International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 11
November 2017

Enabling Multi keyword Search on Secure Encrypted Data using Multi Cloud Approach

Dr. Mamatha G[1], Dr. Ramesh Hegde[2], Shilpitha Swarna[3], Lakshmikanthaiah SM[4]

[1]Professor, Department of ISE, Acharya Institute of Technology, Bangalore

[2]Professor & Head, Department of MCA, Acharya Institute of Technology, Bangalore

3Assistant Professor, Department of MCA, Acharya Institute of Technology

[4]Senior Software Engineer,

Abstract—Cloud computing provides lot of benefits to enterprises to offload their data and software services to cloud saving them lot of money that has to be spent on infrastructure setup cost. Enterprises wanted to offload their data to cloud and save on their infrastructure cost. But when offloading the data security and privacy is a important concern. When the data offloaded to cloud by a enterprise is compromised, the enterprise will lose its business. Similarly when the hospitals and health care organization upload their patients details to cloud and when the data is comprised, it will affect the privacy of the patients. So privacy is a important concern when offloading the data to cloud. Most of solutions for privacy is based on encryption and data to be offloaded is encrypted and stored in cloud. Algorithms like AES, DES etc are used for encrypting the data before offloading to cloud. But the side effect in this encryption mechanism is that, the encrypted data is not order preserving and it is not suitable for searching and ranking. To solve it in work [1], author has proposed homomorphic encryption based mechanism. But the solution suffers from capture attack and relies on semi trust model on cloud. In this paper, we discuss the capture attack in detail and propose a solution based on multi cloud to avoid the same.

Keywords: Cloud computing, MRSE Multi-Keyword Ranked Search, , Privacy Preserving, Confidential Data, homomorphic encryption.

I. INTRODUCTION

Cloud computing provides lot of benefits to enterprises to offload their data and software services to cloud saving them lot of money that has to be spent on infrastructure setup cost. Enterprises wanted to offload their data to cloud and save on their infrastructure cost. But when offloading the data security and privacy is a important concern.

Nowadays there is lot of cloud service providers in the market. The two requirements of security and privacy are the important decision criteria for enterprises in choosing the cloud service providers. For security many encryption protocols were proposed which encrypt the enterprise data and offload to cloud, so that for other cloud users the data is difficult to decrypt and make sense.

Cloud service providers encrypt the enterprise data and save in cloud. But the problem is when some enterprise user want to search for some documents based on keyword, it becomes difficult for him to download all documents, decrypt and search. The cost of downloading the documents and decrypting them to search every time is costly. Also the time needed for search is high in this way.

In the paper work [1], author has proposed a solution called MRSE for multi keyword search on cloud. Among various multi-keyword semantics, authors choosed the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents. Specifically, they use "inner product similarity" [4], i.e., the number of query keywords appearing in a document, to quantitatively evaluate the similarity of that document to the search query in "coordinate matching" principle. During index construction, each document is associated with a binary vector as a sub-index where each bit represents whether corresponding keyword is contained in the document. The search query is also described as a binary vector where each bit means whether corresponding keyword appears in this search request, so the similarity could be exactly measured

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 11
November 2017

by inner product of query vector with data vector. However, directly outsourcing data vector or query vector will violate index privacy or search privacy. To meet the challenge of supporting such multikeyword semantic without privacy breaches, we propose a basic MRSE scheme using secure inner product computation, which is adapted from a secure k-nearest neighbor (kNN) technique [4], and then improve it step by step to achieve various privacy requirements in two levels of threat models.

The problem with this approach is that the search result can be attacked and ranking be modified. Say an example of user searching for hotels nearby a city, the search results comes in rank of closet distance, an attacker in middle can modify the search result to move a selected hotel up in order even though it was not in top 10 or so. In this project the above problem discussed and propose a effective solution against this integrity attack.

II. MRSE AND PROBLEMS

In MRSE, for data privacy the document is encrypted using traditional symmetric key cryptography and then outsourced to cloud for storage. For the case of searchable encryption, a index is computed from the documents and then outsourced to cloud for storage. [5] The index is computed in a way so that it is difficult for any attacker to deduce the contents of documents.

The search query is also encrypted and provided to cloud for search and the search result is also provided in encrypted form to be decrypted at user end.

