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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Privacy Preserving Multi-User, Multi-Keyword Ranked Based Searching Over Encrypted Data

Vijendra Augustine, Prof. P. L. Ramteke

M.E Student, Dept. of CSE, HVPM's COET, Amravati, India

Associate Professor and Head, Dept. of IT, HVPM's COET, Amravati, India

ABSTRACT: Due to the rapid expansion of data, the data owners tend to store their data into the cloud to release the burden of data storage and maintenance. However, as the cloud customers and the cloud server are not in the same trusted domain, our outsourced data may be under the exposure to the risk. Thus, before sent to the cloud, the sensitive data needs to be encrypted to protect for data privacy and combat unsolicited accesses. Unfortunately, the traditional plaintext search methods cannot be directly applied to the encrypted cloud data any more.As the storage and computing requirements of users are increasing, every time data is getting transferred to the remote server in larger chunks, but it's not necessary that the server on which the data is outsourced, is trusted server. In this paper, two of the privacy preserving issues about accessing the cloud data has been identified i.e. acuteness of keywords sent in queries and the data fetched as a result of those queries. Both of them should to be hidden. To keep the privacy of documents, it should get encrypted before outsourcing to the cloud. Among various multi-keyword semantics, we choose the efficient principle of "parameter matching", i.e., as many matches as possible, to capture the similarity between search query and data documents. For preserving privacy, the proposed system uses Advance Encryption Standard (AES) algorithm and MD5 cryptographic technique. Here, the proposed and develop system takes input from user converted in to encrypted format to achieve privacy preservation and the multiple keywords enter for search files uses pattern matching technique and gives proper file if the input keywords gets match properly.

KEYWORDS: Multi-keyword Ranked Search over Encrypted cloud data (MRSE), Privacy Preserving, Document Indexing, Advance Encryption Standard (AES), Secure Cloud Storage.

I. INTRODUCTION

The National Institute of Standards and Technology (NIST) [1] defines the cloud computing as a model for enabling ubiquitous, convenient, on demand network access to a shared pool of configurable computing resources. These resources canbe storage capacities that are controlled, allocated and managed by the Cloud Service Provider (CSP). Therefore, by moving their data to the cloud, users remove the burdenof building and maintaining a local storage infrastructure.As such, they only have to pay their CSP for the allocated resources. Microsoft Windows Azure storage services [2]and Amazon's Simple Storage Service (S3) [3] are good examples. Indeed, these providers offer to their clients thepossibility to store, retrieve and share data with other users in a transparent way.

Due to the rapid expansion ofdata, the data owners tend to store their data into the cloud to release the burden of data storage and maintenance [1]. However, as the cloud customers and the cloud server are not in the same trusted domain, our outsourced data may be under the exposure to the risk. Thus, before sent to the cloud, the sensitive data needs to be encrypted to protect for data privacy and combat unsolicited accesses. Unfortunately, the

traditional plaintext search methods cannot be directly applied to the encrypted cloud data any more. The traditional information retrieval (IR) has already provided multi-keyword ranked search for the data user. In the same way, the cloud server needs provide the data user with the similar function, while protecting data and search privacy. It is meaningful storing it into the cloud server only when data can be easily searched and utilized.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

In this paper, we will solve the problem of multi-keyword latent semantic ranked search over encrypted cloud data and retrieve the most relevant files. We define a new scheme named Latent Semantic Analysis (LSA)-based multi-keyword ranked search which supports multi-keyword latent semantic ranked search. By using LSA, the proposed scheme could return not only the exact matching files, but also the files including the terms latent semantically associated to the query keyword. For example, when the user inputs the keyword "automobile"to search files, the proposed method returns not only the files containing "automobile", but also the files including the term"car". We take a large matrix of term-document association data and construct a semantic space wherein terms and documents are closely associated are placed near one another. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, we propose the idea: the multi-keyword ranked search (MRSE) using "Latent Semantic Analysis".

II. LITERATURE REVIEW & RELATED WORK

In 2015, Lim and Robshaw [8] introduced a new concept of dynamic key infrastructure for grid, to simplify the key management issues. That is, each user is in charge of publishing his IBC–PE to the other entities. He distributes a fixed parameter set through a X.509 certificate to allow users to act as their own trusted authorities for the purpose of delegation and single sign-on. Therefore, they remove the need for a proxy certification. On one hand, this technique avoids the key escrow attack and the need for a secure channel for private key distribution in an ID-based system. Unfortunately, users have to support the cumbersome task of verifying the parameter sets of other entities.

