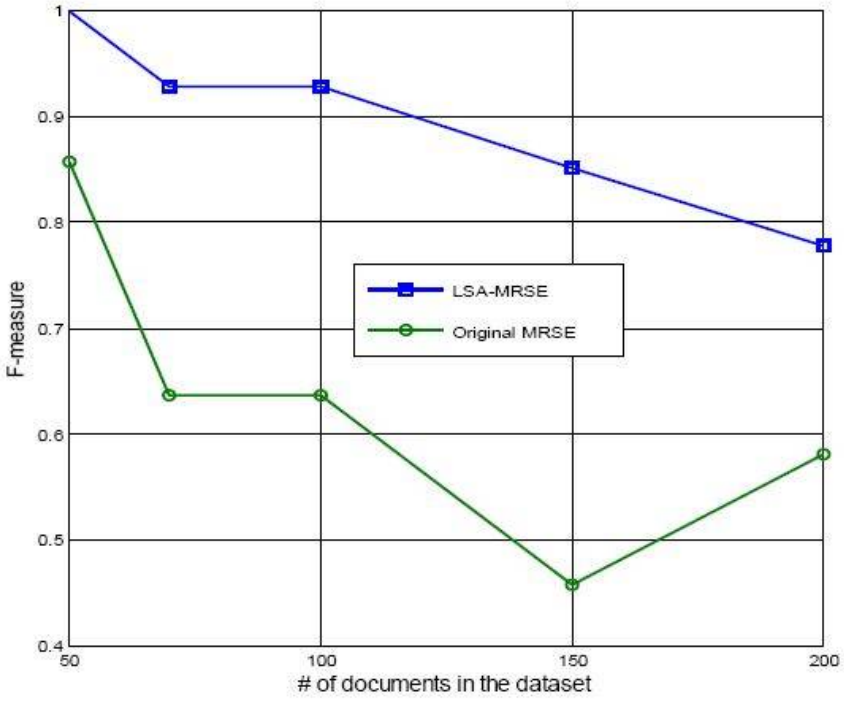

Fig. 2. Comparison of two schemes

XI. EXSISTING AND PROPOSED COMPARISON-

In this section we present comparison result of Single Key word Search Ranked search and Multi Keyword Ranked Search over a Encrypted Data on Cloud as shown in following figures .In this Result Each Ranked search and Multi Keyword Ranked Search Over An Encrypted Data On Cloud. In this Result Existing System is Single Keyword Search System and proposed System is nothing but MRES System.

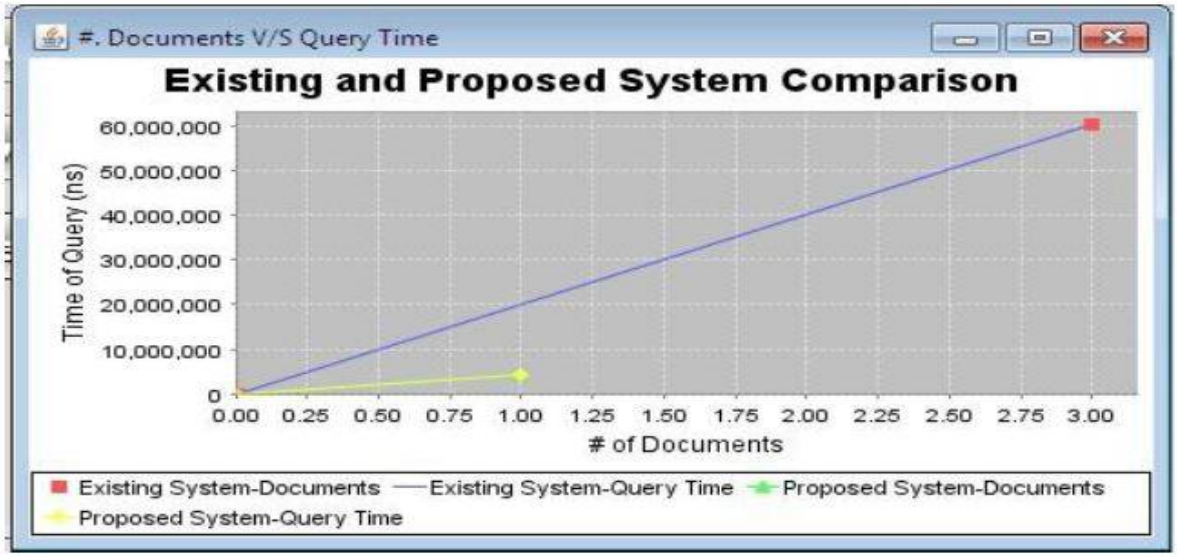

Fig 3: Comparison Graph- No. Of Documents V/S Query Time

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

It is a Comparison graph of Existing System and Our System. The graph is plotted Number of Documents that the respective system's search result returned and Time required to return the documents; in respective System. As shown in the graph Our system requires less time which is less than around 5 ns with most specific result of number of document equal to one which is less than the documents returned by existing system which are three and time required is around 6ns

