ISSN:0975-9646

Secure Multi keyword Retrieval over Encrypted Cloud Data

Ms. Pradnya H. Unde[1], Ms. Arti Mohanpurkar[2]

[1]*Master of Engineering (Computer Engineering), Savitribai Phule Pune University*
Dr. D. Y. Patil SOET, Charoli, Bk Via Lohegaon,
Pune 412105, India

[2]*Head of Computer Engineering Department*
Dr. D. Y. Patil SOET, Charoli, Bk Via Lohegaon,
Pune 412105, India

Abstract: **Now a day, Data owner incentivizes to outsource their data on the cloud to get more flexibility. Cloud computing provides data outsourcing and high quality accommodation. For data security, the data owner provides encryption on their data. The data owners outsource their data on the cloud through which they reduces a cost and computational overhead quandaries. Considering the sizably voluminous number of data users and documents in the cloud, it is obligatory to sanction multiple keywords in the search request and return documents in the order of their suggestion to these keywords. Cognate works on searchable encryption fixate on single keyword search or Boolean keyword search, and virtually not ever sort the look for results. In subsisting system, for the first instance, the conundrum of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE) is define and solve. A set of several privacy desiderata for such a bulwarked cloud data utilization system are defined by proposed work. The server side ranking is according to the order preserving encryption (OPE) ineluctably leaks data privacy. Leakage quandary is solved by utilizing two round searchable encryption scheme and it withal fortifies to top-k multi keyword retrieval. In two rounds searchable encryption scheme involves vector space model and Homomorphic encryption. These propose work results are eliminating the data leakage and data security issues.**

Keywords— Cloud, ranking, cloud computing, privacy preservation, data privacy, vector space model, Homomorphic encryption

I. INTRODUCTION

Cloud migration is the process of partially or consummately deploying an organization's digital assets, accommodations, IT resources or applications to the cloud. Cloud is like network and Cloud computing is mechanism which perform the computing on cloud. Cloud computing provide many benefits to users i.e. scalability ,security, flexibility, pay as utilize which is consider in this implementation. Withal cloud computing provides many accommodation models i.e. platform as an accommodation, infrastructure as an accommodation, software as an accommodation etc.

Each data owner has their own sets of documents, to maintain these documents on their personal computer or locally is arduous process. In other words, maintain and stored the documents locally are sumptuous for storage purpose and it arises computational overhead. Utilizing cloud, capital expenditure can decremented by not having to buy and maintain costly hardware. A cloud accommodation provider can deploy the data to their high performance systems, with no desideratum to maintain and upgrade sumptuous software and systems; instead the employees can be habituated to do some productive work for the organization. Hence data owner incentivize to outsource their sets of documents on cloud to get more flexibility. But afore migration process, the data privacy issue is a raised in front of owner, hence to maintain the security and privacy she used encryption methods.

In cloud computing, data owner may share their outsourced data with a number of users, who might want to only retrieve the data files they are fascinated with. One most popular way is do probing by utilizing keyword base retrieval. Keyword search is data retrieval accommodation which applied on plain text scenarios, in which user retrieve pertinent files from the sets of file predicated on keywords. But this scenario is becomes arduous task when it consider in the case of cipher text, because we can do only circumscribed operations on encrypted data. To ameliorate feasibility and preserve on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most pertinent files that match users interest in lieu of all the files, which denotes that the files should be ranked in the order of pertinence by users interest and only the files with the highest pertinence are sent back to users. Searchable symmetric encryption schemes is enable search on cipher text but it support only Boolean keyword search, without considering the difference of pertinence with the queried keyword of these files in the result. Searchable symmetric encryption schemes are suffer from two problems-Boolean representations and how to strike a balance between security and efficiency. The concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes and to solve in security problem by using two round searchable encryption scheme.

For privacy issue in searchable encryption here we used similarity relevance and schema robustness concept. Server-side ranking is according to the order-preserving encryption (OPE) method, but it inevitably violates data privacy.

Two round searchable encryption technique fulfil the secure multi keyword top-k retrieval over encrypted cloud data.

II. MOTIVATION

There are some of the motivations for the company's decision of migration to the cloud:

•Standardization: Standardization means simplifying the system by dealing with less number of configurations, easily facilitated automation and much simpler support. Along with it, the cloud environments being very flexible allows easy provision in various ways. Also it is very user friendly.

