International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

A Survey on Attribute Based Encryption to Improve Efficiency in Semantic Search Over Cloud Data

Vina M. Lomte[1], Gauri S. Patil[2]

[1]Assistant Professor, Dept. of Computer Engineering, RMD Sinhgad School of Engineering, Pune, Maharashtra, India

[2]M.E. Student, Dept. of Computer Engineering, RMD Sinhgad School of Engineering, Pune, Maharashtra, India

Abstract: ***Mobile cloud computing provides the capability of sharing of the data that is encrypted with different user through cloud Storage. This raises the point of security issues over data confidentiality and authentication access control. The blind storage allows a client to store the number of files on remote server, where remote server is not familiar with the files that are stored in it. In Attribute based encryption (ABE), the encryption is performed on the attributes and allows only those users to access the data if the attributes are available in their ID. This paper proposes the semantic search for encrypted multi-keyword ranked over the cloud data. The multi-keyword ranked scheme can search based on the results obtained in ranked search with efficient accuracy by using k-nearest neighbor technique. It also uses ABE technology with modification that minimizes the limitation of ABE. To make the search efficient blind storage system is used to hide access control issues in searchable encryption technique. The security analysis scheme is used to achieve authentication and confidentiality of the document.***

Keywords: Cloud computing, searchable encryption; attribute based encryption (ABE), multi-keyword ranked search, blind storage, access pattern**.**

1. Introduction

Cloud computing is an emerging technology that has been increasingly used for storing and retrieving large amount of data over the internet. It allows the user to save the copy of their personal data such as photos, documents, multi-media data, etc over the cloud environment. Cloud also allows the co-operate people to store their data at very low cost and provide access storage space for the data. Since most confidential data are stored in the cloud, it needs more confidentiality and privacy for the data owners [1][3]. And hence its security issues must be taken in to the consideration. In order to avoid the loss of data, it is necessary to encrypt the user data before storing the data in the cloud environment. The encrypted data stored in the cloud can be accessed among the number of the data users; in turn the data users need to decrypt the data in the cloud by getting the shared key from the data owner. The key factor of using cloud is to store the user data securely at the possibly low cost. The data owner and cloud server are not in the same trusted domain can put the encrypted data in the cloud at the risk. The cloud server may provide data information to unauthorized entities or even loosed the data confidentiality; hence data in the cloud need to be secured from unauthorized access. The cloud can be deployed in any of these three types as Public, Private and Hybrid clouds. It depend upon the requirement of an organization terms, necessary cloud deployment model can be selected even though a general security term is adopted.

Cloud mainly provides three types of services Software as a service (SAAS), Platform as service (PAAS) and Infrastructure as a service (IAAS).

In concern with the cloud computing, the first scene that comes into picture is the security of data. Every data owner performs encryption over their data before storing it to the cloud as a measure for the security purpose. However, this make searching of any data over encrypted cloud unexpectedly time consuming and complex job. Thus so are proposing the scheme [5] to search through the encrypted data over cloud. . However, allowing cloud service providers (CSPs), which are not in the same authorized trusted domains as enterprise users, to take care of confidential and sensitive data, may raise the security and privacy issues. To keep the sensitive data of the user confidential against entrusted CSPs, a normal way is to apply cryptographic approaches**.** To make this possible the necessary and effective approach is carried out. There are three entities:

- Data owner
- Cloud Service Provider (CSP)
- End user

The data owner will upload the documents over cloud, after applying it will also create document index with keywords. It will share the private key to decrypt the file with end users. The data owner will also create document index and will upload its encrypted version on cloud. Hence in carrying out this, following things are required to be achieved:

- The document privacy that even cloud should not get the actual data
- The efficient search mechanism.

