Keyword Search Techniques over Encrypted Outsourcing Data

Amira Sallam[1], Ahmed Moustafa[2], Ibrahim El-Henawy[3]
[1,2,3]Department of Computer Sciences, faculty of Computers and Informatics, Zagazig University, El-ZeraaSquare, Zagazig, Sharqiyah, Egypt, Postal code: 44519

Abstract

Cloud Computing Technology has been widely spread in many field, many institutions and individuals are needed to store their senistive data in cloud e.g.financial record, health records and personal information but these data must be private so it needed to encrypted before storing in cloud and needed to make search in this data .Thus achieving search over encrypted outsourced data becomes challanging problem.One of the most problem is finding effictivly similarty between keywords based on one technique of similarity measure.

Typically, cloud server need to support with technique which helping to do keyword search over encrypted data correctly, take in consider achieving privacy ,accuracy and efficiency and storage cost .Many reaserches have been doing to solve this problem, we will covring different technique that used in keyword search over encrypted data.

Keywords : ***single encrypted keyword search,multi encrypted keyword search,cloud server,static search***

1 INTODUCTION

Cloud computing technology has been widely spread in many field, institution and individuals storing their senistive data e.g pesrsonal information, bank account informations, private emials and financial record in cloud.By storing such data in cloud, data owners not be worried about data storage or maintenance.Furthermore data owner can get benefit from cloud advantage.Cloud computing have many advantage, it Always-on availability , more cost effective and flexible capacity.But data owner may not fully trust the cloud server.Thus data owners encrypte data before outsource on cloud to achieve privacy and against unauthorized access.Furthermore data owners sometimes need to share some data with many users, and each user may interested in only some not all of these data .Many researches have been doing in keyword search over encrypted data to suit the need of the user.In[12] author proposed practicle technique for the problem of searching on encrypted data without loss of data confidentiality. In [7]author proposed secure index using bloom filter and pseudo-random functions.in[1]author proposed multi keyword search over enrypted data.

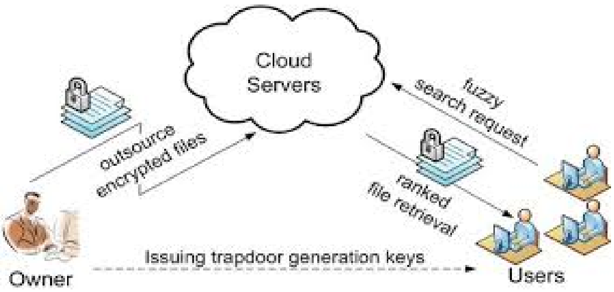

Figure 1: *System Model for keyword search over encrypted outsource data*

In[14]author proposed verifiable keyword search scheme which achieve single fuzzy search which proposed a verify proof to help user to detect if cloud server excute all operation honstly or not.In[9]author proposed two advanced technique to construct fuzzy keyword set(wildcards and n-gram technique. In [11]author proposed split multi-keyword fuzzy and synonym search using inverted index to store keywords which assosiated with corrosponding files also used balanced binary tree as searchable index and inverted index to store keywords and its synonym. .in[2]proposed fuzzy multi keyword search over encrypted data which support dynamic search.In [8]proposed multi keyword fuzzy search.In[13]author proposed scheme that achive semantic search over encrypted outsourced data based on probability of terms using semantic relationship library SRL.Also in[3][5][4] schemes proposed synonym keyword search.System model in search over encrypted data consist of three main items, data owners which have documents that submitted on cloud and keywords, cloud server which make a search over encrypted data and return result and user that sent query request to search for keyword he wanted.Figure 1 show system model in search over encrypted outsourced data.

We covering different techniques which used in searching over encrypted data ,that is categories as the following:

