ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Privacy-Preserving Multi-Keyword Similarity Search Over Encrypted Data

Shubhangi J. Wagh [1], Aishwarya G. Patil [2], Arpita D. Patil [3], Pallavi G. Patil [4]

U.G. Student, Department of Computer Engineering, SSBT COET, Bambhori Jalgaon, Maharashtra, India[1]

U.G. Student, Department of Computer Engineering, SSBT COET, Bambhori Jalgaon, Maharashtra, India [2]

U.G. Student, Department of Computer Engineering, SSBT COET, Bambhori Jalgaon, Maharashtra, India[3]

U.G. Student, Department of Computer Engineering, SSBT COET, Bambhori Jalgaon, Maharashtra, India[4]

ABSTRACT: Due to the appealing features of cloud computing, large amount of data have been stored on the server. Although cloud based services offer many advantages, privacy and security of the sensitive data is a big concern. To less the concerns, it is desirable to outsource confidential data in encrypted form. Encrypted storage protects the data against illegal access, but it complicates some basic, yet important functionality such as the search on the data. To achieve search over encrypted data without compromising the privacy, in proposed system encrypted documents are searched by searching on multiple keywords which are stored at server by data owner with each document. In this project an efficient scheme for multi-keywords similarity search over encrypted data is used. To do so, we utilize AES algorithm to encrypt and decrypt documents and common index.

KEYWORDS: AES(Advanced Encryption Standard), Privacy, Common index, Encrypted data.

I. INTRODUCTION

Nowadays searchable encryption (SE) is very important, especially with the emergence of cloud computing. The keyword-based search is such one widely used data operator in many database and information retrieval applications, and its traditional processing methods cannot be directly applied to encrypted data. Therefore, how to process such queries over encrypted data and at the same time guarantee data privacy becomes a hot research topic. There are many methodologies based on searchable encryption, such as deal with the single keyword search, and works support the multi-keyword search. The single keyword search is not smart enough to support advanced queries and the boolean search is unrealistic since it causes high communication cost. Therefore, more recent works like focus on the multi-keyword search, which is more practical in pay as-you-go cloud paradigm. But most of these methods cannot meet the high search efficiency and the strong data security simultaneously, especially when applying them to big data encryption poses great scalability and efficiency challenges.

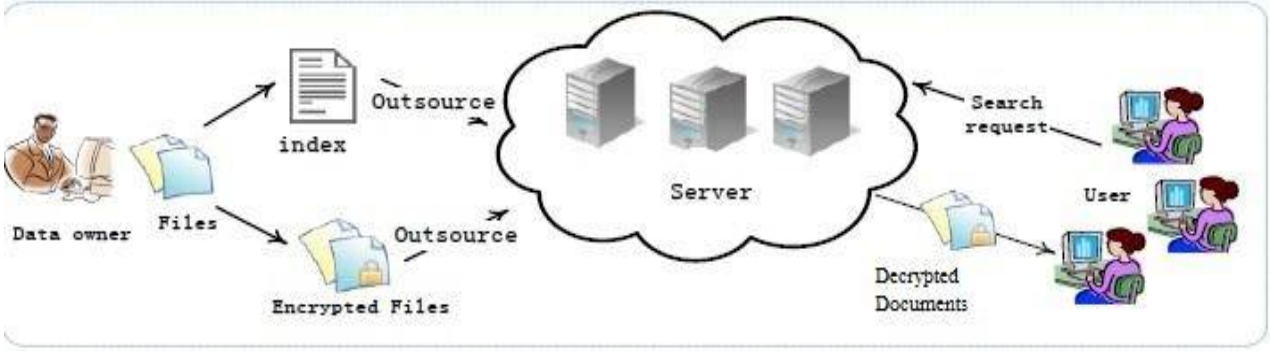

Fig1: System Architecture

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

The figure 1 shows the working of search over encrypted cloud data. The system architecture is considered as three entities, as depicted in fig.1 the data owner, the data user, and the cloud server.

1. Data owner has a collection of data documents D={d1, d2, ...dm}. A set of distinct keywords W={w1, w2,...wm} extracted from the data collection D. Then, the data owner upload the index I, encrypted file and encrypted filename on the server.
2. Data user provides t keyword for searching the document uploaded on the server by data owner in encrypted format. A corresponding trapdoor Tw is generated through search control mechanism. After that decrypted documents are download by data user from server.
3. Server first constructs secure searchable index. Received Tw from the authorized user. Then, the server first decrypt the common secure index, finds similarity and returns the corresponding set of decrypted documents to data user.