2. Related Work

Searchable encryption is a technique that allows the search to be carried out on the over the encrypted cloud data. The technique is classified as searchable public key encryption (SPE) and searchable symmetric encryption (SSE).The first SSE scheme is introduced by song et al [1] provides only single keyword search in the encrypted data. Naveed et al [2] proposed a blind storage system in order to achieve a searchable encryption since it supports only single keyword

search the encrypted data stored in the cloud. Block cipher AES algorithm has been used for carrying the encryption and decryption of the data over the cloud. Cong Wang [3] is a traditional method that has solved the problem of Boolean search of searching technique that come across the effective data utilization. This paper has provided with the strong and secured security guarantee. However it determines relevance score from information retrieval to construct a secure searchable indexing of keyword and develop a one-to-many order preserving mapping technique to securely protect those sensitive score information. Ningcao [4] introduced an efficient and effective similarity measure of coordinating matching that outsource as much as possible data that is concerned to the multiple keywords given by the data user. Hongwei LI [5] has focused to have semantic search efficiently over the encrypted cloud data. Cong Wang [6] has been provided with retrieving of matching files as per its ranked order as per its relevance criteria and it can be carried out based on the indexing. This will allow the searching of the relevant documents using its keyword. Bing Wang [7] introduces the solution for spelling error during the keyword search. The proposed scheme obtains the fuzzy keyword matching through algorithmic design instead of expanding the index file. It also eliminates the need for predefined dictionary and effectively supports multi-keyword fuzzy search of keywords without increasing the complexity or index file. Wenham sun [8] introduced a tree based index structure and multi-dimensional algorithm in order to improve the search efficiency in the practical world. The vector space model with effective cosine similarity and search index has been used in order to support multi-keyword search and ranked efficiency. Hong wel li [9] utilize the relevance score and K-nearest neighbor techniques in order to develop an efficient multi-keyword search that return the ranked search result that are based on accuracy. This multi-keyword search efficiency has been used in the blind storage in cloud system in order to bring it out in the access pattern. Though the SSE helps to increase the computation, but the security of the data is not sure since the shared key is used for both sender and receiver. Jiadi yu [11] overcomes the problem of Boolean keyword search by using two round searchable encryption(TRSE) which retrieves top K data that is related to the given keyword communication overhead is reduced by using vector space model and homomorphism encryption. Diffie helman algorithm is used for encrypting and decrypting the data. The vector space model provides sufficient search accuracy and homomorphism encryption that allows the user to involve in ranking. Curtmola [12] proposed sharing the key in the group of users. It computes more time for sharing of the particular data among the group of user. Thus Baojiang [13] solves the problem of key sharing in order to preserve the confidentiality of the data among the group of users. The traditional method uses single key for retrieving every document sharing in a group, key aggregate searchable encryption method has been proposed in which single user vector gets the single aggregate key is in a group to retrieve all the documents that is outsourced by the data owner this will reduce space in the storage, complexity and provides secure and authorized communication. Zhi-Hua Zhang [14] presents an identity-based authentication scheme for various applications like e-business etc; this paper avoids the issue of revocation and key escrow and sharing problem in the authentication scheme based on the public key certificate. Qin Liu [15] addresses two issues privacy and efficiency and decrease the communication cost. Based on Aggregation and distribution layer (ADL) a middleware layer between the data user and the cloud, the paper presents a scheme, termed efficient information retrieval (IR) for ranked query (EIRQ) in order to reduce querying costs occurred in the cloud.

3. Architectural Model

The architecture of the system is described in detail in figure 1; in this the data owner uploads the data in the cloud with the help of Blind storage. Before uploading the data, it is encrypted. The any document stored in the cloud is always in the encrypted form. The text mining process is a natural language processing used to retrieve the file and Word net tools is used to obtain the files and contents. Natural Language Processing (NLP) process is used to extract the semantic words in file content.