2 SINGLE KEYWORD SEAERCH OVER ENCRYPTED DATA

Some of scheme in single keyword search divid into two categories one of them support dynamic search meaning that data owners can updates uploaded documented and other schemes are static.In [12]author proposed practicle technique for the problem of searching on encrypted data without loss of data confidentiality,each word in document is encrypted indpendently, using two layer encrypted construction, author provid query isolation for searche, searching for the word W returns all the poistions where W found in plaintext,this scheme is static search.in [10]proposed fuzzy single keyword search, used edit distance as a keyword similarity measure, edit distance meaning the number of operation that making to convert one string to the other e.i delete char or add char or substitution,author proposed two technique to construct fuzzy keyword set (wildcard and n-gram),wildcard example, if word is catty so wildcards set based on predefined edit distance 1 S={*catty,c*tty,ca*ty,cat*y,catt*,catty*},* can be any char from a to z,thus storage is reduce.N-gram technique is another technique to construct fuzzy set,n-gram set for example S={catty,atty,ctty,,caty,cay,catt}also this technique reduce the storage,author designed symbol based trie traverse tree as a searchable index that describe in details in non exact keyword search section,by using the two advanced technique show that wildcard technique in this scheme is more efficient than n-gram technique ,coud server returing all files that contain keyword with predefine edit distance.in[14]author proposed a verifiable fuzzy keyword search scheme that used wild card technique to construct fuzzy keyword set which storing in multi way tree ,searching doing by keyword trabdoor which cloud server will transform search request T into sequence of symbols according predefined rule ,cloud server do a search for each trapdoor over index that support fuzzy keyword search based on bloom filter,author in this scheme proposed verify phase depend on bloom filter which bloom filter of node in tree includes childern information of current node by inserting corresponding prefix into bloom filter.in[13]author proposed scheme that achieve siemantic relationship between keywords,he used private cloud and public cloud, expand query keywords set upon SRL that storied in private cloud then use expand query keywords set to retrive the index on public cloud,describe in details in non exact keyword search seaction.All previews schemes not enable to add or delete documents after outsorced to cloud, thus some schemes try to cover this problem.in [7]proposed secure index scheme which develop IND-CKa secure index construction using pseudo-random functions and bloom filter,for each file Bloom filter containing trapdoors of all unique words for searching server, search if any bloomfilter contain trapdoors of query and return matching file also support dynamic search by insert or delete document in bloomfilter Goh,s scheme has linear search time but suffer from false positive.Figure 2 show classification of search over encrypted outsourced data which is described in this paper.

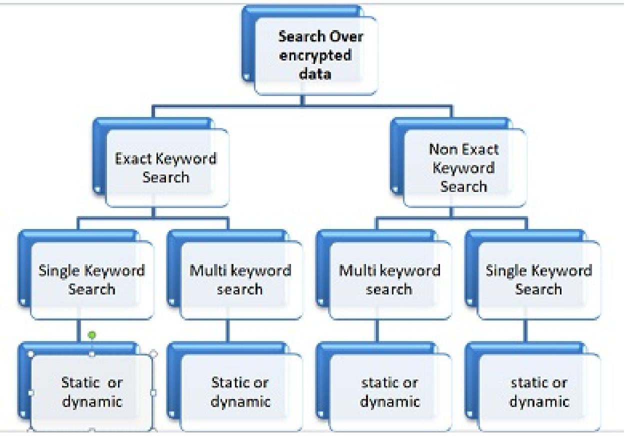

Figure 2: *Keyword Search Classification over encrypted outsource data*

3 MUTLI KEYWORD SEARCH OVER ENCRYPTED DATA

single keyword search not fit with user demond as user some times searching by multi keyword, thus many schemes done to cover this problem.in[1]author achieved mutli keyword search which proposed scheme called MRSE based on secure inner product which used to find similarity i.e number of keywords that found in document to quantitatively measure ,also used binary vector for index construction. in [15] author proposed scheme to deal with secure ranked multi-keyword search in multi-owner model where there are multi- data owners use different secret key to encrypt there keywords,cloud server perform secure search without knowing keywords & trapdoors.Author proposed ”additive order and privacy function” to achieve ranked search result but this scheme did not support updates operation.In[6] author proposed ECSED semantic search scheme based on hierarchy and synonym relationship between words in encrypted documents,achieving semantic search and multi keyword ranked search,author used two type of cloud one for compute similarty measure between document and query search based on distance of them in the concept hierarchy,another one for storing encrypted documents and return ranked results,author used tree as a searchable index.Data owners generate attribute of the document by computing weight for each word in document and select K top wieghts as a attribute of documnet,word weight show that if semantic relationship between words are ignored.For each document,two index vectors are generated, one for matching process and the other for describe the attribute value that satisfied with search request. .in [9] author proposed symbol-based trie traverse search to achieve fuzzy search where multi-away tree construct to store fuzzy keyword set over finite symbol set ,all trapdoor which sharing common prefix may have common nodes, data owner compute trapdoor as a symbol for each keyword in fuzzy set , cloud server divide each trapdoor set into sequence of symbols and comparing trapdoor set in query with stored trapdoor set using depth first search algorithm,this scheme support dynamic search as can add or delete file into leaf node.in [11]author proposed scheme that support fuzzy keyword ,multi-keyword and synonym based keyword search ,used inverted index where keyword is mapped to documents .This inverted index provid with technique for scoring the search result,if the number of query keyword map to large number of individual documents that is considered as relevent document, synonym of keyword is found firstly to construct fuzzy keyword set,for each file a leaf node in balanced binary tree is generated which store index list and file identifier using linked list after that tree is generated using postorder traversal with all leaf nodes that generated before.

