International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 2, Issue 2, February 2015

Implementing an Efficient Procure Technique Over Cipher Data Using Multikeyword Ranked Retrieval Search.

[1] Mr. Vijay S. Gulhane, [2] Mr. Vilas D. Ghonge
[1] Department of IT Sipna College of Engg. & Tech. Amravati , IndiaE-mail : v_gulhane@rediffmail.com,
[2] Mr. Vilas D. Ghonge Department of CSE Sipna College of Engg. & Tech. Amravati , IndiaE-mail :vilasghonge77@gmail.com.

Abstract: **Cloud Computing is an one of the emerging computing technology that uses the internet and control remote servers to maintain data and its application. For the protection of data it is very essential to outsource data and encrypt it, in order to protect it from unauthorized user. Here we are going to propose an efficient, secure and fast data searching technique which will help us to handle data efficiently in cloud storage or server. The indexing scheme is used in order to provide fast data retrieval while ensuring the security. The indexing and ranking scheme for displaying the data will improve the efficiency of searching. The security can be increased by encrypting each of the documents with different keys. Order preserving encryption technique based on ranking has some data leakage problem , In order to avoid this problem multikeyword based data retrieval scheme is to be proposed. It helps to the user to retrieve relevant data or files in which they are interested in. Since the search operation is performed over encrypted data, information leakage can be eliminated and data can be searched and retrieved efficiently.**

Index Terms— **Cloud Computing, Data Retrieval, Encrypted Data, Multikeyword Search.**

I. INTRODUCTION

Cloud computing provides high-performance computing and almost unlimited storage services. It helps to reduce the cost of managing large amount of data. However, the service providers are generally unreliable and the data confidentiality and integrity cannot be protected in Cloud. 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 cipher text , 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] show that they support top-k single keyword retrieval under various scenarios. The authors of [12] made attempts to solve the problem of top-k multikeyword over encrypted cloud data.

I. LITERATURE REVIEW

In the work [1], describes the cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting cryptosystems. In the work [3], they are considering the problem of searching on data that is encrypted using a public key system. Consider user A who sends email to user B encrypted under B'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 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.

EXISTING SYSTEM

To search a file in the internet, we make a query to the internet server. Internet will retrieve the most number of visited files which is called as number of Hits. Till now any Search engine will retrieve the links to the user based on the frequent number of Clicks or Hits made by the user. So ranking proves is achieved using this methodology only. Even some times irrelevant data would be ranked for the user which is of no use. 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. Keyword-based retrieval is a typical data service and widely applied in plaintext scenarios in which users retrieve relevant files in a file set based on keywords. However, it turns out to be a difficult task in cipher text scenario due to limited operations on encrypted data.

LIMITATION OF EXISTING SYSTEM

- Existing system support queries based on a file either matches or does not match a query i.e it is to be based on Boolean queries.
- Data storage on cloud without any proper technique will increase the risk to the data from theft by unauthorized user.

II. PROPOSED SYSTEM

This paper mainly deals with the two concepts, to store the Encrypted data on Cloud and retrieve the Encrypted Cloud data using Multi-Keyword.

In this paper, we 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 scheme. 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 inevitably violates data privacy. We have to propose a MRSE scheme, which fulfills the secure multikeyword top-k retrieval over encrypted cloud data.

ADVANTAGES OF PROPOSED SYSTEM

- This system improves the basic privacy requirement of data.
- It improves the efficiency of a system by low communication and computation overhead.
- Proposed cloud storage system provide basic privacy requirement of data i.e confidentiality, integrity and availability.

III. SYSTEM ARCHITECTURE

Cloud data storage service involves three parts 1.Data Owner Module 2.Data User Module and 3.File Uploaded Module (Server Module) as shown in Fig. Data Owner stores a set of document D on to the cloud server in an encrypted

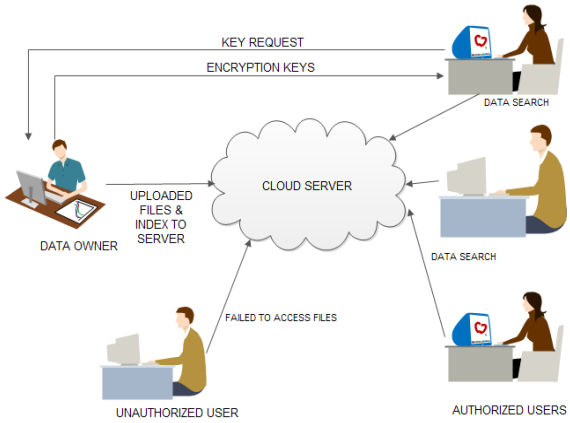

Fig.1. Components and Architecture Of System.

form to avoid the security threats. To enable fast and cost effective data retrieval a search index I is built over an encrypted data. To make a search, a set of keywords K is given by an authorized user. The results are ranked using the ranking algorithm by the cloud server.

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.

IV. RELATED WORK

The Proposed System consist of three Modules

1) Data Owner Module
2) Data User Module
3) File Uploaded Module (Server Module)

1) Data Owner Module

Data Owner stores a set of document D on to the Cloud Server in an encrypted form to avoid security threat .

2) Data User Module

To make a search a set of keywords 'k' is given to an authorized user.

