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. 4, Issue 6, June 2016

Attribute Based Encryption to Improve Efficiency in Semantic Search over Cloud Data

Gauri S. Patil, Vina M. Lomte

M.E. Student, Dept. of Computer Engineering, R.M.D Sinhgad School of Engineering, Pune, Maharashtra, India

Assistant Professor, Dept. of Computer Engineering, R.M.D Sinhgad School of Engineering, Pune, Maharashtra, India

ABSTRACT:In cloud computing, large amount of data is outsourced to cloud for scalable storage. This raises the point of security and privacy of data so as to provide data confidentiality, authentication and control the access pattern. In order to maintain the privacy and security, the data is encrypted before its storage on cloud. This makes the searching of data through cipher texts more complicated. To address this issue, proposed framework develops searchable encryptionprovided for the multi-keyword ranked search over the cloud data. The encryption is based on attributes. The blind storage is used to hide access control issues in searchable encryption techniques. The system also utilizes the conditional random field (CRF) technique, to improve accuracy in data retrieval. The system maintains confidentiality of the document and index, trapdoor privacy, solve trapdoor unlink ability and hiding the access pattern of search user.

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

I. INTRODUCTION

Cloud is used for storing and retrieving the large amount of data. The most of the confidential data are stored in cloud, thus the confidentiality and privacy for the data and data owners should be maintained. And hence its security should be the major issues to be solved. To avoid the loss of data, the data is encrypted and then stored it in to the cloud. As the data is in encrypted form, searching of these data over cloud is time consuming and a complex job.
The main objective is to maintain data privacy such that even cloud should not get the actual data and provide the efficient search mechanism. This paper propose the searching of encrypted data scheme supporting multi-keyword ranked search that is used to enhance the search performance across the encrypted cloud data. The system adopts attribute based encryption. The functionality of the searching of encrypted data should support multi-keyword search that is used to retrieve the data and should carry the similar experienced inoutputs as carrying the search in the gogglewith the use of different keywords so as to easily identify the most relevant search result that is obtained based on relevance score. The blind storage system is used to store the data in order to overcome from data breach. To improve the retrieval of the data from database CRF technique is used.

A. *Motivation*

In concern with the cloud computing, it isnecessary to maintain the security of the data. Thus the data is encrypted and then stored in the cloud. The searching of encrypted data results in indistinguishable difficulties. The traditional approach lacked in case of providing some security levels and functionalities [5] [7] [8]. The main objective of these approach were to provide the data confidentiality but the other problems such as trapdoor unlink ability and hiding the search users number and sequence of access were remained unsolved. The most of existing proposals have failed to provide construction of efficient searchable encryption. Hence proposed framework provide an efficient multi-keyword ranked search (EMRS) scheme for the encrypted data using blindstorage. The system also uses attribute basedencryption (ABE) that allows only those users to decrypt the file if it satisfies the attribute condition.