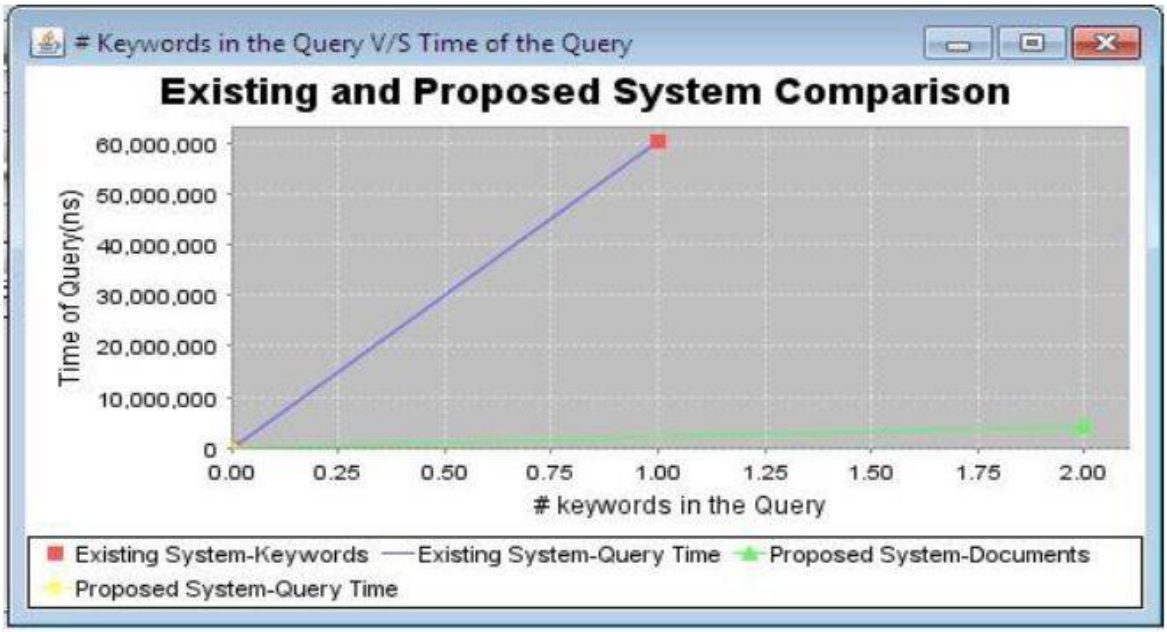

Fig 4: Comparison Graph- No. Of Documents V/S Query Time

It is a Comparison graph of Existing System and Our System. The graph is plotted Number of Documents that the respective system's search result returned and Time required to return the documents; in respective System. As shown in the graph Our system requires less time which is less than around 5 ns with most specific result of number of document equal to one which is less than the documents returned by existing system which are three and time required is around 6ns . Comparison Graph- No. Of Keywords V/S Query Time .It is a Comparison graph of Existing System and Our implemented System. The graph is plotted against Number of Keywords fired in the respective system's search and Time required in respective System. As shown in the graph Our system requires less time which is less than around 5 ns with multiple Keyword Query and existing system requires around 6ns even though a single Keyword query is fired. So Our System Works Better in each and every aspect then existing System.

XII. CONCLUSION

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. We also investigate some further enhancements of our ranked search mechanism, including supporting more search semantics, i.e., TF _IDF, and dynamic data operations. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world data set show our proposed schemes introduce low overhead on both computation and communication.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

REFERENCES

[1] M.Armbrust, "A view of cloud computing",Communications of the ACM,vol.53, no. 4, (2010),pp. 50-58.
[2] D. Boneh, "Public keyencryption with keyword search",Advances in Cryptology-Eurocrypt 2004,Springer, (2004).
[3] R. Curtmola, "Searchable symmetric encryption: improved definitions and efficient constructions",Proceedings of the 13th ACM conference on Computer and communications security,ACM, (2006).
[4] D.X.Song,D. Wagner and A.Perrig,"Practical techniques for searches on encrypted data. in Security and Privacy", 2000. S&P 2000,Proceedings 2000 IEEE Symposium,IEEE, (2000).
[5] C. Wang, "Secure ranked keyword search over encrypted cloud data",Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference,IEEE, (2010).
[6] N. Cao,"Privacy-preserving multi-keyword ranked search over encrypted cloud data",INFOCOM, 2011 Proceedings IEEE,IEEE, (2011).
[7] M.Armbrust, "A view of cloud computing",Communications of the ACM,vol.53, no. 4, (2010),pp. 50-58.
[8] D. Boneh, "Public keyencryption with keyword search",Advances in Cryptology-Eurocrypt 2004,Springer, (2004).
[9] R. Curtmola, "Searchable symmetric encryption: improved definitions and efficient constructions",Proceedings of the 13th ACM conference on Computer and communications security,ACM, (2006).
[10] D.X.Song,D. Wagner and A.Perrig,"Practical techniques for searches on encrypted data. in Security and Privacy", 2000. S&P 2000,Proceedings 2000 IEEE Symposium,IEEE, (2000).
[11] C. Wang, "Secure ranked keyword search over encrypted cloud data",Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference,IEEE, (2010).
[12] N. Cao,"Privacy-preserving multi-keyword ranked search over encrypted cloud data",INFOCOM, 2011 Proceedings IEEE,IEEE, (2011).