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.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Retrieval of Encrypted cloud data using multikeyword

C.Rajeshkumar[1], Dr.K.Rubasoundar[2]

P.G.Scholar, Department of Computer Science, PSR Engineering College, Sivakasi, Tamilnadu, India[1]

HOD, Department of Computer Science, PSR Engineering College, Sivakasi, Tamilnadu, India[2]

ABSTRACT: Cloud Computing represents one of the most significant shifts in information technology many of us are likely to see in our lifetimes. For the protection of data privacy, sensitive data has to be encrypted before outsourcing, which makes effective data utilization a very challenging task. Here searchable symmetric encryption (SSE) used to secure and retrieve the data from the cloud. In this work, we focus on addressing data privacy issues using SSE. The concept was formulating the privacy issues of the stored data and to retrieve the data from the cloud in a sequence manner. We observe that server-side ranking based on order-preserving encryption (OPE) inevitably leaks data privacy. By using OPE we find a Boolean search. To avoid the leakage of data, here we propose a two-round searchable encryption (TRSE) scheme that supports top-k multikeyword retrieval. In TRSE, we employ a vector space model and homomorphic encryption. As a result, information leakage can be eliminated and the stored data is secured. This is proposed scheme guarantees high security and practical efficiency.

KEYWORDS – SAAS, PAAS, IAAS, SSE, TRSE, OPE.

I. INTRODUCTION

The main threat on data privacy roots in the cloud itself [6]. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud at will, lawfully or unlawfully. To ensure privacy, users usually encrypt the data before outsourcing it onto cloud, which brings great challenges to effective data utilization. However, even if the encrypted data utilization is possible, users still need to communicate with the cloud and allow the cloud operates on the encrypted data, which potentially causes leakage of sensitive information. Furthermore, in cloud computing, data owners may share their outsourced data with a number of users, who might want to only retrieve the data files they are interested in. One of the most popular ways to do so is through keyword-based retrieval. It is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance's are sent back to users. A series of searchable symmetric encryption (SSE) schemes have been proposed to enable search on ciphertext. whether a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. To improve security without sacrificing efficiency, schemes presented in [9], [10], [24] show that they support top-k single keyword retrieval under various scenarios. The authors of [25], [26] made attempts to solve the problem of top-k multikeyword over encrypted cloud data.

II. LITERATURE REVIEW

A lot of research is done in cloud data security with number of techniques.

In the work [1], describes the cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. In the work [3], they are considering the problem of searching on data that is encrypted using a public key system. Consider user Bob who sends email to user Alice encrypted under Alice's public key. An email gateway wants to test whether the email contains the keyword \urgent" so that it could route the email accordingly. In the work [5], introduces a new framework for confidentiality preserving

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.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

rank-ordered search and retrieval over large document collections. In the work [9], 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 has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. In the work [10], Query processing that preserves both the data privacy of the owner and the query privacy of the client is a new research problem. It shows increasing importance as cloud computing drives more businesses to outsource their data and querying services. However, most existing studies, including those on data outsourcing, address the data privacy and query privacy separately and cannot be applied to this problem. We believe this work steps towards practical applications of privacy homomorphism to secure query processing on large-scale, structured datasets. As for future work, we plan to extend this work to other query types, including top-k queries, skyline queries and multi-way joins.

III. PROPOSED SYSTEM

A. OBJECTIVE

This chapter deals with the objectives of the project. The objectives are

- To store the Encrypted data on cloud.
- Retrieve the Encrypted Cloud Data Using Multi-Keyword.