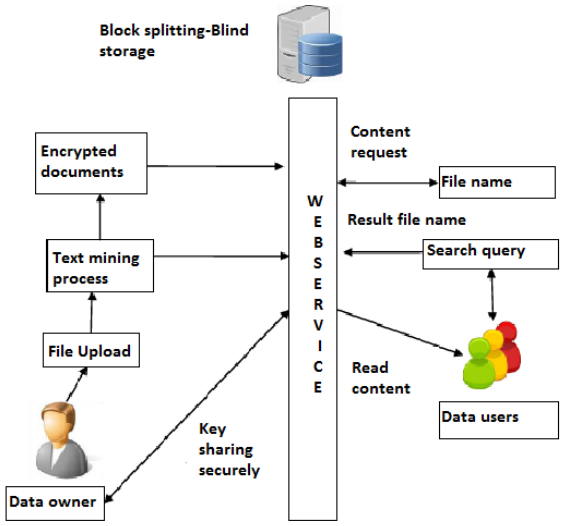

Figure 1: Overall Architecture

Data user tries to search a query in the cloud server. The cloud server will perform the mapping of the keywords and search the related files. The cloud server gives the related filename to user based on the mapping. In order to view the content, the user click the filename, on clicking the file name user actually request to cloud server and in response the server send the user details and file name to the data owner. Then data owner knows all public key of user so as to it can be used to encrypt data by private key of data user public key and encrypted key is send to the server in turn the server will send that key details to user, then user decrypt the key by using the provided private key. After that the data user gets the private key of data owner and then access the data through blind storage in which the data are stored in fixed size of blocks.

4. Multi-Keyword Ranked Search Efficiency

In order to provide the data, user are provided with the necessary keyword that can be obtained by using the necessary keyword search what the user expert can be given with the help of keyword search. The traditional method used

single keyword search for the retrieval of data from cloud which is needed for the search user. The single keyword search gives only some amount of data what the user need based on the obtained keywords but this will not satisfy the search user. Therefore an efficient multi-keyword search has been proposed. To meet the requirement for practical uses and provide better experience, the EMRS (Efficient Multi-keyword Ranked Search)[5] should not only support multi-keyword but also, the efficient multi-keyword ranked search helps the server to search and get the data relevant to the given keywords obtained by necessary keyword search ranking result to the user. It will return only the queried data that the user is needed in order make better a ranked search that has been proposed which will return the data in the ranked form related to the keyword. To overcome the problem of Boolean search [4] proposed multi-keyword ranked search over encrypted data in the cloud environment. The search index depends on term frequency and vector space model with cosine similarity [8] has been used to obtain the highest search result accuracy. The K-nearest neighbor technique and relevance score technique [9] is used for data retrieval used in the blind storage in the cloud that will retrieve the data that depends on the multiple-keyword queried given by the user.

The search over the encrypted data should support following three functions; the searchable encryption scheme should support multi-keyword search used to retrieve the data and should provide the similar output experienced in searching in the goggle search with different keywords in order to fast identify most relevant search result based on relevance score ,the search user considers cloud server to sort the search result that are obtained in relevance based order that are ranked by relevance to search of the documents. The search efficiency should be preserved for the large data that is to be encrypted and stored in cloud.

The main contribution of the proposed work can be given as following:

- Introduce a relevance score in searchable encryption in order to achieve multi-keyword ranked search over the cloud data.
- Modify the data storage system i.e. blind storage system in the EMRS and solve trapdoor unlink ability problem and bring out the access pattern of the user searching the data in the cloud.
- A complete security analysis to show that EMRS reach high security levels.

5. Preliminaries

5.1 Relevance Scoring

In a cloud when large amount of the data is searched, this search results are retrieved and made in the order of the relevancy among the keywords used to search the document. The proposed system uses TF-IDF technique. In TF-IDF, term frequency is given as tf_t, where *f* refers number of term *t* in a document *f*.

5.2 Secured kNN based search

The k-nearest neighbor scheme, encrypt the two vectors confidentially and compute the Euclidean distance of the two vectors. In this scheme, the secret key (S,M_1,M_2) is generated ,where S is the binary vector a splitting indicator to split plain text vector in to two random vectors that can confused the value of plain vector.

5.3 Attribute-based Encryption

In ABE, cipher texts are created with an access structure that defines the access policy. The attribute embedded in the user attribute key decides the access policy of the user, whether the user can decrypt the data.

5.4 Blind Storage System

In order to achieve the searchable encryption scheme, all the documents and index are stored in the blind storage. The blind storage is constructed on the cloud server that allows the user to add, delete, update the document and it hides the access pattern of search user from the cloud server. The documents in blind storage are divided and stored in the fixed sized blocks that are indexed in sequence by random integers.