II. RELATED WORK

Wei Zhang et.al. [1] proposed new protocols such as novel dynamic secret key generation protocol and a new data user authentication protocol. It enabled the cloud server to perform secure search among multiple owner's data, which is in encrypted form with different secret keys. A novel additive order and privacy preserving method is to rank the search results and preserve the privacy of relevance scores between keywords and files. It's more efficient on large data and keyword sets. It does not support secure fuzzy keyword search in a multi-owner data. Zhihua Xia et.al [2] proposed a secure multi-keyword ranked search over encrypted cloud data. In this scheme, data owner can do dynamic update operations like deletion and insertion of documents and sending them to the cloud server. The vector space model and the TF X IDF model are combined to create index file and query generation for encrypted documents. Tree based index structure is created to achieve sub-linear search time and it deals with deletion and insertion of documents. The secure KNN algorithm is used to encrypt the index and query vectors and it is ensuring accurate relevance score calculation between encrypted index and query vectors. In order to avoid statistical attacks, in the index vector phantom terms are added for blinding search results. Zhangjie Fu et.al [3] proposed an efficient Multikeyword fuzzy ranked search scheme based on Wang et.al scheme for multi-keyword fuzzy search. Wang et.al scheme was vulnerable to server out-of-order problems during ranking process, which was addressed by Zhangjie Fu et.al. The traditional techniques were focused on multi-keyword exact search and single keyword fuzzy search. But, those techniques had a very less significance in real-search scenarios encrypted data, compared with multi-keyword fuzzy search. Zhangjie Fu et.al developed a method of keyword transformation based on a unigram, which is tolerable to misspelling of one letter and for other spelling mistakes in search terms. Keyword weight is considered to construct the ranked list of the results to achieve high accuracy. So, the files which are more relevant to the keywords will have greater chances to appear first on the list. Kuzu Mehmet et.al [4] proposed similarity search rather than exact query matching, which supports multi-keyword document retrieval. In this scheme they have employed LSH algorithm for fast nearest neighbor search. The bucket generation process supports to provide sufficient search accuracy. It also implements multi-server and fault tolerant system. Hongwei Li et.al [5] proposed new search scheme called as Fine Grained Multi-keyword Search (FMS). FMS introduced the relevance scores and preference factors upon keywords. It enables the correct keyword search and personalized user experience. Then, they developed an efficient grained multi-keyword search scheme which supports complicated logic search the combination of AND, OR and NO operations of keywords. The classification of sub-dictionaries technique is used to achieve efficiency for building secure index and generation of trapdoor. Finally, they have analyzed the security in terms of confidentiality of documents, privacy protection of index and trapdoor, and unlink ability of trapdoor. They have validated the performance of the proposed schemes using the real-world data set.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

III. PROPOSED ALGORITHM

The proposed system implements following algorithms for Privacy Preserving-Multi-Keywords Similarity Search over Encrypted Data.

Algorithm 1: AES Algorithm
Input: .PDF, .docx, .txt file.
Output: File is encrypted.
Process: AES Cipher
1. begin.
2. byte state[4,Nb]
3. state = in
4. AddRoundKey(state, w[0, Nb-1]).
5. For round = 1 step 1 to Nr-1.
6. Sub Bytes(state) Shift Rows(state) Mix Columns(state) AddRoundKey(state, w[round*Nb, (round+1)*Nb-1]) end for
7. Sub Bytes(state) Shift Rows(state) AddRoundKey(state, w[Nr*Nb, (Nr+1)*Nb-1])
8. out = state.

Algorithm2: Searching Algorithm
Requirements:
E_v : Encrypted Bit Vectors (i.e. list of encrypted common indexes of all files).
$K_{payload}$: Secret Key.
T: Score Limit (i.e in how many documents that queried word is present).
σ_{VB} : Single Common Index.
VB_k: Decrypted Common Index.

Algorithm
For all $\sigma_{VB} \in E_v$

do

$VB_k \leftarrow$ Dec $K_{payload}$ (σ_{VB})

for i $\leftarrow$ 1 to $|VB_k|$ do

if $VB_b[i]$ = “Queried words” then

add i to the candidate identifier list
increment (i)

end if

end for

end for

send id list to data user

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

IV. MODULES DESCRIPTION

The operation of system is given below.

1. Preprocessing Module: Three different entities are involved in proposed system: Server, Data owner and Data user. Data owner and Data user at first has to register with the Server. After that the user and owner has to wait until the administrator approves them. The data owner can upload the documents and multiple keywords in index text area. Third-party data storage and retrieval services are hosted by the server. As the uploaded data may contain sensitive information, files, file name and index of every file is encrypted using AES algorithm when it is outsource on server.

2. Searchable Index: To effectively search documents, server builds a searchable index for all the documents being uploaded on server. Searchable index is build by adding file name and index keywords and then store it in encrypted format.

3. Search Query Processing:

i. User fires query to server for searching the document using single or multiple keywords.
ii. When server receives that query, server first decrypt the secure index using secret key.
iii. Create vector of that indexes and finds similarity one by one between queried words and indexes.
iv. After that matched documents with user query are first decrypted and then send to user.