In this work, introduced the concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve the insecurity problem by proposing a two-round searchable encryption (TRSE) scheme. Novel technologies in the cryptography community and information retrieval (IR) community are employed, including homomorphic encryption and vector space model. We propose the concepts of similarity relevance and scheme robustness. We, thus, perform the first attempt to formulate the privacy issue in searchable encryption, and we show server-side ranking based on order-preserving encryption (OPE) inevitably violates data privacy. We propose a TRSE scheme, which fulfills the secure multikeyword top-k retrieval over encrypted cloud data. Specifically, for the first time, we employ relevance score to support multikeyword top-k retrieval. Thorough analysis on security demonstrates the proposed scheme guarantees high data privacy. Furthermore, performance analysis and experimental results show that our scheme is efficient for practical utilization.

B. SYSTEM ARCHITECTURE

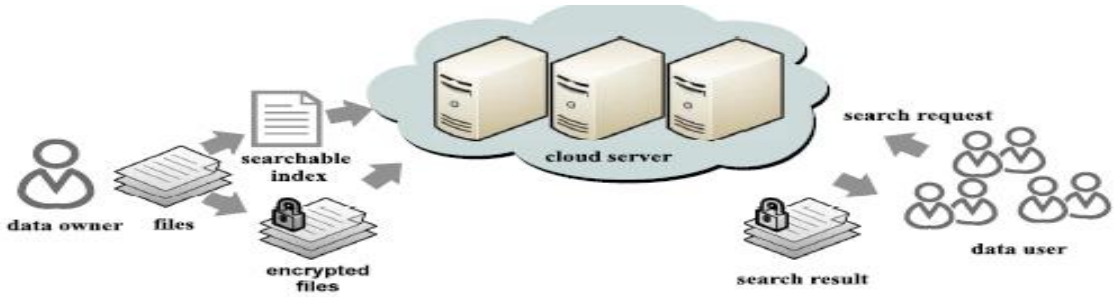

Fig. Components & Architecture of System

The various components present in the architecture of proposed system are actual user stores the data on cloud. Cloud server stores the encrypted data and searching indexes. Data user retrieves the file from the cloud server using top-k multikeyword ranked search.

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.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

C. DATAFLOW DIAGRAM

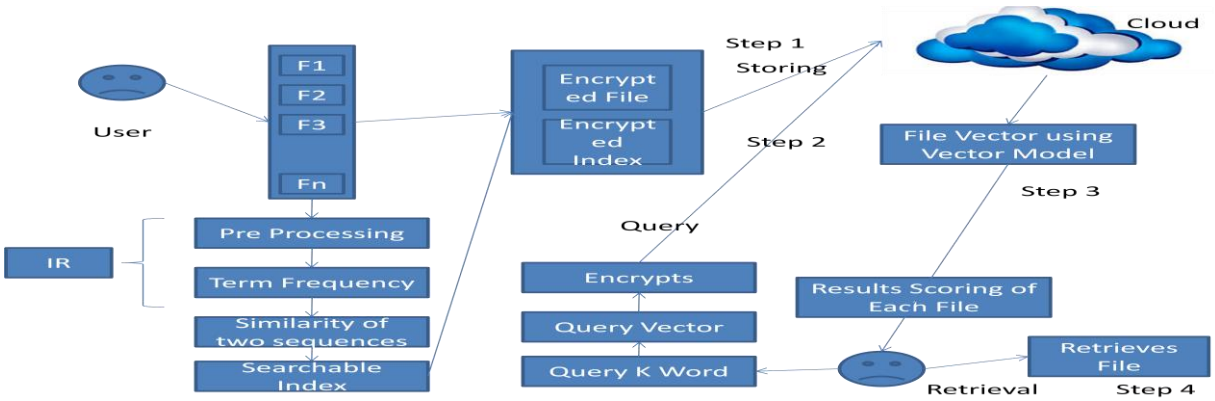

Fig: Dataflow diagram of the data retrieval model

D. MODULES DESCRIPTION

The Proposed Scheme TRSE consists of four main modules

- Setup Phase
- IndexBuild Phase
- TrapdoorGen Phase
- Score Calculate Phase
- Rank Phase

Setup Phase

- Cryptography Community used
- Generates Secret Key
- Encrypts the storing data with the Secret key
- And also generates the Public Key

Index Build Phase

The data owner builds the secure searchable index from the data.