In addition, this paper does not address the arising risk of Man in The Middle attacks [5].In 2005, Lim and Paterson [6] proposed to use IBC in order to secure a grid environment. They describe several scenarios in which IBC simplifies the current grid solutions, like the elimination of the use of certificate, simple proxy generation, easy revocation of proxy certificates and the savings of bandwidth by using the pairing based approach proposed by Boneh and Franklin [9].

In the same way, Li et al. [3] propose to use IBC as an alternative to the SSL authentication protocol in a cloud environment. However, these schemes still suffer from the needed trust hierarchy to ensure a secure working system. Recently, Schridde et al. [4] presented a novel security infrastructure, using IBC, for service-oriented cloud applications to overcome the problems of certificate based solutions. In their proposal, the URLs of the service are used for public keys generation.

In previous systems, effective keyword searching schemes has been developed which utilizes bilinear maps & which are based on the public key encryption technique. This scheme works only for single user and more crucially, queries in this scheme got generated in a very abstract manner, and hence, unable to hide the search pattern [1]. Some schemes are developed in which user must have knowledge about all the valid keywords and their respective positions as mandatory information so as to generate a query [8]. One of the typical uses of cloud storage service is database repository by users. One of the cryptographic primitive is searchable encryption which allows private keyword based searching over encrypted database.

III. PROPOSED WORK

A.*Threat models and Design Goals*

The cloud server is considered as "honest-but-curious" in our model. Particularly, 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 [9].

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

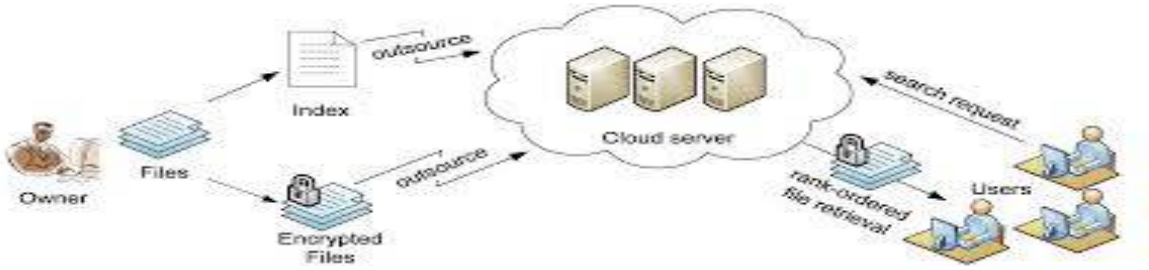

Fig. 1:Encrypted Cloud Data

- **Data owner** has a collection of data documents D □ {*d*1 , *d* 2 , ..., *d m* } .A set of distinctkeywordsW =□ {*w*1 , *w*2 , ..., *wn* } is extracted from the data collection *D* . The data owner willfirstly construct an encrypted searchable index *I* from the data collection *D*. All filesin *D* are encrypted and form a new file collection, *C* .Then, the data owner upload both theencrypted index *I* and the encrypted data collection*C* to the cloud server.
- **Data user** provides *t* keywords for the cloud server. A correspondingtrapdoor through search control mechanisms is generated. In this paper, we assume thatthe authorization between the data owner and the data user is approximately done.
- **Cloud server** received from the authorized user. Then, the cloud server calculates and returns to the corresponding set of encrypted documents. Moreover, to reduce the communication cost, the data user may send an optional number *l* along with thetrapdoor*T* so that the cloud server only sends back top- *l* files that are most relevant to the search query.

In this paper, we purpose to achieve security and ranked search under the above model.The designed goals of our system are following:

- **Latent Semantic Search:** We aim to discover the latent semantic relationship between terms and documents. We use statistical techniques to estimate the latentsemantic structure, and get rid of the obscuring “noise” [11].The proposed scheme tries toput similar items near each other in some space in order that it could return the data userthe files contain the terms latent semantically associated with the query keyword.
- **Multi-keyword Ranked Search:** It supports both multi-keyword query and supportresult ranking.
- **Privacy-Preserving:** Our scheme is designed to meet the privacy requirement andprevent 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 could do some statistical analysis overthe search result. Meanwhile, the same query should generate different trapdoorswhen searched twice. 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, indexby analyzing the statistical information like term frequency.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

B. *Cloud service provider Salesforce Cloud Identity:*

Fig.2: Salesforce Cloud Identity

As Salesforce cloud provides better security with high performance, we are using it as cloud service provider in our system. The number of cloud-based providers and applications continue to grow, users are faced with an ever-growing array of login options and credentials.