V. SIMULATION RESULTS

The simulation results showed that the proposed algorithm performs better than existing system. There are so many searching techniques implemented in the cloud but the drawback with these techniques is that they supports only exact and single keyword search. Moreover typical users searching behavior is also revealed. The proposed system overcomes all these limitations. The main purpose of the proposed system is to preserve the privacy of document and efficiently search the document by using indexed keywords when user fetch the query. Proposed system is fault tolerant i.e it tolerate the typographical mistakes because it uses bucket genration process as follows: "GENERATION" generates bucket as shown - Generation (g, ge, gen, gene, gener……). So proposed system is better than exact query matching and single keyword search scheme.

In our system we take .txt, .pdf, .docx file types for encryption and decryption. Table A shows that system uploads the encrypted filename, file key, file description, encrypted file and encrypted common index on server.

File ID	Encrypted File Name	File key	Small File Description	Encrypted File	Secure Common Index
1	²:-NdÄ}£	834	Introduction of dotnet language	[BLOB - 9.7KiB] Ï	ž£1¤ð“ö¸Ý×S}+ÇSAó u{J¡¿¯Ä.
2	1¼ÿIÖ– mÞ¿}fc°O>?	686	JVM technology evangelist Eva-Andreasson gives an ...	[BLOB - 12.1KiB]	Âôú $rhYF»úÆY
3	7ci)†ã)Æa›€õñ9-	645	Contains question paper of computer network	[BLOB - 570.4KiB]	ÒH¦YóŸYéñ 3 ‡:Lk-m¸¹éXž‹Ü® {
4	{Ž¸ùi°µÔÌ¿ËêÏ	254	Computer network book	[BLOB - 353.1KiB]	±‹vÇv˜E J®#ú;’k“®ìI

Table A: Uploaded data on server

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Table B shows the searching results on encrypted secure common index for file of interest .i.e shown in Table A. For example if user search for file using keyword comp then files which contains this keyword in their common index (i.e. file 3rd and file 4th from Table A) are retrieved in decrypted format.

File Name	View	Original Code	Encrypted Code
C N QP1.PDF	Contains question paper of computer network	Download	Download
Computer network.pdf	Computer network book	Download	Download

Table B: Search Results

REFERENCES

1. Wei Zhang, Yaping Lin, Sheng Xiao, JieWu, Siwang Zhou, "Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing", IEEE Transactions On Computers, Vol.65, No. 5, May 2016.
2. Zhihua Xia, Xinhui Wang, Xingming Sun,Qian Wang, Member, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions On Parallel And Distributed Systems, Vol. 27, No. 2, February 2016.
3. Zhangjie Fu, Xinle Wu, Chaowen Guan, Xingming Sun, Kui Ren, "Toward Efficient Multi-Keyword Fuzzy Search Over Encrypted Outsourced Data With Accuracy Improvement" IEEE Transactions On Information Forensics And Security, Vol. 11, No. 12,December 2016.
4. Kuzu Mehmet, Mohammad Saiful Islam, Murat Kantarcioglu, "Efficient Similarity Search Over Encrypted Data", in Data Engineering (ICDE), 2016 IEEE28th International Conference, IEEE, 2012.
5. Hongwei Li, Yi Yang,Tom H. Luan, Xiaohui Liang, Liang Zhou, Xuemin (Sher-man) Shen, "Enabling Fine-Grained Multi-Keyword Search Supporting Classiffied Sub-Dictionaries over Encrypted Cloud Data", IEEE Transactions On Dependable And Se-cure Computing, Vol. 13, No. 3, May/June 2016.
6. Cong Wang, Ning Cao, Jin Li, Kui Ren and Wenjing Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data", IEEE 30th International Conference on Distributed Computing Systems, 2010, pp. 253-262, doi:10.1109/ICDCS.2010.
7. Yan-Cheng Chang and Michael Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data", in Proc. of ACNS'05, pp. 442-455, 2005.
8. Jan Li, Q. Wang, C. Wang, N. Cao, K. Ren, W. Lou, "Enabling Efficient Fuzzy Keyword Search over Encrypted Data in Cloud Computing", Proc. of IEEE INFOCOM'10 Mini-Conference, pp. 1-5, IEEE, March 2010.
9. Qin Lv, William Josephson, Zhe Wang, Moses Charikar, Kai Li, "Multi-Probe LSH: Efficient Indexing for High Dimensional Similarity Search", in Proceedings of the 33rd international conference on very large databases, pp. 950-961, September 2007.
10. Dan Boneh and Brent Waters, Conjunctive, "Subset and Range Queries on Encrypted Data", in Theory of Cryptography Conference, February 2013.
11. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data" , in IEEE INFOCOM, pp.829837, April 2011.
12. Sun-Ho Lee and Im-Yeong Lee,"Secure Index Management Scheme on Cloud Storage Environment" ,International Journal of Security and Its Applications Vol. 6, No. 3, July, 2012.