3) File Uploaded Module (Server Module)

In order to provide fast & cost effective data search & retrieval a search index is built over an encrypted data onto the Cloud Server. The architecture mainly consist of data owner, data user and cloud Server. There is a registration page for both data owner and data User. They have to sign up by registering their username and password. After sign up data owner and data user get their unique ID. A user can perform search and retrieval over his owner's data files only after getting permission from data owner. There is a login page for data owner/user where they can either sign up or sign in to the system. Data owner enters his username & password and logs into the system. Data Owner also select a set of data files and prepare the word-list. Based upon the word-list a searchable index is prepared. The data files are encrypted using AES and index is encrypted. Then these encrypted index and data files are uploaded to Cloud Server. A data user who logs in by entering his username and password can do search and retrieval. User can enter the search query on provided space. This is send to server. The server calculates scores of each files and send the encrypted score information to data user. On receiving, the data user decrypts the scores and picks out the top-k highest scoring files identifiers, and sends file-request to server. The server responds with requested encrypted files. Users on receiving, can decrypt it and make use of them. There are a set of data owners and some groups of users are associated with each data owner. A data user must be under the authorization of any of the data owner. This ensures that only authorized users are making use of outsourced data. There is an authorization page in system. For the user who sign up for first time, he should select a data owner and send a registration request to concerned owner.

The user is said to be authorized only after he gets permission from data owner. After authorizing a data user, the owner enter that information into database. This user can consume data only from this selected data owner. Data owners generate keys and are transferred to authorized user when such users logs into the system. So only users who are authorized by corresponding data owners are able to access the keys. Keys should be properly managed in order to give flexibility to users. Also these keys have to be protected from attackers. So in database, keys are stored in encrypted form. The key for encrypting/decrypting key-set are called Master key. Only data owners know this master key which ensures privacy. Only authorized user gets the key from owner so that he can perform search and retrieval. Here the files are ranked in the order of relevance by users interest and only the files with the highest priority are sent back to users.

FLOW OF PROJECT

Data owner stores a set of document D onto the cloud server in an encrypted form to avoid security threat .To enable fast and cost effective data retrieval a search index 'I' is built over an encrypted data onto the cloud .To make a search a set of keywords 'k' is given to an authorized user in order to access or retrieve that information. The encryption and decryption process is to be carried out with the help of AES (advanced encryption standard) algorithm. AES provides 128 bit encryption in order to protect the privacy of data which is to be highly secure.

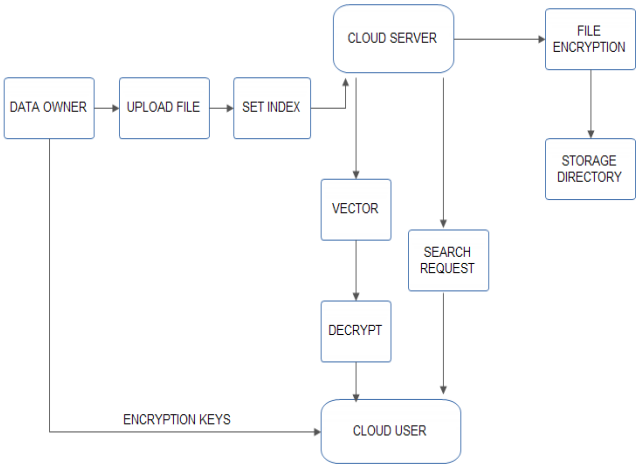

Fig.2. Flow of Project

V. CONCLUSION

In this paper we are trying to propose a system which is able to provide multikeyword ranked retrieval search over encrypted cloud data which improves the basic privacy requirement of data .It also provides efficiency of a system by low communication and computation overhead. Hence it will enable fast similarity search over encrypted data without compromising the security. The multikeyword ranked retrieval search, document ranking and AES encryption technique will provide more efficiency compare to existing ones.

REFERENCES

[1] 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," Proceeding Workshop Storage Security and Survivability, 2007.

[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proceeding IEEE 30th International Conference Distributed Computing Systems (ICDCS), 2010.

[3] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, "Public- Key Encryption with Keyword Search," Proceeding International Conference Theory and Applications of Cryptographic Techniques (Eurocrypt), 2004.

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

[5] H. Hu, J. Xu, C.Ren, and B. Choi, "Processing Private Queries over Untrusted Data Cloud through Privacy Homomorphism," Proceeding IEEE 27th International Conference Data Engineering (ICDE), 2011.

[6] 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.

[7] M. Perc, "Evolution of the Most Common English Words and Phrases over the Centuries," J. Royal Societies Interface, 2012.
[8] M.van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers," Proceeding 29th Annual International Conference Theory and Applications of Cryptographic Techniques, 2010.
[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data," Proceeding IEEE INFOCOM, 2011.
[10] N. Howgrave-Graham, "Approximate Integer Common Divisors," Proceeding Revised Papers from International Conference Cryptography and Lattices (CaLC' 01), pp. 51-66, 2001.
[11] O. Regev, "New Lattice-Based Cryptographic Constructions," J. ACM, vol. 51, no. 6, pp. 899-942, 2004.
[12] S. Gries, "Useful Statistics for Corpus Linguistics," A Mosaic of Corpus Linguistics: Selected Approaches, Aquilino Sanchez Moises