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. 4, Issue 6, June 2016

II. RELATED WORK

The technique searchable encryption allows the user to perform the search over the encrypted data. This technique is classified in to two as searchable public key encryption (SPE) and searchable symmetric encryption (SSE).The searchable technique SSE is introduced first by D. X Song, D. Wagner, and A. Perrig[2] that conducted research on searching of encrypted data using single keyword. H. Li, Y. Dai, L. Tian, and H. Yang [3] presents an encryption of the data based on its identity with authentication that can be used in various applications like e-business etc. this paper avoids the issue of revocation and sharing key problem in the scheme that is based on the public key certificate.
Q. Liu, C.C Tan, J. Wu, and G. Wang [4] address two issues related to the privacy of the data and efficiency in the search that allows minimizing the cost of communication. Based on Aggregation and distribution layer (ADL), the paper presents a secured retrieving of information for ranked query (EIRQ) scheme that is termed to minimize the cost of query in the cloud. D. Cash [5] proposed the first sub-linear SSE scheme supporting the conjunctive keyword search with the general Boolean queries with security and efficiency. This scheme provides better confidentiality but does not address the issue regarding the search user that obtains the relevant data from the server that may give the relationship between the search request and the documents to the server.
Jiadiyu, P. Lu, Y. Zhu, G. Xue, and M. Li [6] has overcome through the problem of Boolean keyword search solved using the two round searchable encryption (TRSE) that retrieves the top K data. The communication overhead of keyword is reduced by using vector space model and homomorphism encryption. For encryption and decryption the Diffiehellman algorithm is used. N Cao, C. Wang, M. Li, K. Ren, and W. Lou [7] proposed the searching of an encrypted data that focuses on searching of data using single keyword orboolean keyword search and get the sorted results of the searches made by the users. This is the first scheme that defines and solve various problem regarding preserving the privacy for using the multiple keyword ranked search performed over encrypted data (MRSE).
M. Naveed, M. Prabhakaranand C.A. Gunter [8] introduced the SSE scheme that is dynamic, simple and more efficient and it discloses the less information to the server than other schemes. This scheme has achieved the security against servers that intends to know the data stored in to it. This scheme utilize the blind storage system in order to protect the various access sequence performed by user but it only supports single keyword search and returns undistinguishable results. B. Wang, S. Yu, W. Lou, and Y.T. Hou[9] gives the solution for the spelling error occurred during the keyword search. The above research work provides the confidentiality of the data but has not addressed some security aspects and functionality regarding improving the search performance. Hence the proposed system provides an EMRS scheme for the encrypted data stored in cloud and uses the blind storage that enhance the search performance by improving the security aspects, functionality and communication overhead as compared to previous work.

Problem Definition: To search over encrypted data becomes worstwhen it is stored on cloud. Apart from this cloud can be curious to know the data or to understand user's frequent searches so may be the data breach is possible. To prevent this problem the multi-keyword ranked search scheme is proposed that allows accurate, efficient and secure searching of encrypted data stored in the cloud.

III. PROPOSED SYSTEM

There arethree main entities in the system: the data owner, cloud server and search user as shown in figure 1.

1. Data owner: The data owner stores the collection of the documents that are encrypted in to the cloud server. The data owner then places the set of keyword dictionary for each document with keywords. The data owner builds the encrypted index so as to allow the users to query that performs the search over these encrypted documents. Both the encrypted documents and index are stored on the cloud server by using the blind storage system [8].

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. 4, Issue 6, June 2016

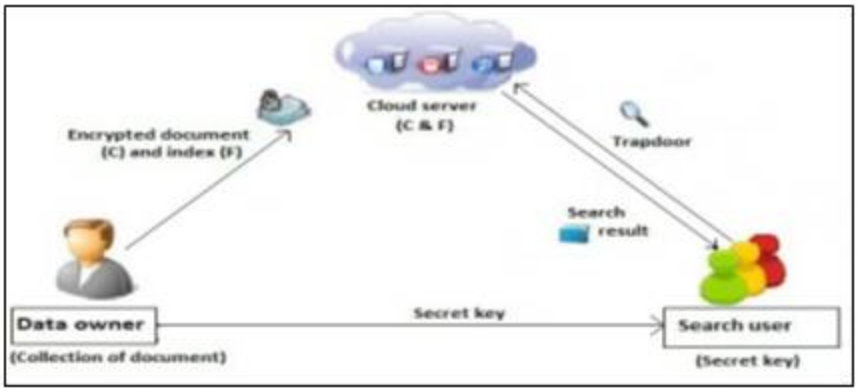

Fig. 1: System Architecture

2. Search user: The search user receives the secret key from the data owner and it chooses a set of related keywords that contains theinterested keyword set and computes a trapdoor including the keyword related token called as stag and the encrypted query vector. The search user sends stag, query vector and optional number k to the cloud server to obtain the most k relevant results.
3. Cloud server: Server access the index in blind storage by using the stag. The relevance score is computed with query vector and send descriptor of top k document relevant to keyword search. The search user uses the descriptor to access document.

A. *Multi-keyword ranked search efficiency*

To provide the data to the users, necessarykeyword can be obtained from search user to perform the search by using the required keyword search that the user expert is to be given.The single keyword [2] does not satisfy the search user needs; therefore we proposed a multi-keyword ranked search [7]. To obtain the requirements and improve the performance the EMRS [6] should support multiple keyword search as well as it helps the server to search and get the data related to the given keywords obtained by necessary keyword search ranking result to the user. The main contribution of the proposed work is to achieve ranked search using multiple keywords across the data and solve trapdoor unlink ability problem and hide access performed by the user.