- Self Service: With self-service, the user has the control and more cost and usage choices, along with increased visibility.
- Virtualization: Virtualization ensures flexibility, increasing the utilization thus being energy efficient. Infrastructure Abstraction and Soft Configuration are characteristics of virtualization.
- Automation: Automation means low human involvement and swift deployment. It also ensures repeatable configuration thus improving compliance.
- Cost Savings: Using cloud, capital expenditure can be decreased by not having to buy and maintain costly hardware. A cloud service provider can deploy the data to their high performance systems, with no need to maintain and upgrade expensive software and systems; instead the employees can be used to do some productive work for the organization.
- Better Collaboration: Good collaboration is the new business success mantra and migration to the cloud makes it much easier to achieve. A more mobile workforce can be achieved who using their own devices can be more productive.
- Improved Network Performance: If organizations are using remote data- centres of their cloud service providers to work on their data, the workload on their on-premise networks can be greatly reduced, thus improving performance of functions using the on-premise internal network.
- Improved Integration and Compatibility: The upcoming big data needs of organizations needs them to be capable of accessing and analysing data stored across

Hence 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 relish the on-demand high-quality applications and accommodations from a shared pool of configurable computing resources. But for bulwarking data privacy, sensitive data has to be encrypted afore outsourcing, which obsoletes traditional data utilization predicated on plain text keyword search. Thus, enabling an encrypted cloud data search accommodation is of paramount consequentiality.

III. LITERATURE SURVEY

Now a days, cloud computing offers a drastically different and affordable approach to IT resource delivery: lease the data and processing capacity you need from a "cloud" (pool) of interconnected, shared computing systems that are maintained by cloud service providers. Toward Secure Multi keyword Top-k Retrieval Over Encrypted Cloud Data, Jiadi Yu et al [1], Top-k Retrieval from a Confidential Index, S. Zerr, D. Olmedilla et al [10] and Fully Homomorphic Encryption over the Integers, M. van Dijk et al [11] discussed about top-k retrieval methods by using confidential index over encrypted data. Homomorphic encryption techniques are used for allowing specific types of computations to be carried out on the corresponding cipher text.

A View of Cloud Computing, M. Armbrust et al [6] discussed about cloud computing benefits such as agility, elasticity, availability, and cost-efficiency are well known, due to cost saving through larger economies of scale and flexible resource allocation schemes provided by different cloud services. Verifiable Privacy – Preserving Multi-keyword Text Search in cloud supporting similarity-Based Ranking, Wenhai Sun et al [2], Secure Ranked Keyword Search over Encrypted Cloud Data, C. Wang et al [9] and Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions, R. Curtmola et al [8] discussed about vector space model working and similarity based ranking multi-keyword text and keyword search scheme. Vector space model is popular technique which provides "tf-idf rule" through which we get accurate ranking result. Privacy Preserving Multi-keyword Ranked Search over Encrypted Cloud Data, Ning Cao et al [3] and [12] discussed about 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, further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of existing schemes is given and establishes a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, 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. Further use "inner product similarity" to quantitatively evaluate such similarity measure. Cloud Migration Research: A Systematic Review, Pooyan Jamshidi et al[4] and [5] discussed about the importance of cloud migration and the relative maturity of this field , a consolidation of existing evidence on legacy to-cloud migration is timely. Amazon S3[7] discussed about Amazon S3 offers a highly durable, scalable and secure solution for backing up and achieving critical data. It has capability to provide protection for stored data. You can also define lifecycle rules to automatically migrate less frequently accessed data to standard IA archive set of objects to Amazon Glacier.

IV. SYSTEM DESIGN

A. System Architecture

There are three main actors present in these activities: cloud server, data owner, and data user. Data owner have her own sets of documents, to maintain these documents locally is become difficult task.

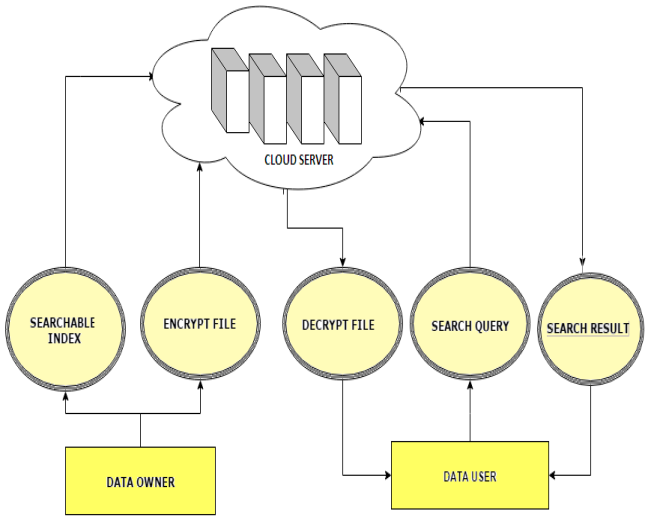

Fig 1. Top ranking result retrieval scenario

Maintain and stored the documents locally are expensive for storage and it arises computational overhead. Hence data owner motivate to outsource their sets of documents on cloud to get more flexibility. But before migration process, the data privacy issue is arises in front of owner, hence to maintain the security and privacy she used encryption methods and outsource the data in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other authorized users.