Salesforce Identity provides Identity and Access management (IAM) for Web and mobile applications through the simplicity, transparency, and trust of the Salesforce Platform. Salesforce Identity helps to improve the usability.

Salesforce features:

1. Single Sign-On & Social Desktop:

Users sign in once into Salesforce Identity and gain one click access to applications.

2. Identity & Access Management:

Administrators centrally manage access to applications designed for desktop, mobiles.

3. Enterprise Directory Integration: For organizations with existing enterprise systems like Active Directory,administrators can utilize automated synchronization of users and Single Sign-On.

4. De-Provisioning:

Administrators have the capacity to quickly stop access to applications and services through automated de-provisioning.

5. Centralized Reporting:

Centralized information on user authentication, access, utilization, and de-provisioning.

IV. SYSTEM FEATURES

To activate ranked search for effective utilization of outsourced cloud data, our system design should simultaneously achieve security and performance guarantees as follows.

1. Secured Multi-keyword Ranked Search: To design search schemes which allow multi-keyword query and provide result similarity ranking for valuable data retrieval, instead of returning undifferentiated results.

2. Privacy: To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements.

3. Effectiveness with high performance: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

4. Authentication is operation of verify the truth of an entity or genuine user. This might involve confirming the identity of a software program or person. Here we will provide user id and password for validation of legitimate user. Confidentiality is a set of rules that limits access or places limitations on certain types of information. To maintain confidentiality of data there will be provision for encryption of data using cryptography tool. [9]

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

V. RESULT ANALYSIS

The implementation of the proposed scheme is done using Asp.Net and C # language in Windows 8 operation system and tests its efficiency. The tests include Search precision on varied privacy level. The search precision of this system is affected by the phantom keywords in proposed method. The results are shown in Fig.3 (a). As terms are added to the index vector to change the relevance score calculation, so that the cloud server cannot detect keywords by checking the Frequency distributions of special keywords.

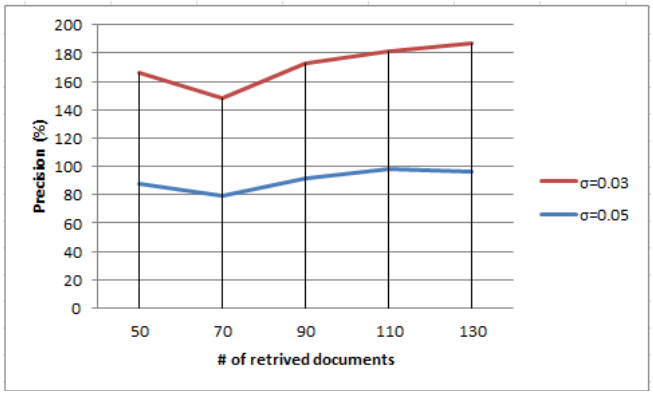

Fig. 3 (a): Precision of searches with different standard deviation

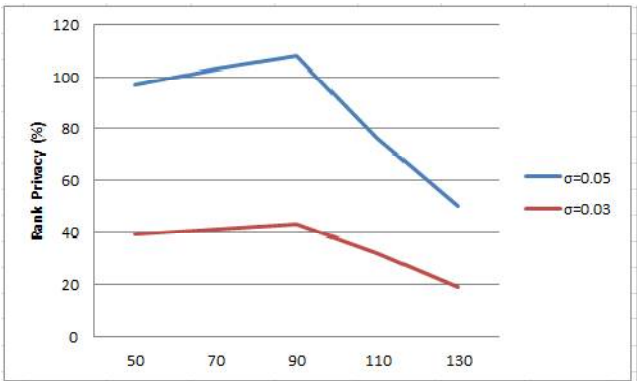

Fig. 3 (b): Rank Privacy of searches with different standard deviation

Where ri is the rank number of document in the retrieved k documents as top documents, and r′ i is its real rank number in the whole ranked results. The larger rank privacy denotes the higher security of the scheme, which is shown in Fig. 3(b). Here, data users can accomplish different requirements on search precision and privacy by adjusting the standard deviation σ. Here the comparison is with a recent work, which achieves good search efficiency. The previous scheme retrieves the search results through exact calculation of document vector and query vector. Thus, top-k search precision of the previous scheme is 91%. But as a similarity-based multi-keyword ranked search scheme, the previous scheme suffers from precision. The average precision of this method is 86%.