6. Security Analysis

The Efficient Multi-Keyword Ranked Search properties are analyzed in terms of the document and index confidentiality, Trapdoor privacy, trapdoor unlinks ability and hiding the access pattern of cloud user.

6.1 Maintain the confidentiality of the documents and index.

Before the documents are stored in the cloud, the documents are encrypted using symmetric encryption. The data user decrypts the document by using an appropriate key. The relevance vector for the document is encrypted in order to maintain the confidentiality of the document and index. The encryption is carried out by using a secret key M_1, M_2 and S. The attributes are encrypted by using ABE technique. Thus cloud server uses index to get a relevance vector without knowing the actual information.

6.2 Preserving privacy of trapdoor

The cloud server should be prevented from getting the actual keywords that is too obtained in trapdoor of the search user that is the keyword information is completely hidden from the cloud server in EMRS. Without the secret key M_1, M2, S and K_φ, the cloud server cannot disturb the privacy of the trapdoor.

6.3 Trapdoor unlink ability

Trapdoor unlink ability is that the cloud server cannot use the information that is associated between any two trapdoor. The association may cause to leak the information of the search

user.

6.4 Hiding the access pattern of the search user

The access pattern is the sequence of the searches the user performed. In the EMRS, the blind storage system is modified in order to hide the access pattern of the search user from the cloud server. Since the document in the cloud is stored in the blocks at a fixed size, the header of block is encrypted with block number j and each descriptor would be different even if they belong to the same document.

7. Proposed System Overview

The data owner stores their data remotely on to the cloud in order to obtain and work with various application and services it demands for. The capability of cloud handling the data so flexibly has attracted number of the business people. Keyword search method is used in order to provide the accurate data to the search users. The result of the multi keyword search depends on the ranked search obtained from the relevance score with accuracy by using K-nearest neighbor techniques. The search efficiency of the data is improved by using the Blind storage system in order to hide the access pattern from the cloud server. The security analysis scheme achieves authentication and confidentiality of the document and the index used for that document. The proposed scheme provide the multi-keyword ranked search scheme that allows accurate, efficient and secure search over encrypted mobile cloud data. Security analysis have given proposed scheme that can effectively achieve confidentiality of documents and index, trapdoor privacy, trapdoor unlink ability, and concealing access pattern of the search user.

7.1. Techniques used in Proposed System

The server of the blind storage system provides two operations; download and upload. The blocks that are downloaded are specified by the list of indices whereas the upload operation is allowed to specify a list of indices of the blocks and data blocks. A blind storage system is build by three polynomial time algorithms on the client side: Bstore.keygen, Bstore.Build and Bstore.Access. Among the three Bstore.Access is an interactive protocol. A security parameter is provided to Bstore.keygen as and input in order to generate the key to perform the encryption and decryption and output as a K_{Bstore} . Note that K_{Bstore} , is required to be independent of the data that to be stored. Bstore. Build takes as input (K_{Bstore} , d_0 , $\{id_i, data_i\})_i^t = 1$ where K_{Bstore} is a key, d_0 is an upper bound on the total number of data blocks that are to be stored in the system, (id_i, $data_i$) are the id and the data of the files that the system to be initialized with, it outputs an array of blocks D to be uploaded to the server. Bstore.Access takes as input a key K_{Bstore}, a file id, an operation specified op € {read, write, update, delete}, and optionally data (if op is write or update).Then it interacts with the cloud server and returns a status message generated by the system and optionally file data. For the update operation, Bstore.Access allows more flexibility first it requires only id as input, and outputs the current size of the file with that ID; then it accepts as input when the size of the file will be after update; then it outputs the current file data, only then requires the new data with which the file will be updated. The attribute based encryption is performed on the attributes and allows only those users to access the data if the attributes are available in their ID. Linear search may not be function on such type of data, so kNN based search needs to be performed. The ranking procedure is performed and needs to the most related document to the user query, this is based on TF-IDF functionality and gives good result.