- To form search index Information Retrieval community (IR) is used.
- Using the collection of files, extracts collection of terms or words.
- Applies similarity between the words.
- Encrypts the search index with similarity using the Secret Key.
- Both the encrypted data and encrypted search index are stored in the cloud.

TrapdoorGen Phase

- The User request for data in the cloud by using query with keywords.
- The keywords are formed as vector query such as 0 and 1

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.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

- The vector query is generated as secure trapdoor using homomorphic encryption.
- The encrypted vector query is passed to the cloud.

Score Calculate Phase

- Receives the request from the user by the cloud
- Cloud performs Scores for each file in the cloud using the requested query
- To perform score applies Vector Model
- Results the Encrypted Result Vector
- Returns back to the user.

Rank Phase

- The user receives the encrypted Vector.
- Applies homomorphic decryption which results the top k scores of related files.

IV. IMPLEMENTATION

TRSE Design

Existing SSE schemes employ server-side ranking based on OPE to improve the efficiency of retrieval over encrypted cloud data. However, server-side ranking based on OPE violates the privacy of sensitive information, which is considered uncompromisable in the security-oriented third party cloud computing scenario, i.e., security cannot be trade-off for efficiency. To achieve data privacy, ranking has to be left to the user side. Traditional user-side schemes, however, load heavy computational burden and high communication overhead on the user side, due to the interaction between the server and the user including searchable index return and ranking score calculation.

Framework of TRSE

The framework of TRSE includes four algorithms: Setup, indexbuild, trapdoorgen; score calculate, and Rank.

Setup

The data owner generates the secret key and public keys for the homomorphic encryption scheme. The security parameter lambda is taken as the input, the output is a secret key SK, and a public key set PK.

IndexBuild(C,PK)

The data owner builds the secure searchable index from the file collection C. Technologies from IR community like stemming are employed to build searchable index I from C, and then I is encrypted into I0 with PK, output the secure searchable index I'.

TrapdoorGen(REQ; PK)

The data user generates secure trapdoor from his request REQ. Vector T' is built from user's multikeyword request REQ and then encrypted into secure trapdoor T' with public key from PK, output the secure trapdoor T'.

ScoreCalculate(T'; I')

When receives secure trapdoor T$, the cloud server computes the scores of each files in I' with T' and returns the encrypted result vector N back to the data user.

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.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Rank(@,SK,k)

The data user decrypts the vector N with secret key SK and then requests and gets the files with top-k scores.

Relevance Scoring

Some of the multikeyword SSE schemes support only Boolean queries, i.e., a file either matches or does not match a query. Considering the large number of data users and documents in the cloud, it is necessary to allow multikeyword in the search query and return documents in the order of their relevancy with the queried keywords.

Vector Space Model

While tf-idf depicts the weight of a single keyword on a file, we employ the vector space model to score a file on multikeyword. The vector space model is an algebraic model for representing a file as a vector.

K- Word (in Bytes)	Time (in MS)
5	0.4
10	0.65
15	0.8
20	1
25	1.02

Table: Computation Time Cost for K-Word Retrieval

Homomorphic Encryption

Homomorphic encryption is a form of encryption which allows specific types of computations to be carried out on ciphertext and obtain an encrypted result which decrypted matches the result of operations performed on the plaintext.

Unpadded RSA

If the RSA public key is modulus m and exponent e, then the encryption of a message x is given by $\mathcal{E}(x) = x^e \bmod m$. The homomorphic property is then

$$\mathcal{E}(x_1) \cdot \mathcal{E}(x_2) = x_1^e x_2^e \bmod m = (x_1 x_2)^e \bmod m = \mathcal{E}(x_1 \cdot x_2)$$

V. EXPERIMENTAL RESULTS

Efficiency Improvement

The main appeal of the modified FHEI that we employ in the TRSE scheme is its conceptual simplicity compared to Gentry's. This simplicity is achieved at the cost of a large key size. Although optimizations like modular reduction and compression can be employed to reduce the size of cipher text, the key size is still too large for the practical system.