4 EXACT KEYWORD SEARCH OVER ENCRYPTED DATA

in[1]proposed privacy multi-keyword searchable scheme called MRSE based on inner product,each document is associated with binary vector as subindex, where each bit represents corrosponding keyword is contained document.Data owner generate binary vector for each document where each binary bit represent keyword that appears in document.Author selecte k-nearest neighbor (kNN) technique to select k nearst index elemnt, used inner product to get similarity measure between index binary vector and query binary vector e.i number of query keyword appearing in document,in the end cloud server return top k ranked identifier files.

5 NON EXACT KEYWORD SEARCH OVER ENCRYPTED DATA

Exact keyword search not fit with user needed which query keyword that user submit to cloud server may have different from original keyword either have typos or synonym keyword ,thus many schemes done to cover this problem,some of them designed to support fuzzy search and other to support synonym search .In[10]author designed new scheme that support fuzzy keyword search,he used two advanced technique to construct fuzzy keyword set i.e (wildcards technique and n grame technique) rather than the previews way where enumerate all possible words based on predefined edit distance to construct fuzzy keyword set resulting in consuming large storage.Author proposed also symbol based trie traverse searching scheme.Data owner extract keyword from document,construct fuzzy keyword set, thus compute trapdoors as a symbol for each element in fuzzy set,building index tree that covering all fuzzy keywords based on symbol set, all paths with same prefix are merged into single trie, encrypted file identifier will be indexed according to (name or addres)and index information stored at the end node.Data owners uploaded symbol based trie traverse tree on cloud.User construct fuzzy set and compute trapdoor for all element in fuzzy set constructing trapdoor set as search request.Upon server recieve search request, server divide each element in trapdoor set into sequense of symbols, thus perform the search over the tree using depth first algorithm, building index time is large and storag cost is large also .in[14]author proposed a verifiable fuzzy keyword search scheme that using wild cards technique to construct fuzzy keyword set ,also author proposed a verifable algorithm to help user to detect if cloud server excute all operations honstly or not.Firstly dataowner extract a distinct keyword from douments and used wild card technique to construct fuzzy keyword set ,thus data owner compute trapdoors for each keyword in fuzzy set ,then he divides hash value into sequence of symbols where each symbole is n bits,data owner built Tree which covering all fuzzy keywords,for each trapdoor first symbol is added into tree as a child node,root node stores two tuple (r0,r1),r0 store symbol value ;r1 store bloom filter and its signature.User construct fuzzy keyword set based on wildcard technique then compute trapdoor for each keyword in fuzzy set and send it to cloud server as search request.Cloud server after recieve search request, he transforms search request into sequence of symbols, he search for each trapdoor over index.If there are matching cloud server return corrosponding file identifiers and bloom filter as a proof.Also in[16]proposed fuzzy multi keyword search using bed tree.in [13]author proposed scheme that achive semantic search over encrypted outsourced data based on probability of terms.Semantic relationship library SRL is constructed to records semantic simalirty values of keywords for expand query keywords formed of expanded query keyword set,in SRL used data mining algorithm(Apriori algorithm) to get similarity degree between terms.author used two types of cloud ,private cloud which SRL uploaded and public cloud which encrypted index uploaded.Thus search operation devided into two steps,first one for private cloud to extend query keyword and send expanded query keywords set to public cloud,secondly for public cloud to find matching file identifiers.in[11]author proposed split multi-keyword fuzzy and synonym search using inverted index to store keywords which assosiated with corrosponding files also used balanced binary tree as searchable index and TF-IDF used for weighting the result, synonym of keyword is found firstly to construct fuzzy keyword set,for each file a leaf node in balanced binary tree is generated which store index list and file identifier using linked list after that tree is generated using postorder traversal with all leaf nodes that generated before.Each internal node contain element in linked list where each list stores keywords.System designed as follow,data owner extract keyword from files to construct keyword set KS ,for each keyword in set KS' get synonym and generate scherchable index for keywords and its synonym.Encrypted index and encrypted files are uploaded in cloud.User input query keywords and get its synonym. Cloud server search for query keyword if it matches exactly with any stored keyword in index, or matching with its synonym, then server will return files contain keyword