Information leakage would affect the data privacy which is unacceptable to data owner. The data user is sanctioned to process multi keyword retrieval over the outsourced data. The data user encrypts the query and sends it to the cloud server that returns the pertinent files to the data user. Afterward, the data user can decrypt and make use of the files.

B. Two round searchable encryption (TRSE)

Existing SSE scheme server side ranking predicated on order-preserving encryption to ameliorate the efficiency of retrieval over encrypted cloud data. User side ranking scheme is challenged by practical use. When the cloud server receives a query consisting of multi keywords, it computes the scores from the encrypted index stored on cloud and then returns the encrypted scores of files to the data user. Next, the data user decrypts the scores and picks out the top-k highest scoring files identifiers to request to the cloud server. The retrieval takes a two-round communication between the cloud server and the data user. Hence the scheme name is TRSE; in which ranking is done at the user side while scoring calculation is done at the server side.

Working of TRSE:

I. The data owner generates the secret key and public keys for the homomorphic encryption scheme.
II. The data owner builds the secure searchable index from the file collection C.
III. The data user generates secure trapdoor from his request REQ. Vector T is built from user's multi keyword request REQ and then encrypted into secure trapdoor with public key from PK, output the secure trapdoor.
IV. When receives secure trapdoor, the cloud server computes the scores of each files in searchable index with secure trapdoor and returns the encrypted result vector B back to the data user.
V. The data user decrypts the vector B with secret key SK and then requests and gets the files with top-k scores.

C. Homomorphic Encryption Scheme

Computational work is burden on user side and server side hence we need an encryption scheme to guarantee the operability and security at the same time on server side. Homomorphic encryption allows specific types of computations to be carried out on the corresponding cipher text. The result is the cipher text of the result of the same operations performed on the plaintext.

D. Vector Space Model

TF-IDF rule is used to find the accurate ranking and similarity measures. Where TF denotes occurrence count of term within a document and IDF is obtained by dividing the total number of documents in collection by number of document containing the term. It gives the top-k retrieval result.

VI. ALGORITHMS

A. *Algorithm for Top Result selection:*

1) Input

Take variable 'k' like a number and list source of selected item

2) Initialization:

Set pointers tk & tid as a null

3) Iteration phase

a. For all i ϵ source do
Insert(tk,(i, index))
b. End for
c. For all tuple ϵ tk do
tid.append(tuple[1])
d. End for

4) Output:

tid

B. *Algorithm for Insertion:*

1) Input

Take list tk to stored the top scoring items
Tuple(i,index)

2) Iteration

a. If length(tk) < k then

Insert(i, index) into tk in ascending order of items

b. Else

For all element ϵ tk do
If i< element[0] then
Continue
Else
Discard tk[0],
insert(i, index) into
tk in ascending order of item
EndIf

EndFor

c. EndIf

VII. MATHEMATICAL MODEL

A. Initialization Phase:

[1.] The data owner calls KeyG(λ) to generate the secret key SK and public key PK for homo graphic encryption scheme. Then the data owner assigns SK to the authorized data users.

[2.] The data owner Extracts the collection of L keywords K = $\{k_1, k_2, \ldots, k_n\}$ and their TF and IDF values out of the collection of n files,
KC = $\{c_1, c_2, \ldots, c_r\}$ for each file c_i ϵ KC , the data owner builds a (L+1) dimensional vector
$u_i = \{Id_i, w_{i,2}, \ldots w_{i,L}\}$.
The searchable index I=$\{u_i \mid 1 \leq i \leq n\}$

[3.] The owner encrypts the searchable index I to secure searchable index I' =$\{u'_i \mid 1 \leq i \leq n\}$ where
$u'_i = \{Id'_i, w'_{i,2}, \ldots w'_{i,L}\}$, Id'_i =Encrypt($REQ_{i,0}$, Id_i)

[4.] The data owner encrypts
KC = $\{c_1, c_2, \ldots, c_r\}$ into KC' = $\{c'_1, c'_2, \ldots, c'_n\}$ with other cryptography schemes, and then outsources KC' and I' to the cloud server.

B. Retrieval Phase:

[1.] The data user generates a set of keyword REQ= $\{k'_1, k'_2, \ldots, k'_n\}$, and then the query vector Q_w= $\{q_1, q_2, \ldots, q_n\}$ is generated in which
$q_i = 1 (1 \leq i \leq L)$ or q_i=0 otherwise. After that Q_w is encrypted in to trapdoor $Q_{w'}$= $\{p_1, p_2, \ldots, p_n\}$ where
p_i = Encrypt (REQ , q) then user sends $Q_{w'}$ to the cloud server.

