Efficient multikeyword search results verification on encrypted outsourced data

B.Abarna
UG Scholar
Department of computer science and engineering
IFET College Of Engineering,Villupuram
abarnamani06@gmail.com
A.Divya
Senior Assistant Professor
Department of computer science and engineering
IFET College Of Engineering,Villupuram
div.anandan88@gmail.com

Abstract: With the advent of cloud computing becomes ubiquitous, more and more data owners are annoyed to outsource their data to cloud servers for great expediency and condensed cost in data management. The shielding data privacy, sensitive data have to be encrypted before outsourcing, which performs conventional data exploitation based on keyword based document retrieval. The majority of existing techniques are focusing on multi-keyword fuzzy search, converting the keyword into bigram set that will be increases the Euclidean distance and the vectors are be hardly ever threshold in the same bit. A keyword transformation based on unigram, it will be reduce the Euclidean distance between the misspelled keyword and correct keyword. Based on this schema, a deterrent based method, multiple data owner are exchange the data will be verifying into the cloud server and user will give a secret information and construction the secret verification data buffer. Finally, with though analysis and extensive experiments, our schema is practically efficient and then achieve high accuracy.

Keyword: multikeyword, outsourcing, verification.

I.INTRODUCTION

In the cloud computing and storage solution provide users and enterprises with various capability to store and process their data in either privately third-party data centers that may be positioned far from the user–range in distance from across a city to across the world. They are three main entities are data owner, cloud server, data user .To minimize the data leakage into cloud service providers (CSP), and data owner will be encrypted their sensitive data, (example: personnel details, health record, Crime information, financial transactions) are outsourcing to the cloud server .Only the authorized data user can be access their data will shared the secret key and then decrypted on it. In this we have challenging problem, in that data can be utilization to the cloud server. For example, in order to search some of their relevant documents are to be encrypted data are to be stored into the cloud, user may be download and then decrypted the data. However Huge cost in terms of data usability. For example, the existing techniques on keyword-based information recovery, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypted locally is clearly not practical. But the multikeyword fuzzy search, we convert the keyword into the bigram set and then Euclidean distance is used to capture the keywords similarity.

The main contributions of this paper are:

- 1.To the best knowledge, this is the first work that deals with the multikeyword fuzzy ranked search over encrypted outsourcing the data with the user data isolation protection.
- 2. A keyword transformation based on unigram, for misspelled of one letter will reduce the Euclidean distance between the correct keyword, the keyword with the same root can be queried from the vector.
- To formalize the top-k ranked keyword search results and the verification problem where multiple data owners are exchange data to the cloud server.

II.LITERATURE SURVEY

[1] CatchYou if You Misbehave: Ranked Keyword Search Results Verification in Cloud Computing, proposed a novel secure and efficient deterrent based Verification scheme. For the keyword search, we use the unigram concept. It will be efficiencyto search the keyword on it.

[2] A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Dataproposed a “Greedy Depth-first Search” algorithm, to construct a special tree like structure and then we have dynamic insertion and deletion operation .it will be execute the parallel search will carried and reduce the time of cost. Disadvantages of cloud service providers (CSP) are data user can be access the sensitive information without the unauthorized user.

[3]As the results, we have to rebuild the index for the keyword while we search on it. We achieve a secure and matching data between the query trapdoor index that will be indexed.Inverted Index Based Multi-Keyword Public-key Searchable Encryption with Strong Privacy Guaranteeto design a novel trapdoor generation algorithm so that the query has to related the indexed are combined together secretly to the cloud server

[4] Multi-User Multi-Keyword Privacy Preserving Ranked Based Search Over Encrypted Cloud DataAfter dismissal single-keyword query, user will get all documents that contain the specified keyword. The advantages are protect the sensitive data by encrypting documents before outsourcing, it is used to the rank based retrieval of the documents, To easily access the encrypted data by multi keyword rank search using keyword index.

[5] Verifiable Auditing for Outsourced Database in Cloud Computing the proposed system can be expanded to support the dynamic database setting by joining the concept of verifiable database with updates in checking process we used thetuple merkle hash tree which is used to provide signature for the each and every individual has attribute to check the correctness and completeness of the data. By using the Bloom filter is used to check whether the data is present or not. The Evdokimov's scheme is used to encrypted the data. The advantage is data user can efficiently perform verifiable auditing for the result returned by the CSP. The disadvantage is the uploading and downloading time is high.

[6]Toexpanded the storage of wild card fuzzy keyword set, our scheme exploits locality sensitive hashing to provide the efficient fuzzy search with contains the constant size of the keyword.Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud will eliminates the pre-defined dictionary and then effectively support the multikeyword search. In this we introduce the unigram concept and then spilt the keyword set, we can search more efficiently.