6 CONCLUSIONS

cloud computing technology has been widely spread in many field ,so many users used it to store their data but this data must be private so it encrypted before stored in cloud.Many reaserches have been doing to keep data privacy and make search on it fast,some of them proposed single keyword search but it not fit with users demand, so multi keyword search is proposed but still not fit with user demand as user may be query for fuzzy keyword either have typos or synonym of stored keyword.We classified searching to single encrypted keyword search ,multi encrypted keyword search , fuzzy search ,synonym search.Some of them support dynamic search where can added or delete document.

References

[1] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. Privacy-preserving multi-keyword ranked search over encrypted cloud data. *IEEE Transactions on parallel and distributed systems*, 25(1):222–233, 2014.

[2] M Chuah and W Hu. Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data. In *Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference on*, pages 273–281. IEEE, 2011.

[3] Zhangjie Fu, Kui Ren, Jiangang Shu, Xingming Sun, and Fengxiao Huang. Enabling personalized search over encrypted outsourced data with efficiency improvement. *IEEE transactions on parallel and distributed systems*, 27(9):2546–2559, 2016.

[4] Zhangjie Fu, Jiangang Shu, Xingming Sun, and Nigel Linge. Smart cloud search services: verifiable keyword-based semantic search over encrypted cloud data. *IEEE Transactions on Consumer Electronics*, 60(4):762–770, 2014.

[5] Zhangjie Fu, Xingming Sun, Zhihua Xia, Lu Zhou, and Jiangang Shu. Multi-keyword ranked search supporting synonym query over encrypted data in cloud computing. In *Performance Computing and Communications Conference (IPCCC), 2013 IEEE 32nd International*, pages 1–8. IEEE, 2013.

[6] Zhangjie Fu, Lili Xia, Xingming Sun, Alex X Liu, and Guowu Xie. Semantic-aware searching over encrypted data for cloud computing. *IEEE Transactions on Information Forensics and Security*, 13(9):2359–2371, 2018.

[7] Eu-Jin Goh et al. Secure indexes. *IACR Cryptology ePrint Archive*, 2003:216, 2003.

[8] P Kalidas and R Chandrasekaran. Searchable encryption and fuzzy keyword search in cloud computing technology. *International Journal*, 3(7), 2013.

[9] Jin Li, Qian Wang, Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou. Fuzzy keyword search over encrypted data in cloud computing. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–5. IEEE, 2010.

[10] Jin Li, Qian Wang, Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou. Fuzzy keyword search over encrypted data in cloud computing. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–5. IEEE, 2010.

[11] S Raghavendra, S Girish, CM Geeta, Rajkumar Buyya, KR Venugopal, SS Iyengar, and LM Patnaik. Split keyword fuzzy and synonym search over encrypted cloud data. *Multimedia Tools and Applications*, 77(8):10135–10156, 2018.

[12] Dawn Xiaoding Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on*, pages 44–55. IEEE, 2000.

[13] Xingming Sun, Yanling Zhu, Zhihua Xia, and Lihong Chen. Privacy-preserving keyword-based semantic search over encrypted cloud data. *International journal of Security and its Applications*, 8(3):9–20, 2014.

[14] Jianfeng Wang, Hua Ma, Qiang Tang, Jin Li, Hui Zhu, Siqi Ma, and Xiaofeng Chen. A new efficient verifiable fuzzy keyword search scheme. *JoWUA*, 3(4):61–71, 2012.

[15] Wei Zhang, Sheng Xiao, Yaping Lin, Ting Zhou, and Siwang Zhou. Secure ranked multi-keyword search for

multiple data owners in cloud computing. In *Dependable Systems and Networks (DSN), 2014 44th Annual IEEE/IFIP International Conference on*, pages 276–286. IEEE, 2014.

[16] Zhenjie Zhang, Marios Hadjieleftheriou, Beng Chin Ooi, and Divesh Srivastava. Bed-tree: an all-purpose index structure for string similarity search based on edit distance. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 915–926. ACM, 2010.