As discussed in Section 5, the user encrypts his trapdoor and sends the cipher text to the cloud server. Therefore, the communication overhead will be very high if the encrypted trapdoor size is too large. To solve this problem and, thus, improve efficiency, a tradeoff 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 have made several attempts to move toward practical fully homomorphic encryption over integers. These

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.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

progresses indicate that the efficiency of the TRSE scheme can be further improved. To allow for ranked keyword search, an ordinary inverted index attaches a relevance score to each posting entry. Our approach replaces the original scores with the ones after one-to-many order-preserving mapping. Specifically, it only introduces the mapping operation cost, additional bits to represent the encrypted scores, and overall entry encryption cost, compared to the original inverted index construction.

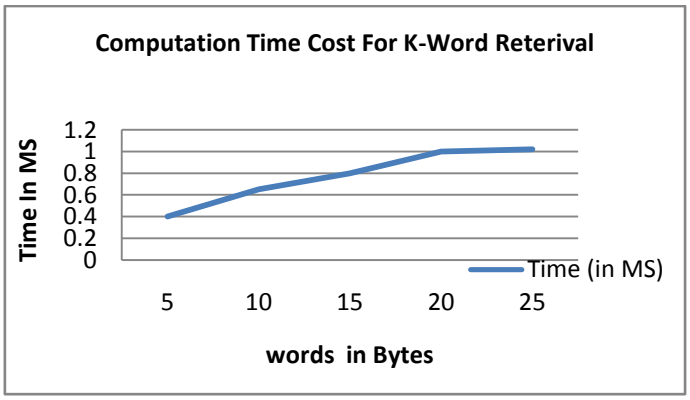

Fig: Computation time cost for k-word retrieval

VI. CONCLUSION

We motivate and solve the problem of secure multikeyword top-k retrieval over encrypted cloud data. We define similarity relevance and scheme robustness. Based on OPE invisibly leaking sensitive information, we devise a server-side ranking SSE scheme. We then propose a TRSE scheme employing the fully homomorphic encryption, which fulfills the security requirements of multikeyword top-k retrieval over the encrypted cloud data. By security analysis, we show that the proposed scheme guarantees data privacy. According to the efficiency evaluation of the proposed scheme over a real data set, extensive experimental results demonstrate that our scheme ensures practical efficiency.

REFERENCES

[1] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.

[2] N. Howgrave-Graham, "Approximate Integer Common Divisors," Proc. Revised Papers from Int'l Conf. Cryptography and Lattices (CaLC' 01), pp. 51-66, 2001.

[3] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, "Public- Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (Eurocrypt), 2004.

[4] O. Regev, "New Lattice-Based Cryptographic Constructions," J. ACM, vol. 51, no. 6, pp. 899-942, 2004.

[5] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A.L. Varna, S. He, M.Wu, and D.W. Oard, "Confidentiality-Preserving Rank-Ordered Search," Proc. Workshop Storage Security and Survivability, 2007.

[6] 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.

[7] 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, 2010.

[8] S. Gries, "Useful Statistics for Corpus Linguistics," A Mosaic of Corpus Linguistics: Selected Approaches, Aquilino Sanchez Moises

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.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Almela, eds., pp. 269-291, Peter Lang, 2010.

[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data," Proc.IEEE INFOCOM, 2011.

[10] H. Hu, J. Xu, C. Ren, and B. Choi, "Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism," Proc. IEEE 27th Int'l Conf. Data Eng. (ICDE), 2011.

[11] M. Perc, "Evolution of the Most Common English Words and Phrases over the Centuries," J. Royal Soc. Interface, 2012.

[12] Jiadi Yu," Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data" IEEE transactions on dependable and secure computing, vol. 10, no. 4, July/August ,2013.