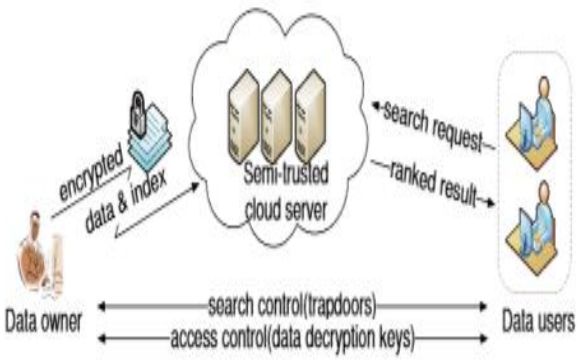


Figure 1: The architecture of MRSE.

The solution suffers from following problems

1. Index construction must be done at user end
2. Search query encryption must be done at user end
3. Search result decryption must be done at user end.
4. Cloud is expected not to launch capture attack on query and infer matching the search terms to document. It is expected to be semi trust.

Problem 1 to 3 becomes critical for access from mobile devices as lot of computing power is needed for these operations.

III. MULTI CLOUD SOLUTION

The proposed solution solves the problems mentioned in above section using multi cloud solution.

In the proposed multi cloud solution, two clouds are used. In one cloud the index construction, encryption and decryption are done at one cloud and storage of encrypted data is at another cloud.

For the sake of explanation, the two cloud as **search cloud** and **storage cloud.**

- Search cloud does the work of index computation, search query encryption and search result decryption.
- Storage cloud store the data and it can retrieved at any time.

Data owner first send the file to search cloud for index computation. Search cloud does the index computation as follows.

1. Extract terms in the document
2. Build term frequency
3. Add term, term frequency, file name to the index.

Data owner then encrypts the file with encryption key, renames the file and sends to search cloud. Search cloud split the encryption file to shares according to Shamir secret sharing and distributes each share[Say total N shares] as separate file to the storage cloud. By this way the data is completely safe at storage cloud.

Data user who wants to search for keyword, provides the keyword to the search cloud, when the data user has the required access for search, the search cloud process the search query in following way.

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 11
November 2017

1. Search the index for the matching keyword

Retrieve the matching file name.

2. From the N shares, it randomly chooses the M shares.

3. Retrieve those M shares file from the storage cloud.

4. Assemble the shares and send to data user.

5. In case of multiple files matching, based on TF frequency order of matching keyword, files retrieved from cloud are sent to the data user.

Once the data user receives the files, he needs the key for decryption of file, which is handled in a offline manner which being out of scope of this paper work. After getting the keys, he decrypts the file.

The advantages in the proposed solution are as follows

1. Index need not be encrypted like in MRSE, so lot of computation is saved.

2. Search query is not encrypted based on index, so lot of bandwidth is saved in sending of encrypted search query to cloud .

3. The time for matching in index is comparatively very less compared to MRSE as the search keyword size is small when compared to encrypted search keyword in MRSE.

4. The storage cloud does not now about shares and whether they belong to same file. Also every time different shares are retrieved by search cloud to reassemble the shares, so it difficult to launch any inference attacks.

5. The file is renamed by owner before sending to search cloud, so even if the search cloud is compromised and the index is taken, the file name in the index are not real name and so no privacy breach.

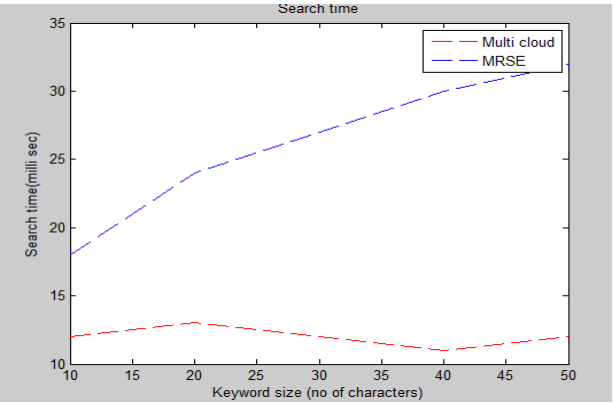
IV. RESULTS

The implemented the proposed system using two clouds. Amazon cloud was used for search cloud and Microsoft azure was used storage cloud.