III. MULTIKEYWORD SEARCH

[1] Bigram vector representation of keyword

One key step to build index is the keyword transformation. The LSH function takes a vector as the input and hash "close" vectors to the same value with high probability. We use the following method to transform a string type keyword to its vector representation so that it can be used in the LSH functions. It is a variant of Q grams where each term is spilt into grams of two characters. For instance netmask will be decomposed in{ne,et,tm,ma,as,sk}strings are compared by counting the number of grams they share. We use a 262 -bit long vector to represent a bigram set. Each element in the vector represents one of the 262 possible bigrams. The element is set to 1 if the corresponding bigram exists in the bigram set of a given keyword. This bigram vector based keyword representation is not sensitive to the position of misspelling, nor is it sensitive to which letter it was misspelled to. "nwtmask", "nvtmask", or "netmagk" will all be mapped to a vector with two-element difference from the original vector. By this representation, a keyword can be misspelled in many different ways but still be represented in a vector that is very close to the correct one, and this closeness (distance) is measured by Euclidean distance, the well-known metric for distance between vector-type data items. This bigram vector representation is robust and inclusive, and key to enabling the use of LSH functions.

[2] Notations

W-The dictionary, namely, the set of keywords,denoted as F={fw1; w2; . . . ; wn}.

F-{f1; f2; . . . ; fn} is used to the collection of the documents are considered to be sequence of keyword element.

n-Total number of the documents in F.

m-The total number of keywords inW.

Wq-The subset of W, representing the keywords inthe query.

C-The encrypted document collection stored in the cloud server.

T- the unencrypted form of index tree for thewhole document collection F.

I - the searchable encrypted tree index generatedfrom T.

Q - The query set of the keyword is Wq.

TD - the encrypted form of Q, it is used the trapdoor for the search request.

IV. SYSTEM MODEL

(a)Privacy Guarantee:

Data owner has collection of documents F= {f1, f2, f3 …} that he wants to outsource to the cloud server in encrypted form while still keeping the capability to search on them for effective utilization (fig.1).Data owner is used to outsourced there data in encrypted format using AES algorithm. RSA is more expensive than AES algorithm it is better to use symmetric key.The data owner firstly builds a secure searchable tree index *I* from document collection *F*. Data owner outsources the encrypted collection *C* and the secure index *I* to the cloud server, and securely distributes the key information of trapdoor creation and document decryption to the allowed data users. Besides, the data owner is responsible for the update operation of his documents stored in the cloud server.

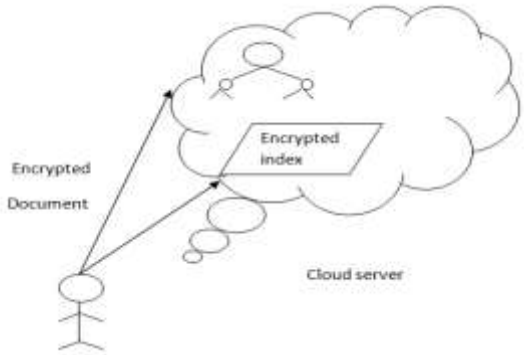

Fig1: encrypted data

(b)Keyword search:

This module is used to help the client to search the file using the multiple keywords concept and get the accurate result list based on the user query. It can regard words one by one. This is known as unigram. For instance, should identify the string "walk" (and possibly "walking", "walker"etc).We use to split the gram as walk means {w, a, l, k} compared by counting the number of grams they share. Data user can access the documents of the data owner. With *t* query keywords; the authorized user can generate a trapdoor *TD* according to search control mechanisms to fetch *k* encrypted documents from cloud server. Then the data user can decrypt the documents and shared the secret key.

(c)Verifying ranked Top-k search results:

These modules ensure the user to search the files that are searched frequently using rank search. It allows the user to download the file using his secret key to decrypt the downloaded data. The proposed scheme is designed to provide not only multikeyword query and accurate result ranking, but also dynamic update on document collections(fig:2).The scheme is designed to prevent the cloud server from learning additional information about the document collection, the index tree, and the query.

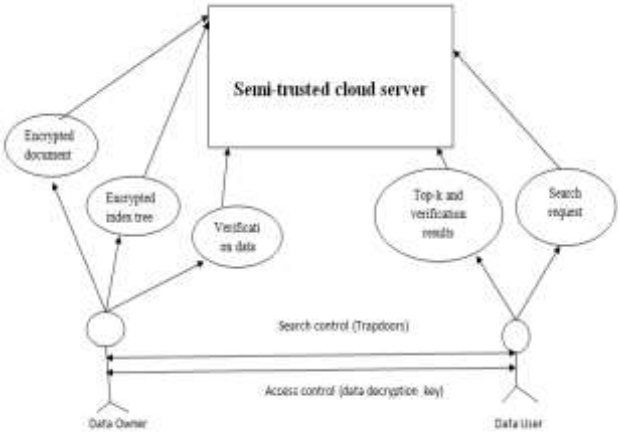

Fig 2: architecture

(d)Search results verification data:

After we get the sample and anchor data are to be concatenate into the string and then encrypted the verification data into the cloud server. When the data user is used to get the search results, if they have any mistrustful data, the data user will construct and submit the request to the verification data to the cloud server. But decrypt the data from the data owner. The data user can reach this goal by simply setting an ID set of his desired data owners. The data ID set should not be uncovered to the cloud server. First, the data user enlarges the file ID set of verification by inserting random Ids(fig:3).A data user wants to get Oi's verification data, we can add other n−1 data owners file ID in the set .Second, The data user attaches a data 0 or 1 to each of the file ID set. Here, if the data user wants to return the data from the data owner's verification data, then he attaches 1 to the corresponding ID, otherwise, 0 is attached. Third, the data user is used to encrypted the attached 0 or 1 with the Paillier encryption.

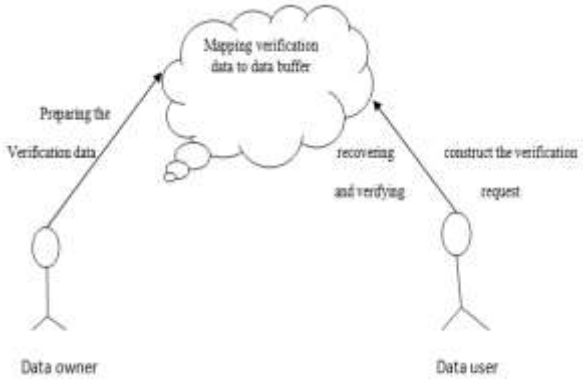

Fig 3: verification data

V. CONCLUSION

In the method of efficient multikeyword search results verification on encrypted outsourced data for efficient data utilization are stored remotely over the encrypted cloud data. Approach of leveraging LSH functions to construct the file index is novel. This project provides an efficient solution to the secure fuzzy ranked search of multiple keywords. Euclidean distance is adopted to capture the similarity between the keywords. Thesecure inner product computation is used to calculate the similarity score so as to enable result ranking. In this model where cloud servers would probably behave dishonestly. Different from previous data verification schemes, this model will propose a novel deterrent-based scheme. During the whole process of verification, the cloud server is not clear of which data owners, or how many number of data owners are exchange anchor data used for verification, he also does not know which data owners‟ data are entrenched in the verification data buffer or how many data owners‟ verification data are actually used for verification results. All the cloud server knows is that, once he behaves dishonestly, he would be discovered with a high probability, and punished critically once exposed.

Additionally, when any suspicious action is detected, data owners can dynamically update the verification data stored on the cloud server. Furthermore, the proposed scheme allows the data users to control the communication cost for the verification according to their preferences, which is especially important for the resource limited data users. Finally, with thorough analysis and extensive experiments, confirm the efficacy and efficiency of the proposed schemes.

REFERENCES

[1] Zhangjie Fu, Xinle Wu, Chaowen Guan, Xingming Sun, "Toward Efficient Multi-Keyword Fuzzy Search Over Encrypted Outsourced Data With Accuracy Improvements," IEEE Transaction on information and forensics and security ., to be published, doi: 10.1109/TIFS.2016.2596138.

[2]W. Zhang, Y. Lin, and Q. Gu, "Catch you if you misbehave: Ranked keyword search results verification in cloud computing," IEEE Transaction. Cloud computing, to be published, doi: 10.1109/TCC.2015.2481389.

[3] J. Wang, X. Chen, X. Huang, I.You, and Y. Xiang, "Verifiable auditingfor outsourced database in cloud computing," *IEEE Transaction. Computing*, vol. 64, no. 11, pp. 3293–3303, Nov. 2015.

[4]B. Wang, W. Song, W. Lou, and Y. T. Hou, "Inverted index basedmulti-keyword public-key searchable encryption with strong privacyguarantee," in *Proc. IEEE INFOCOM*, Apr./May 2015, pp. 2092–2100.

[5]J.Wang, X. Yu, and M. Zhao, "Privacy-preserving ranked multi-keyword fuzzy search on cloud encrypted data supporting range query," *ArabianJ. Sci. Eng.*, vol. 40, no. 8, pp. 2375–2388, 2015.

[6]Z. Fu, J. Shu, X. Sun, and N. Linge, "Smart cloud search services:Verifiable keyword-based semantic search over encrypted cloud data,"*IEEE Trans. Consum. Electron.*, vol. 60, no. 4, pp. 762–770, Nov. 2014.

[7]B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud," in Proc. IEEE INFOCOM, Apr./May 2014, pp. 2112–2120.

[8]Z. Fu, J. Shu, X. Sun, and D. Zhang, "Semantic keyword search based on trie over encrypted cloud data," in *Proc. 2nd Int. Workshop SecurityCloud Comput.*, Kyoto, Japan, Jun. 2014, pp. 59–62.