8. Conclusion and Future Scope

In this paper we have studied the techniques that are used to improve the efficiency in semantic search over the cloud data and propose the multi-keyword ranked search scheme that allows the accurate, efficient and secure search over encrypted mobile cloud data. Security analysis has been carried out to bring the confidentiality of documents and index, trapdoor privacy, trapdoor unlink ability, and hiding the access pattern of the search user. The future scope is to give authentication and access control issues in searchable encryption technique.

References

[1] D. X song, D. Wagner, and A.Perrig, “Practical techniques for searches on encrypted data, “in proc. IEEE symp. Secure. Privacy. May 2000, pp.44-55.
[2] M.Naveed, M.Prabhakaran and C.A.Gunter,”Dynamic searchable encryption via blind storage,” in proc.IEEE symp.secur.privacy, may2014, pp.639-654.
[3] C.Wang, N. Cao, K. Ren, and W. Lou, “Enabling secure and efficient ranked keyword search over outsourced cloud data,” IEEE Trans.Parallel Distrib. Syst., vol.23, no.8, pp. 1467-1479, Aug. 2012.
[4] N Cao, C. Wang, M. Li,K.Ren, and W. Lou, “Privacy preserving Multi-keyword ranked search over encrypted cloud data,” IEEE Trans. Parallel Distrib.Syst., vol, 25, no. 1, pp. 222-223, Jan. 2014.
[5] Hongwei LI, Dongxiao LIU,Y. Dai and Xuemin Shen,”Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage”vol 3,no 1,march 2015.
[6] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, “secure ranked keyword search over encrypted cloud data,” in Proc. IEEE 30th Int. Conf. Distrib.Comput. Syst.(ICDCS), jun.2010, pp. 253-262.
[7] B. Wang, S. Yu, W.Lou, and Y.T.Hou, “privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud,” in Proc. IEEE INFOCOM, Apr./May 2014, pp. 2112-2120.
[8] W. Sun, B .Wang, N .Cao, M. Li, W. Lou, Y. Thomas Hou, H Li, “Verifiable privacy preserving multi-keyword text search in the cloud supporting similarity based ranking”, IEEE Trans.Parallel Distrib. Syst, vol.25, no.11, pp.3025-3034, Nov 2014.
[9] H Li, D Liu, Y dai, T H. Luan and Xuemin “Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage”.

[10] D. Bonch, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. EUROCRYPT, 2004, pp. 506-522.
[11] J. yu, P. Lu, Y. Zhu, G. Xue, and M. Li,"Toward secure multi-keyword top-k retrieval over encrypted cloud data," IEEE Trans.Dependable Secure Compute,vol. 10, no 4,pp.239-250,Jul./Aug. 2013.
[12] R. Curtmola, J. Garay, S. Kamara, R. Ostrpvsky. "Searchable symmetric encryption: improved definitions and efficient constructions," In: Proceeding of the 13th ACM Press, pp. 79-88, 2006.
[13] B. Cui, Z Liu, L Wang, "Key-aggregate searchable encryption for group data sharing via cloud storage," IEEE Transc.on computer. vol.6, no.1.
[14] H. Li, Y. Dai. L. Tian, and H. Yang, "Identity-based authentication for cloud computing," in cloud computing. Berlin, Germany: springer-Verlag, 2009, pp. 157-166.
[15] Q.Liu,C.C Tan, J. Wu, and G.Wang, "Efficient information retrieval for ranked queries in cost-effective cloud environments," in Proc. IEEE INFOCOM, Mar. 2012, pp.2581-2585.

Author Profile

Vina M. Lomte received the B.E. and M.E. Degree in Computer engineering. She is now working with RMDSSOE, Warje, Pune as Asst. Professor. She has experiences of 10 yrs 8 months and her Area of specialization - Web Security & S/W Engg.

Gauri S. Patil received B.E. degree in Computer Science and Engineering in 2014 from AGTI, Dr Daulatrao Aher College of Engineering, Karad and pursuing M.E. in Computer Engineering from RMDSSOE, Warje, Pune.