IV. IMPLEMENTATION STRATEGY

A. *Algorithms Used*

1. Blind Storage*:*
   - Key generation:Input the security parameter and generate the key to perform the encryption and decryption andoutput key K_B, where K_B is independent of data.
   - Build: Input the $(K_B, d_0, \{id_i, data_i\})_i^t = 1$, where K_B is a key, d_0 is an upper bound on number of data blocks stored in system, $(id_i, data_i)$ is id and the data of the files that the system to be initialized. Outputs an array of blocks D to be uploaded to the server.
   - Access: Input key K_B, a file id, an operation specified op € {read, write, update, delete}, and optionally data (if op is write or update).

2. AES Algorithm:

   1. Cipher (Inblock [16], Outblock [16], w [0…..43])
   2. Then BlockToState (inblock, S)
   3. S<— AddRoundKey (S, w [0…3])
   4. Then For (round = 1 to 10)
   5. S<— SubBytes(S)

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. 4, Issue 6, June 2016

6. S<—ShiftRows(S)
7. If (round≠ 10) then S<— MixColumn (S)
8. Then S<—AddRoundKey(S, w [4×round, 4×round + 3])
9. Else StateToBlock(S, Outblock)

3. CRF Technique*:*
    - Pre-processing and features extraction: Input the document D.Documents transferred to tagging sequences and perform pos tagging. Extract feature automatically.Output the words or a phrase.
    - CRF model training: Input the set of feature vectors.Label the keyword such as 'KW_B'.
    - CRF labeling and document keyword extraction: Input the document. Pre-processed the documents and extract its features and predict the type of keyword using CRF.

B. *Proposed Algorithm*

1. Encrypted Database Setup, Index F
    1. Compute the d- dimension relevance vector p= (p1, p2… pd)
    2. For d_i in D, set D in blocks m_b
    3. Choose K_i for Enc ( )
    4. Initialize F
    5. For each keyword ω ϵ W do
    6. set t an empty list
    7. For each document d_i containing the keyword ω do
    8. Get the associated vector P of d_i
    9. Choose the random number x
    10. DscABE$_{vi}$ (id_i ‖ K_i ‖ x)
    11. Append the tuple (Dsc, P) to t
    12. End for
    13. F[ω] = t
    14. End for
    15. Return F

2. Search and Retrieve Document:

Step 1: Receive query vector, stag and k. Parse stag to get range of integers
Step 2: Access index F in blind storage and retrieve index of blocks to get tuple (ABE$_{vi}$(id_i||K_i||x), P).
Step 3: Sorting relevance score, send ABE$_{vi}$(id_i||K_i||x) of k relevant document.
Step 4: If users attribute satisfy access policy then decrypt descriptor using K.
Step 5: Compute $\sigma_{i=} \psi_{K\psi}(id_i)$, to retrieve document.

V. MATHEMATICAL MODEL

- System S= {i, f, D, M, en}
- Initial state (i): Data files
- Final state (f): Success or failure
- Input (D):
- D= {d_1, d_2, d_3…..d_n} documents
- Output (M):
- M= Multi-keyword search
- Algorithm (en):
    1. Initialize the public parameters
       Pk = random gen (setoff (characters; numbers); random user id)
    2. Upload file F

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. 4, Issue 6, June 2016

$E^i(F)$ = Encrypt (F)

3. Generate Trapdoors Td = Trap (F; w_i)}

w_i = set of keywords

4. Generate aggregate key K = Combine (F; tag (level; Pk)
5. Upload to CSP

CSP (F) = E^i (F) + Td

6. If Decrypt (Encrypted File, Key, Trapdoor)

Then Successful

7. Decrypt at user end

F = Decrypt (E^i(F))

VI. RESULT AND DISCUSSION

In the proposed Multi-keyword ranked search scheme, the experiment is performedon cloud platform with minimum 32 GB RAM. The execution time and development cost required for searching the document is less. So the overall execution time is expected to be decrease as the documents retrieval and re-ranking will be faster.The execution timing comparing the proposed system and the existing system is given in figure 2. The existing EMRS scheme is represented in blue bar and the proposed work is given in the red bar. X-axis is file 1 and file 2 and Y-axis is execution time in ms.The comparison of the communication overhead of the existing system and proposed system is given in table 1 and table 1 shows how the proposed system has improved with less communication overhead. The timing in EMRS is given based on the number of words in the documents. So have calculated the execution time based on words. The time is given in milli-second (ms) in table 2.