[2.] For each file vector u_j in I', the cloud server computes the inner product p'_j=u'_j [1:l] . $Q_{w'}$ with modular reduction and then compresses and returns the result vector N'= $\{(Id'_1, s'_1), (Id'_2, s'_2), \ldots, (Id'_n, s'_n)\}$ to the data user.

[3.] The data user decrypts , where
N= $\{(Id'_1, s_1), (Id'_2, s_2), \ldots, (Id'_n, s_n)\}$
s_j=Decrypt(SK, s'_j) . Then TOPK(N,K) is invoked to get the top -k highest scoring files identifiers $\{i_1, i_2, \ldots, i_n\}$ and send it to the cloud server.

[4.] The cloud server returns the encrypted k files $\{c_1, c_2, \ldots, c_k\}$ to the data user.

V. DISCUSSION

Improve efficiency; tradeoffs of the security of search pattern may be needed unless a new encryption scheme that provides more reasonable cipher text size becomes available. Researchers from cryptography community [10], [11] have made several attempts to move toward practical fully homomorphic encryption over integers. These progresses indicate that the efficiency of the TRSE scheme can be further improved.

In a practical cloud computing system, data updates like adding or deleting files lead to a new challenge to the searchable encryption scheme. Since data updates may be frequent, e.g., doctors update patients' medical records every day in a medical system and users update their photo albums weekly or even daily, it is necessary to consider the efficiency of update in searchable encryption design.

Performance Analysis

I. Initialization Phase:

Initialization phase includes Setup phase in which the data owner generates the secret key and public keys for the homomorphic encryption scheme. Index Build in which the data owner builds the secure searchable index from the file collection C. To improve the computing efficiency, the tf-idf values are rounded to integers when building I, which does not affect the retrieve accuracy. Note that encryption needs only the addition operation, so the complexity of encrypting I is O(nl), where n denotes the number of files and l denotes the number of keywords.

II. Retrieval Phase:

The Retrieval phase includes TrapdoorGen in which the data user generates secure trapdoor from his request REQ. Vector T is built from user's multi keyword request REQ and then encrypted into secure trapdoor T with public key from PK, output the secure trapdoor T and needs O(l) time complexity.

ScoreCalculate in which when receives secure trapdoor, the cloud server computes the scores of each files in searchable index with secure trapdoor and returns the encrypted result vector B back to the data user and the complexity of ScoreCalculate is O(nl).

Rank in which the data user is decrypt the vector B with secret key SK and then requests and gets the files with top-k scores. The Rank stage can be subdivided into ResultDecrypt and Top-K.

VI. CONCLUSION

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. But for protecting data privacy, sensitive data has 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. Importance of cloud migration and the relative maturity of this field, a consolidation of existing evidence on legacy to-cloud migration are timely.

Two round searchable encryption scheme employing the fully homomorphic encryption which fulfils the security requirements of multi keyword top-k retrieval over the encrypted cloud data. Existing scheme gives guarantee of high security and practical efficiency.

Future enhancement will check the integrity of rank order in the search result assuming the cloud server is untrusted.

REFERENCES

[1] Jiadi Yu, Peng Lu, Yanmin Zhu and Guangtao Xue, "Toward Secure Multikeyword Top-k Retrival Over Encrypted Cloud Data," IEEE transaction on dependable and secure computing, vol. No. 4 July/august 2013.

[2] Wenhai Sun, Bing Wang, Ning Cao, Ming Li and Wenjing Lou, "Verifiable Privacy – Preserving Multi-keyword Text Search in cloud supporting similarity-Based Ranking," IEEE transaction on parallel and distributed system, vol. no. 11,November 2014.

[3] Ning Cao, Cong Wang, Ming Li and Kui Ren, "Privacy Preserving Multi-keyword Ranked Search over Encrypted Cloud Data," IEEE transcation on parallel and distributed system, vol. 25 no. 1, January 2014.

[4] Pooyan Jamshidi, Aakash Ahmad, and Claus Pahl, "Cloud Migration Research: A Systematic Review", IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 2, Jul-Dec 2013

[5] V. Andrikopoulos, T. Binz, F. Leymann, and S. Strauch, How to Adapt Applications for the Cloud Environment: Challenges and Solutions in Migrating Applications to the Cloud, Computing, vol. 95, no. 6, pp. 493-535, 2013.

[6] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010.

[7] Amazon.com, "Amazon s3 Availability Event: July 20, 2008," http://status.aws.amazon.com/s3-20080720.html, 2008.

[8] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.

[9] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS), 2010.

[10] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k Retrieval from a Confidential Index," Proc. 12th Int'l Conf. Extending Database Technology: Advances in Database Technology (EDBT), 2009.

[11] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers," Proc. 29th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques, H. Gilbert, pp. 24-43, 2010.

[12] 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.