For files are different sizes, files are uploaded according to proposed solution in different batches and then searched for keywords of different length. We measured the following parameters.

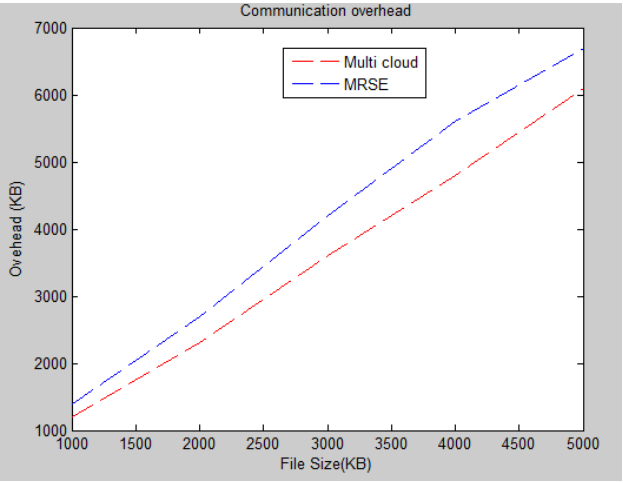
1. Search time
2. Communication overhead
3. Data upload time

Search time is measured for keywords of different sizes for the proposed multi cloud solution and the MRSE and the results are plotted below



From the results, we see that search time is less in the multi cloud compared to MRSE. The reason being encryption of search query and index searching with encrypted search query is avoided.

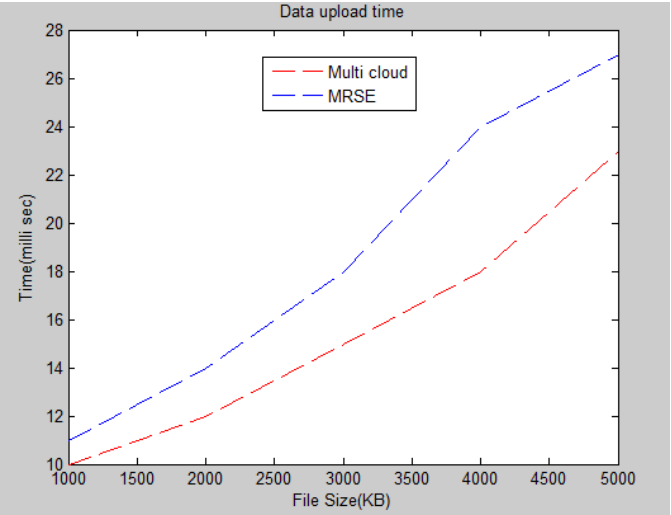
Communication overhead is measured in terms of number of bytes consumed during data upload and search operations. It is measured for different size of file uploads.



From the results communication overhead is less in the multi cloud solution. The reason being index is not upload into network and the search query size is less.

Data upload time is measured for different file size and the result is plotted below

International Journal of Engineering Technology Science and Research
IJETSR
www.ijetsr.com
ISSN 2394 – 3386
Volume 4, Issue 11
November 2017



From the results, the data upload time is comparatively less in multi cloud, because during the data upload the encrypted index construction and upload of encrypted index to cloud is avoided in the proposed multi cloud solution.

V. CONCLUSION AND ENHANCEMENTS

In this paper, the performance issues in the MRSE scheme and proposed multi cloud solution to enable secure search on encrypted data with low complexity. The paper also prove through performance results that the communication overhead is reduced and search time is reduced in the proposed approach. The Shamir share computation can be moved to data owner device if the device is having sufficient computing power, in this case the solution if full proof against un-trusty search cloud.

REFERENCES

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50–55, 2009.

[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS. Springer, Heidelberg.

[3] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, no. 4, pp. 35–43, 2001.

[4] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.

[5] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of S&P, 2000.

[6] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, 2003, http:// eprint.iacr.org/2003/216.

[7] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.

[8] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.

[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYPT, 2004.

[10] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficiently searchable encryption," in Proc. of CRYPTO, 2007.

[11] M. Abdalla, M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange, J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, "Searchable encryption revisited: Consistency properties, relation to anonymous ibe, and extensions," J. Cryptol., vol. 21, no. 3, pp. 350–391, 2008.

]12] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.