Method	Number of Rounds	Size of Messages (bytes)
Existing EMRS	One/Two	9000
Proposed EMRS using CRF	One/Two	8560

TABLE 1: Communication Overhead

Keyword Scanned	EMRS(ms)	Proposed EMRS using CRF(ms)
6000	0.43	0.413
8000	0.5	0.486

TABLE 2: Execution Time

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. 4, Issue 6, June 2016

VII. CONCLUSION AND FUTURE WORK

The proposed system improves the efficiency in semantic search over the encrypted data stored in the cloud and proposes the multi-keyword ranked search scheme that improves the search in terms of accuracy, efficiency and security over encrypted cloud data. It also maintain the confidentiality of documents and index, trapdoor privacy, solvethe trapdoor unlink ability, and hiding the access pattern of the search user. The system at current work on single cloud, in future it can be extended up to sky computing so as to provide better security in multi-user systems.

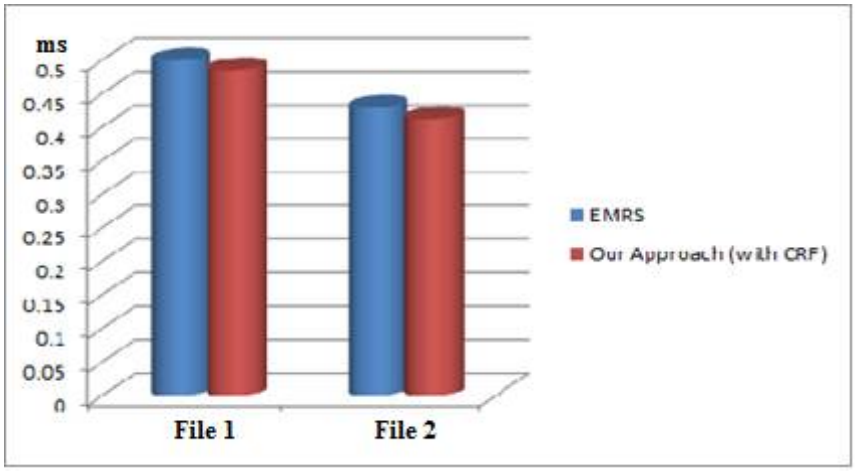

Figure 2: Execution Time

VIII. ACKNOWLEDGEMENT

I take this opportunity to express my sincere gratitude to my guide and head of department, Prof. Vina M. Lomte, Department of Computer Engineering, RMDSSOE, Pune University, for her kind cooperation and capable guidance.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. C. Rou, and M. Steiner, "Highly-scalable searchable symmetric encryption with support for Boolean queries", in Proc. *CRYPTO*, 2013, pp. 353373.

[6] 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.

[7] 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.

[8] M.Naveed, M.Prabhakaran and C.A.Gunter, "Dynamic searchable encryption via blind storage", in proc.IEEE symp.secure.privacy, may2014, pp.639-654.

[9] 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./May2014, pp. 2112-2120.

[10] 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.

[11] D. Bonch, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. EUROCRYPT, 2004, pp. 506-522.

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. 4, Issue 6, June 2016

BIOGRAPHY

Miss. Gauri S. Patilreceived B.E. degree in Computer Science and Engineering in the year 2014 from AGTI, Dr. DaulatraoAher College of Engineering, Karad and pursuing M.E. in Computer Engineering from RMD S.S.O.E., Warje, Pune.

Prof. Vina M. Lomte is the HOD of Computer Dept. at RMD S.S.O.E., Pune, having more than 10+ years of experience in the field of teaching and research. The domains of her research are Software Testing, Software Engineering and Web Security.