The implementation of the Encryption & Decryption, Secure index construction is successfully completed with desirable performance. After firing single-keyword query, user will get all documents that contain the specified keyword. Timing function has been added which will calculate encryption & indexing 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)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

VI. CONCLUSION

In this project, we have studied the new framework for the problem for multi-keyword ranked search over encrypted cloud data (MRSE), to establish a variety of privacy requirements for the data which is stored in the cloud computing environment. As the data owner prefers to store the data in the cloud on the pay-per-use basis. Among various multi-keyword semantics, the efficient similarity measure is Indexing the document is used i.e., as many matches are possible, according to the multi-keyword rank search to effectively capture the relevance of the outsourced documents for query keywords from the cloud. It makes to support for the quantitatively evaluate such measure of the similarity. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, MRSE framework is proposed by using secure inner product computation and apply the cryptographic technique and send data over the network in the encrypted manner. The proposed system is Multi-keyword ranked keyword searching which is working much more efficient, proper and user friendly than earlier old fashioned and manual single keyword based searching. We can enhance this system by including more GUI based searching that provides benefit as any new user can use the system in much more user friendly manner. Providing such features enable the users to use more facilities of the system. And belongs to today's advance computer and internet technology the concept of cloud storage needs to be improved to tackle large and scalable storage requirement.

REFERENCES

[1] Ning Caoy, Cong Wangz, Ming Liy, Kui Renz, and Wenjing Louy,"Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data"
[2] Weifeng Su, Jiying Wang, and Frederick H. Lochovsky, Member, "Record Matching over Query Results from Multiple Web Databases", *IEEE Computer Society*.
[3] Y.Srikanth,M.Veeresh Babu, PNarasimhulu,"Combined Keyword Search over Encrypted Cloud Data Providing Security and Confidentiality".
[2] Ayad Ibrahim, Hai Jin, Ali A.Yassin, DeqingZou, "Secure RankOrdered Search of Multi-Keyword Trapdoor over Encrypted Cloud Data",*IEEE Asia-Pacific Services Computing* Conference 2012.
[3] Yanjiang Yang, "Towards Multi-User Private Keyword Search for Cloud Computing" *IEEE 4th International Conference on Cloud Computing*,2011.
[4] Qin Liu, Guojun Wag, Jie Wu, "An Efficient Privacy Preserving Keyword Search Scheme in Cloud Computing" *IEEE International Conference on Computational Science and Engineering*. 2009.
[5] Ning Cao et al., "Privacy-Preserving MultiKeyword Ranked Search over Encrypted Cloud Data", *IEEE Transactions on parallel and distributed systems*, vol. 25, no. 1, jan 2014.
[6] Shih-Ting Hsu et al., "A Study of Public Key Encryption with Keyword Search", *International Journal of Network Security*, Vol.15, No.2, PP.71-79, Mar. 2013.
[7] Yong Zhang, Jian-lin Li, "Research and Improvement of Search Engine Based on Lucene",*IEEE Transaction* 2009.
[8] QIAN Liping, WANG Lidong, "An Evaluation of Lucene for Keywords Search in Large-scale Short Text Storage",*IEEE Transaction* 2010.
[9] "FAULT-TOLERANCE AND TRANSFORMATION ANALYSIS IN SELF-TUNING SERVERS" *International Journal of Current Engineering and Technology*, Vol.4,No.3 Jun-14 E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
[10] Govind S.Pole, Madhuri Potey, "A Highly Efficient Distributed Indexing system based on large cluster of commodity machines",*IEEE Transaction* 2012.
[11]Bruce Schneier,John Kelseyy, Doug Whitingz David, Wagnerx Chris Hall, "Twofish: A 128 Bit Block Cipher".
[12] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data", *Proc. IEEE Symp. Security and Privacy*, 2000.